Mapping Aggregation of Hemoglobin S Using Allostery Changes Induced by Point Mutations

Patrick Weinkam[a,*] and Andrej Sali[a,b,*]

[a] Department of Bioengineering and Therapeutic Sciences, [b] Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, San Francisco, CA 94158, USA.

[*]Corresponding authors: pweinkam@salilab.org and sali@salilab.org (415 514-4227)

Abstract

Conformational dynamics of proteins, including allostery and polymerization, can generally be probed by point mutations and ligand binding. Here, we predict the impact of mutations on allostery using a generalized version of our previously developed method that now accounts for multiple ligand binding sites. First, we predict the impact of 866 annotated mutations on hemoglobin's oxygen binding equilibrium. We then characterize a subset of 30 mutations that occur in the presence of the sickle cell mutation and whose effects on polymerization have been measured. Seven of these HbS mutations occur in three predicted druggable binding sites that might be exploited to directly inhibit polymerization; one of these sites is not apparent in the crystal structure but is predicted based on our simulations. In addition, we discuss the potential relationship between mutation-induced conformational change within a single tetramer and the corresponding impact on polymerization. Our model predicts that the relative stability between conformational substates is consistent with previous descriptions of hemoglobin and allostery in general. The predictions provide a hypothesis for why hemoglobin evolved to have multiple subunits and a persistent, low frequency sickle cell mutation resulting in HbS.

Keywords

energy landscape; funnel; polymerization; molecular dynamics; machine-learning

Introduction

For decades, hemoglobin has been a model system for studying proteins. The discovery that mutated hemoglobin (HbS) plays a role in sickle cell anemia was the first time that a specific protein was linked to a genetic disease[1]. Hemoglobin A (HbA) was used in the first mechanistic descriptions of allostery[2,3], which led to the characterization of hundreds more allosteric proteins[4]. Since these pioneering studies, there has been much progress regarding the allosteric mechanism of HbA[5-7] and the polymerization mechanism of HbS[8-14]. In fact, the two processes can be coupled because HbS polymers primarily consist of deoxygenated hemoglobin. By exploiting this coupling, researchers discovered many ligands for hemoglobin that alter the allosteric transition and in turn can reduce polymerization[15,16] (hematologists disagree whether this strategy is appropriate). Nevertheless, there are no effective therapies for sickle cell disease. Stem cell transplantation has yielded promising results[17] but is not currently cost effective for most patients. Instead, treatments focus on stimulating

the production of fetal hemoglobin[18,19], providing supplemental oxygen, and treating symptoms such as pain. To contribute towards the discovery of HbS aggregation modulators, we apply our previously developed allostery model[20,21] and predict which surface sites on HbS should be targeted to inhibit polymerization.

An allosteric transition involves an equilibrium between the effector bound and unbound states, each of which follows a different energy landscape (Figure 1). An energy landscape describes the relative stabilities of all conformations for a system in a specific chemical environment[22,23]. The effector bound and unbound landscapes have two conformational substates, one that binds the effector and another that binds the effector less well. A substate may contain diverse conformations, or microstates, which are separated by energy and/or entropy barriers. A challenge for any protein dynamics model is to account for how perturbations, such as point mutations and ligand binding, can affect these complex and hierarchal energy landscapes[24,25].

Our allostery model is a dual basin structure-based Gō model[26-29], and can be used to deconstruct a protein's energy landscape into relevant substates and microstates. Atomic contacts from the effector bound and unbound crystal structures are used to define major minima in the energy landscape, which is then sampled using constant temperature molecular dynamics. The two major minima in the landscape correspond to conformational substates whose relative stability can be varied with a single input parameter (r^{AS}). All atoms within a distance r^{AS} of the effector in the crystal structure are assigned to the allosteric site. Atomic contacts in the allosteric site have a single energetic minimum corresponding to the effector bound or unbound structure, while the remaining contacts have dual minima corresponding to both bound and unbound structures. While r^{AS} is a fully adjustable parameter, we demonstrated that using a value of 12 Å allows relatively accurate predictions of the change in ligand binding free energy due to mutation[20,21]. These predictions depend on energy landscape features derived from crystal structure topology, including contact density patterns[30].

If a protein's energy landscape is affected by different perturbations, we can create separate landscapes to model conformational changes that occur under the different conditions. For instance, different solvent pH and salt concentrations can result in different side chain ionization states and therefore distinct energy landscapes. A landscape perturbation due to changes in solvent conditions is also termed chemical frustration[31,32]. Such perturbations can cause a protein to experience different allosteric[7,33] and folding mechanisms[27,32,34,35]. For hemoglobin to efficiently transport oxygen, its

energy landscape is influenced by pH induced perturbations[36], binding of different effector ligands, and polymerization of hemoglobin monomers. For systems with such complicated mechanisms, we may gain insight by monitoring conformational changes resulting from different energy landscapes, which then allows us to characterize the effect of each perturbation.

Phenomenological models of allosteric mechanisms include Monod-Wyman-Changeux (MWC)[2], Koshland-Nemethy-Filmer (KNF)[3], population shift[37,38], and induced fit[39,40]. These mechanisms differ by the degree of cooperativity observed during the allosteric transition. A highly cooperative mechanism occurs when effector binding induces a concerted change in many residues, corresponding to the KNF and induced fit mechanisms. In our allostery model, we can model such a cooperative mechanism by creating a landscape with a large allosteric site (i.e., a large r^{AS}). For smaller values of r^{AS}, our allostery model results in relatively weak coupling between residues, which is consistent with the MWC and population shift mechanisms.

Here, our goal is to predict whether or not hemoglobin has small molecule binding sites that might be used to inhibit polymerization (Figure 1). To identify such sites, we use mutations as probes of conformational dynamics. Mutations can have effects similar to drug binding, phosphorylation, and other post-translational modifications[41]. In all of these cases, the system's energy landscape is perturbed by adding, deleting, or modifying a few atoms, which may covalently or non-covalently interact with the rest of the system. While experimental data can indicate the impact of a mutation on polymerization, the mutation's impact may be due to direct interference with the polymer interface or by affecting local structure and dynamics within the single hemoglobin tetramer. Such local structural transitions could be related to allosteric conformational change. We therefore use our machine-learning method[21] to predict the impact of mutations on the oxygen binding affinity. We then identify a subset of 30 mutations that have large, measured effects on polymerization. Seven mutations occur in three predicted druggable binding sites that might be exploited to directly inhibit polymerization. Ligands that bind at such an interface may inhibit polymerization more effectively than ligands that primarily affect the allosteric transition. We also discuss the model's implications regarding phenomenological views of allostery as well as hemoglobin's evolution.

Methods

Our approach is described in Figure 1. Briefly, we create energy landscapes for binding of oxygen and DPG. We then sample the energy landscapes by constant temperature molecular dynamics simulations. The resulting structural ensembles are analyzed using machine-learning[21] to predict the impact of a mutation on the binding affinity for either oxygen or DPG. These separate predictions are combined into a single prediction of the mutation effect on the oxygen binding equilibrium, using a calculation of atomic density around the effector binding sites. We then focus on a subset of these predictions corresponding to HbS mutations whose impact on polymerization has been measured. We discuss the relationship between allostery and polymerization using a simplified equilibrium: $R \rightleftharpoons T \rightleftharpoons (T)_{30}$ where the "R" corresponds to the oxy substate (i.e. the oxy quaternary structure) and "T" corresponds to the deoxy substate (i.e. the deoxy quaternary structure). The factor of 30 is used as a rough approximation of the average aggregate size that forms at the polymerization critical concentration and is approximately the number of HbS molecules in the nucleus required to initiate polymerization[42] (Results). We then identify three predicted druggable binding sites that might be exploited to directly inhibit polymerization. Each step listed above is indicated by an arrow in Figure 1 and explained in detail in the following sections.

Allostery model simulations for oxygen and DPG binding

Simulations of ligand-induced dynamics were performed using our web server at http://salilab.org/allosmod/[20,21]. The allostery model simulations of HbA are based on several effector bound and unbound landscapes with allosteric effectors including the 4 oxygen molecules and 1 DPG molecule. Constant temperature molecular dynamics simulations are used to sample the landscapes. 30 simulations are ran for each effector bound and unbound landscape at 3 different r^{AS} (6, 9, and 12 Å for oxygen and 9, 15, and 18 Å for DPG). In each simulation, the system is first equilibrated starting from a perturbed structure that is an interpolation between the input crystal structures (2DN2 and 2DN1 or 1B86) and then simulated for 6 nanoseconds using three femtosecond time steps and velocity rescaling every 200 steps. The simulations may undergo sampling that represents much more than 6 nanoseconds of real time because the energy function is smooth and allows fast conformational (energetic) equilibration.

Assigning substates and microstates: QI_{diff}

Substates and microstates within the simulation structural ensembles are assigned using QI_{diff} calculated over one or more residues: QI_{diff} is defined as $(Q^{e+} - Q^{e-}) / (1 - \Delta Q)$ where Q is the overall fold similarity[43] to the effector bound (e+) or the unbound (e-) crystal structure and ΔQ is the structural similarity (Q) between the effector-bound and unbound crystal structures:

$$Q^t = \frac{1}{N} \sum_{i<j+1}^{N} \exp[-(r_{ij} - r_{ij}^t)^2 / 2\sigma_{ij}^2] .$$

The summation is over all pairs of sequentially non-adjacent residues. N is the number of contacts, r_{ij}^t is the side chain center of mass distance between residues i and j in structure t, and σ_{ij} is 2 Å.

$QI_{diff}(i)$ is a distance similarity metric that describes the local environment of residue i, which is positive if a residue's configuration is closer to the effector bound structure than to the effector unbound structure and negative otherwise[20]. We assign the effector bound or unbound substate with QI_{diff} calculated using all residues in all oxygen binding sites. Microstates are assigned with QI_{diff} calculated using all residues at each oxygen binding site. The oxygen binding site is defined as all residues with side chain centers of mass less than 11 Å from oxygen in the crystal structure.

Machine-learning based on simulation trajectories: allostery mutation effects

Mutation effect predictions utilize a machine-learning algorithm that relates experimentally characterized mutations to structures from our allostery model simulations[21]. Each mutation effect prediction requires 37 features, including those based on molecular mechanics energy functions, entropy calculations, stereochemical effects, mutation properties, and predictions of coupling between sites. These features capture local properties of the mutation and global properties of the entire system. To train the method, we use a boosted decision tree regression algorithm, available in the "Toolkit for Multivariate Data Analysis" as part of Root[44], to relate a set of experimentally measured mutation effects to the corresponding 37 features. For hemoglobin, we train the decision tree on 9 unrelated proteins (152 mutations). While the 9 proteins differ in protein function and experimental data types, mutation effects are defined for hemoglobin to be the $\Delta\Delta G$ of the oxygen dissociation reaction: $\Delta\Delta G'^{oxy} = \Delta G^{mut} - \Delta G^{wt} = RT \log(K_d^{wt} / K_d^{mut})$, which directly measures the equilibrium shift between the oxy and deoxy conformational substates[21]. Importantly, K_d^{wt} and K_d^{mut} are measured in the same solvent conditions. For oxygen binding, K_d is $P50^n$ where P50 is the midpoint of the oxygen

dissociation reaction and n is the Hill coefficient, which is set to 2.7. For DPG binding, $\Delta\Delta G'^{DPG}$ = -RT log(K_d^{wt} / K_d^{mut}) where K_d is the dissociation constant of DPG. The negative sign allows direct comparison of the effects of DPG (an inhibitor of oxygen binding) and oxygen.

Spatial density calculation: assessing the contribution of each binding site to the overall impact of mutation on oxygen binding ($\Delta\Delta G^{oxy}$)

In systems with multiple effector binding sites, mutations may have complicated effects on the allosteric transition because of the coupling between the sites. The spatial density calculation allows mutation effects to transmit further in less dense regions of the protein than in more dense regions. We therefore combine separate predictions ($\Delta\Delta G'^{oxy}$ and $\Delta\Delta G'^{poly}$) into a single prediction of a mutation's effect on the oxygen binding equilibria ($\Delta\Delta G^{oxy}$) using a Boltzmann-like average based on the spatial density (SD):

$$\Delta\Delta G^{oxy} = \sum_{lig} \left(e^{(-SD_{lig}/0.0005)} \Big/ \Sigma_i e^{(-SD_i/0.0005)} \right) \times ME_{lig}$$

where both summations (i and lig) iterate over the 5 binding sites, ME_{lig} is the mutation effect calculated from the trajectory with the effector lig ($\Delta\Delta G'^{oxy}$ or $\Delta\Delta G'^{DPG}$), SD is the spatial density, and 0.0005 allows smooth interpolation between multiple ME_{lig} values. SD is based on the atomic density of the region between the ligand and mutation site, calculated from the ligand bound crystal structure:

$$SD = (1 - F) \left(N_{atoms} \Big/ (4/3) R_g^3 \right) + F \times 0.105$$

$$F = 0.5\,(1 + tanh\,(0.2(r_{ligand} - 30)))$$

where r_{ligand} is the center of mass distance between the mutated side chain and the ligand, N_{atoms} is the number of non-hydrogen atoms in the region defined by the intersection of the 2 spheres with radius r_{ligand} centered on either the mutation site or the ligand, and R_g is the radius of gyration of the atoms in that region. Heme atoms are counted in N_{atoms} if the heme is directly in-between the ligand and mutation site, which allows for mutants on the oxygen-proximal side of the heme to have more of an effect on binding than mutants on the oxygen-distal side. F is a sigmoidal function is parameterized to ensure a smooth transition of the spatial density at long distances to a value 0.105,

which is the maximum density at 30 Å calculated using 10 crystal structures. Residues 30 Å from the binding site are more affected by intra-protein contacts than ligand binding and therefore the spatial density calculation has no meaning for residues greater than 30 Å from the binding sites.

Comparing impact of mutation on oxygen binding ($\Delta\Delta G^{oxy}$) to polymerization

The prediction of a mutation's impact on the oxygen binding equilibria ($\Delta\Delta G^{oxy}$) can be used to hypothesize its impact on polymerization. Based on the simplified allostery and polymerization equilibria proposed above, a mutation's impact on allostery is approximately coupled to polymerization: $\Delta\Delta G^{poly} = RT\log\left(\left[T^{wt}\right]^{30}\Big/\left[T^{mut}\right]^{30}\right)$. This equation allows us to relate allostery to polymerization by expressing the concentration of unbound hemoglobin quaternary structure (T) as a function of the free energy difference between oxy and deoxy substates (ΔG^{oxy}) and the allostery mutation effect ($\Delta\Delta G^{oxy}$). The hypothetical impact of allostery on polymerization is plotted for discrete values of ΔG^{oxy}:

$$\Delta\Delta G^{poly} = 30RT\log\left(\left(1+\exp\left(-\Delta G^{oxy}/k_BT\right)\right)\Big/\left(1+\exp\left(-\left(\Delta G^{oxy}+\Delta\Delta G^{oxy}\right)/k_BT\right)\right)\right)$$

The factor of 30 in the equation can change dramatically in different solvent conditions (Results). The hypothetical $\Delta\Delta G^{poly}$ equation remains valid because curves plotted at different ΔG^{oxy} values cover a range that is similar to the curves plotted using factors between 10 and 80 (instead of 30). The polymer size is not likely much less than 10, and while polymers can get very large, the curves change exponentially less when increasing the factor above 80. This latter fact may be irrelevant because the equation only needs to be valid near the polymerization critical concentration, at which monomers are in equilibrium with polymers of limited size.

Measured impact of mutations on polymerization

Polymerization data were collected from several studies based on 4 techniques. Each experiment directly or indirectly measures changes in the concentration of HbS critical for polymerization: 1) solubility midpoint measured by ultracentrifugation (c^{sat})[10,11,13], 2) hemoglobin concentration at which oxygen binding affinity drops rapidly (c^*)[10,12-14], 3) solubility at a high ionic strength of 2 M phosphate (s)[9], and 4) ionic strength at which the solubility is 10^{-5} M (i)[8]. Based on the simplified equilibria

described above, polymerization mutation effects are defined as $\Delta\Delta G^{poly} = 30\lambda RT \log(X^{wt} / X^{mut})$, where X is one of the experimental data and λ is a correction factor. Importantly, X^{wt} and X^{mut} are measured in the same solvent conditions. The correction factor λ is used to account for different data types. The factor is set by minimizing the difference between mutation effects measured for chemically similar mutations at the same site: 1 for c^{sat}, 1 for c*, 0.13 for s, and 1.17 for i. For a value of λ, $\Delta\Delta G^{poly}$ for one experimental method correlates with the $\Delta\Delta G^{poly}$ for any other experimental method when comparing chemically similar mutations. The following section provides an error estimate resulting from our approximations.

Prediction error

Accuracy must be considered to assess the significance of the predictions. There are two noisy sources of information: 1) predictions of the allostery mutation effect ($\Delta\Delta G^{oxy}$) and 2) measurements of polymerization mutation effect ($\Delta\Delta G^{poly}$). Noise in $\Delta\Delta G^{poly}$ is due variable experimental methods and solvent conditions while noise in $\Delta\Delta G^{oxy}$ is due to inaccuracy of our predictions, part of which is attributed to experimental error caused by varying solvent conditions. We demonstrate that the prediction error is small enough to justify the conclusions, as follows. Varying experimental conditions can contribute to experimental error due to systematic changes in hemoglobin's structure and dynamics[36] and can be estimated from linear relationships between free energy and either pH or temperature[45]. The average difference between experiments is about 0.3 pH units and 8 ^{o}C, which would yield errors in $\Delta\Delta G^{oxy}$ of 0.3 k_BT and 0.4 k_BT, respectively. If solvent conditions are not known, we can estimate experimental error in $\Delta\Delta G^{oxy}$ using measurements of chemically similar mutations at the same site. The average unsigned error in experimental measurements of $\Delta\Delta G^{oxy}$ between all pairs of similar mutations is 0.2 k_BT for whole blood samples and 0.9 k_BT for purified hemoglobin. The large error for purified hemoglobin is due to the low oxygenation midpoint (P50) that scales similarly to the Hill coefficient (n) resulting in noisy dissociation constants ($P50^n$): $27^{2.7}$ and $5^{2.9}$ for whole blood and purified HbA, respectively[46]. Purified hemoglobin data is not used in the analysis to avoid an increase in our prediction error. Similarly, we estimate that the average unsigned error in $\Delta\Delta G^{poly}$ is 1.4 k_BT for all pairs of similar mutations. In summary, the experimental error may be as much as 0.9 k_BT for $\Delta\Delta G^{oxy}$ and 1.4 k_BT for $\Delta\Delta G^{poly}$. These estimated errors are similar to the averaged unsigned error in our $\Delta\Delta G^{oxy}$ predictions, which was 1.3 k_BT in our previous study[21] and 0.8 k_BT in the current study. The difference between $\Delta\Delta G^{oxy}$ and the hypothetical curves describing the relationship

between allostery and polymerization is greater than 1 k_BT for $\Delta\Delta G^{oxy}$ and greater than 5 k_BT for $\Delta\Delta G^{poly}$. The signal is therefore significantly greater than the noise.

We calculate the likelihood of accurately predicting mutation effects using an error score empirically derived from our previous study[21]. The features that increase error, from least to most, are: 1) wild type residue is charged, 2) mutation to a charged residue, 3) mutation increases side chain size by 3 or more atoms, and 4) mutation is less than 8 Å from binding site. The error score is a sum of factors pertaining to these features: 0.2, 0.5, 1.0, and 2.0. A score of less than 1.3 implies a mutation effect prediction that should be on average less than 1 k_BT from the correct value. Mutations with scores of greater than 1.3 are omitted from analysis to avoid large outliers.

Predicting druggable binding pockets

We predict druggable pockets by applying the program FPocket[47] to snapshots from the oxygen bound and unbound simulations (600 each). The FPocket druggability score was obtained by machine-learning optimization against a dataset of drug-bound and nondrug-bound crystal structures. Here, we create a residue specific score, d_i, which is the druggability of the most druggable pocket that has a vertex (FPocket uses to identify pockets) within r^{cutoff} of the residue's side chain center of mass. r^{cutoff} is 11 Å when identifying pockets for HbS and 6 Å when monitoring the oxygen binding pocket. We also calculate the probability that $d_i > 0.5$ in the simulation snapshots: $P_{d>0.5} = \sum_{d>0.5} P_i$ where the summation occurs over all snapshots with $d_i > 0.5$ and P_i is the Boltzmann weighted probability of each structure. P_i is given by $\exp(-E_i / \sigma_i) / \sum_i \exp(-E_i / \sigma_i)$ where σ_i is the standard deviation of the energy. This residue-proximal druggability score can be used to identify clusters of residues near a highly druggable pocket. Residues flanking a binding pocket have similar d_i distributions.

Results

Substates and Microstates in the Oxygen Binding Equilibrium

We model and sample several distinct oxygen bound and unbound landscapes that differ by the chosen allosteric site radii (r^{AS}). The bound (unbound) landscape involves implicit modeling of ligand binding by biasing the allosteric site structure with contacts from the bound (unbound) crystal structure. However, when sampling the unbound landscape, some oxygen binding sites populate an oxygen bound-like structure. These structural changes could occur even in the absence of oxygen and may even affect polymerization. We therefore monitor the possibility of ligand binding using the binding site structure, which is potentially bound if more similar to the oxygen bound crystal structure than the unbound crystal structure (using pairwise distance similarity metric QI_{diff}), and vice versa if not bound. For hemoglobin, 16 microstates exist based on whether or not oxygen binding occurs at the 4 binding sites. The population of the microstates calculated from the oxygen bound (red) and unbound (blue) simulations is sensitive to the input parameter r^{AS} (Figure 2). With r^{AS} less than 12 Å, the microstates can be grouped into similarly populated oxy and deoxy substates. These simulations are omitted from further analysis because hemoglobin structural ensembles should differ in the oxygen bound and unbound states.

The model predicts that the oxy substate is more stable than the deoxy substate. With r^{AS} equal 12 Å, the oxy substate is 74% populated in the oxygen bound simulation, while the deoxy conformational substate is 66% populated in the unbound simulation (substates are defined in Figure 2). In comparison, low r^{AS} results in similarly populated substates, which indicates that the unequal populations at high r^{AS} are not an artifact of substate assignment. The unequal substate populations at high r^{AS} conflict with the equivalent relative stabilization energy in landscapes with the same r^{AS}, in which the oxy crystal structure should be favored in the oxygen bound landscape by the same amount as the deoxy crystal structure in the unbound landscape. In this case, entropy drives the stability of the oxy substate because there are more ways to satisfy oxygen bound conformations than unbound conformations. This stability difference is consistent with previous molecular dynamics studies[48,49] and predictions from a Gaussian network model that the carbon monoxide bound structure (similar to that with oxygen bound) is entropically more stable than the unbound structure[50].

The simulations also predict varying populations of microstates. The fully oxygen bound microstate is dominant in the oxygen bound simulation, while the fully unbound microstate is never dominant (Figure 2). In transition from these microstates to either the single-oxygen bound microstate or the triple-oxygen bound microstate, the α subunits are more likely than the β subunits to be oxygenated or deoxygenated, respectively. This result is consistent with a study of the unbound crystal that found

a stronger preference for oxygenation of the α subunits compared to the β subunits[51]. The simulations may provide an explanation for this observation. We predict that the α subunit oxygen binding site conformational ensembles differ only slightly between the oxygen bound and unbound simulations (Figure 3). In contrast, the β subunit oxygen binding sites populate distinct conformational ensembles in the oxygen bound and unbound simulations, in agreement with a previous study[48]. Therefore, the β subunits may be most important for determining the oxygen binding state while the α subunits undergo relatively modest changes in conformation and oxygen binding.

Predicting the impact of mutation on oxygen binding: $\Delta\Delta G^{oxy}$

Allosteric coupling in hemoglobin involves one DPG binding site and four oxygen binding sites. Allostery occurs because the four oxygen binding sites are coupled to the tertiary/quaternary structure and therefore oxygen binding at one site positively modulates binding at another site. This positive cooperativity is broken by a single DPG ligand that has strong binding affinity to a pocket in the unbound conformation. The DPG binding pocket is more highly solvated than most binding pockets and contains residues not supported by a dense network of interactions (Figure 4D), which makes DPG binding susceptible to perturbations, such as mutations.

We predict impact of mutations on the binding of oxygen or DPG. As described previously[21], we define a mutation effect as the free energy change of ligand binding due to mutation ($\Delta\Delta G'^{oxy}$ and $\Delta\Delta G'^{poly}$ in Figure 1). Each mutation effect prediction requires features that either describe the mutation itself or are calculated from the simulations of the allostery model of HbA (Methods). A given mutation effect is relatively accurately predicted using either the oxygen binding or DPG binding simulations (Figure 4A-B). The results show that mutations far from the DPG binding site (greater than 20 Å) are well predicted using the oxygen bound simulations and the remaining mutations are well predicted using the DPG bound simulations (average unsigned error of 0.70 k_BT and 0.96 k_BT, respectively). This trend may indicate that a mutation impacts the DPG binding site at further distances than for an oxygen binding site. If so, there must be a physical explanation for how mutations affect one binding site more than another. We propose such an explanation, as follows.

The atomic density surrounding ligand binding sites can be used to assess the contribution of each binding site to the overall impact of a mutation on oxygen binding. We use a spatial density calculation, which allows mutation effects to transmit further in less dense regions of the protein than

in more dense regions (Methods), to combine separate predictions ($\Delta\Delta G'^{oxy}$ and $\Delta\Delta G'^{poly}$) into a single prediction of a mutation's impact on the oxygen binding equilibria ($\Delta\Delta G^{oxy}$). In support of this approach, we analyzed mutation effects on binding sites from 10 different proteins[21]. We observed that while significant mutation effects on ligand binding (> 2 k_BT) generally occur at sites within 8 Å of the ligand, significant mutation effects on hemoglobin oxygen binding regularly occur at sites much further than 8 Å from oxygen or DPG. Our spatial density calculation exploits the fact that the atomic density of the region within 20 Å of the DPG binding site is similar to the density of the region within 8 Å of a typical binding site (Figure 4D). Using the spatial density calculation, our combined predictions yield an average unsigned error of 0.84 k_BT and a 0.76 correlation with experiment (Figure 4C).

Impact of mutations on allostery and polymerization

Allostery can be coupled to polymerization because polymers primarily consist of hemoglobin in the unbound quaternary structure. We can approximate the allostery and polymerization equilibria to be: $R \rightleftharpoons T \rightleftharpoons (T)_{30}$ where the "R" or "relaxed" species corresponds to the oxy substate (i.e. the oxy quaternary structure) and the "T" or "tensed" species corresponds to the deoxy substate (i.e. the deoxy quaternary structure). The factor of 30 is used as a rough approximation of the average aggregate size that forms at the polymerization critical concentration and is approximately the number of HbS molecules in the nucleus required to initiate polymerization[42] (Methods). The first, rate limiting step in HbS aggregation is homogeneous polymerization. The next step is heterogeneous polymerization, which involves new polymers nucleating from existing polymers[5,52]. Varying solvent conditions and temperature can shift the ratio of the polymerization types[53]. Consequently, the size of the aggregates at the polymerization critical concentration can vary significantly, as indicated by the total aggregation rate that can depend on up to the 80th power of HbS concentration[53]. We account for this heterogeneity by varying another parameter (the relative stability of R and T), which can be used to describe a larger range of conditions than can be achieved by varying the aggregate size (Methods).

While allostery and polymerization is coupled in the presence of oxygen, there is uncertainty regarding the conformational changes that occur in the absence of oxygen, i.e. the conditions of some polymerization experiments. Studies of gel encapsulated hemoglobin in deoxygenated conditions demonstrate carbon monoxide rebinding kinetics that suggest multiple conformations (not

a single T state structure), including at least one that binds oxygen with a higher affinity than the T state[54]. In fact, this state (T-high) is thought to be on pathway between the R and T structures[54]. This data is consistent with other results that suggest a transiently stable R-like conformation in deoxy conditions: 1) the oxy substate is predicted to be more stable than the deoxy substate in other computational studies[48-50] in addition to the work presented here and 2) the existence of a deoxyhemoglobin crystal structure in the quaternary oxygen bound conformation with a well-structured oxygen binding pocket[55]. Contrary to these points, a previous study has fit a two state model to polymerization data, which suggests that the allosteric transition is unnecessary to interpret the data[56]. For polymerization experiments in deoxy conditions, the conformational changes induced by mutation within a single tetramer are not well characterized. For example, if adding a small oxygen molecule can shift the equilibrium between the R and T structures sufficiently for allostery to become important for polymerization, so could in principle a single point mutation. We therefore use the hypothetical, simplified equilibrium mentioned above to describe how the impact of a mutation on allostery is related to its corresponding impact on polymerization.

We calculate several hypothetical curves that demonstrate how a mutation's impact on allostery can affect polymerization (Methods). Figure 5 shows the hypothetical curves at different relative stabilities between the oxy and deoxy substates, representing varying oxygen pressure or concentrations of allosteric effectors. We also plot a set of experimentally measured HbS mutations[57] using the measured polymerization mutation effects ($\Delta\Delta G^{poly}$) and the predicted allostery mutation effects ($\Delta\Delta G^{oxy}$). Here, we assume that all HbS mutations are uncoupled to β-Glu6Val. Most of the data strongly deviates from the hypothetical curves, which suggests that allosteric conformational changes are not dominant during polymerization. Because most of the experiments are performed in deoxygenated conditions, conformational changes to the oxy substate may be limited. Therefore, a mutation may be more likely to induce a local structure change rather than a quaternary structure change. Nonetheless, we can identify sites of mutations that perturb polymerization and then predict potentially druggable binding pockets near these sites, many of which are close to a protein-protein interface in the HbS crystal[58].

Druggable binding pockets that can be used to affect polymerization

We predict druggable pockets by applying the program FPocket[47] to the structures from our allostery model simulations. Due to structural changes, pockets grow and disappear during the simulation

trajectories, thereby causing pocket assignment issues. We therefore create a residue specific score d_i, which is the druggability of the most druggable pocket within a cutoff distance of residue i (Methods). We can predict druggable pockets using d_i distributions calculated from simulation trajectories.

We identify three druggable pockets that are near mutations that perturb polymerization. These pockets are also near polymer interfaces in the HbS crystal structure (Figure 6). Pocket 3 has a bimodal d_i distribution indicating distinct conformations sampled in the simulation (Figure 7A). In comparison, pockets 1 and 2 have unimodal distributions and are more likely to form druggable pockets. Pockets 1 and 2, however, may not be better targets than pocket 3, which is: 1) more druggable in deoxyhemoglobin than oxyhemoglobin, 2) adjacent to sites of mutation predicted to decrease oxygenation, and 3) adjacent to sites of mutation predicted to significantly decrease polymerization. Therefore, a ligand may bind specifically to pocket 3 in deoxyhemoglobin and inhibit a polymerization interface.

Pocket prediction using simulation trajectories can be advantageous in the case of dynamic proteins such as hemoglobin. For instance, ligand binding pockets can form transiently even if no pockets exist in the crystal structure[59]. Even in the absence of significant structural change, ligand binding prediction based on structural ensembles can be more accurate than predictions based on a single structure[60,61]. In hemoglobin, pockets populate slightly expanded conformations in the simulation trajectories compared to the crystal structure (Figure 6). These expanded conformations could be favored upon ligand binding and in turn perturb an interface in the HbS polymer. In the simulation trajectories, pocket 3 populates such an expanded conformation and is predicted to be much more druggable than the corresponding location in the crystal structure (Figure 7B). In contrast, pockets 1 and 2 maintain druggability in the simulation trajectories and the crystal structures.

Discussion

The original descriptions of allostery in general as well as hemoglobin in particular are phenomenological and rooted in experimental data. The dominant states of hemoglobin, with and without oxygen (R or T, respectively), were determined by X-ray crystallography. The KNF mechanism, now discredited for hemoglobin, hypothesized that oxygen binding induces a concerted change from T to R[3]. The MWC mechanism hypothesizes that there is a structural equilibrium

between R and T and that oxygen binding promotes the R state, effectively allowing oxygen binding in both the oxygen bound and unbound structures. Subsequent more intricate models account for structural details such as salt bridges that contribute to pH dependence of hemoglobin oxygen binding (known as the Bohr effect)[62-64].

More recent descriptions of allostery rely not only on the crystal structures but also broader conformational ensembles[37-40]. In fact, several structural states of hemoglobin other than R and T have been identified[65]. While the MWC and KNF mechanisms do not explicitly describe structural ensembles, they are generally consistent with the population shift and induced fit mechanisms, respectively. A clear divergence from the purely structural view of allostery is entropy driven allostery[66]. This mechanism describes dynamic coupling between sites in which the average structure remains unchanged, but includes different excursions from the average structure in the bound and unbound states.

Phenomenological mechanisms explicitly relate structural details to experimental observables. For instance, the MWC and related mechanisms[2,63,64] explain measured subunit oxygen binding affinities using the quaternary structure. Another mechanism, called TTS, decouples subunit tertiary structure from the quaternary structure[5], thereby allowing subunits to adopt the R or T conformations without a quaternary structure change. The TTS mechanism is consistent with our approach of assigning one of two substates to the binding site structure, which in our model moderately correlates with the subunit tertiary structure (correlation coefficient of 0.5 if using QI_{diff}).

Here, we rely on a model of hemoglobin allostery defined by its energy landscape rather than a phenomenological mechanism. The inputs to our model are the effector bound and unbound crystal structures and the parameter r^{AS} that controls how strongly effector binding influences the energy landscape. The output is a set of energy landscapes with minima corresponding to the effector bound and bound crystal structures as well as the corresponding simulated trajectories. These trajectories describe the interconversion between the input structures and sometimes describe conformations that are distinct from the input structures. The trajectories are then used to predict: 1) the impact of a mutation on oxygen or DPG binding, 2) the relative populations of substates and microstates, and 3) the magnitude of coupling between sites.

Our simulations describe weak coupling between hemoglobin's oxygen binding sites. Such behavior has also been reported in molecular dynamics simulations[48,49,67-69], elastic network models[50,70], and an experimental study of gel encapsulated hemoglobin that separately mapped tertiary and quaternary structure changes[71]. Weak coupling involves oxygen binding at one site triggering both tertiary and quaternary structural changes, which in turn result in changes of the size and shape of the other oxygen binding pockets. In our simulations, oxygen bound pockets are predicted to be slightly smaller and less druggable[47] than unbound pockets. Binding site dynamics is not homogeneous, however, as indicated by the stronger coupling of the quaternary structure with the oxygen binding sites in the β subunits than in the α subunits (Figure 3). This result is consistent with experiments that report a larger impact on oxygen binding for mutations in the β subunits than for mutations in the α subunits (average magnitude of 1.5 ± 0.6 k_BT and 0.5 ± 0.5 k_BT, respectively). Therefore, modulation of hemoglobin's allosteric transition may well be achieved by perturbing the interface between the two β subunits. Interestingly, such a perturbation has resulted from evolution that positioned the DPG binding pocket at the β subunit interface.

Hemoglobin has evolved to have a complex allosteric mechanism, yet also permits point mutations at many sites. Our predictions suggest that naturally occurring mutations[57] can be tolerated due to hemoglobin's structural symmetry (Figure 8). We predict α subunit mutations to typically inhibit oxygen binding, which can counteract β subunit mutations that we predict to typically promote oxygen binding (Figure 8A). Natural HbS mutations (those that occur in patients with β-Glu6Val)[57] display an even stronger predicted trend of increased oxygen binding than the β subunit mutations (Figure 8A). These HbS mutations may improve hemoglobin's oxygen delivery if they shift the equilibrium to the oxy state, sufficiently reduce polymerization, and do not increase oxygen affinity too much so that sufficient oxygen is not released.

One might expect the sickle cell mutation to be selected out of the human population, but the HbS allele persists probably because it allows malaria resistance[72]. In fact, the HbS allele occurs in 18% of some populations that suffer from a high frequency of malaria. Unfortunately, the detailed mechanism of resistance is not known. Some insight is gained by considering the impact of 5 naturally occurring HbS mutations on polymerization. These 5 mutants tend to increase polymerization even though most HbS mutants decrease polymerization (Figure 8B). If caused by selective pressure, this result suggests that malaria resistance is a direct result of polymerization, which either promotes red blood cell destruction or kills parasites more directly[73,74]. Because we predict two of these HbS mutations to

also recover oxygen binding inhibited by polymerization (Figure 5), evolution may be improving hemoglobin's ability to transport oxygen while simultaneously increasing hemoglobin's tendency for polymerization.

Interpreting the role of a mutation on hemoglobin dynamics is a challenge because so many processes can occur simultaneously. A mutation may impact: 1) oxygen binding, 2) binding of DPG or other effectors, 3) the allosteric conformational equilibrium between oxy and deoxy substates, 4) pH or temperature induced conformational changes, and 5) polymerization. Here, we use a model that describes the allosteric conformational transition, yet without an explicit model of ligand binding. This model is appropriate if a mutation affects the transition between oxy and deoxy substates instead of perturbing specific protein-ligand interactions. The relative accuracy of our predictions (Figure 4C) supports this statement for hemoglobin. Also, naturally occurring mutations occur far from the oxygen binding sites (98 % greater than 5 Å and 92 % greater than 8 Å) and therefore are unlikely to affect oxygen binding without affecting the allosteric transition. However, a more detailed model would be necessary to predict a mutation's impact on the binding affinity within the oxy substate. Similarly, a detailed model of polymerization is necessary to characterize the role of any mutation in aggregation. Our model of allostery is a convenient bridge between models of ligand-binding and polymerization that may further contribute to understanding of hemoglobin function.

Conclusion

Hemoglobin exists in a conformational equilibrium, involving allostery and polymerization, that is affected by mutations and conditions such as oxygen and DPG concentration, pH, and temperature[45]. Using separate landscapes for each ligand-induced conformational change, the prediction allows us to further interpret experimental data. In particular, we identify 3 binding sites that can potentially be used to inhibit HbS aggregation by destabilizing a polymerization interface. These sites might be more effective than sites that are distant to polymerization interfaces because ligand binding effects dissipate at long distances due to weak coupling in hemoglobin. Due to the high concentration of hemoglobin in the blood, direct inhibition of polymerization may be necessary to limit the amount of drug required to treat patients. In conclusion, mutations can serve as natural probes of function and may help identify ligand binding pockets that can be used to perturb allosteric proteins such as hemoglobin.

Acknowledgments

We are grateful for helpful discussions with Charles Homcy, Matt Jacobson, Natalia Khuri, Seung Joong Kim, and Riccardo Pellarin. The work was supported by a grant from the NIH (R01 GM083960).

References

(1) Pauling, L.; Itano, H. A. *Science* **1949**, *110*, 543.
(2) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88.
(3) Koshland, D. E.; Nemethy, G.; Filmer, D. *Biochemistry* **1966**, *5*, 365.
(4) Huang, Z. M.; Zhu, L. A.; Cao, Y.; Wu, G.; Liu, X. Y.; Chen, Y. Y.; Wang, Q.; Shi, T.; Zhao, Y. X.; Wang, Y. F.; et.al. *Nucleic Acids Research* **2011**, *39*, D663.
(5) Henry, E. R.; Bettati, S.; Hofrichter, J.; Eaton, W. A. *Biophys. Chem.* **2002**, *98*, 149.
(6) Knapp, J. E.; Pahl, R.; Srajer, V.; Royer, W. E. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 7649.
(7) Cui, Q.; Karplus, M. *Protein Sci* **2008**, *17*, 1295.
(8) Benesch, R. E.; Yung, S.; Benesch, R.; Mack, J.; Schneider, R. G. *Nature* **1976**, *260*, 219.
(9) Benesch, R. E.; Kwong, S.; Benesch, R.; Edalji, R. *Nature* **1977**, *269*, 772.
(10) Benesch, R. E.; Kwong, S.; Edalji, R.; Benesch, R. *J. Biol. Chem.* **1979**, *254*, 8169.
(11) Nagel, R. L.; Johnson, J.; Bookchin, R. M.; Garel, M. C.; Rosa, J.; Schiliro, G.; Wajcman, H.; Labie, D.; Moopenn, W.; Castro, O. *Nature* **1980**, *283*, 832.
(12) Monplaisir, N.; Merault, G.; Poyart, C.; Rhoda, M. D.; Craescu, C.; Vidaud, M.; Galacteros, F.; Blouquit, Y.; Rosa, J. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 9363.
(13) Dellano, J. J. M.; Manning, J. M. *Protein Sci.* **1994**, *3*, 1206.
(14) Li, X. F.; Himanen, J. P.; de Llano, J. J. M.; Padovan, J. C.; Chait, B. T.; Manning, J. M. *Biotechnology and Applied Biochemistry* **1999**, *29*, 165.
(15) Fylaktakidou, K. C.; Duarte, C. D.; Koumbis, A. E.; Nicolau, C.; Lehn, J. M. *Chemmedchem* **2011**, *6*, 153.
(16) Abdulmalik, O.; Ghatge, M. S.; Musayev, F. N.; Parikh, A.; Chen, Q. K.; Yang, J. S.; Nnamani, I.; Danso-Danquah, R.; Eseonu, D. N.; Asakura, T.; et.al. *Acta. Crystallogr. D.* **2011**, *67*, 1076.
(17) Fitzhugh, C. D.; Unno, H.; Hathaway, V.; Coles, W. A.; Link, M. E.; Weitzel, R. P.; Zhao, X. C.; Wright, E. C.; Stroncek, D. F.; Kato, G. J.; Hsieh, M. M.; Tisdale, J. F. *Blood* **2012**, *119*, 5671.
(18) Platt, O. S.; Orkin, S. H.; Dover, G.; Beardsley, G. P.; Miller, B.; Nathan, D. G. *J. Clin. Invest.* **1984**, *74*, 652.
(19) Sankaran, V. G.; Menne, T. F.; Xu, J.; Akie, T. E.; Lettre, G.; Van Handel, B.; Mikkola, H. K.; Hirschhorn, J. N.; Cantor, A. B.; Orkin, S. H. *Science* **2008**, *322*, 1839.
(20) Weinkam, P.; Pons, J.; Sali, A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4875.
(21) Weinkam, P.; Chen, Y. C.; Pons, J.; Sali, A. *J. Mol. Biol.* **2013**, *425*, 647.
(22) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902.
(23) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
(24) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598.
(25) Zhuravlev, P. I.; Materese, C. K.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 8800.
(26) Ueda, Y.; Taketomi, H.; Go, N. *Biopolymers* **1978**, *17*, 1531.
(27) Schug, A.; Whitford, P. C.; Levy, Y.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 17674.
(28) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins* **2009**, *75*, 430.
(29) Li, W.; Wolynes, P. G.; Takada, S. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3504.
(30) Weinkam, P.; Zong, C. H.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12401.
(31) Weinkam, P.; Romesberg, F. E.; Wolynes, P. G. *Biochemistry* **2009**, *48*, 2394.
(32) Weinkam, P.; Zimmermann, J.; Romesberg, F. E.; Wolynes, P. G. *Acc. Chem. Res.* **2010**, *43*, 652.
(33) Yifrach, O.; Horovitz, A. *Biochemistry* **1995**, *34*, 5303.
(34) Krishna, M. M.; Maity, H.; Rumbley, J. N.; Englander, S. W. *Protein Sci.* **2007**, *16*, 1946.
(35) Cho, S. S.; Weinkam, P.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 118.
(36) Benesch, R. E.; Benesch, R.; Yu, C. I. *Biochemistry* **1969**, *8*, 2567.

(37) Kumar, S.; Ma, B. Y.; Tsai, C. J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9*, 10.
(38) Motlagh, H. N.; Hilser, V. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4134.
(39) Bosshard, H. R. *News in Physiological Sciences* **2001**, *16*, 171.
(40) James, L. C.; Tawfik, D. S. *Trends in Biochemical Sciences* **2003**, *28*, 361.
(41) Salazar, C.; Hofer, T. *J. Mol. Biol.* **2003**, *327*, 31.
(42) Hofrichter, J.; Ross, P. D.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 4864.
(43) Cho, S. S.; Levy, Y.; Wolynes, P. G. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 586.
(44) Hoecker, A.; Speckmayer, P.; Stelzer, J.; Therhaag, J.; von Toerne, E.; Voss, H. *PoS* **2007**, *ACAT*, 040.
(45) Samaja, M.; Melotti, D.; Rovida, E.; Rossibernardi, L. *Clinical Chemistry* **1983**, *29*, 110.
(46) Wajcman, H.; Galacteros, F. *Hemoglobin* **2005**, *29*, 91.
(47) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. *Bmc Bioinformatics* **2009**, *10*, e1.
(48) Hub, J. S.; Kubitzki, M. B.; de Groot, B. L. *PLoS Comput. Biol.* **2010**, *6*, e1000774.
(49) Yusuff, O. K.; Babalola, J. O.; Bussi, G.; Raugei, S. *J. Phys. Chem. B* **2012**, *116*, 11004.
(50) Xu, C. Y.; Tobi, D.; Bahar, I. *J. Mol. Biol.* **2003**, *333*, 153.
(51) Mozzarelli, A.; Rivetti, C.; Rossi, G. L.; Eaton, W. A.; Henry, E. R. *Protein Sci.* **1997**, *6*, 484.
(52) Ferrone, F. A.; Rotter, M. A. *Journal of Molecular Recognition* **2004**, *17*, 497.
(53) Christoph, G. W.; Hofrichter, J.; Eaton, W. A. *Biophys. J.* **2005**, *88*, 1371.
(54) Samuni, U.; Roche, C. J.; Dantsker, D.; Juszczak, L. J.; Friedman, J. M. *Biochemistry* **2006**, *45*, 2820.
(55) Wilson, J.; Phillips, K.; Luisi, B. *Journal of Molecular Biology* **1996**, *264*, 743.
(56) Sunshine, H. R.; Hofrichter, J.; Ferrone, F. A.; Eaton, W. A. *Journal of Molecular Biology* **1982**, *158*, 251.
(57) Giardine, B.; Borg, J.; Higgs, D. R.; Peterson, K. R.; Philipsen, S.; Maglott, D.; Singleton, B. K.; Anstee, D. J.; Basak, A. N.; Clark, B.; et.al. *Nature Genetics* **2011**, *43*, 295.
(58) Harrington, D. J.; Adachi, K.; Royer, W. E. *J. Mol. Biol.* **1997**, *272*, 398.
(59) Bowman, G. R.; Geissler, P. L. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11681.
(60) Fan, H.; Irwin, J. J.; Webb, B. M.; Klebe, G.; Shoichet, B. K.; Sali, A. *J. Chem. Inf. Mod.* **2009**, *49*, 2512.
(61) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. *Bioinformatics* **2011**, *27*, 3276.
(62) Perutz, M. F. *Nature* **1970**, *228*, 726.
(63) Szabo, A.; Karplus, M. *J. Mol. Biol.* **1972**, *72*, 163.
(64) Lee, A. W. M.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 7055.
(65) Dey, S.; Chakrabarti, P.; Janin, J. *Proteins* **2011**, *79*, 2861.
(66) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. *Nat Struct Mol Biol* **2006**, *13*, 831.
(67) Gelin, B. R.; Lee, A. W. M.; Karplus, M. *J. Mol. Biol.* **1983**, *171*, 489.
(68) Mouawad, L.; Perahia, D. *J. Mol. Biol.* **1996**, *258*, 393.
(69) Fischer, S.; Olsen, K. W.; Nam, K.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5608.
(70) Eom, K.; Baek, S. C.; Ahn, J. H.; Na, S. *J Comput Chem* **2007**, *28*, 1400.
(71) Jones, E. M.; Balakrishnan, G.; Spiro, T. G. *J Am Chem Soc* **2012**, *134*, 3461.
(72) Piel, F. B.; Patil, A. P.; Howes, R. E.; Nyangiri, O. A.; Gething, P. W.; Williams, T. N.; Weatherall, D. J.; Hay, S. I. *Nature Communications* **2010**, *1*, 104.
(73) Luzzatto, L.; Nwachuku.Es; Reddy, S. *Lancet* **1970**, *1*, 319.
(74) Friedman, M. J. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 1994.

Figures

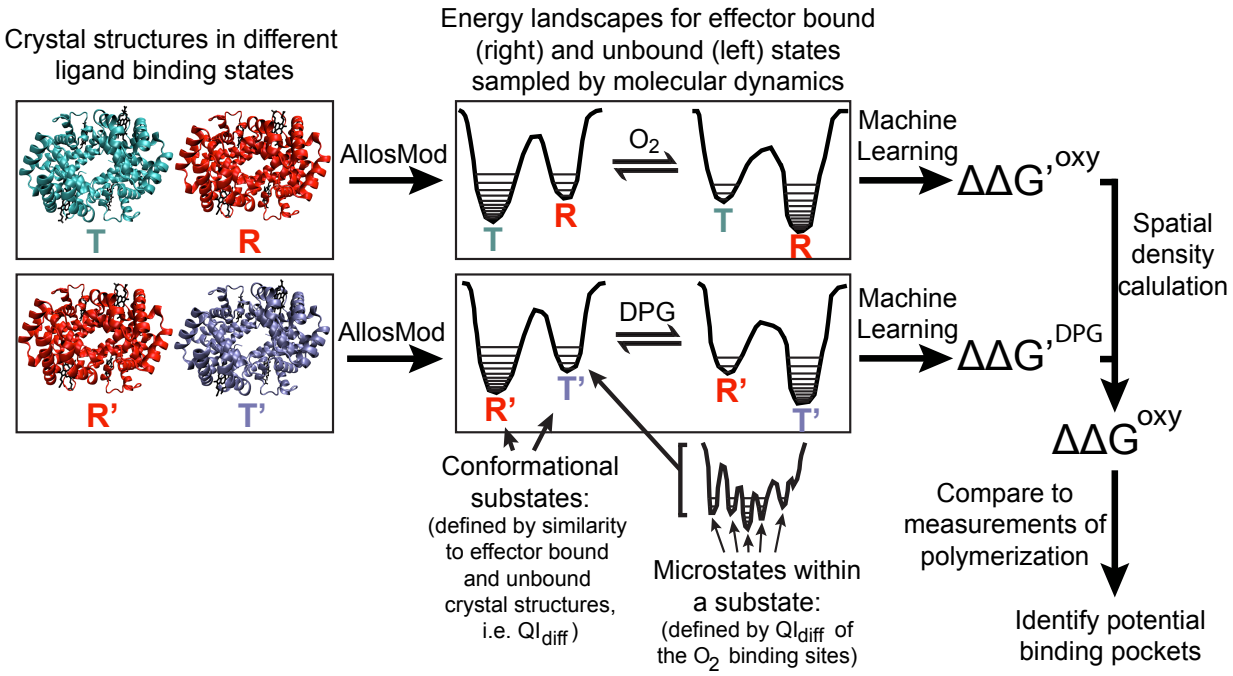

Figure 1 An overview of the current work. Each arrow represents a different Methods section.

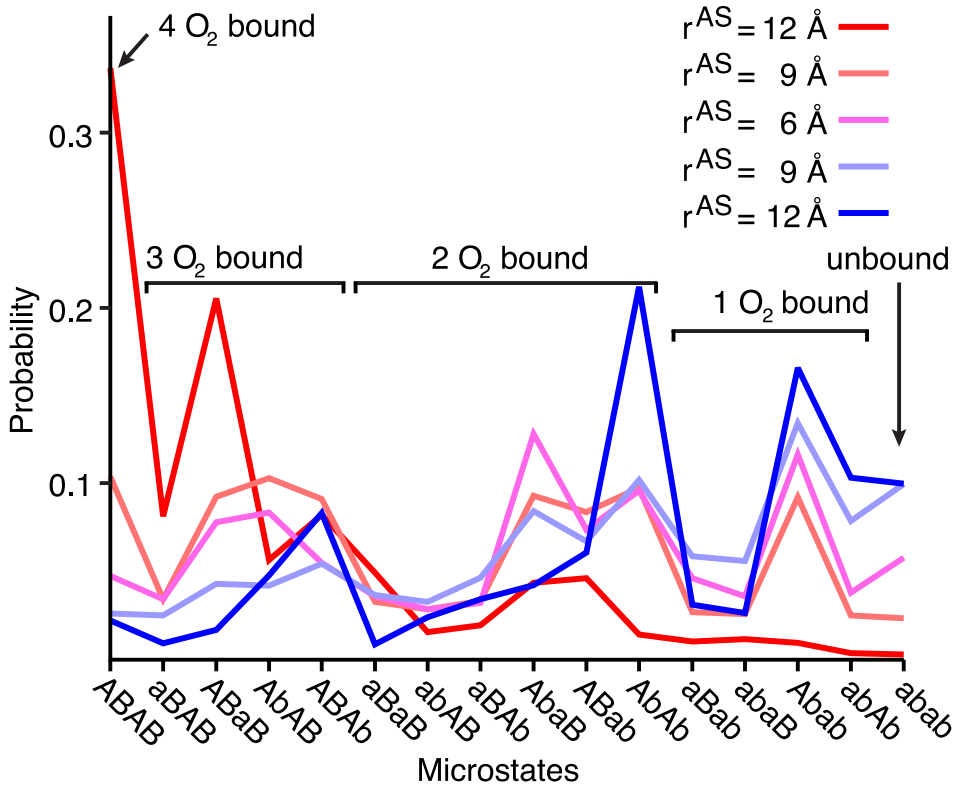

Figure 2 Probability distributions of hemoglobin's 16 microstates. Each line represents the probability distribution calculated from sampling of different energy landscapes, which have different allosteric site radii (r^{AS}) and ligand binding states (oxygen bound are shades of red and unbound are shades of blue). Microstates are defined by the conformations of the oxygen binding sites using QI_{diff}. Microstates are labeled ABAB corresponding to $\alpha_1\beta_1\alpha_2\beta_2$ where capital or lowercase letters imply oxygen bound ($QI_{diff} > 0$) or unbound ($QI_{diff} < 0$), respectively. The 5 left most and 5 right most microstates indicate the oxy and deoxy conformational substates, respectively.

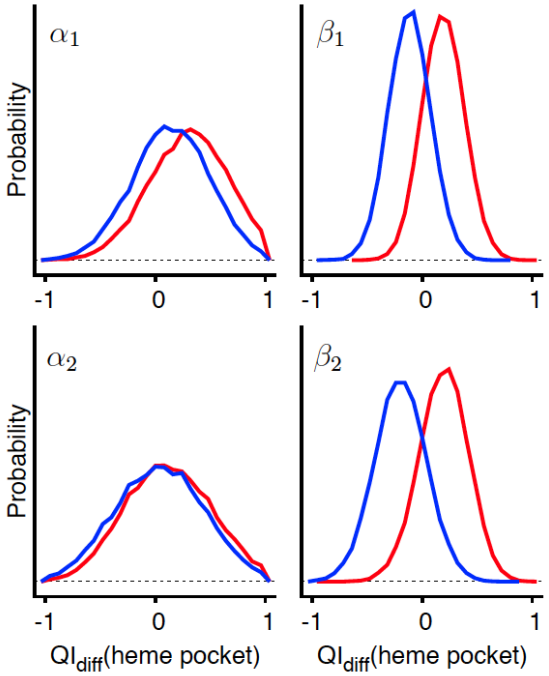

Figure 3 Probability distributions of QI_{diff} for the 4 oxygen binding sites calculated from the oxygen bound (red) and unbound (blue) simulations (r^{AS} = 12 Å). QI_{diff} is 1 if the oxygen binding site is more similar to the oxygen bound crystal structure than the unbound crystal structure and vice versa for QI_{diff} equal to -1. The probability overlap of the bound to unbound distributions is 86% for the α subunits and 41% for the β subunits, which suggests the β subunit binding sites are more highly coupled to the quaternary structure than the α subunit binding sites.

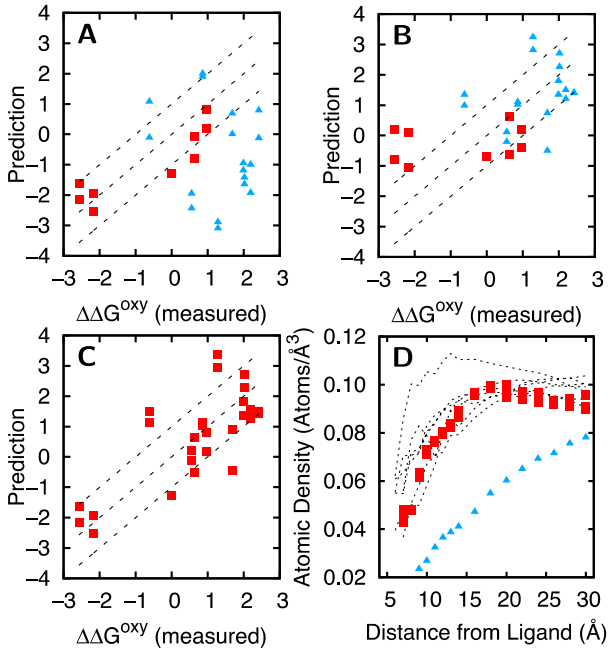

Figure 4 We predict the impact of mutations (in units of k_BT) on: A) oxygen binding ($\Delta\Delta G'^{oxy}$) and B) DPG binding ($\Delta\Delta G'^{DPG}$). Mutations further than 20 Å from the DPG binding site (red squares) are well predicted using the simulations with oxygen binding and the remaining mutations (blue triangles) are well predicted using the simulations with DPG binding. C) The predictions in A and B are combined into a single prediction of $\Delta\Delta G^{oxy}$ using a spatial density calculation (Methods). D) Atomic density of the non-hydrogen atoms around ligand binding sites in several proteins. Red squares and blue triangles represent densities around hemoglobin's oxygen binding sites and DPG binding site, respectively. Dashed lines are for other protein's binding sites[21].

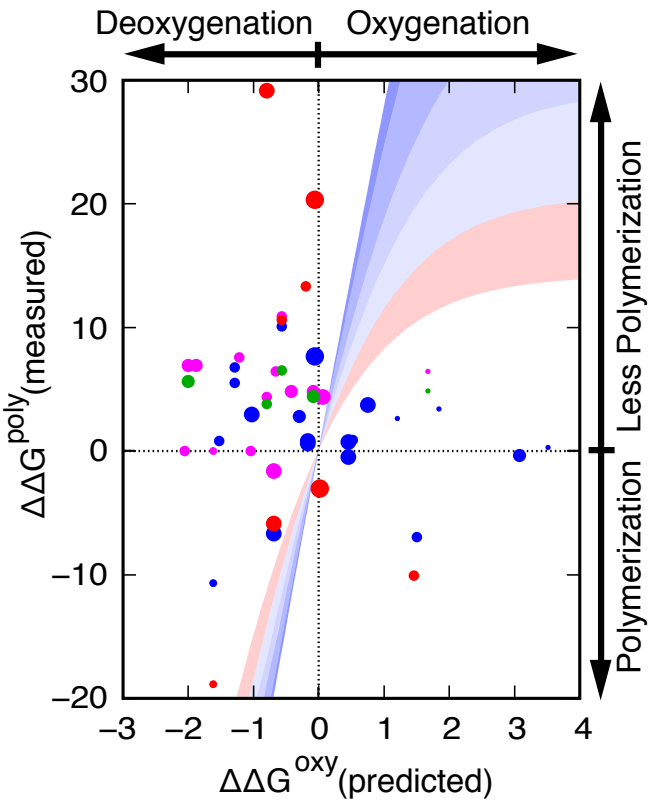

Figure 5 We characterize hemoglobin polymerization by comparing the contribution of a mutation's impact on allostery and polymerization. We plot a mutation's measured impact on polymerization ($\Delta\Delta G^{poly}$) and the corresponding predicted impact on allostery ($\Delta\Delta G^{oxy}$) as well as hypothetical curves that approximate how allostery could be coupled to polymerization (in units of k_BT). The point radius is inversely proportional to the predicted error (Methods) and the color represents different measurements of polymerization: red is c^{sat}, blue is c*, green is i, and magenta is s (Methods). The hypothetical curves (shades of blue to red) correspond to different ratios between the oxy and deoxy substates: dark blue from 95% deoxy to 88% deoxy, then to 73% deoxy, then to 62% deoxy, then to 50% deoxy, and light red to 38% deoxy.

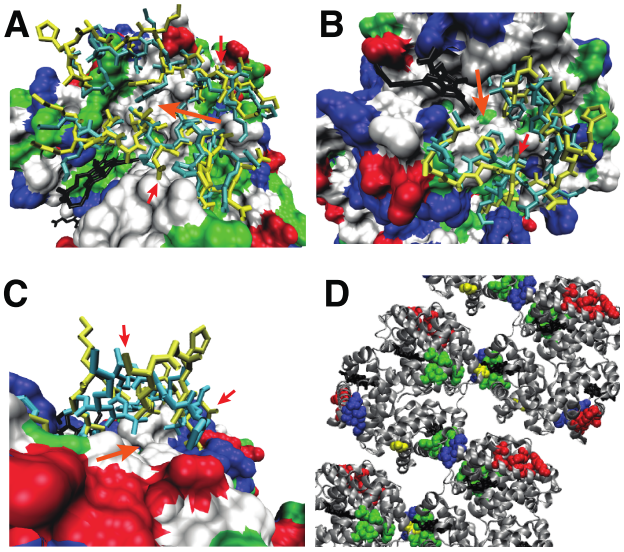

Figure 6 Pockets are identified with large, orange arrows: A) pocket 1 near αLys11, αAsn68, and αGlu116, B) pocket 2 near βGlu26 and βLeu88, and C) pocket 3 near αAsp47 and αGlu54. These residues (small, red arrows) are sites of mutation predicted to directly interfere with polymerization. Most of the protein is shown in surface representation with hydrophobic residues in white, polar residues in green, negatively charged residues in red, and positively charged residues in blue. The remaining protein is shown as cyan sticks (unbound crystal structure) and yellow sticks (simulation snapshot). D) The pockets are shown in the HbS crystal structure (2HBS)[58]. Pocket 2 (green) is located directly at a polymer interface while pocket 1 (red) and pocket 3 (blue) are located adjacent to polymer interfaces. Note that pocket 2 is on the β subunits while pockets 1 and 3 are on the α subunits.

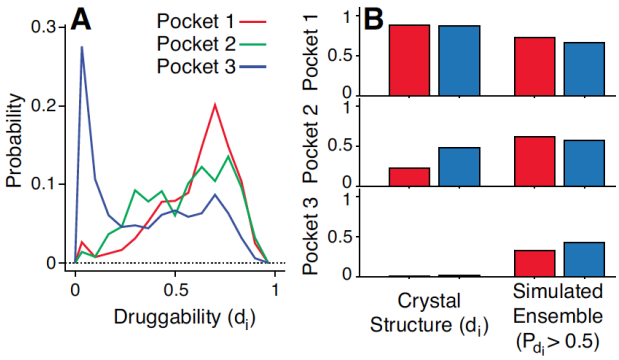

Figure 7 A residue specific druggability score (d_i) is calculated using the crystal structure and simulation snapshots. A) Each curve is the probability distribution of d_i calculated using snapshots from the oxygen bound simulation. At least one other residue in each pocket has a similar distribution B) The d_i calculated using the crystal structure is compared to the probability that a simulation snapshot will have a d_i greater than 0.5. Red bars indicate oxygen bound structures and blue bars indicate unbound structures.

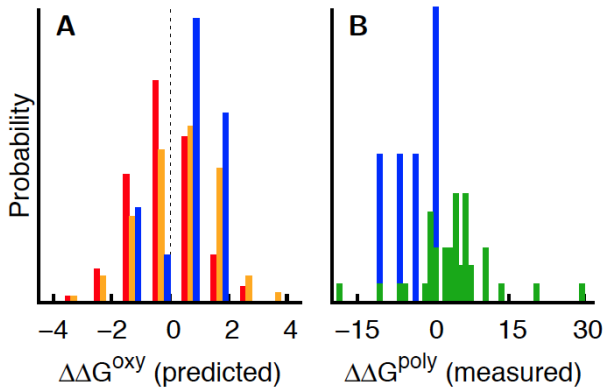

Figure 8 A) We predict the impact of all naturally occurring (non-engineered) hemoglobin mutations in the HbVar database[57] on oxygen binding ($\Delta\Delta G^{oxy}$ in k_BT): 319 HbA mutations in the α subunits (red), 423 HbA mutations in the β subunits (yellow), and 13 HbS mutations (blue). B) We report the measured impact of HbS mutations on polymerization ($\Delta\Delta G^{poly}$ in k_BT): 5 data points for naturally occurring mutations (blue) and 41 data points for engineered mutations (green).

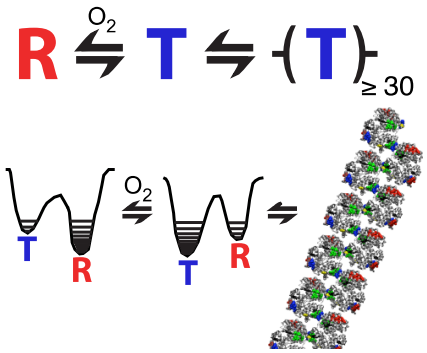

Table of contents figure