

Mapping Polymerization and Allostery of Hemoglobin S Using Point Mutations

Patrick Weinkam, and Andrej Sali

J. Phys. Chem. B, Just Accepted Manuscript • DOI: 10.1021/jp4025156 • Publication Date (Web): 19 Aug 2013

Downloaded from <http://pubs.acs.org> on August 26, 2013

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

High quality. High impact.

The Journal of Physical Chemistry B is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society.
However, no copyright claim is made to original U.S. Government works, or works
produced by employees of any Commonwealth realm Crown government in the course
of their duties.

1
2
3
4
5
6
7
8
9
10

11 **Mapping Polymerization and Allostery of Hemoglobin S Using Point
12 Mutations**

13

14 Patrick Weinkam^{a,*} and Andrej Sali^{a,b,*}

15

16

17

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

21

22

23 *Corresponding authors: 1700 4th Street, Byers Hall 503B, University of California, San Francisco,
24 San Francisco, CA 94158; tel 415-514-4227; web <http://salilab.org>; emails pweinkam@salilab.org and
25 sali@salilab.org.

26

27

28

29

30 **Keywords:** Energy landscape; funnel; Gō model; molecular dynamics; machine-learning

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

1 2 Abstract

3
4 Hemoglobin is a complex system that undergoes conformational changes in response to oxygen,
5 allosteric effectors, mutations, and environmental changes. Here, we study allostery and
6 polymerization of hemoglobin and its variants by application of two previously described methods: (i)
7 AllosMod for simulating allostery dynamics given two allosterically related input structures and (ii) a
8 machine-learning method for dynamics- and structure-based prediction of the mutation impact on
9 allostery (Weinkam et al. J. Mol. Biol. 2013), now applicable to systems with multiple coupled binding
10 sites such as hemoglobin. First, we predict the relative stabilities of substates and microstates of
11 hemoglobin, which are determined primarily by entropy within our model. Next, we predict the impact
12 of 866 annotated mutations on hemoglobin's oxygen binding equilibrium. We then discuss a subset of
13 30 mutations that occur in the presence of the sickle cell mutation and whose effects on
14 polymerization have been measured. Seven of these HbS mutations occur in three predicted
15 druggable binding pockets that might be exploited to directly inhibit polymerization; one of these
16 binding pockets is not apparent in the crystal structure but only in structures generated by AllosMod.
17 For the 30 mutations, we predict that mutation-induced conformational changes within a single
18 tetramer tend not to significantly impact polymerization; instead, these mutations more likely impact
19 polymerization by directly perturbing a polymerization interface. Finally, our analysis of allostery
20 allows us to hypothesize why hemoglobin evolved to have multiple subunits and a persistent low
21 frequency sickle cell mutation.

22 23 24 Introduction

25
26 For decades, hemoglobin has been a model system for studying proteins. The discovery that a
27 particular mutated hemoglobin (HbS) plays a role in sickle cell anemia was the first time that a
28 specific protein was linked to a genetic disease¹. Hemoglobin A (HbA) was used in the first
29 mechanistic descriptions of allostery^{2,3}, which led to the characterization of hundreds more allosteric
30 proteins⁴. Since these pioneering studies, there has been much progress regarding the allosteric
31 mechanism of Hb⁵⁻⁷ and the polymerization mechanism of HbS⁸⁻¹⁴. In fact, the two processes are
32 coupled because HbS polymers primarily consist of deoxygenated hemoglobin. By exploiting this
33 coupling, researchers discovered many ligands for hemoglobin that stabilize the oxygenated versus
34 deoxygenated state and in turn can reduce polymerization^{15,16}, although hematologists disagree on
35 whether or not decreasing oxygen dissociation is appropriate for treating sickle cell disease. In fact,
36 there are no still effective therapies for this disease (those that can improve patient health without
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

serious side effects). One of the problems is that a high concentration of hemoglobin in the blood presumably requires a high concentration of the modulator. While stem cell transplantation has yielded promising results¹⁷, it is currently prohibitively expensive for many patients. Instead, treatments focus on stimulating the production of fetal hemoglobin using hydroxyurea (which has sometimes fatal side effects)^{18,19}, providing supplemental oxygen, and treating symptoms such as pain. To contribute towards the discovery of HbS aggregation modulators, we predict which surface sites on HbS could be targeted to inhibit polymerization, by applying our previously developed allosteric model^{20,21} to many experimentally characterized hemoglobin variants.

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 hierarchical energy landscapes^{24,25}.

Our allosteric model is a dual basin structure-based Gō model²⁶⁻²⁹, 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 input crystal structures, including contact density patterns³⁰.

We can create separate landscapes to model conformational changes that occur as a result of perturbations, including ligands, mutations, and environmental changes. 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 has evolved to be influenced by pH³⁶, effector ligands, and polymerization of hemoglobin monomers. For such complex systems, we may gain insight by monitoring conformational changes resulting from different energy landscapes.

Phenomenological models of allosteric mechanisms for proteins in general 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, we analyze HbS allostery and polymerization in the context of point mutations, with the goal to facilitate prediction of small molecule binding sites that might be used to inhibit HbS polymerization (Figure 1). To identify such sites, we use experimental annotations of mutations to probe polymerization. A mutation at a given site can be a useful probe because it often has effects similar to drug binding, phosphorylation, and other post-translational modifications at the same site⁴¹; in all of these cases, including mutations, 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. In addition, we predict conformational substate stabilities, by using our AllosMod model of allostery dynamics²⁰, followed by relating our findings to phenomenological views of allostery. We then predict the impact of mutations on the oxygen binding affinity, by using our generalized machine-learning method²¹, which in turn allows us to speculate about the evolutionary advantage of the HbS mutation. Lastly, we explore the coupling between allostery and polymerization, which occurs because polymers are primarily composed of deoxygenated hemoglobin. A simplified model of the thermodynamic equilibrium between the R, T, and polymeric states allows us to relate the predicted impact of a mutation on allostery to its corresponding impact on polymerization. However, this simplified model turned out not to fit the data within the estimated experimental and computational

errors. Therefore, under the conditions of the experiments, mutations are more likely to perturb polymerization primarily *via* a direct impact on polymerization interfaces rather than on the R to T equilibrium.

Methods

Our approach is described in Figure 1. Briefly, we create energy landscapes for binding of oxygen and 2,3-diphosphoglycerate (DPG). We then sample the energy landscapes by constant temperature molecular dynamics simulations. The resulting structural ensembles are analyzed using machine-learning²¹ 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 “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 by experiment. We discuss the relationship between allostery and polymerization, by assuming a simplified equilibrium between oxyhemoglobin, deoxyhemoglobin, and the aggregate states. Finally, we 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 HbA ligand-induced dynamics *in vacuo* were performed using our web server at <http://salilab.org/allosmod>^{20,21}. These simulations were 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 were used to sample the landscapes. 30 simulations were 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 was 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. Appropriate velocity rescaling ensures energy conservation within the system and is necessary to avoid artifacts from energy dissipation. Artifacts rarely occur because AllosMod’s smooth energy landscape facilitates rapid equilibration. The simulations are likely to be

effectively longer than 6 nanoseconds of real time, primarily because of the absence of damping by the solvent and electrostatic interactions occurring in the real system⁴².

Assigning substates and microstates: QI_{diff}

Substates and microstates within the simulation structural ensembles are assigned based on QI_{diff} that is calculated over one or more residues. QI_{diff} is defined as $(Q^{e+} - Q^{e-}) / (1 - \Delta Q)$, where Q is the overall fold similarity⁴³ 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:

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

The summation is over all contacting pairs of sequentially non-adjacent residues; two residues are in contact when the average side chain atom positions are less than 11 Å apart. N is the number of contacts, r_{ij}^t is the distance between average side chain atom positions for residues i and j in structure t , and σ_{ij} is 2 Å.

$QI_{diff}(i)$ is a distance metric that describes the local environment of residue i ; it is positive if a residue configuration is closer to the effector bound structure than to the effector unbound structure and negative otherwise²⁰. 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 average side chain atom positions less than 11 Å from any oxygen atom in the crystal structure.

Machine-learning based on simulation trajectories: allosteric mutation effects

Prediction of a mutation effect is made by a machine-learning algorithm that is based on relating experimentally characterized mutations to structures from our allosteric model simulations²¹. Each mutation effect prediction reflects 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 Root⁴⁴, 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: $\Delta\Delta G^{\text{oxy}} = \Delta G^{\text{mut}} - \Delta G^{\text{wt}} = RT \log(K_d^{\text{wt}} / K_d^{\text{mut}})$, which directly measures the equilibrium shift between the oxy and deoxy conformational substates²¹. Note that 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, $\Delta\Delta G^{\text{DPG}} = -RT \log(K_d^{\text{wt}} / K_d^{\text{mut}})$ 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 ($\Delta\Delta G^{\text{oxy}}$)

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 ($\Delta\Delta G^{\text{oxy}}$ and $\Delta\Delta G^{\text{poly}}$) into a single prediction of a mutation’s effect on the oxygen binding equilibria ($\Delta\Delta G^{\text{oxy}}$) using a Boltzmann-like average based on the spatial density (SD):

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

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 ($\Delta\Delta G^{\text{oxy}}$ or $\Delta\Delta G^{\text{DPG}}$), 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:

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

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

where r_{ligand} is the distance between the mutated average side chain's atom position 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 that is parameterized to ensure a smooth transition of the spatial density at long distances to a value of 0.105, which is the maximum density at 30 Å calculated using 10 representative crystal structures other than hemoglobin. Residues 30 Å from the binding site are more likely to be affected by intra-protein contacts than ligand binding, imposing the range on our spatial density calculations.

Relating the impacts of a mutation on allostery ($\Delta\Delta G^{\text{oxy}}$) and polymerization ($\Delta\Delta G^{\text{poly}}$)

To facilitate a discussion of coupling between allostery and polymerization, we need a simple model of thermodynamic equilibrium involving all relevant hemoglobin states. The prediction of a mutation's impact on the oxygen binding equilibria ($\Delta\Delta G^{\text{oxy}}$) could be used to hypothesize its corresponding impact on polymerization, based on the following thermodynamic equilibrium at a given oxygen partial pressure and other conditions: $R \rightleftharpoons T \rightleftharpoons \{T\}_2 \rightleftharpoons \{T\}_3 \rightleftharpoons \dots \rightleftharpoons \{T\}_i$, where "R" corresponds to the oxy substrate (i.e., the oxy quaternary structure, with or without oxygen), "T" corresponds to the deoxy substrate (i.e., the deoxy quaternary structure), and T_i corresponds to an aggregate with i monomers; note that the stoichiometry is omitted for visual clarity. When concentrations of monomers and polymers are equal (i.e., at the polymerization critical concentration), we approximate the equilibrium as $R \rightleftharpoons T \rightleftharpoons \{T\}_n$, where n is the size of the effective aggregate. The effective aggregate represents, as an approximation, a minimal aggregate that forms at the critical concentration; in principle, the effective aggregate should not be stable below the critical concentration, and larger polymers should be stable above the critical concentration. The left side of the equation can describe a mutation's impact on the allosteric conformational equilibrium and right side of the equation can describe a mutation's impact on the stability of the effective aggregate, which represents experimental measurement of a shift in the critical concentration. The effective aggregate size can be as large as the polymerization nucleus; aggregates smaller than the nucleus are not stable. Given the simplified equilibrium, a mutation that impacts allostery also impacts polymerization as follows:

1 $\Delta\Delta G^{\text{poly}} = RT \log \left([T^{\text{wt}}]^n / [T^{\text{mut}}]^n \right)$. This equation allows us to relate allostery to polymerization by
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

expressing the concentration of unbound hemoglobin quaternary structure (T) as a function of the free energy difference between the oxy and deoxy substates for the wild type (ΔG^{oxy}) and the impact of mutation on the equilibrium between the oxy and deoxy substates ($\Delta\Delta G^{\text{oxy}}$). The hypothetical impact of allostery on polymerization is plotted for discrete values of ΔG^{oxy} and for a given value of n :

$$\Delta\Delta G^{\text{poly}} = nRT \log \left(\frac{(1 + \exp(-\Delta G^{\text{oxy}} / k_B T))}{(1 + \exp(-(ΔG^{\text{oxy}} + ΔΔG^{\text{oxy}}) / k_B T))} \right)$$

The estimates of the nucleus size vary from 10 to 100 and can change dramatically in different solvent conditions, in part due to changes of the polymerization mechanism^{45,46}. Hemoglobin aggregation is a multistep process that begins with homogeneous polymerization and is followed by heterogeneous polymerization, which involves new polymers nucleating from existing polymers^{5,47}. Varying solvent conditions and temperature can shift the ratio of the polymerization types⁴⁶.

The $\Delta\Delta G^{\text{poly}}$ equation can be used to create limiting models of the coupling between allostery and polymerization. These limiting models represent all ways the equation could possibly fit the data. Therefore, if the deviation between the data and limiting models is greater than experimental and computational error, there is no set of parameters that would allow the equation to fit the data. Limiting models are obtained by maximizing the range of $\Delta\Delta G^{\text{poly}}$ as a function of ΔG^{oxy} ; we do this by fixing n to an arbitrary value between 10 and a very large number, and selecting discrete values of ΔG^{oxy} . In other words, varying ΔG^{oxy} has a stronger effect on the $\Delta\Delta G^{\text{poly}}$ equation than varying n . The conclusions presented here are not affected by changing n so long as we also select discrete values of ΔG^{oxy} .

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 $\Delta\Delta G^{\text{poly}} = n\lambda RT \log(X^{\text{wt}} / X^{\text{mut}})$, where

X is one of the experimental data, λ is a correction factor, and n is the effective aggregate size. As discussed in the previous section, we can vary n to an arbitrary value greater than 10 without affecting the coarse conclusions presented here. X^{wt} and X^{mut} are measured in the same solvent conditions. The correction factor λ attempts 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 given value of λ , $\Delta\Delta G^{\text{poly}}$ for one experimental method correlates with the $\Delta\Delta G^{\text{poly}}$ for any other experimental method when comparing chemically similar mutations.

Prediction error

Experimental error is the difference between a measured value of quantity and its true value (Oxford Dictionary definition); similarly, computational error is the difference between a prediction and its true value. These errors must be considered to assess the significance of the experimental results and predictions. Next, we estimate the experimental and computational errors, demonstrating that the errors are small enough to justify the conclusions.

Experimental error in $\Delta\Delta G^{\text{oxy}}$ is approximated with the precision of a set of experimental measurements. To maximize the use of data and thus coverage, for any given $\Delta\Delta G^{\text{oxy}}$ estimate, this set includes measurements at different, uncertain, or even unknown pH values, temperatures, ionic strengths, and effector concentrations. Of these, temperature and pH can be the most impactful on hemoglobin oxygen binding and can contribute to experimental error when not precisely specified, due to systematic changes in hemoglobin structure and dynamics³⁶. They can each be estimated from the linear relationships between free energy and either pH or temperature⁴⁸, respectively. The average difference between experiments is about 0.3 pH units and 8 °C, which would yield errors in $\Delta\Delta G^{\text{oxy}}$ of 0.3 $k_B T$ and 0.4 $k_B T$, respectively. For the other variable conditions, we can estimate experimental error of $\Delta\Delta G^{\text{oxy}}$ using measurements of chemically similar mutations at the same site, which presumably reflect primarily the varying experimental conditions. We define a pair of mutants to be chemically similar if they are alike in terms of size (difference in the number of heavy atoms is two or less) and charge (positive, negative, or neutral); thus, chemically similar groups are (Asp, Glu), (Arg, His, Lys), (Tyr, Phe, Trp), and (Val, Ala, Ile, Leu, Met, Ser, Thr, Asn, Gln, Cys). The average unsigned difference between all pairs of $\Delta\Delta G^{\text{oxy}}$ for chemically similar mutations is 0.2 $k_B T$ for whole blood samples and 0.9 $k_B T$ for purified hemoglobin. The larger error for purified hemoglobin may

reflect the increased sensitivity of measurements to relevant environmental conditions. An example is the low oxygenation midpoint (P50) that scales similarly to the Hill coefficient (n), thus potentially resulting in larger uncertainty in measuring dissociation constants ($P50^n$) for purified HbA [$(5 \text{ tor})^{2.9}$] than for whole blood HbA [$(27 \text{ tor})^{2.7}$] (ref. ⁴⁹). Also, purified hemoglobin experiments are typically performed using phosphate buffer, which is known to affect hemoglobin dynamics differently than whole blood. Purified hemoglobin data is therefore not used in the analysis. Similarly, the average unsigned difference between all pairs of $\Delta\Delta G^{\text{poly}}$ for chemically similar mutations is $1.4 \text{ k}_\text{B}T$. In summary, the experimental error as defined here may be as much as $0.9 \text{ k}_\text{B}T$ for $\Delta\Delta G^{\text{oxy}}$ and $1.4 \text{ k}_\text{B}T$ for $\Delta\Delta G^{\text{poly}}$.

Computational error is estimated as the average difference between $\Delta\Delta G^{\text{oxy}}$ predictions and experimentally measured values, which was $1.3 \text{ k}_\text{B}T$ in our previous study of allosteric proteins in general²¹ and is $0.8 \text{ k}_\text{B}T$ in the current study. In comparison, the experimentally measured range of $\Delta\Delta G$'s for a mutation's impact on hemoglobin oxygen binding is $6 \text{ k}_\text{B}T$ (see Results)²¹. Also, the difference between $\Delta\Delta G^{\text{oxy}}$ and the hypothetical curves describing the relationship between allostery and polymerization is greater than $1 \text{ k}_\text{B}T$ for $\Delta\Delta G^{\text{oxy}}$ and greater than $5 \text{ k}_\text{B}T$ for $\Delta\Delta G^{\text{poly}}$. Thus, experimental measurements and computational predictions may be sufficiently accurate to be useful.

We calculate the likelihood of accurately predicting mutation effects using an error score empirically derived from our previous study²¹. 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 \AA 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 $1 \text{ k}_\text{B}T$ from the correct value. Mutations with scores of greater than 1.3 are omitted from analysis to avoid large outliers.

48 Predicting druggable binding pockets

We predict druggable pockets by applying the program FPocket⁵⁰ to snapshots from the oxygen bound and unbound AllosMod simulations (600 each). The FPocket druggability score was obtained by machine-learning optimization against a dataset of holo and apo crystal structures. Here, we create a residue specific score, d_i , which is the druggability of the most druggable pocket that has a vertex, which FPocket uses to identify pockets, within r^{cutoff} of the residue's average side chain atom

position. 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-based 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 populations of microstates calculated from the oxygen bound (red) and unbound (blue) simulations are 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. Because experiments indicate the relative stability between the oxy and deoxy substate populations should change upon oxygen binding, the simulations with r^{AS} less than 12 Å are omitted from further analysis.

The model predicts that the oxy substate is more stable than the deoxy substate. With r^{AS} equal to 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

1 comparison, low r^{AS} results in similarly populated substates, which indicates that the unequal
2 populations at high r^{AS} are not an artifact of substate assignment. The unequal substate populations
3 at high r^{AS} conflict with the equivalent relative stabilization energy in landscapes with the same r^{AS} , in
4 which the oxy crystal structure should be favored in the oxygen bound landscape by the same
5 amount as the deoxy crystal structure in the unbound landscape. In this case, entropy drives the
6 stability of the oxy substate because there are more ways to satisfy oxygen bound conformations
7 than unbound conformations. This stability difference is consistent with previous molecular dynamics
8 studies^{51,52} and predictions from a Gaussian network model that the carbon monoxide bound
9 structure (similar to that with oxygen bound) is entropically more stable than the unbound structure⁵³.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 the unbound (bound) microstate to the single-oxygen (triple-oxygen) bound microstate, the α subunits are more likely than the β subunits to be oxygenated (deoxygenated). 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 subunits⁵⁴. The simulations may provide an explanation for this observation. We find 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 study⁵¹. 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: $\Delta\Delta G^{\text{oxy}}$

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 relatively susceptible to perturbations, such as mutations.

We predict impact of mutations on the binding of oxygen or DPG. As described previously²¹, we define a mutation effect as the free energy change of ligand binding due to mutation ($\Delta\Delta G^{\text{oxy}}$ and $\Delta\Delta G^{\text{poly}}$ in Figure 1). Each mutation effect prediction requires features that either describe the mutation itself or are calculated from the simulations of the allosteric 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 mutation effects far from the DPG binding site (greater than 20 Å) are well predicted using the oxygen bound simulations and the remaining effects are well predicted using the DPG bound simulations (average unsigned error of 0.70 $k_B T$ and 0.96 $k_B T$, 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 ($\Delta\Delta G^{\text{oxy}}$ and $\Delta\Delta G^{\text{poly}}$) into a single prediction of a mutation's impact on the oxygen binding equilibria ($\Delta\Delta G^{\text{oxy}}$). In support of this approach, we previously analyzed mutation effects on binding sites from 10 different proteins²¹. While significant mutation effects on ligand binding (> 2 $k_B T$) generally occur at sites within 8 Å of the ligand, we observed here that 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 $k_B T$ and a 0.76 Pearson correlation coefficient with experiment (Figure 4C).

Impact of mutations on allosteric and polymerization

Allostery is coupled to polymerization because polymers consist of hemoglobin in the unbound quaternary structure. We can approximate the equilibria between oxyhemoglobin, deoxyhemoglobin, and aggregates using a simplified equation (Methods). This approach effectively provides limiting hypotheses about how the impact of a mutation on allostery could be coupled to its corresponding impact on polymerization.

1
2
3 However, there is uncertainty regarding the propensity of conformational fluctuations that occur in the
4 absence of oxygen (*i.e.*, the conditions of some polymerization experiments). Studies of gel
5 encapsulated hemoglobin in deoxygenated conditions demonstrate carbon monoxide rebinding
6 kinetics that suggest multiple conformations (not a single T state structure), including at least one that
7 binds oxygen with a higher affinity than the T state⁵⁵. In fact, this state (T-high) is thought to be on
8 pathway between the R and T structures⁵⁵. This data is consistent with other results that suggest a
9 transiently stable R-like conformation in deoxy conditions: 1) the oxy substate is predicted to be more
10 stable than the deoxy substate in other computational studies⁵¹⁻⁵³ in addition to the work presented
11 here and 2) the existence of a deoxyhemoglobin crystal structure in the oxygen bound conformation
12 with a well-structured but empty oxygen binding pocket⁵⁶. In contrast, a two state model (T and T
13 aggregate) does fit polymerization data, suggesting that the allosteric transition is unnecessary to
14 interpret the data⁵⁷. For polymerization experiments in deoxy conditions, the conformational changes
15 induced by mutation within a single tetramer are not well characterized. For example, if adding a
16 small oxygen molecule can shift the equilibrium between the R and T structures sufficiently to impact
17 polymerization, so could in principle a single point mutation.
18
19

20
21 We therefore use the simplified equilibrium outlined in Methods to hypothesize about the coupling
22 between mutation impacts on allostery and polymerization (hypothetical curves in Figure 5). The
23 curves are plotted at different relative stabilities between the oxy and deoxy substates, representing
24 varying oxygen pressure or concentrations of allosteric effectors, for an arbitrary value of n = 30
25 (Methods). We also plot a set of experimentally measured HbS mutations⁵⁸, using the measured
26 polymerization mutation effects ($\Delta\Delta G^{\text{poly}}$) and the predicted allostery mutation effects ($\Delta\Delta G^{\text{oxy}}$). Here,
27 we assume that all HbS mutations are uncoupled to β -Glu6Val. The difference between the data and
28 the hypothetical curves is larger than the computational and experimental errors. For example, many
29 mutations are predicted to decrease hemoglobin oxygen affinity, while the corresponding experiments
30 indicate the mutations inhibit polymerization. Thus, our simplified equilibrium cannot fit the data for
31 most mutations, suggesting that (i) most of the mutations disrupt a polymerization interface directly
32 without an impact on oxygen binding and/or (ii) the population of the oxy substate of the mutant is
33 small because most of the experiments are performed in deoxygenated conditions. The first
34 suggestion is supported by the location of many mutations on the protein surface near interfaces
35 between tetramers in the crystal structure. The second suggestion is supported by the dominance of
36 the deoxy substate of the native hemoglobin under the conditions of experiment. However, a clear
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 picture requires additional thermodynamic and structural data for the mutants under a broader set of
2 experimental conditions.
3
4

5
6
7 *Druggable binding pockets that can be used to affect polymerization*
8
9

10 We can identify sites of mutations that perturb polymerization and then predict potentially druggable
11 binding pockets near these sites, many of which are close to a protein-protein interface in the HbS
12 crystal⁵⁹. We predict druggable pockets by applying the program FPocket⁵⁰ to the structures from our
13 allosteric model simulations. Due to structural changes, pockets grow and disappear during the
14 simulation trajectories, thereby making pocket identification non-trivial. We therefore create a residue
15 specific score d_i , which is the druggability of the most druggable pocket within a cutoff distance of
16 residue i (Methods). We can predict druggable pockets using d_i distributions calculated from
17 simulation trajectories.
18
19

20 We identify three druggable pockets near mutations that perturb polymerization. These pockets are
21 also near tetramer-tetramer interfaces in the HbS crystal structure (Figure 6). Pocket 3 has a bimodal
22 d_i distribution indicating distinct conformations sampled in the simulation (Figure 7A). In comparison,
23 pockets 1 and 2 have unimodal distributions and are more likely to form druggable pockets. Pockets 1
24 and 2, however, may not be better targets than pocket 3, which is (i) more druggable in
25 deoxyhemoglobin than oxyhemoglobin, (ii) adjacent to sites of mutations predicted to decrease
26 oxygenation, and (iii) adjacent to sites of mutations predicted to significantly decrease polymerization.
27 Therefore, a ligand may bind specifically to pocket 3 in deoxyhemoglobin and inhibit a polymerization
28 interface.
29
30

31 Pocket prediction using simulation trajectories can be advantageous in the case of dynamic proteins
32 such as hemoglobin. For instance, ligand binding pockets can form transiently even if no pockets
33 exist in the crystal structure⁶⁰. Even in the absence of significant structural change, ligand binding
34 prediction based on structural ensembles can be more accurate than predictions based on a single
35 structure^{61,62}. In hemoglobin, pockets populate slightly expanded conformations in the simulation
36 trajectories compared to the crystal structure (Figure 6). These expanded conformations could be
37 favored upon ligand binding and in turn perturb an interface in the HbS polymer. In the simulation
38 trajectories, pocket 3 occurs in such an expanded conformation and is predicted to be much more
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 druggable than the corresponding location in the crystal structure (Figure 7B). In contrast, pockets 1
2 and 2 maintain druggability in the simulation trajectories and the crystal structures.
3
4
5

6 Discussion

7

8
9
10 The original descriptions of allosteric regulation in general as well as hemoglobin in particular were
11 phenomenological and rooted in experimental data. The dominant states of hemoglobin, favored with
12 and without oxygen (R and T, respectively), were determined by X-ray crystallography. The KNF
13 mechanism, shown to be inadequate for hemoglobin, hypothesized that oxygen binding induces a
14 concerted change from T to R³. The MWC mechanism hypothesized that there is a structural
15 equilibrium between R and T, and that oxygen binding promotes the R state, effectively allowing
16 oxygen binding in both the oxygen bound and unbound structures. Subsequent more intricate models
17 accounted for structural details such as salt bridges that contribute to pH dependence of hemoglobin
18 oxygen binding (known as the Bohr effect)⁶³⁻⁶⁵.
19
20
21
22
23
24
25
26
27

28 More recent descriptions of allosteric regulation rely not only on the crystal structures but also broader
29 conformational ensembles³⁷⁻⁴⁰. In fact, several structural states of hemoglobin other than R and T
30 have been identified⁶⁶. While the MWC and KNF mechanisms do not explicitly describe structural
31 ensembles, they are generally consistent with the population shift and induced fit mechanisms,
32 respectively. A clear divergence from the purely structural view of allosteric regulation is entropy driven
33 allosteric regulation⁶⁷. This mechanism describes dynamic coupling between sites in which the average structure
34 remains unchanged, but includes different excursions from the average structure in the bound and
35 unbound states.
36
37
38
39
40
41
42
43

44 Phenomenological mechanisms explicitly relate structural details to experimental observables. For
45 instance, the MWC and related mechanisms^{2,64,65} explain measured subunit oxygen binding affinities
46 using the quaternary structure. Another mechanism, called TTS, decouples subunit tertiary structure
47 from the quaternary structure⁵, thereby allowing subunits to adopt the R or T conformations without a
48 quaternary structure change. The TTS mechanism is consistent with our approach of assigning one
49 of two substates to the binding site structure, which in our model moderately correlates with the
50 subunit tertiary structure (correlation coefficient of 0.5 if using QI_{diff}).
51
52
53
54
55
56
57
58
59
60

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

10
11
12
13
14
15
16
17
18
19 Our simulations describe weak coupling between hemoglobin's oxygen binding sites. Such behavior
20 has also been reported in molecular dynamics simulations^{51,52,68-70}, elastic network models^{53,71}, and
21 an experimental study of gel encapsulated hemoglobin that separately mapped tertiary and
22 quaternary structure changes⁷². Weak coupling involves oxygen binding at one site triggering both
23 tertiary and quaternary structural changes, which in turn result in changes of the size and shape of
24 the other oxygen binding pockets. Binding site dynamics is not homogeneous, however, as indicated
25 by the stronger coupling of the quaternary structure with the oxygen binding sites in the β subunits
26 than in the α subunits (Figure 3). This result is consistent with experiments that report a larger impact
27 on oxygen binding for mutations in the β subunits than for mutations in the α subunits (average
28 magnitude of $1.5 \pm 0.6 \text{ k}_\text{B}T$ and $0.5 \pm 0.5 \text{ k}_\text{B}T$, respectively). Therefore, modulation of hemoglobin's
29 allosteric transition may well be achieved by perturbing the interface between the two β subunits.
30 Interestingly, such a perturbation has resulted from evolution that positioned the DPG binding pocket
31 at the β subunit interface.

32
33
34
35
36
37
38
39
40
41
42
43
44 Hemoglobin has evolved to have a complex allosteric mechanism, yet also permits point mutations at
45 many sites. Our predictions suggest that naturally occurring mutations⁵⁸ can be tolerated due to
46 hemoglobin's structural symmetry (Figure 8). We predict α subunit mutations to typically inhibit
47 oxygen binding, which can counteract β subunit mutations that we predict to typically promote oxygen
48 binding (Figure 8A). Natural HbS mutations (those that occur in patients with β -Glu6Val)⁵⁸ display an
49 even stronger predicted trend of increased oxygen binding than the β subunit mutations (Figure 8A).
50 These HbS mutations may improve hemoglobin's oxygen delivery if they shift the equilibrium to the
51 oxy state, thus sufficiently reducing polymerization, but do not increase oxygen affinity too much so
52 that sufficient oxygen is still released.

One might expect the sickle cell mutation to be selected out of the human population, but the HbS allele persists, possibly at least in part because it allows malaria resistance⁷³. 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 naturally occurring HbS mutations on polymerization. These 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^{74,75}. 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 and inter-dependently. A mutation may impact (i) oxygen binding, (ii) binding of DPG or other effectors, (iii) the allosteric conformational equilibrium between oxy and deoxy substates, (iv) pH or temperature induced conformational changes, and (v) polymerization. Here, we use a model that describes the allosteric conformational transition, yet without an atomic structural modeling of interactions between hemoglobin and any of its ligands. Nevertheless, the relative accuracy of our predictions (Figure 4C) supports the model; moreover, most naturally occurring mutations are unlikely to affect oxygen binding without affecting the allosteric transition, because they are located far from the oxygen binding sites (98 % greater than 5 Å and 92 % greater than 8 Å). A more detailed model is, however, 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. Thus, our model of allostery may be a convenient stepping stone to these more explicit models.

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 temperature⁴⁸. 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

1 used to inhibit HbS aggregation by destabilizing a polymerization interface. These sites might be
2 more effective than sites that are distant to polymerization interfaces because ligand binding effects
3 tend to dissipate at long distances. In conclusion, mutations can serve as natural probes of function
4 and may help identify ligand binding pockets that can be used to perturb allosteric proteins such as
5 hemoglobin.
6
7
8

12 Acknowledgments

13

16 We are grateful for helpful discussions with Charles Homcy, Matt Jacobson, William Eaton, Natalia
17 Khuri, Seung Joong Kim, and Riccardo Pellarin. We also thank the reviewers for many critical
18 comments that improved the manuscript during the journal review process. The work was supported
19 by an NIH grant R01 GM083960.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

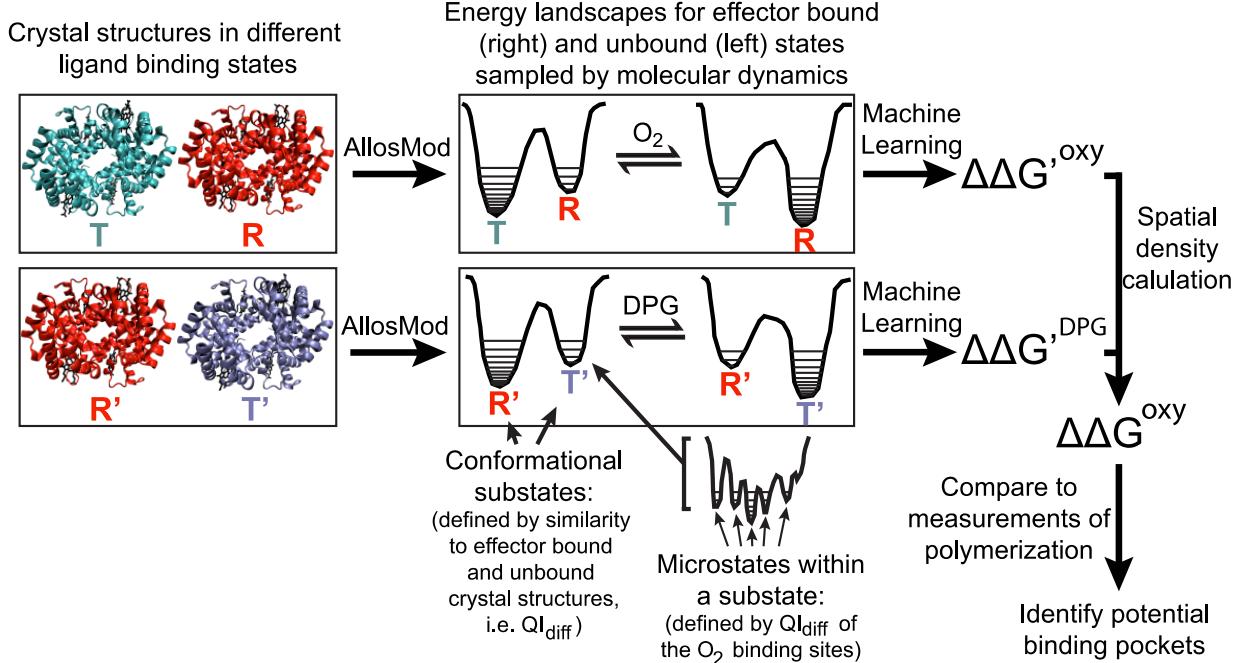
1. Pauling, L.; Itano, H. A., Sickle Cell Anemia a Molecular Disease. *Science* **1949**, *110*, 543-548.
2. Monod, J.; Wyman, J.; Changeux, J. P., On Nature of Allosteric Transitions - a Plausible Model. *J Mol Biol* **1965**, *12*, 88-118.
3. Koshland, D. E.; Nemethy, G.; Filmer, D., Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5*, 365-368.
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., ASD: A Comprehensive Database of Allosteric Proteins and Modulators. *Nucleic Acids Res* **2011**, *39*, D663-D669.
5. Henry, E. R.; Bettati, S.; Hofrichter, J.; Eaton, W. A., A Tertiary Two-State Allosteric Model for Hemoglobin. *Biophys Chem* **2002**, *98*, 149-164.
6. Knapp, J. E.; Pahl, R.; Srajer, V.; Royer, W. E., Allosteric action in real time: Time-resolved Crystallographic Studies of a Cooperative Dimeric Hemoglobin. *Proc Natl Acad Sci U S A* **2006**, *103*, 7649-7654.
7. Cui, Q.; Karplus, M., Allostery and Cooperativity Revisited. *Protein Sci* **2008**, *17*, 1295-307.
8. Benesch, R. E.; Yung, S.; Benesch, R.; Mack, J.; Schneider, R. G., Alpha-Chain Contacts in Polymerization of Sickle Hemoglobin. *Nature* **1976**, *260*, 219-221.
9. Benesch, R. E.; Kwong, S.; Benesch, R.; Edalji, R., Location and Bond Type of Intermolecular Contacts in Polymerization of Hemoglobin-S. *Nature* **1977**, *269*, 772-775.
10. Benesch, R. E.; Kwong, S.; Edalji, R.; Benesch, R., Alpha-Chain Mutations with Opposite Effects on the Gelation of Hemoglobin-S. *J Biol Chem* **1979**, *254*, 8169-8172.
11. Nagel, R. L.; Johnson, J.; Bookchin, R. M.; Garel, M. C.; Rosa, J.; Schiliro, G.; Wajcman, H.; Labie, D.; Moopenn, W.; Castro, O., Beta-Chain Contact Sites in the Haemoglobin-S Polymer. *Nature* **1980**, *283*, 832-834.
12. Monplaisir, N.; Merault, G.; Poyart, C.; Rhoda, M. D.; Craescu, C.; Vidaud, M.; Galacteros, F.; Blouquit, Y.; Rosa, J., Hemoglobin-S Antilles - a Variant with Lower Solubility Than Hemoglobin-S and Producing Sickle-Cell Disease in Heterozygotes. *Proc Natl Acad Sci U S A* **1986**, *83*, 9363-9367.
13. Dellano, J. J. M.; Manning, J. M., Properties of a Recombinant Human Hemoglobin Double Mutant - Sickle Hemoglobin with Leu-88(Beta) at the Primary Aggregation Site Substituted by Ala. *Protein Sci* **1994**, *3*, 1206-1212.
14. Li, X. F.; Himanen, J. P.; de Llano, J. J. M.; Padovan, J. C.; Chait, B. T.; Manning, J. M., Mutational Analysis of Sickle Haemoglobin (Hb) Gelation. *Biotechnology and Applied Biochemistry* **1999**, *29*, 165-184.
15. Fylaktakidou, K. C.; Duarte, C. D.; Koumbis, A. E.; Nicolau, C.; Lehn, J. M., Polyphosphates and Pyrophosphates of Hexopyranoses as Allosteric Effectors of Human Hemoglobin: Synthesis, Molecular Recognition, and Effect on Oxygen Release. *Chemmedchem* **2011**, *6*, 153-168.
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., Crystallographic Analysis of Human Hemoglobin Elucidates the Structural Basis of the Potent and Dual Antisickling Activity of Pyridyl derivatives of vanillin (vol D67, pg 920, 2011). *Acta Crystallogr D* **2011**, *67*, 1076-1076.
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., Infusion of Hemolyzed Red Blood Cells within Peripheral Blood Stem Cell Grafts in Patients with and without Sickle Cell Disease. *Blood* **2012**, *119*, 5671-5673.
18. Platt, O. S.; Orkin, S. H.; Dover, G.; Beardsley, G. P.; Miller, B.; Nathan, D. G., Hydroxyurea Enhances Fetal Hemoglobin Production in Sickle Cell Anemia. *J Clin Invest* **1984**, *74*, 652-656.

- 1 19. Sankaran, V. G.; Menne, T. F.; Xu, J.; Akie, T. E.; Lettre, G.; Van Handel, B.; Mikkola, H. K.;
2 Hirschhorn, J. N.; Cantor, A. B.; Orkin, S. H., Human Fetal Hemoglobin Expression is Regulated by
3 the Developmental Stage-Specific Repressor BCL11A. *Science* **2008**, *322*, 1839-1842.
- 4 20. Weinkam, P.; Pons, J.; Sali, A., Structure-based Model of Allostery Predicts Coupling Between
5 Distant Sites. *Proc Natl Acad Sci U S A* **2012**, *109*, 4875-4880.
- 6 21. Weinkam, P.; Chen, Y. C.; Pons, J.; Sali, A., Impact of Mutations on the Allosteric
7 Conformational Equilibrium. *J Mol Biol* **2013**, *425*, 647-661.
- 8 22. Bryngelson, J. D.; Wolynes, P. G., Intermediates and Barrier Crossing in a Random Energy-
9 Model (with applications to protein folding). *J Phys Chem* **1989**, *93*, 6902-6915.
- 10 23. Dill, K. A., Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133-7155.
- 11 24. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G., The Energy Landscapes and Motions of
12 Proteins. *Science* **1991**, *254*, 1598-1603.
- 13 25. Zhuravlev, P. I.; Materese, C. K.; Papoian, G. A., Deconstructing the Native State: Energy
14 Landscapes, Function, and Dynamics of Globular Proteins. *J Phys Chem B* **2009**, *113*, 8800-8812.
- 15 26. Ueda, Y.; Taketomi, H.; Go, N., Studies on Protein Folding, Unfolding, and Fluctuations by
16 Computer-Simulation .2. 3-Dimensional Lattice Model of Lysozyme. *Biopolymers* **1978**, *17*, 1531-
17 1548.
- 18 27. Schug, A.; Whitford, P. C.; Levy, Y.; Onuchic, J. N., Mutations as Trapdoors to Two Competing
19 Native Conformations of the Rop-Dimer. *Proc Natl Acad Sci U S A* **2007**, *104*, 17674-17679.
- 20 28. Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N., An All-
21 Atom Structure-based Potential for Proteins: Bridging Minimal Models with All-atom Empirical
22 Forcefields. *Proteins: Struct, Funct, Bioinf.* **2009**, *75*, 430-441.
- 23 29. Li, W.; Wolynes, P. G.; Takada, S., Frustration, Specific Sequence Dependence, and
24 Nonlinearity in Large-amplitude Fluctuations of Allosteric Proteins. *Proc Natl Acad Sci U S A* **2011**,
25 *108*, 3504-3509.
- 26 30. Weinkam, P.; Zong, C. H.; Wolynes, P. G., A Funneled Energy Landscape for Cytochrome c
27 Directly Predicts the Sequential Folding Route Inferred from Hydrogen Exchange Experiments. *Proc
28 Natl Acad Sci U S A* **2005**, *102*, 12401-12406.
- 29 31. Weinkam, P.; Romesberg, F. E.; Wolynes, P. G., Chemical Frustration in the Protein Folding
30 Landscape: Grand Canonical Ensemble Simulations of Cytochrome c. *Biochemistry* **2009**, *48*, 2394-
31 2402.
- 32 32. Weinkam, P.; Zimmermann, J.; Romesberg, F. E.; Wolynes, P. G., The Folding Energy
33 Landscape and Free Energy Excitations of Cytochrome c. *Acc Chem Res* **2010**, *43*, 652-60.
- 34 33. Yifrach, O.; Horovitz, A., Nested Cooperativity in the Atpase Activity of the Oligomeric
35 Chaperonin GroEL. *Biochemistry* **1995**, *34*, 5303-5308.
- 36 34. Krishna, M. M.; Maity, H.; Rumbley, J. N.; Englander, S. W., Branching in the Sequential
37 Folding Pathway of Cytochrome c. *Protein Sci* **2007**, *16*, 1946-56.
- 38 35. Cho, S. S.; Weinkam, P.; Wolynes, P. G., Origins of Barriers and Barrierless Folding in BBL.
39 *Proc Natl Acad Sci U S A* **2008**, *105*, 118-23.
- 40 36. Benesch, R. E.; Benesch, R.; Yu, C. I., Oxygenation of Hemoglobin in Presence of 2,3-
41 Diphosphoglycerate . Effect of Temperature pH Ionic Strength and Hemoglobin Concentration.
42 *Biochemistry* **1969**, *8*, 2567-2571.
- 43 37. Kumar, S.; Ma, B. Y.; Tsai, C. J.; Sinha, N.; Nussinov, R., Folding and Binding Cascades:
44 Dynamic Landscapes and Population Shifts. *Protein Sci* **2000**, *9*, 10-19.
- 45 38. Motlagh, H. N.; Hilser, V. J., Agonism/Antagonism Switching in Allosteric Ensembles. *Proc Natl
46 Acad Sci U S A* **2012**, *109*, 4134-4139.
- 47 39. Bosshard, H. R., Molecular Recognition by Induced Fit: How Fit is the Concept? *News in
48 Physiological Sciences* **2001**, *16*, 171-173.

- 1 40. James, L. C.; Tawfik, D. S., Conformational Diversity and Protein Evolution - a 60-Year-Old
2 Hypothesis Revisited. *Trends in Biochemical Sciences* **2003**, 28, 361-368.
3 41. Salazar, C.; Hofer, T., Allosteric Regulation of the Transcription Factor NFAT1 by Multiple
4 Phosphorylation Sites: A Mathematical Analysis. *J Mol Biol* **2003**, 327, 31-45.
5 42. Hinsen, K.; Kneller, G. R., Solvent Effects in the Slow Dynamics of Proteins. *Proteins: Struct.,*
6 *Funct., Bioinf.* **2008**, 70, 1235-1242.
7 43. Cho, S. S.; Levy, Y.; Wolynes, P. G., P Versus Q: Structural Reaction Coordinates Capture
8 Protein Folding on Smooth Landscapes. *Proc. Natl. Acad. Sci. USA* **2006**, 103, 586-591.
9 44. Hoecker, A.; Speckmayer, P.; Stelzer, J.; Therhaag, J.; von Toerne, E.; Voss, H., TMVA:
10 Toolkit for Multivariate Data Analysis. *Pos* **2007**, ACAT, 040.
11 45. Behe, M. J.; Englander, S. W., Sickle Hemoglobin Gelation - Reaction Order and Critical
12 Nucleus Size. *Biophys J* **1978**, 23, 129-145.
13 46. Christoph, G. W.; Hofrichter, J.; Eaton, W. A., Understanding the Shape of Sickled Red Cells.
14 *Biophys J* **2005**, 88, 1371-1376.
15 47. Ferrone, F. A.; Rotter, M. A., Crowding and the Polymerization of Sickle Hemoglobin. *Journal of Molecular Recognition* **2004**, 17, 497-504.
16 48. Samaja, M.; Melotti, D.; Rovida, E.; Rossibernardi, L., Effect of Temperature on the P50 Value
17 for Human-Blood. *Clinical Chemistry* **1983**, 29, 110-114.
18 49. Wajcman, H.; Galacteros, F., Hemoglobins with High Oxygen Affinity Leading to
19 Erythrocytosis. New Variants and New Concepts. *Hemoglobin* **2005**, 29, 91-106.
20 50. Le Guilloux, V.; Schmidtke, P.; Tuffery, P., Fpocket: An Open Source Platform for Ligand
21 Pocket Detection. *Bmc Bioinformatics* **2009**, 10, e1.
22 51. Hub, J. S.; Kubitzki, M. B.; de Groot, B. L., Spontaneous Quaternary and Tertiary T-R
23 Transitions of Human Hemoglobin in Molecular Dynamics Simulation. *PLoS Comput Biol* **2010**, 6,
24 e1000774.
25 52. Yusuff, O. K.; Babalola, J. O.; Bussi, G.; Raugei, S., Role of the Subunit Interactions in the
26 Conformational Transitions in Adult Human Hemoglobin: An Explicit Solvent Molecular Dynamics
27 Study. *J Phys Chem B* **2012**, 116, 11004-11009.
28 53. Xu, C. Y.; Tobi, D.; Bahar, I., Allosteric Changes in Protein Structure Computed by a Simple
29 Mechanical Model: Hemoglobin T <-> R2 Transition. *J Mol Biol* **2003**, 333, 153-168.
30 54. Mozzarelli, A.; Rivetti, C.; Rossi, G. L.; Eaton, W. A.; Henry, E. R., Allosteric Effectors do not
31 Alter the Oxygen Affinity of Hemoglobin Crystals. *Protein Sci* **1997**, 6, 484-489.
32 55. Samuni, U.; Roche, C. J.; Dantsker, D.; Juszczak, L. J.; Friedman, J. M., Modulation of
33 Reactivity and Conformation within the T-Quaternary State of Human Hemoglobin: the Combined use
34 of Mutagenesis and Sol-Gel Encapsulation. *Biochemistry* **2006**, 45, 2820-2835.
35 56. Wilson, J.; Phillips, K.; Luisi, B., The Crystal Structure of Horse Deoxyhaemoglobin Trapped in
36 the High-Affinity (R) State. *J Mol Biol* **1996**, 264, 743-756.
37 57. Sunshine, H. R.; Hofrichter, J.; Ferrone, F. A.; Eaton, W. A., Oxygen Binding by Sickle-Cell
38 Hemoglobin Polymers. *J Mol Biol* **1982**, 158, 251-273.
39 58. Giardine, B.; Borg, J.; Higgs, D. R.; Peterson, K. R.; Philipsen, S.; Maglott, D.; Singleton, B. K.;
40 Anstee, D. J.; Basak, A. N.; Clark, B.; et.al., Systematic Documentation and Analysis of Human
41 Genetic Variation in Hemoglobinopathies using the Microattribution Approach. *Nature Genetics* **2011**,
42 43, 295-302.
43 59. Harrington, D. J.; Adachi, K.; Royer, W. E., The High Resolution Crystal Structure of
44 Deoxyhemoglobin S. *J Mol Biol* **1997**, 272, 398-407.
45 60. Bowman, G. R.; Geissler, P. L., Equilibrium Fluctuations of a Single Folded Protein Reveal a
46 Multitude of Potential Cryptic Allosteric Sites. *Proc Natl Acad Sci U S A* **2012**, 109, 11681-11686.

- 1 61. Fan, H.; Irwin, J. J.; Webb, B. M.; Klebe, G.; Shoichet, B. K.; Sali, A., Molecular Docking
2 Screens Using Comparative Models of Proteins. *Journal of Chemical Information and Modeling* **2009**,
3 49, 2512-2527.
- 4 62. Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X., MDpocket: Open-Source Cavity
5 Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011**, 27, 3276-
6 3285.
- 7 63. Perutz, M. F., Stereochemistry of Cooperative Effects in Haemoglobin. *Nature* **1970**, 228, 726-
8 734.
- 9 64. Szabo, A.; Karplus, M., Mathematical-Model for Structure-Function Relations in Hemoglobin. *J
10 Mol Biol* **1972**, 72, 163-197.
- 11 65. Lee, A. W. M.; Karplus, M., Structure-Specific Model of Hemoglobin Cooperativity. *Proc Natl
12 Acad Sci U S A* **1983**, 80, 7055-7059.
- 13 66. Dey, S.; Chakrabarti, P.; Janin, J., A Survey of Hemoglobin Quaternary Structures. *Proteins*
14 **2011**, 79, 2861-2870.
- 15 67. Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G., Dynamically Driven Protein Allostery.
16 *Nat Struct Mol Biol* **2006**, 13, 831-838.
- 17 68. Gelin, B. R.; Lee, A. W. M.; Karplus, M., Hemoglobin Tertiary Structural-Change on Ligand-
18 Binding - Its Role in the Co-Operative Mechanism. *J Mol Biol* **1983**, 171, 489-559.
- 19 69. Mouawad, L.; Perahia, D., Motions in Hemoglobin Studied by Normal Mode Analysis and
20 Energy Minimization: Evidence for the Existence of Tertiary T-like, Quaternary R-like Intermediate
21 Structures. *J Mol Biol* **1996**, 258, 393-410.
- 22 70. Fischer, S.; Olsen, K. W.; Nam, K.; Karplus, M., Unsuspected Pathway of the Allosteric
23 Transition in Hemoglobin. *Proc Natl Acad Sci U S A* **2011**, 108, 5608-5613.
- 24 71. Eom, K.; Baek, S. C.; Ahn, J. H.; Na, S., Coarse-Graining of Protein Structures for the Normal
25 Mode Studies. *J Comput Chem* **2007**, 28, 1400-1410.
- 26 72. Jones, E. M.; Balakrishnan, G.; Spiro, T. G., Heme Reactivity is Uncoupled from Quaternary
27 Structure in Gel-Encapsulated Hemoglobin: A Resonance Raman Spectroscopic Study. *J Am Chem
28 Soc* **2012**, 134, 3461-3471.
- 29 73. Piel, F. B.; Patil, A. P.; Howes, R. E.; Nyangiri, O. A.; Gething, P. W.; Williams, T. N.;
30 Weatherall, D. J.; Hay, S. I., Global Distribution of the Sickle Cell Gene and Geographical
31 Confirmation of the Malaria Hypothesis. *Nature Communications* **2010**, 1, 104-110.
- 32 74. Luzzatto, L.; Nwachuku, E. S.; Reddy, S., Increased Sickling of Parasitised Erythrocytes as
33 Mechanism of Resistance against Malaria in Sickle-Cell Trait. *Lancet* **1970**, 1, 319-321.
- 34 75. Friedman, M. J., Erythrocytic Mechanism of Sickle-Cell Resistance to Malaria. *Proc Natl Acad
35 Sci U S A* **1978**, 75, 1994-1997.

1
2
3
Figures
4
5
6
7
8



33 **Figure 1** An overview of the current work. Each arrow represents a different Methods section.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

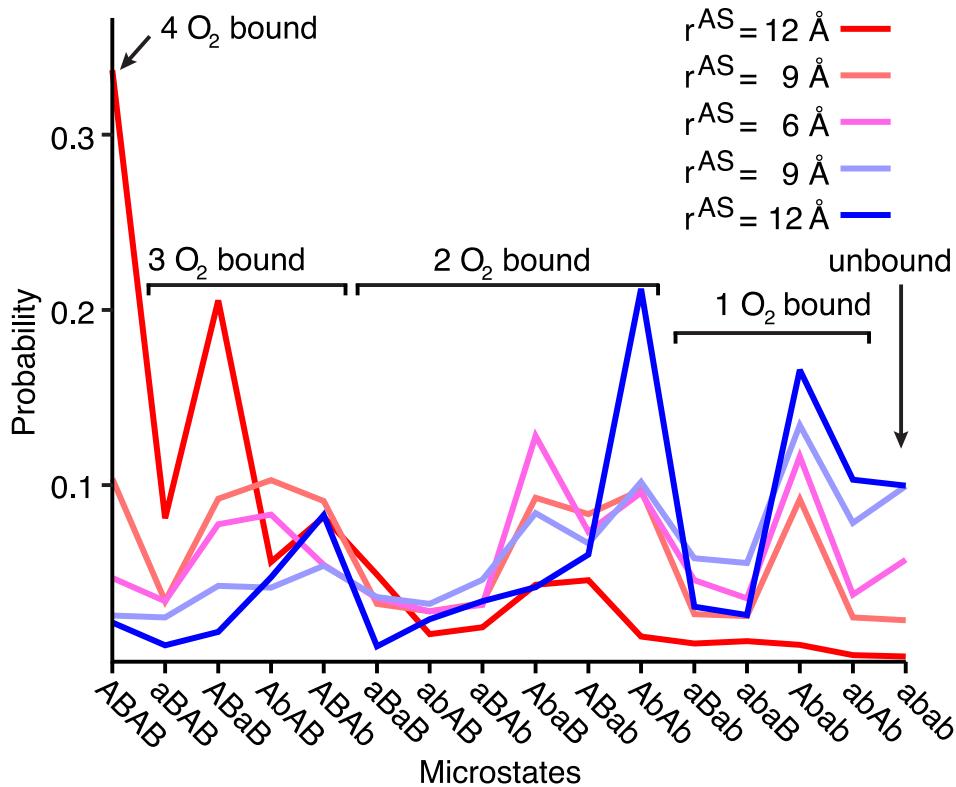


Figure 2 Probability distributions of hemoglobin's 16 microstates. Each line represents the probability distribution calculated from sampling of different energy landscapes, which have different allosteric site radii (r^{AS}) and ligand binding states (oxygen bound are shades of red and unbound are shades 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 ($\text{QI}_{\text{diff}} > 0$) or unbound ($\text{QI}_{\text{diff}} < 0$), respectively. The 5 left most and 5 right most microstates indicate the oxy and deoxy conformational substates, respectively.

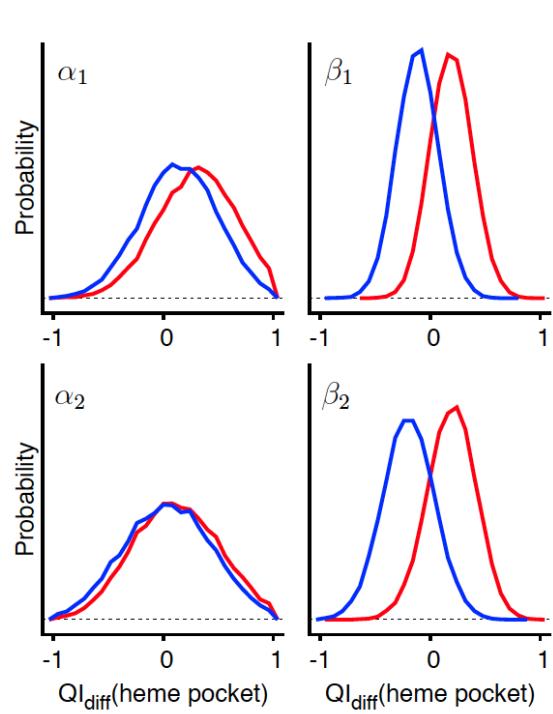


Figure 3 Probability distributions of QI_{diff} for the 4 oxygen binding sites calculated from the oxygen bound (red) and unbound (blue) simulations ($r^{AS} = 12 \text{ \AA}$). QI_{diff} is 1 if the oxygen binding site is more similar to the oxygen bound crystal structure than the unbound crystal structure 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.

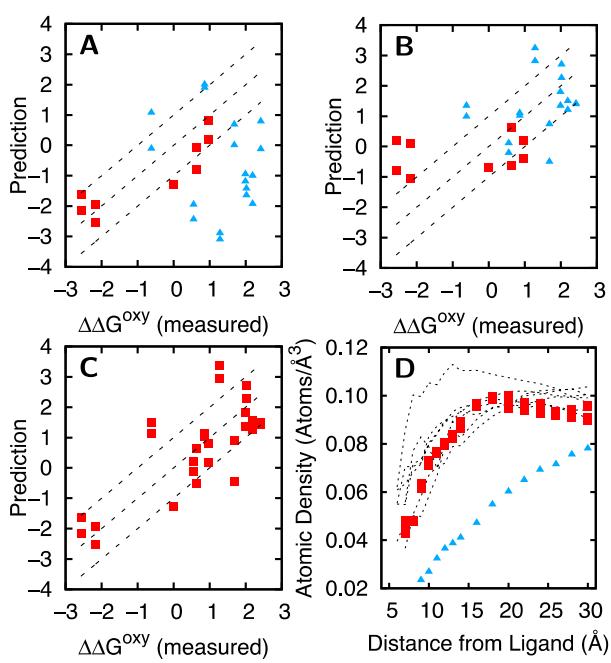


Figure 4 We predict the impact of mutations (in units of $k_B T$) on: A) oxygen binding ($\Delta\Delta G^{\text{oxy}}$) and B) DPG binding ($\Delta\Delta G^{\text{DPG}}$). Mutations further than 20 Å from the DPG binding site (red squares) are well predicted using the simulations with oxygen binding 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 $\Delta\Delta G^{\text{oxy}}$ 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 sites²¹.

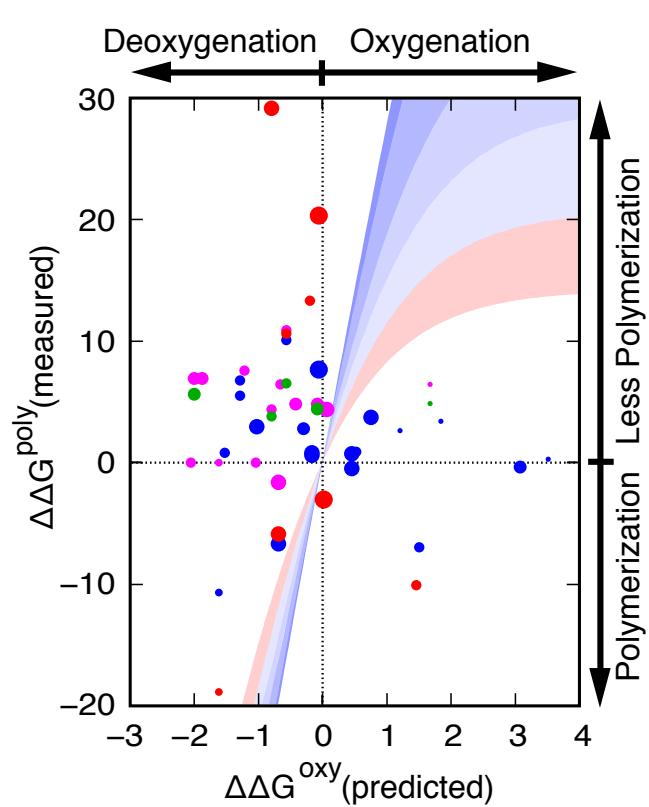


Figure 5 We characterize hemoglobin polymerization by comparing a mutation's impact on allostery and polymerization. We plot a mutation's measured impact on polymerization ($\Delta\Delta G^{\text{poly}}$) and the corresponding predicted impact on allostery ($\Delta\Delta G^{\text{oxy}}$) as well as hypothetical curves that approximate how allostery could be coupled to polymerization (in units of $k_B T$). 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.

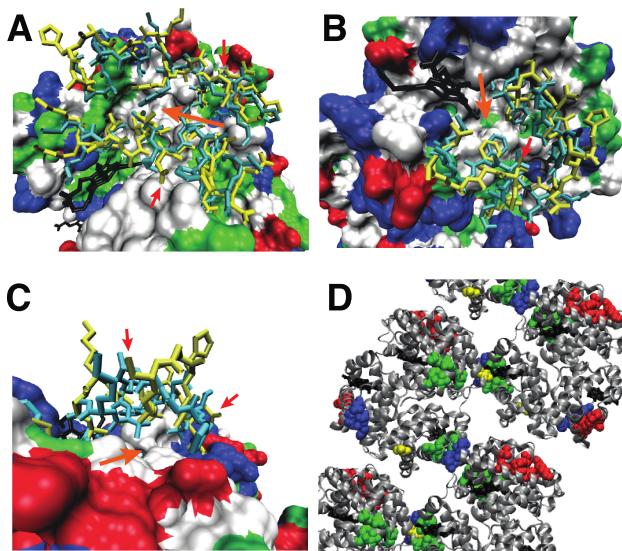


Figure 6 Pockets are identified with large, orange arrows: A) pocket 1 near α Lys11, α Asn68, and α Glu116, B) pocket 2 near β Glu26 and β Leu88, and C) pocket 3 near α Asp47 and α Glu54. These residues (small, red arrows) are sites of mutation predicted to directly interfere with polymerization. Most 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)⁵⁹. 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.

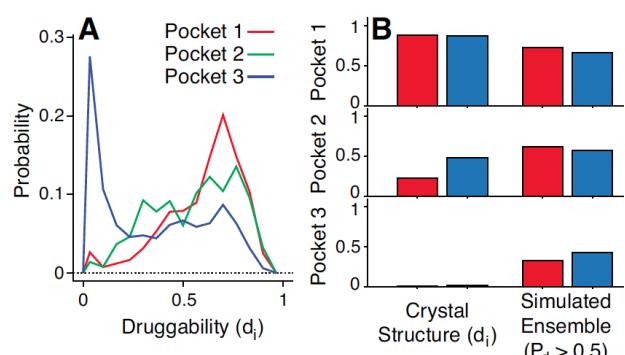


Figure 7 A residue specific druggability score (d_i) is calculated using the crystal structure and simulation snapshots. A) Each curve is the probability distribution of d_i calculated using snapshots from the oxygen bound simulation. At least one other residue in each pocket has a similar distribution 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.

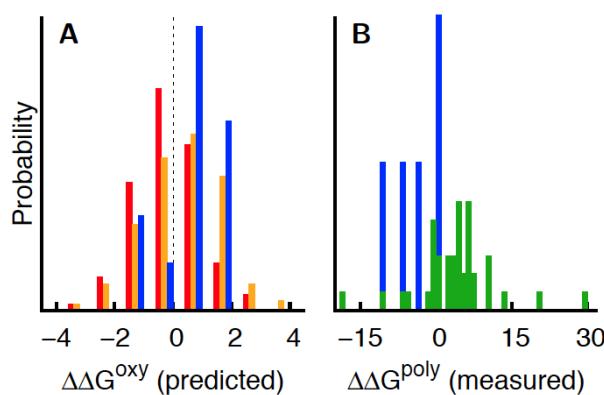
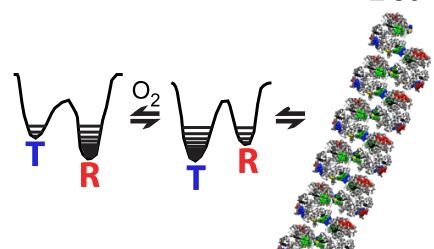
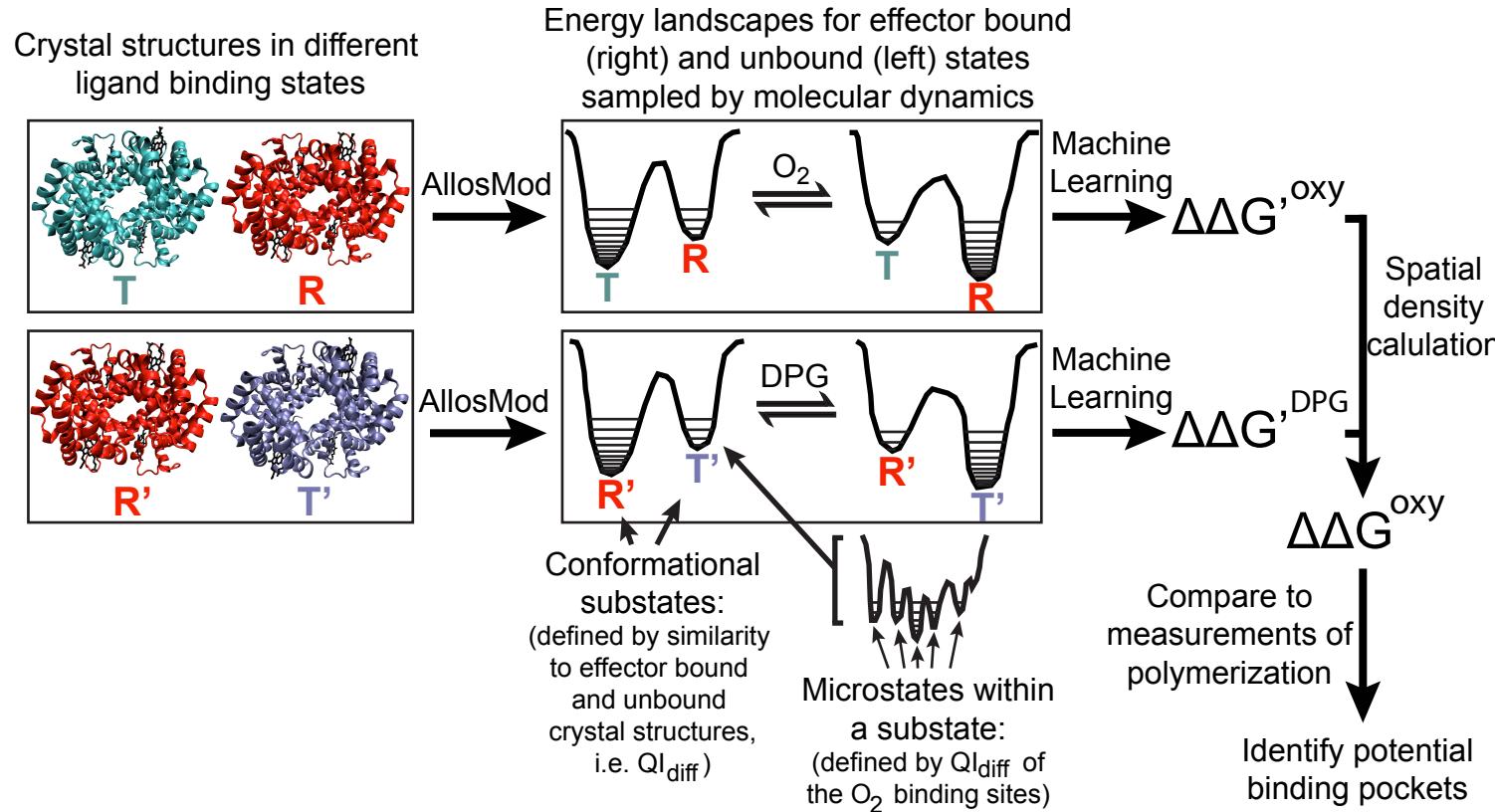


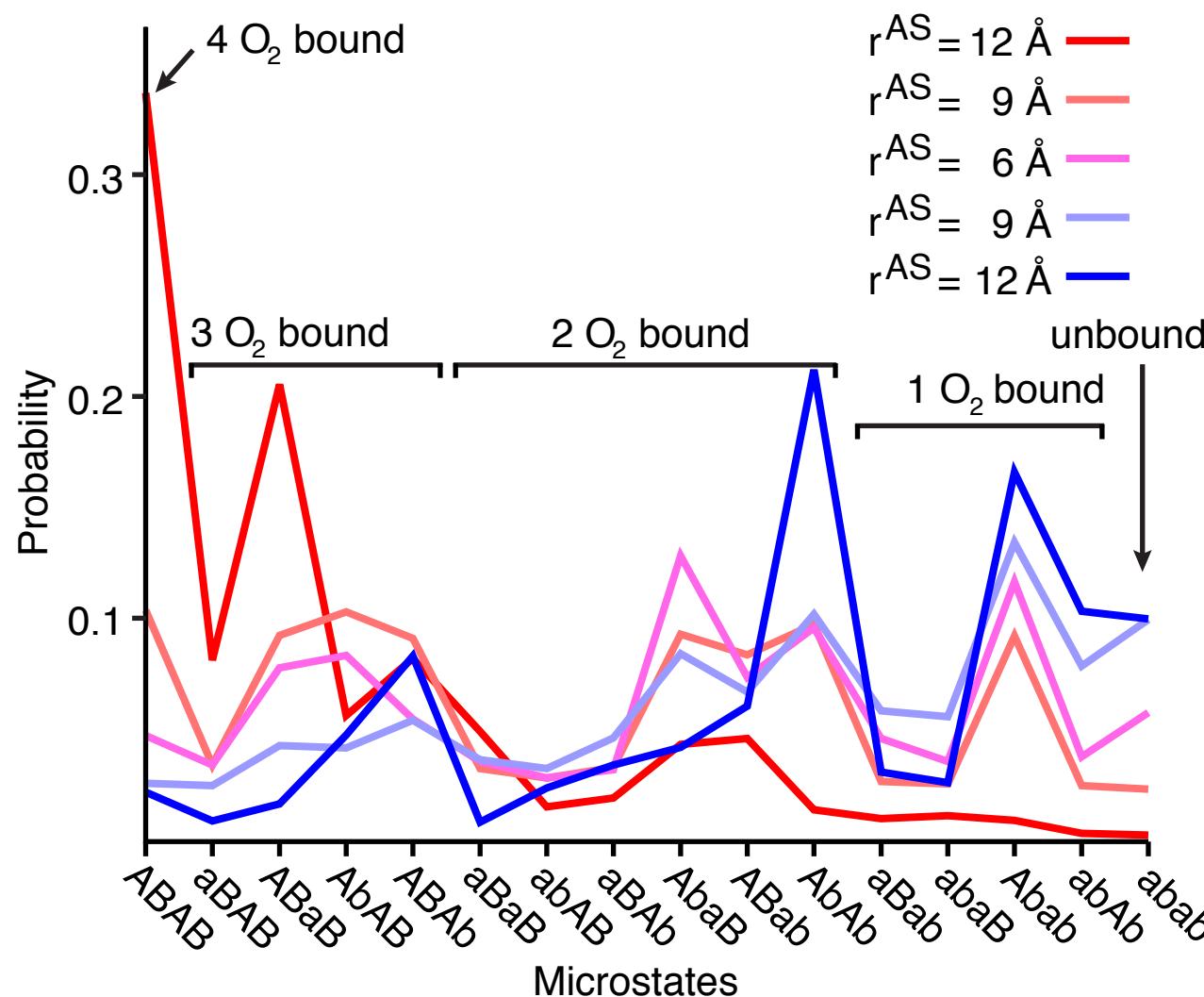
Figure 8 A) We predict the impact of all naturally occurring (non-engineered) hemoglobin mutations in the HbVar database⁵⁸ on oxygen binding ($\Delta\Delta G^{\text{oxy}}$ in $k_B T$): 319 HbA mutations in the α subunits (red), 423 HbA mutations in the β subunits (yellow), and 13 HbS mutations (blue). B) We report the measured impact of HbS mutations on polymerization ($\Delta\Delta G^{\text{poly}}$ in $k_B T$): 5 data points for naturally occurring mutations (blue) and 41 data points for engineered mutations (green).



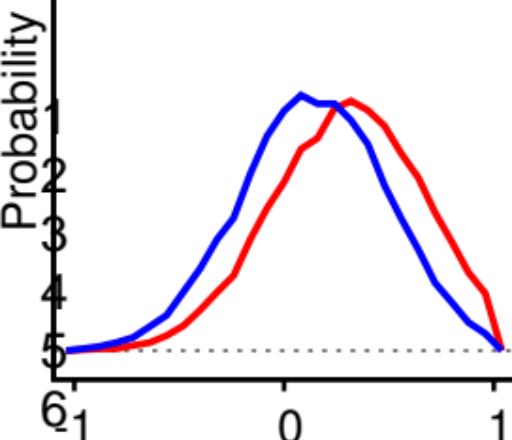
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table of contents figure

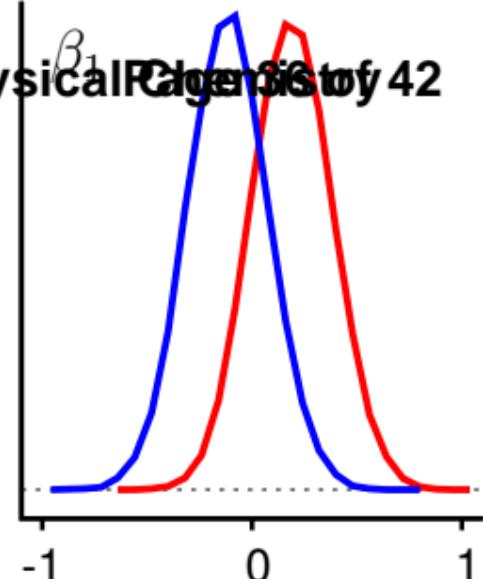




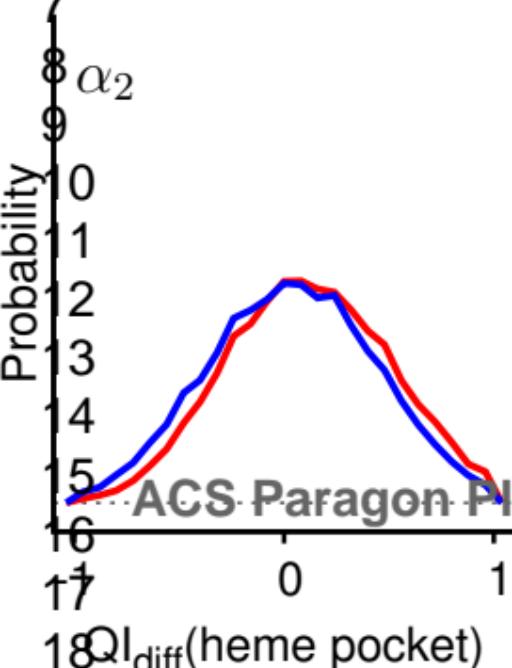
α_1



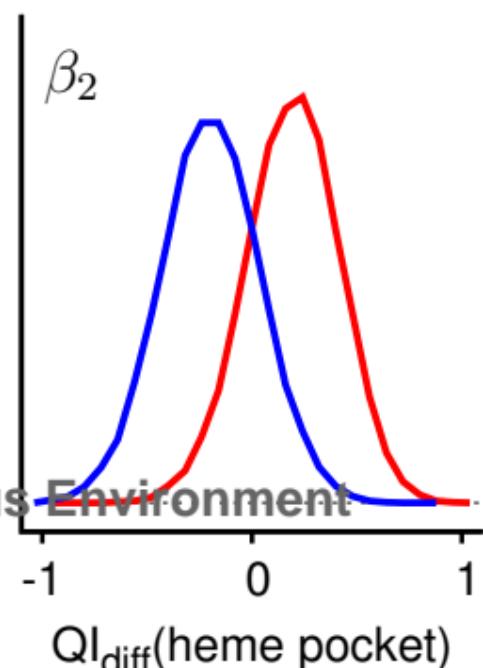
β_1



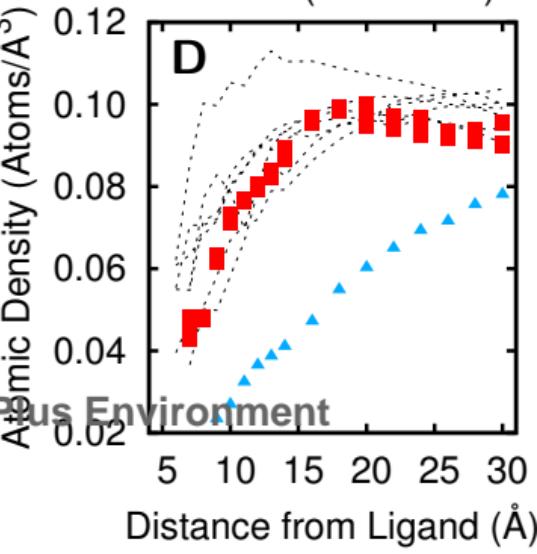
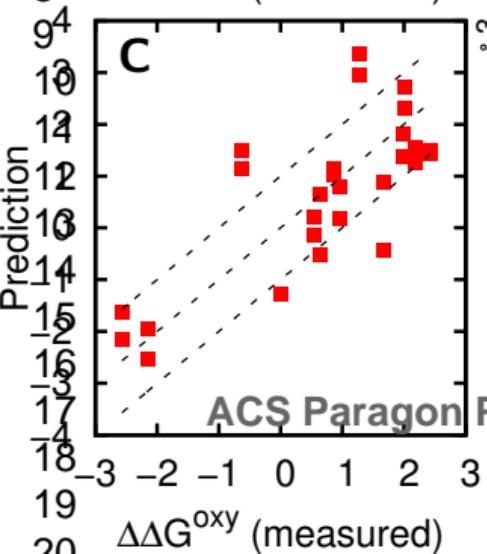
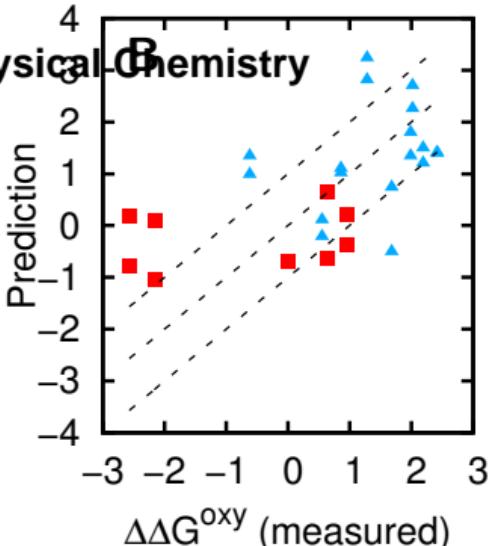
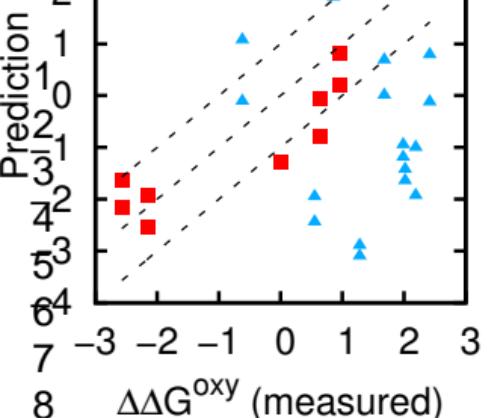
α_2

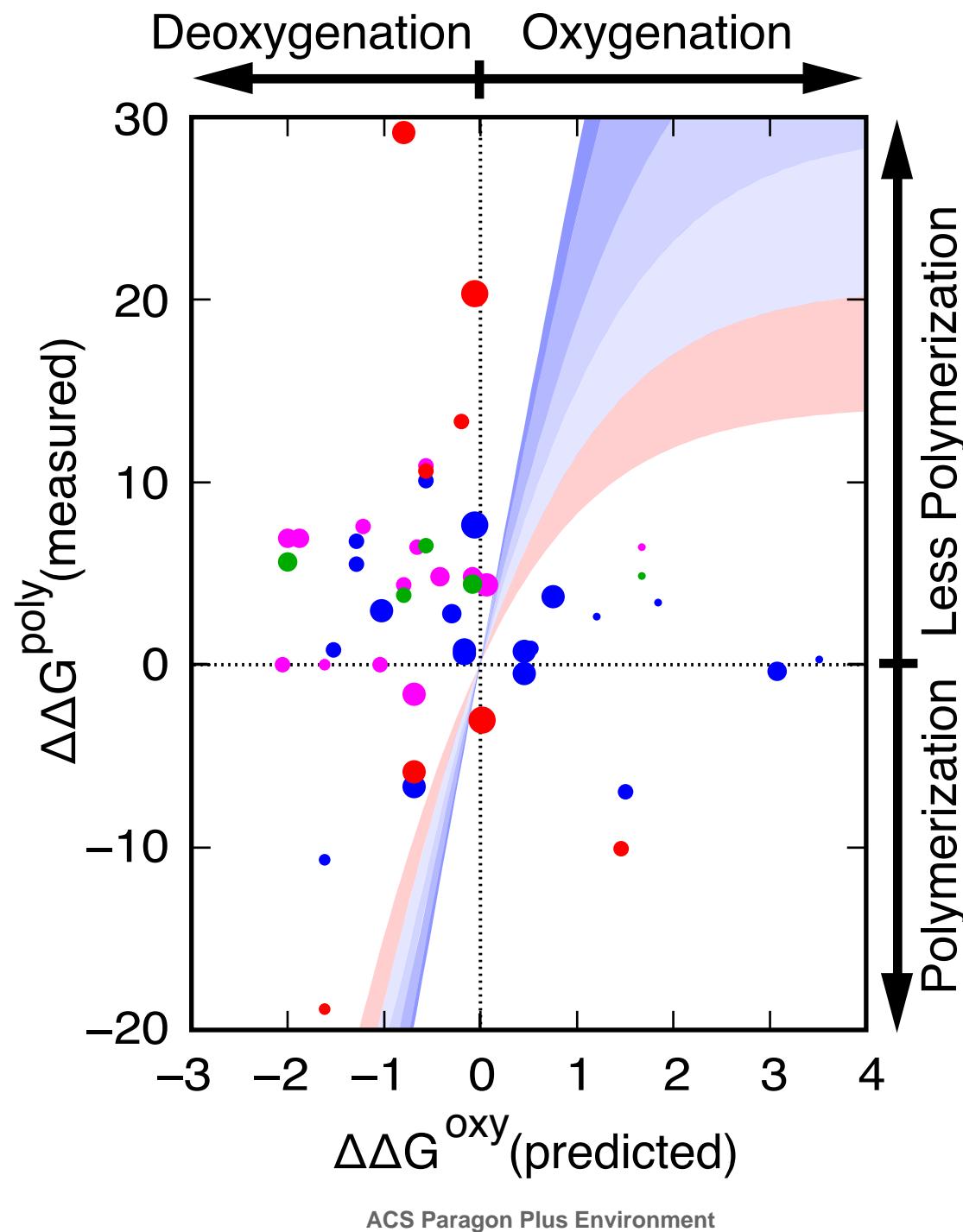


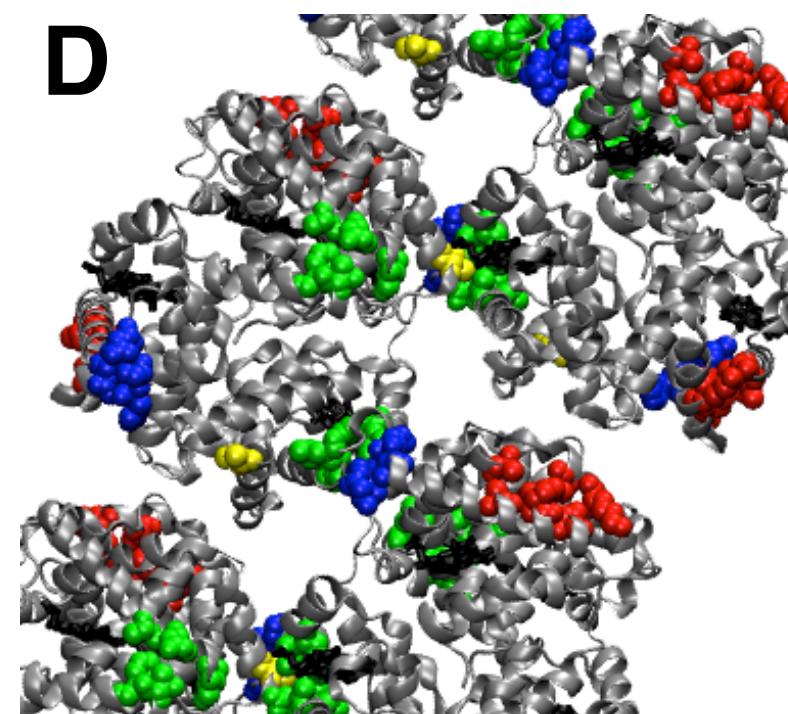
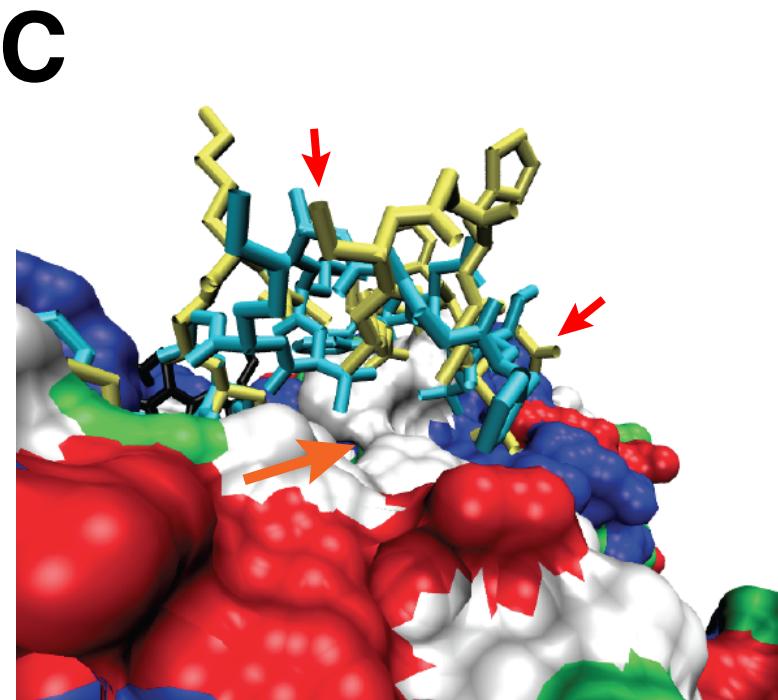
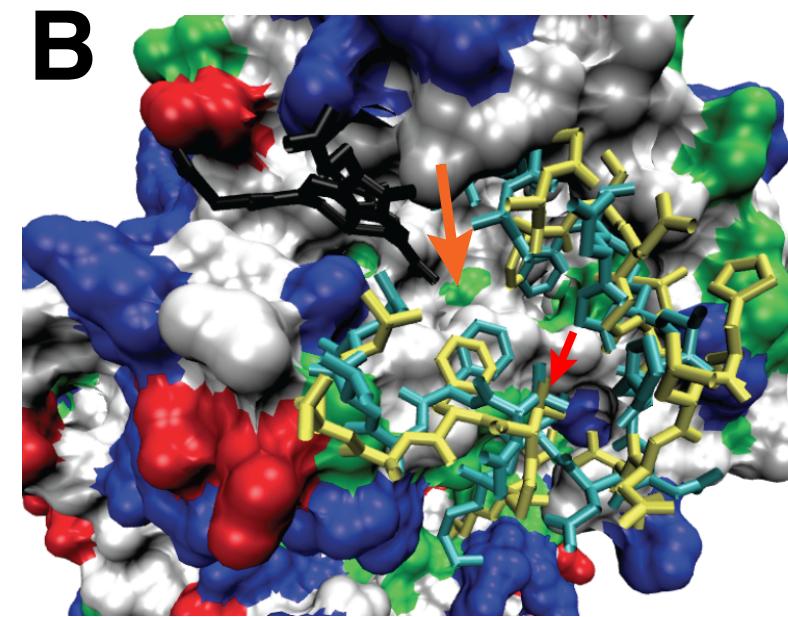
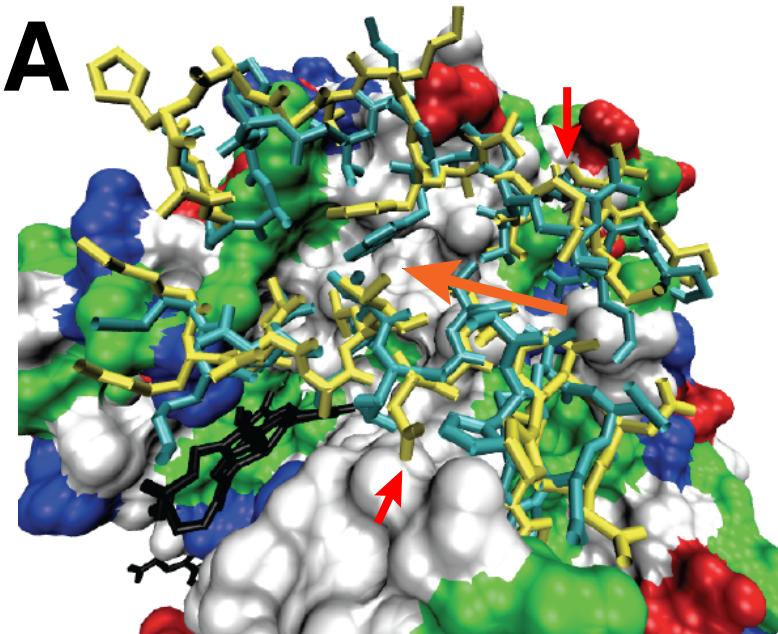
β_2

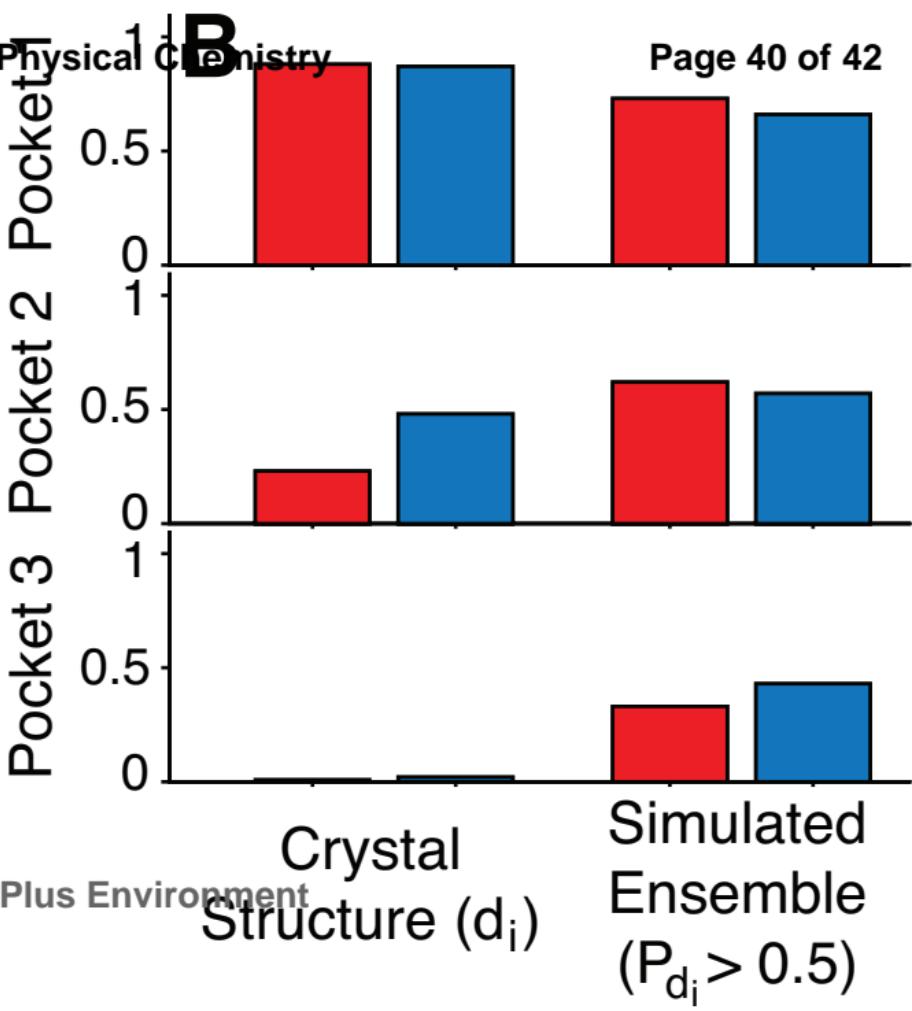
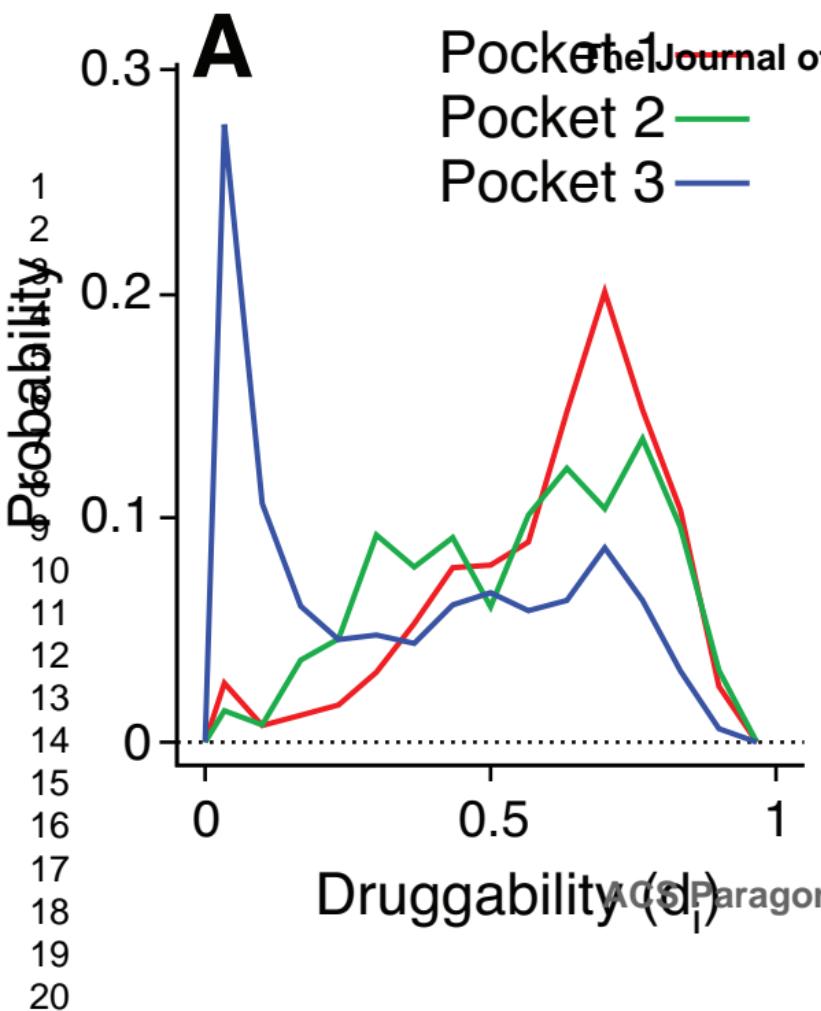


ACS Paragon Plus Environment



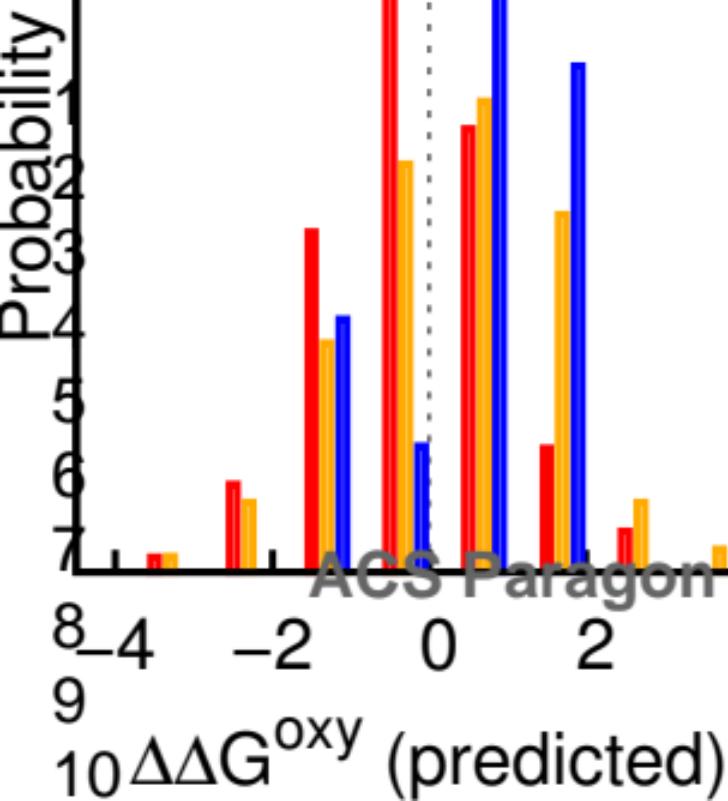




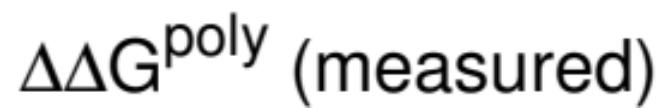


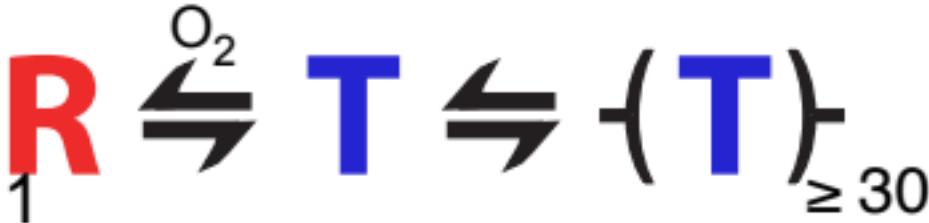
A

Page 41 of 42 Journal of Physical Chemistry



B





2

3

1

1

1

7

7

8

Q

