Received: 20 December 2021 | Revised: 2 April 2022 | Accepted: 20 April 2022

DOI: 10.1002/prot.26352

RESEARCH ARTICLE

PROTEINS WILEY

Multiscale MD simulations of wild-type and sickle hemoglobin aggregation

Maryam O. Olagunju[1] | Jennifer Loschwitz[1,2] | Olujide O. Olubiyi[1,3,4] | Birgit Strodel[1,2]

[1]Institute of Biological Information Processing, Structural Biochemistry, Forschungszentrum Jülich, Jülich, Germany

[2]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[3]Department of Pharmaceutical and Medicinal Chemistry, College of Pharmacy, Afe Babalola University, Ado-Ekiti, Nigeria

[4]Institute of Drug Research and Development, Bogoro Centre, Afe Babalola University, Ado-Ekiti, Nigeria

Correspondence
Birgit Strodel, Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany.
Email: b.strodel@fz-juelich.de

Abstract

Sickle cell disease is a hemoglobinopathy resulting from a point mutation from glutamate to valine at position six of the β-globin chains of hemoglobin. This mutation gives rise to pathological aggregation of the sickle hemoglobin and, as a result, impaired oxygen binding, misshapen and short-lived erythrocytes, and anemia. We aim to understand the structural effects caused by the single Glu6Val mutation leading to protein aggregation. To this end, we perform multiscale molecular dynamics simulations employing atomistic and coarse-grained models of both wild-type and sickle hemoglobin. We analyze the dynamics of hemoglobin monomers and dimers, study the aggregation of wild-type and sickle hemoglobin into decamers, and analyze the protein–protein interactions in the resulting aggregates. We find that the aggregation of sickle hemoglobin is driven by both hydrophobic and electrostatic protein–protein interactions involving the mutation site and surrounding residues, leading to an extended interaction area and thus stable aggregates. The wild-type protein can also self-assemble, which, however, results from isolated interprotein salt bridges that do not yield stable aggregates. This knowledge can be exploited for the development of sickle hemoglobin-aggregation inhibitors.

KEYWORDS
Glu6Val mutation, MD simulation, protein aggregation, protein–protein interactions, sickle cell disease

1 | INTRODUCTION

Sickle cell disease (SCD) is a genetic disorder that affects the red blood cells (RBC) and it results from a single-point mutation in the β-globin gene that substitutes glutamic acid at the sixth position of the β-globin chain of adult hemoglobin (HbA) to valine in sickle hemoglobin (HbS).[1] Hemoglobin is a hemoprotein found in the RBC whose core function is to transport oxygen from the lungs to the tissues and carbon dioxide from the tissues back to the lungs. It is formed by four polypeptide chains, specifically, two α chains and two β chains (Figure 1).[2] What is clinically known as SCD is caused by a combination of physicochemical events at the molecular level that gives rise to hemoglobin dehydration, pathologic aggregation, altered RBC structure, and ultimately compromised RBC function. The levels of the endogenous substrate, the 2,3-diphosphoglycerate (2,3-DPG), have been found to increase during SCD crises, with the result that 2,3-DPG interacts with hemoglobin to increase its polymerization.[3,4] A combination of these phenomena and others are responsible for reducing the solubility of HbS, ultimately resulting in the phenotypically observed sickling process. The consequences of these changes include impeded transport and binding of oxygen, damage to the RBC morphology, and RBC interaction with endothelial surfaces, premature damage to the erythrocytes,[5,6]

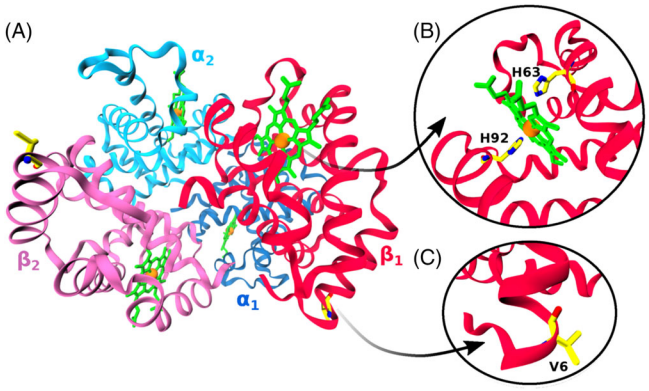

FIGURE 1 (A) The quaternary structure of HbS consists of two α subunits (here denoted α_1 and α_2 for ease of distinction, shown in shades of blue) and two β (β_1 and β_1, shades of red) subunits. Each globin subunit carries one heme (green), including an Fe^{2+} ion (orange). (B) The hemes are linked to the globin by covalent bonds between their irons and the N_ϵ of histidines His87 of the α chains and His92 of the β chains, known as the proximal histidines. On the other side of the hemes, the distal histidines are located, which are His58 in the α chains and His63 in the β chains. (C) The single mutation Glu6Val happens on the surface of the β chains near their N-terminus. The His and Val residues are shown as sticks and are colored by atom name (C: yellow; N: blue; O: red). This figure is reproduced with permission[25] and uses PDB entry 5E6E[26] as HbS structure

agonizing vaso-occlusive crisis, an overall poor health condition, and death.[7–10] It should be noted that there are disparities in clinical symptoms (e.g., frequency and severity of pain crises) exhibited among genetically identical SCD patients and thereby suggesting that apart from HbS mutation and concentration, environmental factors might also play important roles in disease development.[11,12] However, despite being studied extensively and being among the first molecular diseases to be understood up to the genetic level,[13,14] only a few drugs exist for disease management, including L-glutamine, hydroxyurea, and a recently FDA-approved drug called voxelotor.[15,16] A treatment alternative is the highly expensive bone marrow transplant, which, however, is not readily available to patients in developing nations where a significant majority of SCD patients are found.[17–22] For instance, prevalence is concentrated in sub-Saharan Africa and parts of Southeast Asia, with more than 75% of the cases are believed to be in Nigeria, Democratic Republic of the Congo, and India.[23,24]

With pathological processes in SCD directly linked to the aggregation of HbS, having a working understanding of the structural and dynamical processes underlying protein aggregation is crucial.[14] First, this provides an understanding of the aggregation process in detail, which can then be exploited in rationally developing therapeutic strategies, including peptide-based inhibitors that target HbS aggregation.[25] HbS aggregation, or polymerization, occurs via a double nucleation mechanism,[27–29] starting with an homogenous nucleation phase where HbS aggregates randomly. This is followed by heterogeneous nucleation, where the rate of polymerization increases and new nuclei form on the already existing polymer strands derived from primary nucleation.[30–32] It was suggested that the Glu6Val substitution in HbS encourages aggregation due to hydrophobic attraction between the gained valine and a hydrophobic pocket involving Phe85 and Leu88 of the adjacent HbS β globin (Figure 2). This substitution provides both the shape and physicochemical requirements necessary to kick-start the first stages of HbS polymerization. However, it should be noted that also HbA is able to aggregate (Figure 2A). Both HbA and HbS form linear aggregates involving the formation of axial contacts between α and β chains. Only in the case of HbS, these linear aggregates grow into double filaments, facilitated by lateral β-Val6-β'-Phe85/Leu88 contacts (where the prime indicates that Phe85 and Leu88 belong to another HbS molecule than Val6). The double filaments further assemble into ≈200 Å thick fibers, which eventually accumulate in highly complex, pathological HbS fiber networks.[33]

One of the first molecular dynamics (MD) studies of HbA and HbS, carried out for 62.5 ps, compared the flexibility of α and β chains in both HbA and HbS.[34] It was revealed that the β chains are generally more flexible in comparison to the α chains and that in HbS, the N-terminal region and helices D and F of the β chains exhibited a greater flexibility than those of HbA. It was implied that the HbS aggregation process might be due to this increased flexibility.[34] This study also revealed that the stability of the subunits in both HbA and HbS is due to three factors, namely hydrogen bonding, hydrophobic interactions, and conformational energy of association. In a recent simulation study, the binding free energy of HbA was determined through MD simulations and umbrella sampling.[35] The binding free energy of HbA was found to be -4.4 ± 0.5 kcal/mol, which is significantly higher than the binding free energy reported from a previous study for HbS (-14 ± 1 kcal/mol).[36] Furthermore, it was revealed that less than the 20% of the interactions in the contact interfaces are hydrophobic and that although there are similar electrostatically favored interactions found in both HbA and HbS, the potential energy associated with β-Glu6 is largely repulsive, while mildly attractive potential energies are associated with β-Val6.[37] It was concluded that i) the presence of β-Val6 is less important for the HbS polymerization process than the absence of β-Glu6, and ii) even though hydrophobic interactions play a role in the aggregation process of HbS, electrostatic interactions are found to be more predominant as opposed to what is generally believed that aggregation of HbS is driven majorly by hydrophobic interactions.[35,36] This confirmed the findings from a simulation study dating back to 1990 that employed alchemical free energy calculations and concluded that the contribution of hydrophobic interactions to HbS aggregation could be in fact negligible.[38]

The aim of this study is to provide an understanding of the structural and conformational basis of HbS aggregation, in particular the role of the Glu6Val mutation, in the aggregation process using MD simulations. A difference from previous simulation studies[35,36,38] is that we allow the hemoglobin molecules to freely associate, where they can form both axial and lateral contacts. We report the results obtained from both all-atom and coarse-grained MD simulations performed for both HbS and HbA. We first analyze the conformational flexibility of both proteins and test the applicability of Martini as coarse-grained force field[39] for modeling hemoglobin. We next apply Martini to simulate the aggregation of both HbA and HbS. The

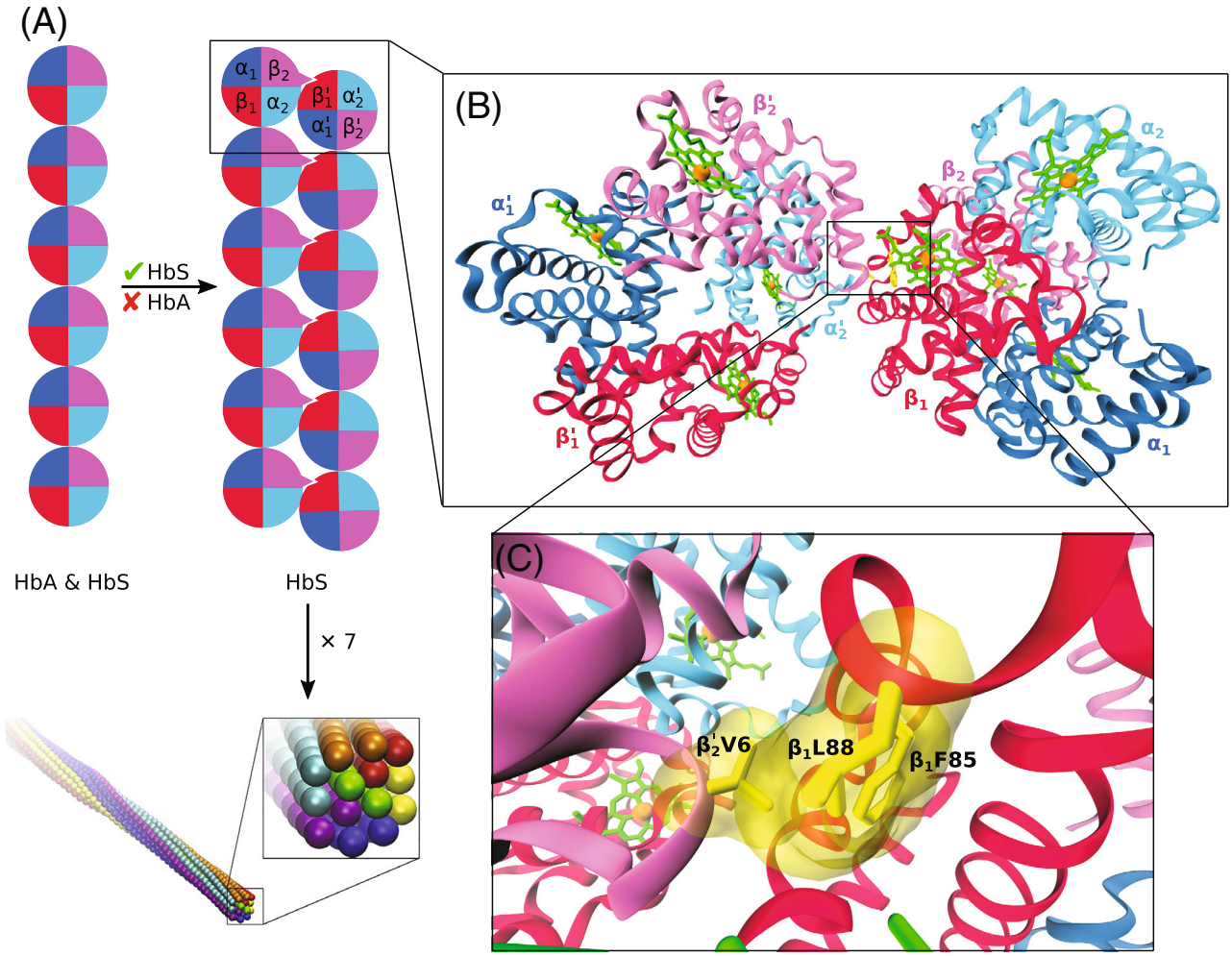

FIGURE 2 (A) Schematic representation of how the Glu6Val mutation modifies normal hemoglobin polymerization of HbS heterotetramers, involving linear Hb aggregates formed by both HbA and HbS (left) into double HbS filaments (right). The hemoglobin tetramer is represented as a circle, such that one quarter corresponds to one protein subunit using the same coloring as in Figure 1. The β-Glu6Val mutation is indicated as a protrusion from the circle in the β_2 subunit and the hydrophobic pocket as a nick in the neighboring β'_1 subunit. Seven double filaments aggregate further to form fibers (bottom). (B) A dimer formed by two HbS aggregates is shown. (C) This aggregation is mediated by β_2-Val6 interacting with the hydrophobic pocket formed by β'_1-Phe85 and β'_1-Leu88. The side chains of these three residues are shown as yellow sticks and also transparent van der Waals surfaces to better indicate the space these residues occupy. Panels B and C were produced using the crystal structure deposited in PDB entry 2HBS.[26] The figure is reproduced with permission from ref. [25]

protein–protein interactions in the resulting aggregates are elucidated and their stability is further examined by all-atom simulations of the aggregates that were back mapped from the coarse-grained to the atomistic level. These simulations enable us to quantify the relative strengths of molecular contacts and identify protein–protein interaction hotspots between HbS molecules, which in future studies can be prioritized for aggregation inhibitor design.

2 | METHODOLOGY

2.1 | Model systems

The crystal structures of HbA (PDB code 4HHB)[40] and HbS (PDB code 2HBS)[26] were used as starting structures for the MD simulations. Hemoglobin consists of four polypeptide chains, namely, two α chains and two β chains.[2] The α chains consist of 141 amino acid residues and the β chains involve 146 residues per chain. In the following, we refer to these four chains as α_1, β_1, α_2, and β_2. Each of the four chains contains a heme group at the center to which molecular oxygen binds (Figure 1). The HbS crystal structure 2HBS is in fact a homodimer, containing another four chains denoted α'_1, β'_1, α'_2, and β'_2. Throughout this study, α_1 through β_2 will be called a monomer, whereas chains α_1–β_2 plus α'_1–β'_2 will be referred to as a dimer.

2.2 | All-atom MD simulations

To investigate the structural stability of HbA and HbS as monomers and HbS also as a dimer, all-atom MD (AA-MD) simulations were

initiated from heme-containing crystal structures using CHARMM22* as force field for the proteins[41] and the TIP3P model for water.[42] The proteins were first studied in their monomeric states. To this end, the monomers were inserted into a dodecahedron box, corresponding to a distance of at least 1.2 nm between the protein and the nearest box face. The systems were then solvated with about 52 000 water molecules and ions were added to both neutralize the system and reach an NaCl concentration of 100 mM. Using the steepest descent algorithm, initial energy minimization was performed on the systems. This was followed by MD equilibration of the systems for 100 ps to reach a pressure p of 1 bar and a temperature T of 300 K. Three equilibration runs per system using different initial velocities were performed, which were then submitted to the production runs in the NpT ensemble (with N being the number of atoms) for 300 ns per run. During the production runs, the temperature was regulated using the Nosé–Hoover temperature coupling method,[43] while the pressure was controlled using the Parrinello–Rahman barostat.[44] The particle mesh Ewald method[45] was used for the calculation of electrostatic interactions in conjunction with periodic boundary conditions that were applied in all three directions of space. A cutoff of 1.2 nm was applied to the short-range Coulomb interactions calculated in real space as well as the van der Waals interactions. The LINCS algorithm was used to constrain all bond lengths,[46] and the equations of motions were solved using the leapfrog algorithm with a time step of 2 fs. The same steps used for the monomer setup were repeated for the HbS dimer, and the single production run for the dimer was carried out for 300 ns.

2.3 | Coarse-grained MD simulations

Using the final structure of the AA-MD simulation of the HbS dimer, a coarse-grained MD (CG-MD) simulation was performed to test the applicability of Martini[39] for modeling hemoglobin. The atomistic structure was converted to the CG model using the *martinize.py* script (version 2.6.). The CG topology for heme was taken from the Martini website, and the CG structure of heme was generated as described by De Jong et al.[47] The Martini force field (version 2.2) was used to model HbS and the surrounding water.[39] The protein was inserted into a simulation box using the same box dimensions as in the corresponding AA-MD simulation. Energy minimization using the steepest descent algorithm in a vacuum was first performed, followed by solvation and adding ions to obtain an NaCl concentration of 100 mM. Another round of energy minimization using the steepest descent algorithm was performed; afterward, a protein position-restrained equilibration MD run was performed for a total of 200 ps. The production run was carried out for 300 ns with a time step of 20 fs. During the production run, the temperature was regulated using velocity rescaling with canonical sampling,[48] while the pressure was kept constant at 1 bar using the Parrinello–Rahman barostat.[44] The bonds were constrained using the LINCS algorithm,[46] and the secondary structure was kept in order using elastic networks as implemented in Martini.[49]

In order to study the aggregation of HbS into larger aggregates, we performed CG-MD simulations starting from 10 HbS monomers. We inserted these HbS monomers randomly into a cubic box of dimension 41 nm × 41 nm × 41 nm, with a minimum distance of 8–10 nm between any two HbS monomers. Energy minimization using the steepest descent algorithm in a vacuum was first performed, followed by solvating the system and adding ions to obtain an NaCl concentration of 100 mM. Another energy minimization using the steepest descent algorithm was performed, followed by MD equilibration for 250 ps in the NVT ensemble (with V being the volume of the system) and for 500 ps under NpT conditions. A 30 μs production run was then performed, using a time step of 30 fs. To serve as a control, the aggregation of HbA was also simulated using the same simulation protocol as just described for HbS.

2.4 | All-atom MD simulations of CG-to-AA mapped dimers

HbA and HbS dimers that formed during the CG-MD simulations studying aggregation were extracted and converted to all-atom models through back-mapping.[50] Each back-mapped HbA and HbS dimer was then subjected to AA-MD simulations for 250 ns. For these simulations, the same simulation protocol as described in section 2.2 was applied.

2.5 | Simulation and analysis software

All MD simulations were carried out using the GROMACS software package, version 2018.[51] The analysis of the simulations was also realized using various tools of the GROMACS package as well as with the help of the MDAnalysis package.[52] More details of the analyses will be given when providing the results. Visualization of the proteins was done with Visual Molecular Dynamics (VMD),[53] while the data were plotted using Xmgrace.

2.6 | MM/PBSA analysis

To quantify the HbS-HbS interactions in the all-atom simulation of the HbS dimer crystal structure, we calculated the binding free energies ΔG_{bind} using the method based on molecular mechanics combined with Poisson–Boltzmann and surface area continuum solvation (MM/PBSA), as implemented in *g_mmpbsa* (https://rashmikumari.github.io/g_mmpbsa/).[54] Three hundred snapshots sampled at 1 ns intervals in that MD simulation were subjected to this analysis. Within the MM/PBSA scheme, the binding free energy is defined as

$$\Delta G_{\text{bind}} = \langle G_{\text{dimer}} - G_{\text{HbS-1}} - G_{\text{HbS-2}} \rangle \quad (1)$$

where $\langle \cdot \rangle$ represents the ensemble average. The free energy for each of these three entities is given as

$$G = E_{bonded} + E_{Coul} + E_{LJ} + G_{polar} + G_{nonpolar} - TS \quad (2)$$

where E_{bonded}, E_{Coul}, and E_{LJ} indicate the bonded, electrostatic, and Lennard–Jones interactions, which are obtained from the force field. G_{polar} and $G_{nonpolar}$ are the polar and nonpolar contributions to the solvation free energy, and the last term is the absolute temperature, T, multiplied by the configurational entropy, S. The entropy can be estimated by a normal-mode analysis of the vibrational frequencies, yet this term is neglected by *g_mmpbsa*. The polar energy term G_{polar} is obtained by solving the Poisson–Boltzmann equation, whereas the nonpolar term $G_{nonpolar}$ is estimated from a linear relation to the solvent accessible surface area (SASA). The parameters for the calculation of ΔG_{bind} were set $D_{solv} = 80$ for the dielectric constant of the solvent (corresponding to water), $D_{solute} = 2$ for the dielectric constant of the solute (corresponding to a globular protein), $\gamma = 0.0226778$ kJ/(mol·Å^2) for the surface tension, *sasrad* = 1.4 Å as probe radius for the SASA calculation. ΔG_{bind} was further decomposed into its contributions stemming from the interactions within the binding site and between the relevant globin-chain pairings as well as the contributions by the individual residues. To estimate the standard errors of the mean, we applied the bootstrap method with 2000 bootstrap steps.

3 | RESULTS AND DISCUSSION

3.1 | Conformational dynamics of monomeric hemoglobin

In order to assess the structural dynamics of HbA and HbA as monomers, we calculated the evolution of the root mean square deviation (RMSD) of the Cα atoms relative to the corresponding crystal structure. The RMSD results of the three runs per system were averaged and in Figure 3A, the evolution of the mean for HbS and HbA together with the standard deviations is shown. HbS attains equilibrium somewhat faster, within 60 ns, than HbA, which underwent a conformational change in one of the simulations at about 100 ns. Nonetheless, for both proteins, the RMSD stabilized at about 0.35 nm in the last 200 ns of the simulations. These low RMSD values imply that the two proteins did not deviate much from their crystal structures, indicating their stability in the simulations. The Cα RMSD values of the α and β globin chains feature even lower deviations from their respective starting structure, which entails that the four chains relaxed their positions with respect to each other, explaining the higher RMSD values for the whole proteins. For both HbA and HbS, the α chains have smaller RMSD values than the β chains, which is in agreement with the observations made in previous MD simulations of HbA and HbS.[34,55,56] Furthermore, in both systems, the β_2 chains are the most flexible.

To obtain an understanding of the origin of the structural fluctuations, we determined the root mean squared fluctuation (RMSF) of the Cα atoms (Figure 3B). In agreement with the RMSD results, a larger structural flexibility is observed for the β chains, especially for β_2 in HbA. This observation is in good agreement with a previously reported, yet much shorter MD simulation of 12 ns done for HbS.[56] The two systems show similar RMSF patterns for the α and β chains, with most fluctuations occurring in the loop regions which are known from experiments to be the most flexible regions, especially those in the β-globin chains. The highest fluctuations are noticed for the loop involving residues Val67-Gly83 of β_2 of HbA. The fact that the HbS system underwent less fluctuations in that region cannot be directly correlated to the mutation in HbS, as position six of the β globin is neither in direct nor indirect contact with the mobile loop region of the β chains. Moreover, from experiments, it is known that both proteins exhibit the same unfolding kinetics,[57] that is, whether there is a Glu or a Val at position six of the β globins does not affect the stability of hemoglobin. In neither HbA nor HbS is the stability of the helices affected by the loop motions; they are all stable as the analysis of the secondary structure confirms.

Considering that the highest level of structural changes in the β_2 chain occurred in the vicinity of the two residues involved in Fe^{2+} binding, His63 and His92 (Figure 1), we decided to investigate

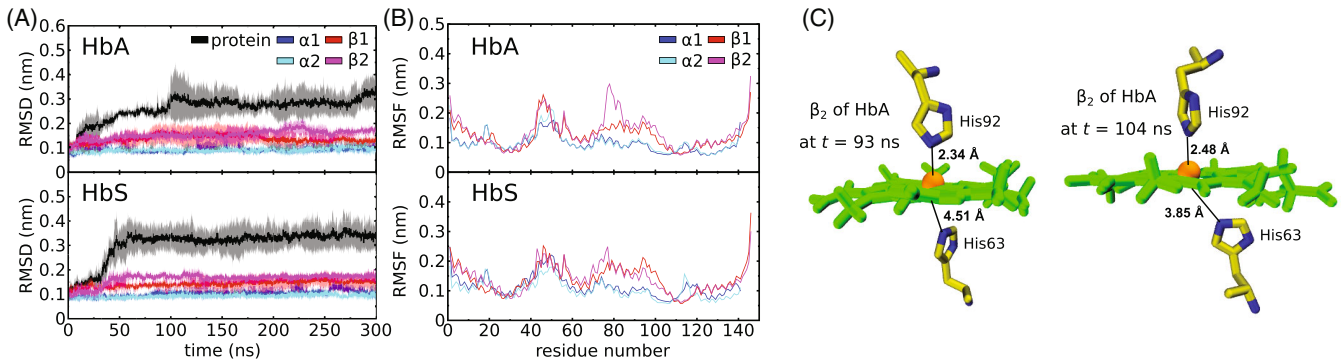

FIGURE 3 Results of the HbA and HbS monomer simulations. (A) The evolution of the Cα RMSD of HbA (top) and HbS (bottom) is shown (black lines) as well as of the corresponding individual chains (α_1: blue, α_2: cyan, β_1: red, β_2: magenta). For all RMSD calculations, the alignment was with respect to the unit for which the RMSD was calculated (i.e., the whole protein or one of the globin chains). Averages over three simulations per protein are shown; the shaded areas indicate the standard deviations. (B) The corresponding RMSF values (averaged over three MD runs) of the Cα atoms of the individual chains are shown for HbA (top) and HbS (bottom). (C) Snapshots taken from one of the three HbA trajectories at $t = 93$ ns and $t = 104$ ns demonstrate the variability of the distance between Fe^{2+} and the distal histidine residue that was monitored for the β_2 chain

whether these motions might indirectly affect the His-Fe^{2+} interactions. To this end, for the HbA and HbS simulations with the highest structural fluctuations, the minimum distances between the corresponding His residues and Fe^{2+} were calculated for all globin chains of both HbA and HbS (Figure S1). In general, the His-Fe^{2+} distances were maintained over the whole trajectories. For both proteins, the distances to His63 in the β chains were on average higher by 0.15–0.2 nm than the other His-Fe^{2+} distances. This finding might be of physiological importance for the gas binding during which a single gas molecule (e.g., O_2) inserts between a histidine and the Fe^{2+} ion. The allosteric transition between the T and R states of hemoglobin is also directly coupled to this function. The strongest changes in His-Fe distances were recorded for His63 in the β_2 globin chain of HbA. Thus, this distance is indeed affected by the motion of the neighbored loop. Snapshots of this His63-Fe^{2+} interaction taken at 93 ns (long distance) and 104 ns (short distance) show that for the long distance, His63 moved away from the heme group (Figure 3C). However, this motion is fast and reversible.

3.2 | Conformational dynamics of dimeric hemoglobin

Next, we tested the structural fluctuations of the HbS dimer, using both AA-MD and CG-MD simulations. First, the results from the 300 ns AA-MD simulation will be reported. As for the monomer, we calculated the Cα RMSD with respect to the crystal structure (Figure 4A). During the first 200 ns, the dimer was very stable with

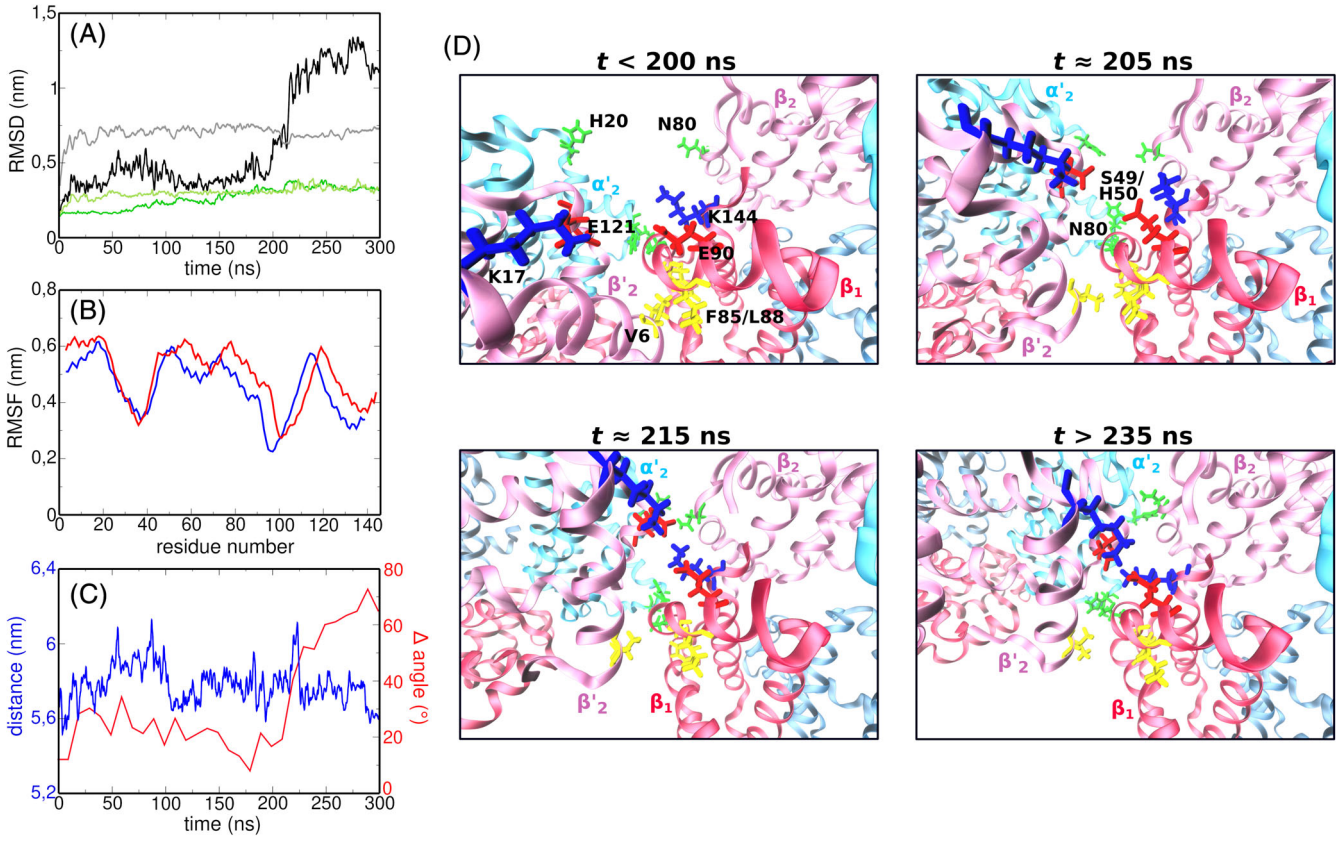

FIGURE 4 Results of the HbS dimer simulations. (A) The evolution of the RMSD of the whole HbS dimer (black) and of the composing HbS-1 and HbS-2 proteins (different shades of green) obtained from the AA-MD simulation is shown. The RSMD of the dimer in the CG-MD simulation is shown in gray. (B) The average RMSF values for the α and β chains (red and blue, respectively) support the structural stability of the globins in the AA-MD simulation. (C) The evolution of the distance between the centers of mass of HbS-1 and HbS-2 in the AA-MD simulation (blue) confirms the stability of the HbS dimer. However, the change in the angle between the lines fitted through the atomic coordinates of HbS-1 and HbS-2 (red) reveals rotations of HbS-1 and HbS-2 with respect to each other during the AA-MD simulation. (D) Snapshots from the AA-MD simulation show the rotation motion and the change in interprotein contacts accompanying it. In the first 200 ns, the interactions are dominated by the hydrophobic contacts involving β_1-Phe85/Leu88 and β'_2-Val6. In addition, there is also a contact between β_1-Asn80 and α'_2-Ser49/His50, which is the only interaction that was present throughout the 300 ns simulation. At $t \approx 205$ ns, the side chains of β_1-Glu90/Lys144 and of β'_2-Lys17/Glu121 reoriented, causing electrostatic attractions. This gives rise to a rotation of HbS-1 and HbS-2 with respect to each other, leading to stable electrostatic contacts between β_1 and β'_2 at $t \approx 215$ ns. This reorientation is completed by a further polar contact between β_2-Asn80 and α'_2-His20, while the initial hydrophobic contact β_1-Phe85/Leu88–β'_2-Val6 is broken in the rest of the simulation. The same perspective is used for the four snapshots, highlighting the changes in orientation of chains α'_1–β'_2 with time. The coloring is the same as in Figure 1; hydrophobic residues are shown in yellow, polar ones in green, and positively and negatively charged ones in blue and red, respectively

RMSD values below ≈0.5 nm. After 200 ns, however, the RMSD increased to above 1 nm. To test whether the structural changes underlying this rise in RMSD are caused by structural instabilities in either or both of the two HbS proteins composing the dimer (denoted as HbS-1 and HbS-2 in the following), we calculated their individual RMSDs after separately aligning HbS-1 and HbS-2 to their respective starting structures. Both HbS-1 and HbS-2 are found to be stable as their RMSD values do not fluctuate and do not rise beyond 0.3 nm. To further characterize the structural fluctuations within the different chains, we calculated the Cα RMSF, which we present as average over α_1, α_2, α'_1, and α'_2 for the α chains and β_1, β_2, β'_1, and β'_2 for the β chains (Figure 4B). In contrast to the fluctuation profiles reported above for the HbS monomer, in the HbS dimer, the flexibility of the α and β chains is very similar, which indicates a higher structural stability of the β-globin chains in the dimerized structure relative to the monomeric HbS. This can be explained by the contacts that are present between β_1 and β'_2 in the dimer (Figure 2), which limit the motions of these residues. We analyzed the secondary structure of the dimer and observed the preservation of all helices throughout the trajectory.

It can, thus, be concluded that both proteins composing the dimer are more stable than in monomeric HbS, which indicates that the rise of the RMSD for the HbS dimer must result from reorientations of HbS-1 and HbS-2 with respect to each other. To characterize this motion, we calculated the distance between the centers of mass of HbS-1 and HbS-2 (Figure 4C, blue line), which shows that the two proteins did not drift away from each other. Only small distance fluctuations occurred, which, however, cannot explain the sudden RMSD increase at ≈200 ns. To further probe the cause of the RMSD increase, we tested whether the orientation between HbS-1 and HbS-2 changed during the simulation. To this end, we employed the Tcl script *fit_angle.tcl* (provided at http://www.ks.uiuc.edu/Research/vmd/) using its function *sel_sel_angle_frames*. This calculation started with the least square fitting a line through the coordinates of the atoms of HbS-1 and HbS-2, respectively. The angle between the resulting two lines was then determined for all frames of the trajectory. In Figure 4D (red line), the deviation of this angle from its starting value in the crystal structure is shown. In the first 200 ns, only small changes in the orientation of HbS-1 and HbS-2 with respect to each other occurred. However, after 200 ns, a considerable angular motion of about 60° took place. This angle change coincides with the rise in the overall RMSD, allowing us to conclude that this increase results from a change in the orientation of the two HbS proteins composing the dimer. The origin of this reorientation is due to the formation of more stable interprotein contacts, as discussed below in section 3.3.

Since our aim is to simulate HbS aggregation, which can only be accomplished at the CG level given the considerable system size, we first probed the stability of the HbS dimer in a CG-MD simulation. For comparability with the AA-MD simulation, this was run for 300 ns and also analyzed in terms of Cα RMSD (Figure 4A). The dimer modeled at the CG level was found to be stable. The RMSD rose quickly to 0.6 nm within the first 10 ns and fluctuated between 0.6 and 0.8 nm for the rest of the simulation. This is below the RMSD that was obtained in the AA-MD simulation of the HbS dimer, as the two HbS proteins composing the dimer did not rotate with respect to each other as happened in the AA-MD simulation. This suggests that the Martini force field is a suitable choice for simulating the aggregation of hemoglobin.

3.3 | Protein–protein contacts in the HbS dimer

In order to understand the reorientation motion that occurred in the AA-MD simulation of the HbS dimer, we analyzed the residue–residue contacts between HbS-1 and HbS-2. Since the reorientation took place at about 200 ns, the average interchain contact maps were computed for the first 200 ns and last 100 ns of this simulation. The same kind of analysis was applied to the CG-MD simulation of the HbS dimer, which serves as a reference here.

For the HbS dimer in the AA-MD simulation, the interacting chains for the first 200 ns were identified as β_1-β'_2 and β_1-α'_2, while in the last 100 ns, additional contacts were formed in the chain combinations β_2-β'_2 and β_2-α'_2. To further dissect these interactions, they were resolved at the residue level. Residue *i* is said to be in contact with residue *j* when they are within a distance of 0.5 nm of each other. Table S1 shows the comparison between the amino acid residues interacting during the first 200 ns and last 100 ns. These contacts are generally quite similar, especially those involving the β_1 chain, and involve lateral contacts that have been previously reported experimentally to be critical in the aggregation process of HbS,[26,58] such as β_1-Thr87/Leu88 with β'_2-Ser9, β_1-Thr84 with β'_2-Val6, and β_1-Thr87 with β'_2-Ala10. These are contacts involving and surrounding the mutation site β-Val6, lending support to the importance of that mutation for the HbS aggregation. Another contact, which was present all the time, is between β_1-Asn80 and α'_2-Ser49/His50.

As the contact information does not provide insight into the strength of these interactions, we calculated the non-bonded interaction energies consisting of both Lennard-Jones and electrostatic interactions using the *rerun* option of the GROMACS *mdrun* program for all the residue pairs identified by the protein–protein contact analysis. For the residue pairs with a time-averaged interaction energy below −2 kcal/mol, the results are shown in Figure 5A. The strongest intermolecular interactions are mostly electrostatic in nature, involving salt bridges, such as the interactions β_1-Glu90–β'_2-K17 and β_1-Lys144–β'_2-Glu121. The major difference noticed in the interactions between the first 200 ns and last 100 ns is the presence of hydrophobic interactions in the first 200 ns involving β'_2-Pro5 interacting with β_1-Ala70 and β'_2-Val6 forming contacts with β_1-Ala70/Phe85/Leu88 of the other HbS protein, yet the absence of these interactions in the last third of the simulation. The electrostatic interactions, on the other hand, are only weakly present in the first 200 ns but dominate in the last 100 ns. This indicates that the Val6 gained from mutation remained near the hydrophobic cavity as present in the crystal structure during the first 200 ns but moved away from that site in the latter part of the simulation. This is a consequence of the rotation of the two HbS proteins with respect to each other (Figure 4C), which is initiated by electrostatic interactions of β_1-Glu90/Lys144 with β'_2-Lys17/Glu121 that are only weak initially but, due to structural

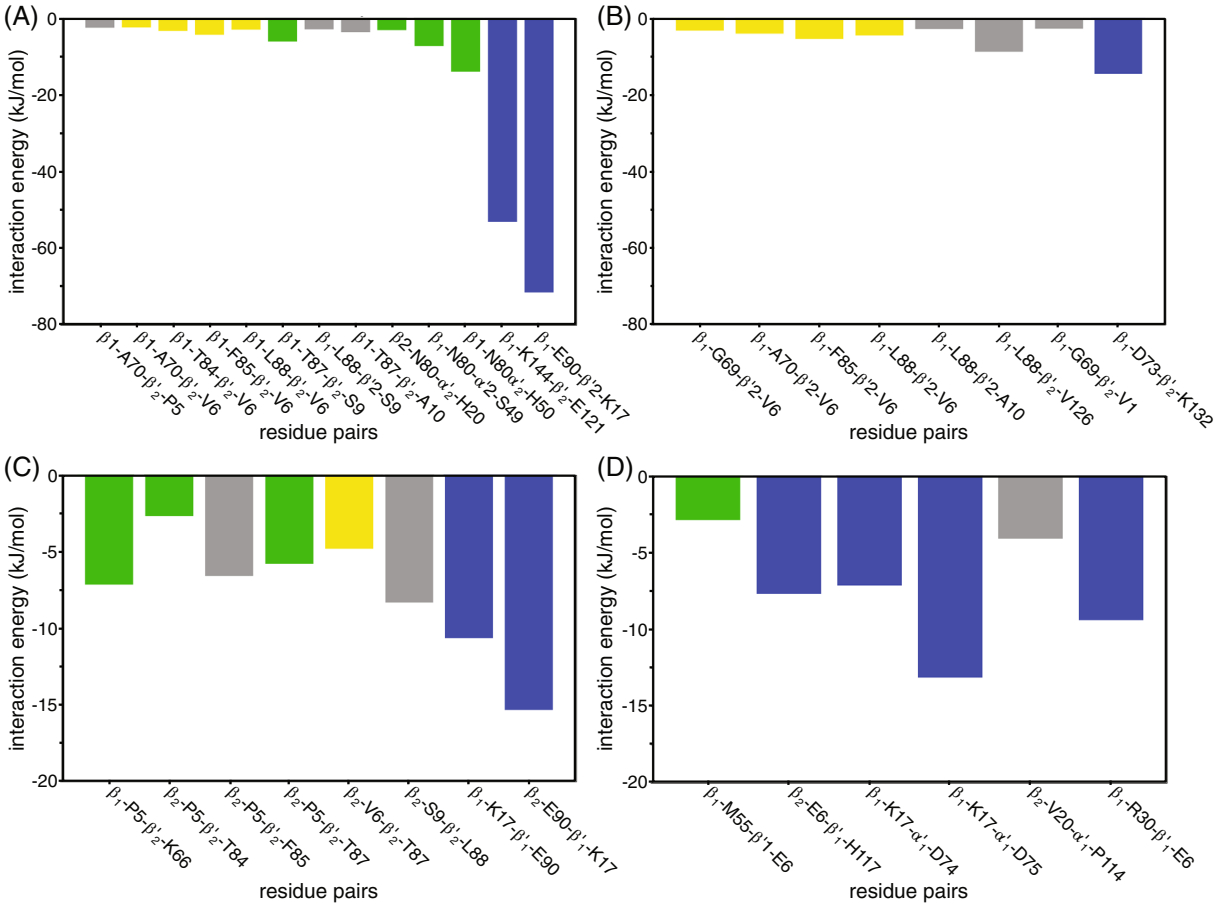

FIGURE 5 Time-averaged residue–residue interaction energies between hemoglobin molecules. In the upper panels, the energies between the two HbS proteins composing a dimer from (A) the AA-MD simulations and (B) the CG-MD simulations are shown. In the lower panels, the interaction energies between any two hemoglobin molecules in the aggregation simulations leading to a decamer in the CG-MD simulations of (C) HbS and (D) HbA are presented. Interaction energies involving Val6 of HbS are highlighted in yellow, while electrostatic and polar interactions are colored in blue and green, respectively. Purely hydrophobic interactions as well as interactions between hydrophobic and polar amino acids are shown by gray bars

fluctuations of the long side chains being involved, gain traction in the course of the simulation. At about 205–215 ns, these interactions are fully established. This is accompanied by a further polar contact that formed, namely between β_2-Asn80 and α'_2-His20, while the other polar contact, β_1-Asn80–α'_2-Ser49/His50, is the only interaction that survived the whole 300 ns simulation. The rotational motion of HbS-1 and HbS-2 with respect to each other along with the relevant residue–residue contacts is illustrated by snapshots taken from the AA-MD simulation in Figure 4D. It should be noted that the importance of salt bridges in the lateral HbS-HbS contact formation was also emphasized by Galamba and Pipolo; based on umbrella sampling MD simulations they identified, among others, the β_1-Glu90–β'_2-K17 salt bridge as a strong interaction.[36]

The gain of interprotein Coulomb and Lennard-Jones interactions may be counteracted by the loss of interactions with the surrounding solvent. To assess the interplay between protein–protein and protein–solvent interactions, we calculated the binding free energy, ΔG_{bind} using the MM/PBSA method and decomposed it into relevant contributions (see Equations (1) and (2), Table S2 and Figure S2). It should be noted that the absolute energy values that result from this method when applied to HbS dimers are too approximate to warrant an in-depth analysis. They are neither comparable to experimental binding free energies of HbS dimerization[59,60] nor to those obtained from more accurate, yet computationally more expensive umbrella sampling calculations.[35,36] Moreover, a previous simulation study that used MD simulations and MM/PBSA to calculate ΔG_{bind} for HbS dimerization produced a value an order of magnitude smaller than the experimental and umbrella sampling values.[56] Therefore, here, we only use the MM/PBSA results to study the change in ΔG_{bind} contributions with time in order to unravel the underlying interaction changes. In a recent work by our group, where we applied the same MM/PBSA method for the binding of small molecules to a protein, we were very successful in identifying strongly and weakly binding ligands, as confirmed by a wet-lab binding assay, by using relative ΔG_{bind} energies (and ignoring the absolute values).[61]

Considering the major structural rearrangement that occurred in the HbS dimer at ≈200 ns, the assessment of ΔG_{bind} was performed separately for the time spans 0–200 ns and 200–300 ns. Comparison of the resulting ΔG_{bind} values reveals that the reorientation in the HbS dimer at ≈200 ns is driven by the formation of more stable

residue–residue contacts than those present before, leading to a decrease in ΔE_{Coul}. This confirms the conclusions drawn above that the creation of salt bridges involving globin chains β_1 and β'_2 is the main driving force behind that reorientation within the dimer. This comes at the energetic cost of solvation, as ΔG_{polar} increases; however, the gain from ΔE_{Coul} is larger than the loss from ΔG_{polar}. The dissection of the ΔG_{bind} values into their per-residue contributions unravels that negatively charged residues are often complex stabilizing, whereas positively charged residues contribute with positive energy values to ΔG_{bind}. The comparison between the per-residue contributions at 0–200 ns and 200–300 ns reinforces our inference that the hydrophobic residues around β'_2-Val6 are no longer of relevance after the dimer adjusted its geometry to create salt-bridge interactions between chains β_1 and β'_2. We, thus, conclude that the lateral contacts in the HbS dimer crystal reported by Harrington et al.[26] are strong enough to sustain for a certain simulation time of the HbS dimer in solution at room temperature, yet other, electrostatic contacts are also possible and, due to their strength, can cause reorientations in the dimer.

For the HbS dimer in the CG-MD simulation, two residues *i* and *j* were considered to be in contact when the distance between any two beads from respective residues was under 0.75 nm. The interacting chains discovered in the first 200 ns and last 100 ns are the same, which correlates with the small RMSD values observed throughout this simulation (Figure 4A). Table S3 shows the list of interprotein contacts found in the CG system. The interchain contacts are similar to those found in the AA-MD simulation, such as β'_2-Val6 interacting with the hydrophobic residues Ala70, Phe85, and Leu88 of the β_1 chain. However, there are also few differences. For example, β_1-Gly69 interacts with β'_2-Val1/Val6, β_1-Leu88 with β'_2-Val126, and β_1-Asp73 with β'_2-Lys132. Apart from the latter interaction, the predominant interactions are found to be mainly of hydrophobic nature (Figure 5B) and are mainly at and around the mutation site β'_2-Val6 and its preferred interaction region at β_1-Phe85/Leu88. This implies that these hydrophobic interprotein contacts are strong enough that the crystal structure of the HbS dimer remains stable in the CG-MD simulations. Moreover, based on Figure 5B, it seems that electrostatic interactions are generally less dominant in the Martini force field, as the single salt bridge present between HbS-1 and HbS-2 in the CG-MD simulation of the HbS dimer is only somewhat stronger than the hydrophobic interactions. Therefore, there is no electrostatic driving force for reorientations between HbS-1 and HbS-2 as witnessed in the AA-MD simulations.

3.4 | Hemoglobin aggregation simulations

The aggregation of HbS was studied using CG-MD simulations, by allowing 10 HbS monomers to aggregate freely in a cubic simulation box for 30 μs. At the end of the simulation, a decamer had formed, which adopted an elongated shape (Figure 6, left). An oligomer size analysis was carried out to study the growth of the aggregate as a function of time (Figure 7A). Monomers *i* and *j* are said to associate when the distance between any bead in *i* is under 0.75 nm from any bead in *j*. It should be noted that this kind of analysis is quite robust against the chosen cutoff distance as a reduction of this cutoff to 0.65 nm did not change the result. We decided to use 0.75 nm to have a consistent contact definition throughout this study, as this cutoff was applied to the HbS dimer, too. Monomers *i* and *j* were said to dissociate when the minimum distance between them is more than 1 nm. This higher cutoff compared to the one applied for defining association allows for reorientations between two HbS proteins in the process of assembly, which might temporarily increase the distance between them, without them being counted as dissociation events. Figure 7A shows that the aggregation either involves the attachment of a monomer or of a transiently formed dimer to the growing oligomer, which reaches the decamer state at about 20 μs. To check whether the individual monomers underwent noteworthy structural changes during the aggregation process, the Cα RMSD was recorded for all HbS proteins (Figure 7B). The resulting values are all in the range of 0.4–0.6 nm and stable within 5 μs, revealing that, apart from some initial changes, the individual HbS proteins did not undergo noteworthy structural rearrangements following the oligomer formation process.

In order to identify important protein–protein contacts involved in the self-assembly process, interprotein contact maps were calculated for the interfaces between any two proteins (called dimers in the following) present in the decamer. A total of 10 such dimers were identified (Figure 6, left). To separate the weak and thus unimportant contacts from the strong ones, we calculated the interaction energies

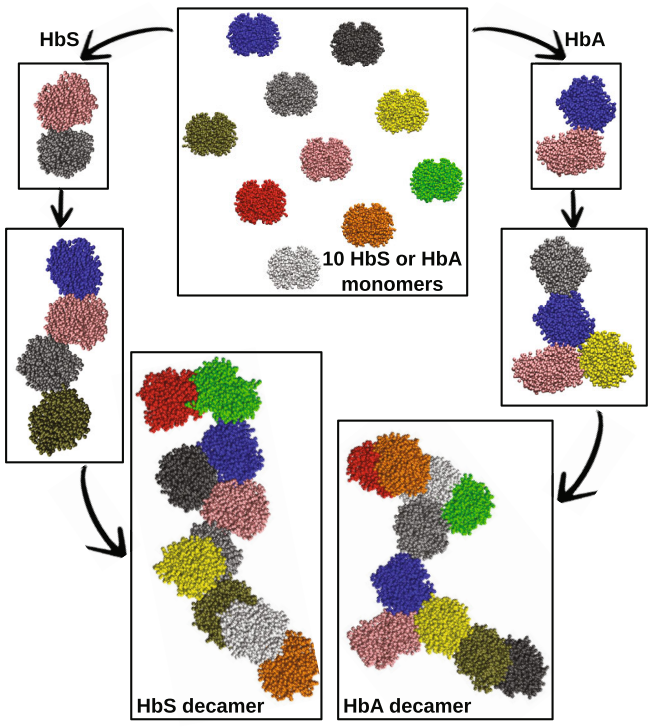

FIGURE 6 Hemoglobin aggregation pathways obtained from CG-MD simulations. The simulations were started from 10 HbS or HbA monomers. The aggregation proceeded in a stepwise fashion until an HbS (left) or HbA (right) decamer formed

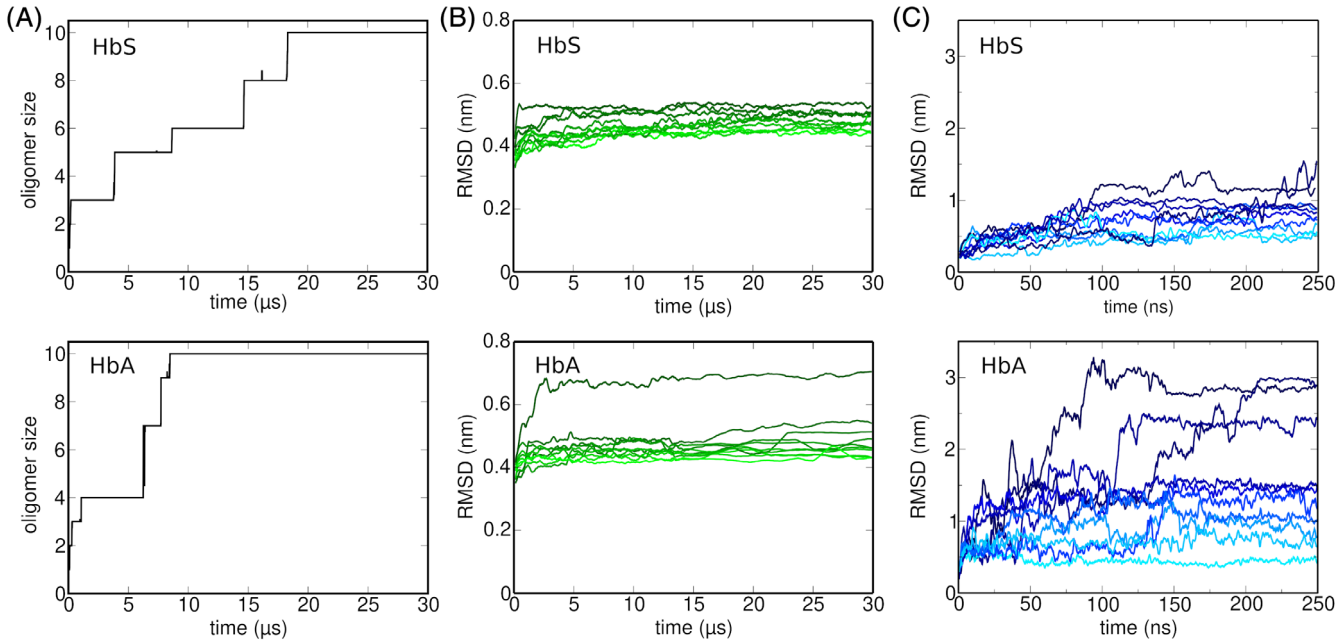

FIGURE 7 Results of the aggregation simulations of HbS (top) abd HbA (bottom). (A) The aggregation state in terms of the number of hemoglobin proteins in the oligomer that formed is provided. (B) The evolution of RMSD of the individual proteins during the aggregation simulations confirms protein stability. (C) However, the RMSD of the back-mapped dimers composing the decamer during the AA-MD simulations reveals that some of the HbA dimers are not stable. In (B) and (C) the different RMSD curves for the different proteins or dimers are shown in different shades of green and blue, respectively

for all contacts identified. The interaction energies of these strong contacts are displayed in Figure 5C. Several amino acid contacts reported from previous studies were found to be present in our simulation. Interestingly, β_2-Val6, which is the mutated residue believed to be the main cause of sickling of red blood cells, is found forming lateral contacts with hydrophobic amino acid residue β'_2-Thr87 of the interacting HbS protein. As mentioned above, these lateral contacts are also present in the crystal structure of the HbS dimer[26] and were also reported in the studies carried out on the molecular interactions in the crystal structure of HbS by Padlan and Love.[58] However, there are differences in the location of the amino acid residues on the β-globin chains. In the previous HbS crystallographic studies, these contacts are found to exist between β_1 and β'_2 or β_2 and β'_1 chains. Pro5, which is also an important amino acid residue forming contacts in the HbS fiber, is seen here on the β_2 chain to interact with a series of amino acid residues, such as Thr84, Phe85, and Thr87 of the β'_2 chain of the other monomer. This is complemented by β_2-Ser9 interacting with β'_2-Leu88. All these residues are in the direct neighborhood of β_2-Val6 and β'_2-Thr87, respectively. It can, thus, be concluded that the lateral interactions in HbS aggregation are not only mediated by β-Val6 but also a cooperation between this and surrounding residues. The energy plots in Figure 5 show that these interactions involve both hydrophobic and polar interactions.

As for the HbS dimer in the AA-MD simulation, electrostatic interactions are also found to play a role, especially those involving β-Glu90, which prefers to form contacts with β-Lys17. Here, the combinations β_1-Lys17–β'_1-Glu90 and β_2-Glu90–β'_1-Lys17 are encountered. This contact can be formed together with the neighboring contacts involving β-Val6, as the snapshot of the HbS dimer at $t = 215$ ns in Figure 4D shows. Finally, the interaction β_1-Pro5–β'_2-Lys66 should be mentioned, as it is of considerable strength and has not been observed for the HbS dimer. In summary, the Martini force field identified the lateral contacts known from the HbS dimer crystal structure as driving force behind HbS aggregation, in addition to a neighboring electrostatic interaction that is not present in the crystal structure.

With the objective to test how robust our simulation results are with respect to the mutation β-Glu6Val, we performed the same kind of CG-MD simulation for the aggregation of HbA. Figure 7A shows the oligomer size as a function of time for HbA, which is very similar to that of HbS. The only difference is that the maximum oligomerization state of 10 was reached earlier, already at about 10 μs. Similar to that of the HbS, the HbA decamer also adopts an elongated, yet more curved shape (Figure 6). It should be noted that HbA is known to be able to form linear aggregates, which is facilitated by the formation of axial contacts between the α and β globins of neighboring hemoglobin molecules (Figure 2A).[33] The structural changes of the individual HbA proteins during the simulation were assessed by calculating the Cα RMSD from their respective starting structure (Figure 7B). The RMSD values are mostly similar to those of the HbS proteins during the aggregation simulation. Only one HbA protein deviated more from the starting structure, reaching RMSD values of about 0.7 nm.

As for the HbS aggregation, we identified important protein–protein contacts involved in the self-assembly process of HbA by analyzing the interprotein contacts present in the decamer. Most of the dimers that formed during the HbA aggregation are found to have similar interacting regions. The strongest contacts are present

between chain α_1 and β'_1, chain β_1 and β'_1, as well as chain β_2 and α'_1. Comparison with the chain–chain interactions that drove the aggregation of HbS reveals a smaller involvement of chain β_2 and an engagement of α_1 instead in HbA aggregation. Among the strongest molecular interactions, the electrostatic ones prevail, also in terms of their number (Figure 5D). An example is the positively charged residue β_1-Lys17 that interacts with the negatively charged Asp74 and Asp75 of α'_1. Importantly, Glu6, which is the only amino acid residue that differentiates HbA from HbS, is seen here as part of the β'_1 chain of one monomer interacting with several amino acid residues, including Arg30 and Met55 of the β_1 chain of another monomer. The presence of β-Glu and the absence of a valine residue at that place, thus cause the aggregation of HbA to proceed via a different protein–protein interaction surface than in the case of HbS. This observation is supported by the fact that the only noteworthy hydrophobic/hydrophilic contact was formed between β_2-Val20 and α'_1-Pro114, an interaction that was not observed during HbS aggregation. Moreover, while the CG-MD simulation of HbS aggregation predicted this process to be mainly driven by lateral contacts, in the case of HbA, the aggregation is largely facilitated by axial contacts.

3.5 | Stability of back-mapped dimers

While it is reassuring that the Martini force field is able to distinguish between different interactions giving rise to HbS and HbA aggregation, we further verified that the protein–protein contacts sampled in the CG-MD simulations are indeed stable by transferring the 10 HbS and 10 HbA dimers that compose the respective decamer back to the all-atom level and performed for each of them a 250 ns AA-MD simulation. The resulting Cα RMSD plots (Figure 7C) reveal that the HbS dimers are more stable than the HbA dimers. This indicates that more stable protein–protein contacts are present in the HbS dimers, which in reverse renders HbS more prone to aggregation in contrast to HbA. For the HbS systems, eight of the 10 dimers have RMSD values below 1.0 nm and the remaining two show deviations of maximal 1.5 nm. For HbA, on the other hand, 70% of the dimers feature RMSD values above 1.0 nm, and three of the dimers even reach RMSD values clearly above 2.0 nm.

The two most stable and two least stable back-mapped HbS dimers (final RMSD $\approx$ 0.5 and >1 nm, respectively) and the three least stable HbA dimers (final RMSD >2 nm) were analyzed in more detail to understand the sources of stability and instability, respectively. We first assessed whether the dimers dissociated by calculating the distance between the centers of mass of the two hemoglobin proteins composing the dimer in question (Figure S3A). While none of the HbS or HbA dimers dissociated within the 250 ns simulation time, the distance between HbA-1 and HbA-2 is generally larger than that between HbS-1 and HbS-2 and even reached beyond 7 nm. In the HbS dimers, on the other hand, this interprotein distance is mostly below 6 nm and thus similar to the corresponding distance in the HbS dimer crystal. These different distances follow from the distinct interprotein contact areas in HbA and HbS dimers, involving axial contacts in the case of HbA and lateral contacts in the HbS dimers. Next, we analyzed whether the hemoglobin proteins change their orientation with respect to each other by calculating the change in angle between them. As done before, we fitted a line through the atoms belonging to HbS-1 and HbS-2 (or HbA-1 and HbA-2), respectively, and computed the angle between these lines. In Figure S3B, the change in this angle is plotted. In the two HbS dimers with the lowest RMSD values, the orientation of the two HbS molecules with respect to each other is stable and the angle changes are generally below 25°. In the unstable HbS dimers, on the other hand, the relative orientation of the two HbS proteins is less conserved, as angles of 50° and above are reached. However, in two of the most unstable HbA dimers, this angle even rises beyond 150°, reinforcing the conclusion that no stable interprotein contact surface has formed here.

In order to identify the protein–protein contacts that lead to either stable or unstable hemoglobin dimers, interprotein contact maps were calculated and for the contacts identified, the interaction energies were determined. For the strong HbS contacts, the time- and dimer-averaged interaction energies are provided in Figure S3C. Similar residue–residue contacts are found as identified in the AA-MD simulation of the HbS dimer started from the crystal structure (Figure 5A) and as encountered during HbS aggregation in the CG-MD simulation (Figure 5D), which underscores the importance of these lateral interprotein contacts for both the aggregation process and the stability of the resulting aggregates. The contacts majorly involve or surround the mutation site Val6, such as β_1-Ala70/Thr84/Leu88/Phe85 interacting with β'_2-Pro5/Val6/Ser9. Moreover, as observed in the HbS dimer in the AA-MD simulation and CG-MD aggregation simulation, where electrostatic interactions appeared to play a major role, β_1-Glu90 is found to strongly interact with β'_2-Lys17. Another noteworthy protein–protein interaction is observed between β_1-Asp73 and the polar amino acid residue β'_2-Thr4. This interaction emerged from its neighborhood to the mainly hydrophobic interactions involving the mutation site Val6. The number of interactions in the least stable back-mapped HbS dimers is notably smaller. Especially, the interactions involving Val6 of the β chains and its surrounding residues are less pronounced or even missing, which cannot be compensated by the additional contact involving β_2-His2 and β'_2-Lys120 and the very strong electrostatic interaction between β_1-Glu90 and β'_2-Lys17. It can be inferred that the latter gives rise to the stably low interprotein distance in the least stable HbS dimers (Figure S3A), yet it is not sufficient to keep the orientation between the two proteins the same.

For the HbA back-mapped dimers, the contacts in the least stable dimers are completely different than those that drove the aggregation process (Figure S3C versus Figure 5D). This finding is different from the observations made for HbS and indicates that the interprotein contacts encountered during HbA aggregation do not give rise to a characteristic aggregation pattern as seen for HbS. Another difference is that the contacts are not formed between two β chains underlying lateral contacts but between α_1 and β'_1 or β'_2 corresponding to axial contacts. As for the HbA aggregation process, the strongest intermolecular interactions in the HbA back-mapped dimers are mostly

electrostatic in nature, such as the interactions α_1-Lys60–β'_1-Glu90/Asp94 and α_1-Lys82–β'_1-Asp47. Another prominent interaction involves the polar contact α_1-Ser49/Ser52–β'_1-Asn80. Hydrophobic contacts, which are individually weaker than an electrostatic or polar contact, yet in HbS involve several residues at once and are in sum of noteworthy magnitude and involve a larger contact area, are completely missing in the HbA dimers. This implies that the electrostatic/polar interprotein contacts in the HbA dimers are strong enough to prevent the HbA proteins to dissociate from each other, yet they are too local to avoid protein reorientations in the dimer.

4 | CONCLUSIONS

We studied different aspects of the human sickle hemoglobin beginning with the conformational dynamics of its monomeric units and finishing off with the aggregation into decamers and an analysis of the underlying protein–protein interactions. At each point, the wild-type hemoglobin was studied alongside to provide a reference system for interpreting the observations from the sickle hemoglobin structure and, crucially, for shedding light on the structural effects of the disease-causing Glu6Val mutation. Our investigation revealed that this mutation kicks off effects that may not be directly obvious. We uncovered that the mutation leads to an increase in the overall structural rigidity of the sickle hemoglobin monomeric and dimeric assemblies. The β-globin chains in particular were observed to exhibit differences in flexibility between the sickle and the wild-type hemoglobin, with the former's β-globin chain being more stable. The involvement of this particular chain in reported aggregate contacts indicates that even a slight stabilization of this chain could contribute to the difference between pathologic aggregation observed in sickle hemoglobin and the absence of it in the wild type. Our analysis also revealed a stabilization of the His63-Fe^{2+} coordination as a result of the Glu6Val mutation that may play a role in the reduced O_2 binding by sickle hemoglobin.

From the aggregation simulations, we identified some previously reported residue contacts and new ones that are likely involved in the early phase of the HbS polymerization process. In particular, the aggregation simulations resulting in an HbS decamer and the in-depth analysis of the HbS dimers composing the decamer reinforce the importance of the lateral contact formed between β-Val6 and β'-Phe85/Leu88 of the interacting HbS protein. Importantly, this hydrophobic interaction gives rise to a number of further hydrophobic and polar residue–residue contacts, involving β-Thr4/Pro5/Ser9 and β'-Ala70/Asp73/Thr84/Thr87. In addition, there is a particularly stable electrostatic interaction that is in the direct neighborhood surrounding the contact area involving β-Val6, namely the β-Lys17–β'-Glu90 contact. Only the sum of these interactions creates a stable contact area around the β-Val6–β'-Phe85/Leu88 interaction. This observation is not in full agreement with the previous conclusion that the presence of β-Val6 is less important for the HbS polymerization than the absence of β-Glu6.[37] From comparing our HbS and HbA simulation results, we deduce that both the absence of β-Glu6 and the presence of β-Val6 are important for HbS aggregation. The absence of β-Glu6 allows the contact between β-Lys17 and β'-Glu90 to be formed, as in HbA the β-Glu6 residue being close to β-Lys17 prevents the latter to get close to β'-Glu90. A similar conclusion was reached by Galamba[35] and would explain why HbA is able to polymerize in a similar fashion as HbS at high salt (1.5 M potassium phosphate) concentrations,[27–29,62] which screens the repulsion between β-Glu6 and β'-Glu90. The presence of β-Val6, on the other hand, allows a network of contacts to be established around the well-known β-Val6–β'-Phe85/Leu88 interaction. With regard to HbA, we conclude that it is also able to form aggregates, yet involving mainly axial contacts that are not sufficient in terms of number and strength to lead to stable, long-lived HbA aggregates.

Taken together, we have presented important structural information about sickle hemoglobin structure and aggregation, which, we believe, will be important in the search for novel aggregation inhibitors for the treatment of sickle cell disease. Here, in addition to targeting the hydrophobic HbS–HbS interaction involving β-Val6, one could additionally aim at interrupting the electrostatic β-Lys17–β'-Glu90 contact.

ACKNOWLEDGMENTS

M.O.O. gratefully acknowledges the funding she received for this project from DAAD via the Research Grants - Doctoral Programmes in Germany. The authors thank Dr. Bogdan Barz for fruitful discussions at the beginning of this project. Simulations were performed with computing resources granted by RWTH Aachen University under project rwth0518. Financial support by the Deutsche Forschungsgemeinschaft through funds INST 208/704-1 FUGG to purchase the hybrid computer cluster used in this study is gratefully acknowledged. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTERESTS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Birgit Strodel https://orcid.org/0000-0002-8734-7765

REFERENCES

1. Ingram V. Abnormal human haemoglobins: I. the comparison of normal human and sickle-cell haemoglobins by “fingerprinting”. *Biochim Biophys Acta*. 1958;28:539-545.
2. Perutz MF. Stereochemistry of cooperative effects in haemoglobin: haem–haem interaction and the problem of allostery. *Nature*. 1970;228(5273):726-734.
3. Poillon WN, Kim BC. 2,3-Diphosphoglycerate and intracellular pH as interdependent determinants of the physiologic solubility of deoxyhemoglobin S. *Blood*. 1990;76:1028-1036.

4. Poillon WN, Kim BC, Labotka RJ, Hicks CU, Kark JA. Antisickling effects of 2,3-diphosphoglycerate depletion. *Blood*. 1995;85:3289-3296.
5. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. *Lancet*. 2010;376(9757):2018-2031.
6. Li X, Dao M, Lykotrafitis G, Karniadakis GE. Biomechanics and biorheology of red blood cells in sickle cell anemia. *J Biomech*. 2017;50:34-41.
7. Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz R, Wallace RB. Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc Natl Acad Sci U S A*. 1983;80(1):278-282.
8. Hahn EV, Gillespie EB. Sickle cell anemia: report of a case greatly improved by splenectomy. Experimental study of sickle cell formation. *Arch Intern Med*. 1927;39(2):233-254.
9. Mohandas N, Evans E. Adherence of sickle erythrocytes to vascular endothelial cells: requirement for both cell membrane changes and plasma factors. *Blood*. 1984;64:282-287.
10. Chien S, Usami S, Bertles JF. Abnormal rheology of oxygenated blood in sickle cell anemia. *J Clin Investig*. 1970;49(4):623-634.
11. Ballas SK, Mohandas N. Sickle red cell microrheology and sickle blood rheology. *Microcirculation*. 2004;11(2):209-225.
12. Niscola P, Sorrentino F, Scaramucci L, De Fabritiis P, Cianciulli P. Pain syndromes in sickle cell disease: an update. *Pain Med*. 2009;10(3):470-480.
13. Galkin O, Pan W, Filobelo L, Hirsch RE, Nagel RL, Vekilov PG. Two-step mechanism of homogeneous nucleation of sickle cell hemoglobin polymers. *Biophys J*. 2007;93(3):902-913.
14. Eaton WA, Bunn HF. Treating sickle cell disease by targeting HbS polymerization. *Blood*. 2017;129(20):2719-2726.
15. Gardner RV. Sickle cell disease: advances in treatment. *Ochsner J*. 2018;18(4):377-389.
16. Vichinsky E, Hoppe CC, Ataga KI, et al. A phase 3 randomized trial of voxelotor in sickle cell disease. *N Engl J Med*. 2019;381(6):509-519.
17. Bhatia M, Sheth S. Hematopoietic stem cell transplantation in sickle cell disease: patient selection and special considerations. *J Blood Med*. 2015;6:229.
18. Mentzer WC, Heller S, Pearle PR, Hackney E, Vichinsky E. Availability of related donors for bone marrow transplantation in sickle cell anemia. *Am J Pediatr Hematol Oncol*. 1994;16(1):27-29.
19. Dallas MH, Triplett B, Shook DR, et al. Long-term outcome and evaluation of organ function in pediatric patients undergoing haploidentical and matched related hematopoietic cell transplantation for sickle cell disease. *Biol Blood Marrow Transplant*. 2013;19(5):820-830.
20. Bernaudin F, Socie G, Kuentz M, et al. Long-term results of related myeloablative stem-cell transplantation to cure sickle cell disease. *Blood*. 2007;110(7):2749-2756.
21. Vermylen C, Cornu G, Ferster A, et al. Haematopoietic stem cell transplantation for sickle cell anaemia: the first 50 patients transplanted in Belgium. *Bone Marrow Transplant*. 1998;22(1):1-6.
22. Gluckman E, Cappelli B, Bernaudin F, et al. Sickle cell disease: an international survey of results of HLA-identical sibling hematopoietic stem cell transplantation. *Blood*. 2017;129(11):1548-1556.
23. Lopez AD, Williams TN, Levin A, et al. Remembering the forgotten non-communicable diseases. *BMC Med*. 2014;12(1):1-19.
24. Williams TN. Sickle cell disease in sub-Saharan Africa. *Hematol Oncol Clinics*. 2016;30(2):343-358.
25. Olubiyi OO, Olagunju MO, Strodel B. Rational drug Design of Peptide-Based Therapies for sickle cell disease. *Molecules*. 2019;24(24):4551.
26. Harrington DJ, Adachi K, Royer WE Jr. The high resolution crystal structure of deoxyhemoglobin S. *J Mol Biol*. 1997;272(3):398-407.
27. Adachi K, Asakura T. Demonstration of a delay time during aggregation of diluted solutions of deoxyhemoglobin S and hemoglobin CHarlem in concentrated phosphate buffer. *J Biol Chem*. 1978;253(19):6641-6643.
28. Adachi K, Asakura T. Nucleation-controlled aggregation of deoxyhemoglobin S. possible difference in the size of nuclei in different phosphate concentrations. *J Biol Chem*. 1979;254(16):7765-7771.
29. Adachi K, Asakura T. Kinetics of the polymerization of hemoglobin in high and low phosphate buffers. *Blood Cells*. 1982;8(2):213-224.
30. Ferrone FA, Hofrichter J, Eaton WA. Kinetics of sickle hemoglobin polymerization. I. Studies using temperature-jump and laser photolysis techniques. *J Mol Biol*. 1985;183(4):591-610.
31. Ferrone FA, Hofrichter J, Eaton WA. Kinetics of sickle hemoglobin polymerization. II. A double nucleation mechanism. *J Mol Biol*. 1985;183(4):611-631.
32. Mirchev R, Ferrone FA. The structural link between polymerization and sickle cell disease. *J Mol Biol*. 1997;265:475-479.
33. Perutz MF, Fermi G, Abraham DJ, Poyart C, Bursaux E. Hemoglobin as a receptor of drugs and peptides: x-ray studies of the stereochemistry of binding. *J Am Chem Soc*. 1986;108(5):1064-1078.
34. Prabhakaran M, Johnson ME. Molecular dynamics of sickle and normal hemoglobins. *Biopolymers*. 1993;33(5):735-742.
35. Galamba N. On the nonaggregation of Normal adult hemoglobin and the aggregation of sickle cell hemoglobin. *J Phys Chem B*. 2019;123(50):10735-10745.
36. Galamba N, Pipolo S. On the binding free energy and molecular origin of sickle cell hemoglobin aggregation. *J Phys Chem B*. 2018;122(30):7475-7483.
37. Ghatge MS, Ahmed MH, Omar ASM, et al. Crystal structure of carbonmonoxy sickle hemoglobin in R-state conformation. *J Struct Biol*. 2016;194(3):446-450.
38. Kuczera K, Gao J, Tidor B, Karplus M. Free energy of sickling: a simulation analysis. *Proc Natl Acad Sci U S A*. 1990;87(21):8481-8485.
39. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput*. 2008;4(5):819-834.
40. Fermi G, Perutz MF, Shaanan B, Fourme R. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J Mol Biol*. 1984;175(2):159-174.
41. MacKerell AD Jr, Feig M, Brooks CL. Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc*. 2004;126(3):698-699.
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-935.
43. Hoover WG. Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A*. 1985;31(3):1695-1697.
44. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*. 2005;7182(1981):7182-7190.
45. Darden T, York D, Pedersen L. Particle mesh Ewald: an N×log (N) method for Ewald sums in large systems. *J Chem Phys*. 1993;98(12):10089-10092.
46. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. *J Chem Phys*. 1997;18(12):1463-1472.
47. De Jong DH, Liguori N, Van Den Berg T, Arnarez C, Periole X, Marrink SJ. Atomistic and coarse grain topologies for the cofactors associated with the photosystem II core complex. *J Phys Chem B*. 2015;119(25):7791-7803.
48. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007;126(1):014101.
49. Periole X, Cavalli M, Marrink SJ, Ceruso MA. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput*. 2009;5(9):2531-2543.
50. Wassenaar TA, Pluhackova K, Böckmann RA, Marrink SJ, Tieleman DP. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J Chem Theory Comput*. 2014;10(2):676-690.

51. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19-25.
52. Michaud-Agrawal N, Denning E, Woolf T, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011;32:2319-2327.
53. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-38.
54. Kumari R, Kumar R, Lynn A. g_mmpbsa–a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Comp Sci*. 2014;54(7):1951-1962.
55. Saito M, Okazaki I. A 45-ns molecular dynamics simulation of hemoglobin in water by vectorizing and parallelizing COSMOS90 on the earth simulator: dynamics of tertiary and quaternary structures. *J Comput Chem*. 2007;28(6):1129-1136.
56. Abroshan H, Akbarzadeh H, Parsafar GA. Molecular dynamics simulation and MM–PBSA calculations of sickle cell hemoglobin in dimer form with Val, Trp, or Phe at the lateral contact. *J Phys Org Chem*. 2010;23:866-877.
57. Artmann GM, Burns LE, Canaves JM, et al. Circular dichroism spectra of human hemoglobin reveal a reversible structural transition at body temperature. *Eur Biophys J*. 2004;33:490-496.
58. Padlan E, Love W. Refined crystal structure of deoxyhemoglobin S. II. Molecular interactions in the crystal. *J Biol Chem*. 1985;260:8280-8291.
59. Wang Y, Ferrone FA. Dissecting the energies that stabilize sickle hemoglobin polymers. *Biophys J*. 2013;105:2149-2156.
60. Ross PD, Hofrichter J, Eaton WA. Thermodynamics of gelation of sickle cell Deoxyhemoglobin. *J Mol Biol*. 1977;115:111-134.
61. Loschwitz J, Jäckering A, Keutmann M, et al. Novel inhibitors of the main protease enzyme of SARS-CoV-2 identified via molecular dynamics simulation-guided in vitro assay. *Bioorg Chem*. 2021;111:104862.
62. Chen K, Ballas SK, Hantgan RR, Kim-Shapiro DB. Aggregation of Normal and sickle hemoglobin in high concentration phosphate buffer. *Biophys J*. 2004;87:4113-4121.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Olagunju MO, Loschwitz J, Olubiyi OO, Strodel B. Multiscale MD simulations of wild-type and sickle hemoglobin aggregation. *Proteins*. 2022;90(11):1811-1824. doi:10.1002/prot.26352