Understanding the aggregation of sickle hemoglobin (HbS) and identification of HbS aggregation inhibitors

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Maryam Olayemi Olagunju
from Lagos, Nigeria

Jülich, April 2022

from the institute of Theoretical and Computational Chemistry
at the Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof.Dr. Birgit Strodel
Co-supervisor: Dr. Oliver Weingart

Date of the oral examination: 30.05.2022

Statement of Author

I, Maryam Olayemi Olagunju, hereby certify that I wrote the thesis titled "Understanding the aggregation of sickle hemoglobin (HbS) and identification of HbS aggregation inhibitors", without any unauthorized external assistance and used only sources acknowledged in the work. All texts excerpted verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. My thesis contains no material published elsewhere or extracted in whole or part from a thesis submitted for a degree at this or any other university. Where the results are produced in collaboration with others, my contributions are clearly stated.

Acknowledgment

First, I would like to express my gratitude to Prof.Dr.Birgit Strodel who is my supervisor, for funding my doctoral research and giving me the opportunity to work on my doctoral research under her supervision. She was always open to scientific discussions and providing me with useful advice throughout my Ph.D. studies. I am particularly grateful for her support during the challenging times of my PhD. Thank you Birgit!

I also appreciate and acknowledge Prof.Olujide Olubiyi for his supervision, scientific contributions and for always willing to answer all my questions throughout my doctoral training and for always believing in me. My thanks also goes to Dr.Bogdan Barz for his contributions at the beginning of this project.

This project was funded by the German academic exchange (DAAD) and the Forchungszentrum Jülich which I want to thank. Special thanks to the RWTH Aachen and Heinrich Heine University, Düsseldorf for providing computational resources used for this project.

I thank my fellow group members especially Jennifer for always willing to share her bash scripts with me and explain complex problems to me, Hebah, Mohammed and Suman for their friendship and support, Anna for always helping me with bureaucratic matters and translating my letters from German to English, Wibke, Lara and Feng. All of you have been so kind to me and I'm glad that we crossed paths.

I would also like to thank all the Nigerian friends I met in Jülich and Aachen. You made my time unforgettable.

I would like to thank my darling friends, Olubunmi Obagaye, Genevieve Edobor, Efe Abighe, Damilola Watti, Fisayo Watti, Molakun Bankole, Abiona Tade, Sharon Abraham and Abdallah Mohammed.

I appreciate my Mum and Dad, Mr and Mrs. Olagunju, my beautiful siblings, Olagunju Olayinka, Olagunju Boluwatife and Olagunju Babatunde for their continued financial, emotional, mental support and prayers. Thank you all for constantly encouraging me towards achieving my goals and believing in me. I love you all!

I dedicate this thesis to the memory of my dear friend, Olubunmi Obagaye. I love you!

Above all, I am grateful to God for His mercies and goodness to me. Thank You!

Author Contributions

Chapter 3.1

Article: Multiscale MD simulations of wild-type and sickle hemoglobin aggregation.

Authors: Maryam Olagunju, Jennifer Loschwitz, Olujide Olubiyi & Birgit Strodel

Contributions: **Maryam Olagunju**: investigation, methodology, data analysis, visualization and writing - Original Draft, **Jennifer Loschwitz**: data analysis, **Olujide Olubiyi**: supervision, writing - review & editing, **Birgit Strodel**: supervision, research design, visualization and writing - review & editing.

Chapter 3.2

Article: Computer-aided drug design-directed experimental identification of novel inhibitors of sickle hemoglobin polymerization.

Authors: Maryam Olagunju, Olujide Olubiyi & Birgit Strodel

Contributions: research design, investigation, methodology, data analysis, visualization and writing - Original Draft, **Olujide Olubiyi**: research design, supervision, writing - review & editing, **Birgit Strodel**: supervision, research design, writing - review & editing.

List of Publications

Published Manuscripts

- Olubiyi, Olujide O., **Olagunju, Maryam. O.**, Birgit Strodel (2019). Rational Drug Design of Peptide-Based Therapies for Sickle Cell Disease. **Molecules, 24(24), 4551.**
- Olubiyi, O. O., **Olagunju, M.**, Keutmann, M., Loschwitz, J., and Strodel, B. (2020). High throughput virtual screening to discover inhibitors of the main protease of the coronavirus SARS-CoV-2. **Molecules, 25(14), 3193.**
- Loschwitz, J., Jäckering, A., Keutmann, **M., Olagunju,** M., Eberle, R. J., Coronado, M. A., Olubiyi, O.O. and Strodel, B. (2021). Novel inhibitors of the main protease enzyme of SARS-CoV-2 identified via molecular dynamics simulation-guided in vitro assay. **Bioorganic Chemistry, 111, 104862.**
- Loschwitz, J., Jäckering, A., Keutmann, M., **Olagunju, M.,** Olubiyi, O. O., and Strodel, B. (2021). Dataset of AMBER force field parameters of drugs, natural products and steroids for simulations using GROMACS. **Data in Brief, 35, 106948.**

Manuscript Accepted for Publication

- **Maryam O. Olagunju.,** Jennifer Loschwitz, Olujide O. Olubiyi and Birgit Strodel. Multiscale MD simulations of wild-type and sickle hemoglobin aggregation. **Proteins: Structure, Function, and Bioinformatics. (accepted on 20.04.2022)**

Manuscript under Preparation

- **Olagunju Maryam**, Olujide Olubiyi and Birgit Strodel. Computer-aided drug design-directed experimental identification of novel inhibitors of sickle hemoglobin polymerization.

Abstract

Sickle cell disease commonly referred to as (SCD) is a blood disorder that is caused by the replacement of glutamic acid (E) with valine (V) at the sixth position of the β-globin chain of hemoglobin. This in turn leads to the aggregation of hemoglobin, this aggregation results in clinically observed symptoms known as "sickle cell crisis". Hemoglobin sickle (HbS) is a variant of hemoglobin found in people with SCD. SCD is well investigated, but until recently, only three medications were approved by the drug authorities for the management of SCD and they do not fully address the underlying cause of the disease. Voxelotor, which is an HbS aggregation inhibitor, was recently approved by the FDA but it is highly expensive. Another alternative to the treatment of SCD is the highly expensive bone marrow transplant. Most of these treatments are not readily available in developing nations where the highest number of SCD patients are found. Therefore, there is an urgent need to develop cheap, effective and readily available drugs for the treatment of SCD. In this thesis, using computer aided drug design methods (CADD) we aim to understand the aggregation of HbS caused by the E6V mutation, identify protein-protein interaction hot-spots that can be prioritized in aggregation inhibitor designs and finally identify prospective HbS aggregation inhibitors. First, using multiscale molecular dynamic simulations, the conformational dynamics of both wild-type and sickle hemoglobin at both monomeric and dimeric level were elucidated to assess their stability and highlight the effect of the E6V mutation on each structure. Next, we studied their aggregation into decamers and analyzed the protein-protein interactions of the aggregates in details. The conclusions from these investigations revealed that the β-globin chains are less flexible in HbS than in HbA and the aggregation of HbS is not only driven by protein-protein interactions that are hydrophobic in nature but also electrostatic interactions are also important. Protein-protein contactS specific to HbS were identified in the first phase of the project, and these contacts were further exploited in the next phase to design inhibitors of HbS-aggregation. Using the knowledge obtained from the initial simulations, we performed high throughput virtual screening, using a library of of compounds including approved drugs, investigational drugs, natural products, and D-enantiomeric peptides followed by MD simulations in search of compounds that can bind to HbS and thereby inhibit its

aggregation. From this investigation, we identified 16 promising organic molecules and 7 D-enantiomeric peptides. The organic molecules identified computationally, will be tested experimentally using both cell based *in vitro* assays to assess their HbS inhibitory properties.

Contents

Chapter 1

Introduction

1.1 Declaration

This chapter was excerpted in parts from a review published in Molecules 24(24), 4551 (2019) by the authors Olujide O. Olubiyi, Maryam O. Olagunju, Birgit Strodel: Rational Drug Design of Peptide-Based Therapies for Sickle Cell Disease. All the images are reprinted with Copyright © 2019 by the authors.

1.2 Sickle Cell Disease: Cause, Management, Geographic Distribution and Diagnosis

Sickle cell disease (SCD), is a result of a point mutation involving the replacement of glutamic acid at position 6 of the β-globin chain of hemoglobin to valine that leads to the polymerization of hemoglobin [1]. In manifestation, SCD represents a symptom complex that involves dehydration of the Glu6 to Val6 mutated hemoglobin, which is called sickle hemoglobin or hemoglobin S (HbS), and elevated 2,3-diphosphoglycerate (2,3-DPG) levels whose interaction with hemoglobin reduces HbS solubility and promotes polymerization, also called sickling [2, 3]. This ultimately leads to hampered O_2 binding and transport,impaired erythrocyte morphology and interaction with endothelial surfaces [4, 5], premature erythrocyte rupture and anemia, painful vaso-occlusive crisis, a general poor health, and, in many cases, death [6–10].

Geographic Distribution: SCD is the most common genetic disease worldwide. Of the millions of people with SCD, more than 75% are believed to live in Nigeria, Democratic Republic of Congo, and India (Figure 1.1) [4, 11, 12]. These countries are additionally responsible for about 80% of global newborns having the causative Glu6 to Val6 mutation [13]. It is estimated that at least 150,000 babies are born with SCD annually in Nigeria. Estimating accurately the total number

of newborns born with SCD in Nigeria is quite arduous due to the lack of federal new-born screening programs [14] The geographic distribution of SCD is similar to the spread of malaria; countries where SCD is very common is associated with high rates of malaria [15, 16] because the sickle gene in the heterozygote form, protects against death from prevalent malaria infections caused by *Plasmodium falciparum* [17, 18]. Medical indicators of SCD vary among these locations, with individuals from India and the Arabian Peninsula and Senegal suffering from less severe diseases than those from other parts of Africa. This pattern suggests that there is a substantial disparity in medical manifestations between certain populations that may be due to genetic variation. Due to immigration, there is also an existence of SCD in Europe, North, South and Central America, and they represent primarily the more severe African types [18].

Figure 1.1: Countries with the highest incidence of sickle cell disease. Nearly 90% of individuals with SCD live in these three countries marked in red [14]. Copyright © 2022 UpToDate, Inc. and/or its affiliates. All Rights Reserved

.

Diagnosis: Detection of SCD early increases the chances of survival. In the United States, Jamaica and European countries, it has been proven that early detection of sickle cell diseases during newborn screenings have helped in the reduction of the mortality rates. However, in most sub-Saharan African countries, there are only a few centres where new born screening is carried out and offer comprehensive health care at an early age because of the high cost associated with these programs [19, 20]. Other patients are diagnosed only when they present with a complication of the disease [14].

Available diagnostic tests: Patients who are suspected of having SCD based on their clinical symptoms are diagnosed using hemoglobin electrophoresis. Depend-

ing on the availability, High Performance Liquid Chromatography can also be used for the diagnosis of SCD. Point-of-care (POC) diagnostic testing for SCD is under development; as soon as they are available, these tests could be helpful in places where it is difficult or impossible to transfer blood samples to a central laboratory [14].

Newborn screening: SCD is better treated if diagnosed early in life. In addition, early diagnosis provides an opportunity to start well-established prevention measures, such as prophylactic penicillin, routine childhood vaccination, education of parents and caregivers about rapid medical management of fever, and detection of splenic sequestration. In the United States, newborn screening for SCD is routinely performed in every state since these strategies demonstrated benefits to reducing morbidity and improving outcomes [14].

Childhood presentation: In sub-Saharan Africa, most SCD patients are not diagnosed by newborn screening and display symptoms during childhood, at an average age of two years. Very few cases are diagnosed earlier (e.g., in infancy). Previously undiagnosed SCD commonly presents as dactylitis in the first and splenic sequestration in the second year of life. Therefore, it is recommended that children who exhibit either dactylitis or splenic sequestration should have a complete blood count and hemoglobin analysis performed to diagnose SCD [14].

Disease management and treatment options: Despite growing understanding of the polymerization of HbS and its effects on red blood cells (RBCs), until very recently, only two drugs, hydroxyurea and L-glutamine were approved by the United States (US) Food and Drug Administration (FDA) for the management of SCD [21]. Hydroxyurea is the most widely employed drug treatment of sickle cell anemia in different age groups [22–27]. While its clinically observed efficacy has been attributed to different effects at the cellular level [28], the most important mechanism of action relates to its ability to induce the production of fetal hemoglobin (HbF), which does not polymerize, and to increase the total concentration of hemoglobin [29, 30]. Hydroxyurea remains a viable treatment option for SCD, and the concern of toxicities associated with its administration has largely been limited to side effects that resolve with medication discontinuation [30–35]. There have, however, been certain reports of associated malignancies [36–41], but further investigations are needed to categorically confirm these [42].

L-glutamine is the second approved drug treatment [21, 43]. While its mechanism of action is not known, and only suggested to involve a reduction of oxidative stress via elevation of the levels of reduced glutathione [44, 45], it is clear that it has no effect on hemoglobin S aggregation and hemoglobin production [46–50]. Another option for the treatment of SCD is hemopoietic stem transplantation, but its general applicability is limited by technical and cost considerations, and thus, out of the reach of

SCD sufferers in third-world countries [51–56] A number of research attempts have been made to design interventions aimed at modulating the structural properties, aggregation tendencies, and defective O_2 transport properties of sickle hemoglobin. For example, allosteric modulators and covalent modifiers of HbS that stabilize the non-polymer forming R-state Hb conformation have been reported and include the recently FDA approved voxelotor (GBT 440) [57] and derivatives of vanillin [58, 59].

In 2019, an oral polymerization inhibitor, Voxelotor, a first-in-class therapy was approved for treatment of SCD in patients 12 years and older by the FDA [60]. In 2021, it was later approved for a broader use in patients from 4 years of age and older. The mechanism of action of Voxelotor involves increasing the affinity of HbS for oxygen, thereby preventing HbS aggregation [57, 61]. Compounds like senicapoc, a Gardos channel blocker, were also reported with the ability to prevent RBC dehydration [62]; clinical assessment in SCD, however, failed to find a correlation between improvements in hemolysis and vaso-occlusive crisis [63]. Selective inhibition of phosphodiesterase 9A by IMR-687 was recently reported to reduce both sickling and vaso-occlusion, which is believed to result from the induction of cGMP (cyclic guanosine monophosphate) and HbF [64]. Compounds which directly interact with HbS and disrupt the intermolecular contacts crucial to HbS polymerization have also been investigated, and they include small organic compounds [59], amino acid-based compounds, as well as herbal preparations (e.g., Nix-0699 [65, 66]). Other drug discovery efforts have focused on biochemical processes downstream of HbS polymerization rather than seeking to explore specific peculiarities of the aggregation process. A recent review by Eaton and Bunn argued in favor of research attention directed at the HbS polymerization process, especially because the aggregation kinetics as well as the circulatory transit time make it possible to achieve clinical improvement with only a small fraction of HbS aggregation inhibited [67, 68].

1.3 Hemoglobins: Structure, Function, and Aggregation

The function of the red blood cells and their hemoglobin is to carry oxygen (O_2) from the lungs to all the body tissues and to carry carbon dioxide (CO_2) back to the lungs. This function is enabled by the structural characteristics of hemoglobin (Hb), allowing it to bind O_2 and CO_2. Both HbA, which refers to the wild-type hemoglobin present in individuals without sickling disorder, and HbS exist as tetramers consisting of two α subunits and two β subunits arranged into a pseudotetrahedral symmetry (Figure 1.2A). With the two 141-residue α-globin chains and the two

146-residue β-globin chains, and each globin chain carrying one heme group, the full HbA/HbS assembly contains 574 amino acids and four heme molecules.It is from these four heme molecules and the four globin chains that hemoglobin derives its name.

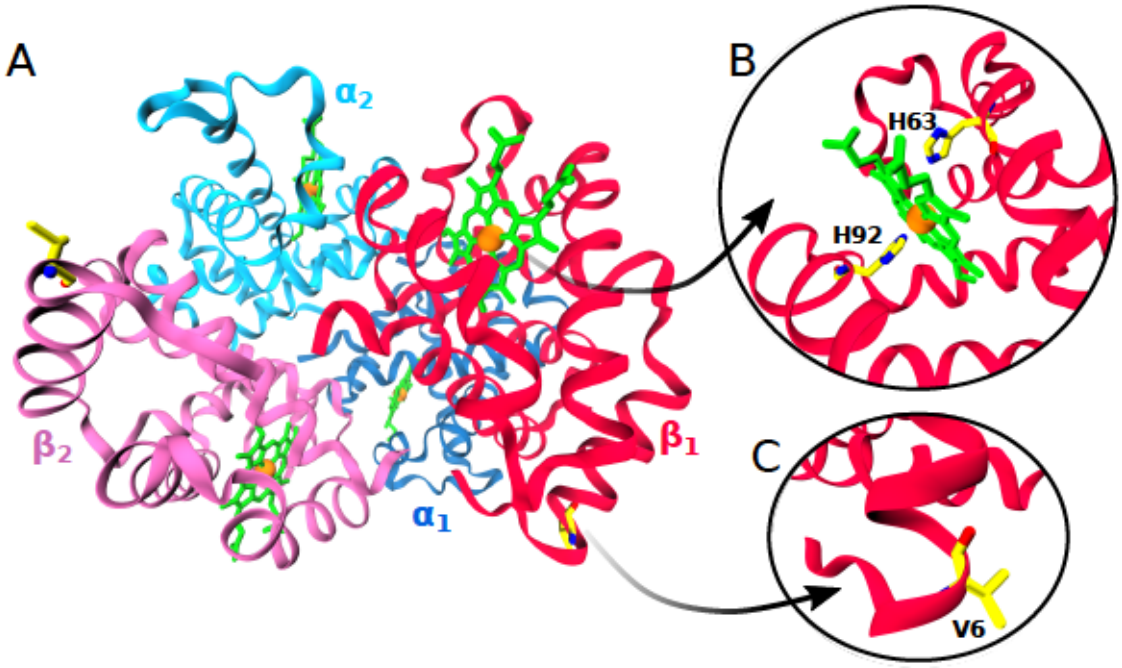

Figure 1.2: (A) The quaternary structure of HbS consisting of two α subunits (here denoted α_1 and α_2 for ease of distinction, shown in shades of blue) and two β (β_1 and β_2, shades of red) subunits. Each globin subunit carries one heme (green), including an Fe^{2+} ion (orange). (B) The hemes are linked to the globin by covalent bonds between their irons and 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 was produced using PDB entry 5E6E [69]

The quaternary structure of hemoglobin is maintained by relatively weak but precisely coordinated non-covalent interaction forces, including van der Waals interactions, hydrogen bonds, and salt bridges between the different globin chains.In total, there are 30 helices in the hemoglobin structure: The two α-globin chains feature a total of 14 helices between them, while the β-globins have 16 helices. Each globin chain is covalently linked to a heme molecule via their proximal histidine residue (His87 in the α-globin chains and His92 in the β-globin chains). The heme, in turn, consists of a protoporphyrin part and a centrally coordinated iron ion (Figure 1.2B). The local environment of the globin molecules maintains the coordinated iron ion in its reduced form, in which state it can form a total of six bonds. Four of the six coordination sites of the ferrous ion are covalently bonded to the protoporphyrin ring, another to the imidazole side chain of the histidine residues, while the sixth coordination site allows for binding and unbinding of dissolved gases (Figure 1.3A). It is this last coordination site that is responsible for O_2 binding. Following Fe^{2+} binding, bound oxygen establishes hydrogen bonding with the imidazole side chain of His58 in the α-globins, and His63 of the β-globins, the distal histidine. In this state, heme adopts a relaxed, conformationally unstrained arrangement structurally representing the "R" conformation and functionally the oxygenated hemoglobin [69].

In the deoxygenated form, the distal histidine side chains have a propensity to swing out of the heme pocket, thus allowing a compression of the surrounding helices with respect to each other, which in turn causes Fe^{2+} to move out of the porphyrin plane [70]. This gives rise to a tensed conformation ("T" conformation) with the heme adopting a dome-like arrangement (Figure 1.3B).This structural change precipitates a series of further changes in the remaining body of the HbS protein, which, under deoxygenation and dehydration conditions, provokes a pathologic cascade that ultimately leads to clinical manifestations.

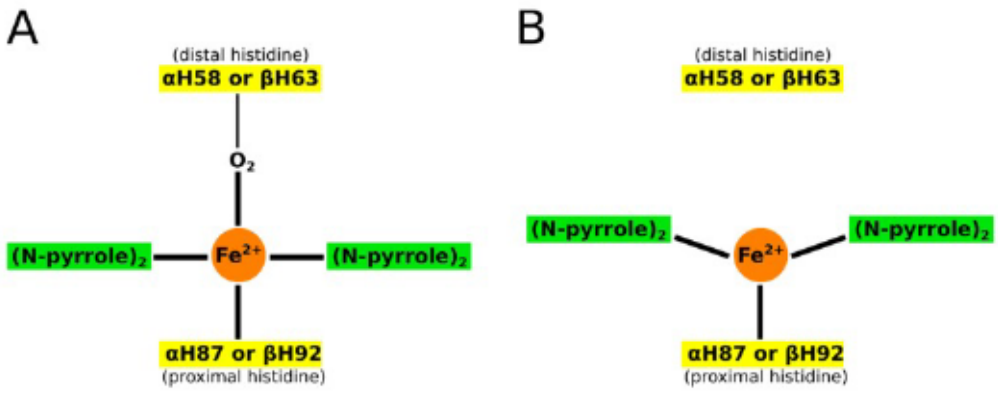

Figure 1.3: Schematic representation of the main structural differences between the (A) R and (B) T conformations of hemoglobin.

It deserves noting that the Glu6Val mutation involves an amino acid replacement

at the HbS surface (Figure 1.2C) and, as such, only affects protein–protein interactions involving surface residues [71–84], without any effect on amino acids located at the core [78, 82]. The side chain of Val6 in the β-globin structure (the donor β-globin) of HbS forms a hydrophobic key, which fits well into an essentially hydrophobic cavity formed by Ala70, Phe85 and Leu88 of the β-globin of an adjacent HbS molecule (Figure 1.4). It should be noted that both HbA and HbS form linear aggregates involving the formation of axial contacts between Hb molecules. Only in the case of HbS, these linear aggregates grow into double filaments, facilitated by lateral βVal6-β'Ala70/β'Phe85/β'Leu88 contacts (where the prime indicates that Ala70, Phe85 and Leu88 belong to another hemoglobin than Val6). The double filaments further assemble into ≈200 Å thick fibers, which eventually accumulate in highly complex, pathological HbS fiber networks [85]. These aggregates affect the functionality of the red blood cell by destroying their structural pliability into stiffened and deformed erythrocytes. Differences at the cellular level, for instance, originating from different degrees of cellular dehydration or oxidative stress, may further complicate the HbS polymerization, such that each patient's clinical manifestations are, to some extent, unique [86–88].

1.4 HbS as a Target for Drug Design

1.4.1 HbS aggregation is an Inefficient Process

Efforts to rationally design antisickling agents have often viewed the sickle hemoglobin both as the drug target as well as the starting point for lead discovery. Such efforts are indeed not new; the 1970s through the 1980s witnessed a good deal of research interest into the molecular nature of the HbS molecule, as well as the search for compounds capable of disrupting its polymerization. A prevailing doubt about the suitability of the HbS molecule as target for drug development has to do with the perceived limitation imposed by its high content level in man (about 450g) [67], suggesting that an intolerably high dose of antisickling compound would be required to achieve clinically useful degrees of inhibition [91]. This perception was mostly based on an aggregation model built on the assumption of a highly efficient nucleation dependent HbS polymerization process believed to involve two nucleation stages, beginning with a rate-limiting homogeneous nucleation, followed by a highly efficient heterogeneous nucleation phase [92, 93]. For aggregation to occur, the delay time associated with the homogeneous nucleation should necessarily be shorter than re-oxygenation circulation time, which is the time required for the hemoglobin to pass through the blood vessels and be re-oxygenated [94]. In light of recent find-

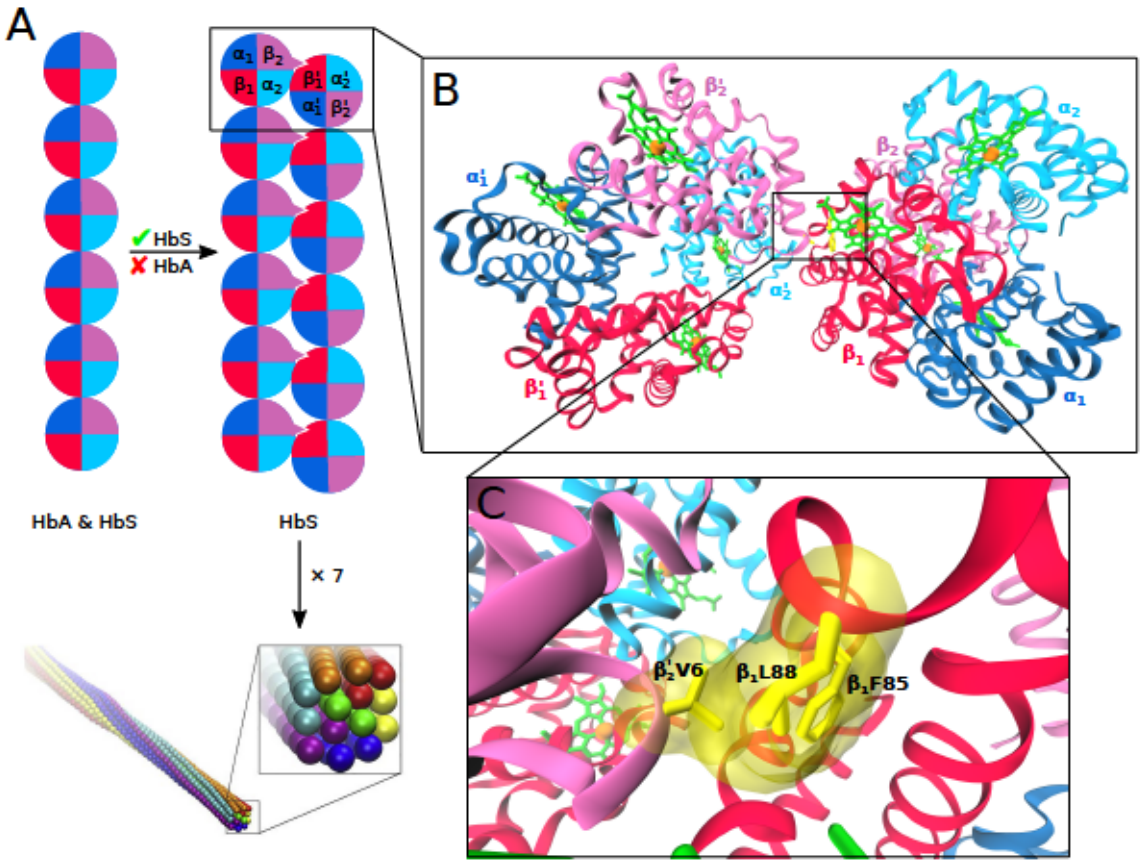

Figure 1.4: (A) Schematic representation of how the Glu6Val mutation modifies normal hemoglobin polymerization of HbS heterotetramers, involving linear Hb aggregates formed by both HbA and HbS (left) into double HbS filaments (right). The hemoglobin tetramer is represented as a circle, such that one quarter corresponds to one protein subunit using the same coloring as in Figure 1.2. The βGlu6Val 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, reproduced with permission from reference [89]). (B) A dimer formed by two HbS aggregates is shown. (C) This aggregation is mediated by β_2Val6 interacting with the hydrophobic pocket formed by β'_1Phe85 and β'_1Leu88. 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 from PDB entry 2HBS [90]. The figure was reproduced and reprinted with permission from Olubiyi et al., 2019 Copyright ©2019).

ings [95, 96], there is increasing need to revisit what is accepted with respect to HbS polymerization kinetics. In a recent study employing high resolution differential interference contrast (DIC) microscopy (55 nm resolution at 1 Hz, the highest resolution currently available for HbS aggregation kinetics), monomer incorporation into HbS polymers was found to be a highly inefficient process, with only 30,000 out of one million HbS monomers incorporated per second [96]. This translates to a 3% efficiency for HbS polymerization as against the previously reported monomer

incorporation efficiency of more than 95% [97, 98]. This observation is supported by the finding of Wang and Ferrone, who, based on light scattering experiments, revealed that the overall thermodynamics into double filaments (Figure 1.4A) is marginally unfavorable, with the axial contacts being 1.8 kcal/mol weaker than the lateral contacts [99]. At such a low polymerization efficiency, HbS monomer binding and unbinding events are only marginally in favor of polymer growth, such that small disturbances (for instance, resulting from inhibitor binding) are sufficient to push the equilibrium towards polymer disassembly. Castle et al. calculated the magnitude of binding disturbance required and estimated it to be a 1.2 kcal/mol change in HbS monomerpolymer interaction in 5% of the available HbS molecules that is required to halt the polymerization process (see reference [96] for the calculation). This agrees qualitatively with the earlier estimated ≈1.5 kcal/mol hydrophobic free energy contribution resulting from Val6 binding within the Phe85/Leu88 pocket [100]. With about 30 picogram (pg) of hemoglobin per RBC [101, 102], disruption of polymerization in less than 1.5 pg HbS per cell should in principle be sufficient to frustrate aggregation, especially considering that only between 40 and 60% of the RBCs typically undergo sickling [103]. This reasoning does not only bring HbS polymerization within the purview of non-covalent inhibition, but it also rationalizes why antisickling effects have been observed for various small molecular weight inhibitors [104–106]. For instance, screening for non-covalent antisickling agents that reverse HbS polymerization by altering RBC shape and volume (towards more spherical structures with larger volumes) discovered antisickling properties for gramicidin A and monensin A at concentrations of 200 pM and 2 nm, respectively [107]. Another example is the aggregation inhibition by HbF, which is required to be present in a just a little fraction (0.2) of total hemoglobin of SCD patients to achieve clinical resolution of symptoms [108, 109].This antisickling effect of HbF serves as the mechanistic basis for SCD treatment with HbF-inducing hydroxyurea. Like HbF, addition of HbA to polymerizing HbS has also been shown to inhibit HbS aggregation [110].

1.4.2 Antisickling Effect and HbS Conformation

Targeting sickle hemoglobin for inhibitor design does not only aim to directly inhibit its aggregation into multi-stranded polymers, but also includes approaches that either result in the stabilization of the R conformation of the HbS molecule, or the destabilization of the T conformer [111, 112]. Compounds whose antisickling properties are based on this concept include vanillin and pyridyl derivatives of vanillin, 5-hydroxymethylfurfural (5-HMF), and the recently approved voxelotor (GBT440) [58, 105, 112–116]. They bind to the N-terminal valine (and possibly lysine) residues of the α-globin chains of HbS (Figure 4) [104], forming a reversible

Schiff-base adduct which stabilizes the R-state and/or destabilizes the T-state, increasing hemoglobin solubility, and thus inhibiting HbS aggregation. Iqbal et al. employed an electrochemistry-based technique to investigate HbS polymerization in the presence of vanillin and 5-HMF [92]. At HbS concentrations of 100 mg/mL, aggregation inhibition was obtained for vanillin concentrations corresponding to 0.5:1, 1:1, and 10:1 mole ratios relative to HbS. A similar pattern was obtained for 5-HMF, except for an interesting observation that the 0.5:1 inhibitor/HbS ratio was found to slightly promote aggregation. At 1:1 inhibitor/HbS concentration, both compounds achieved roughly 70% aggregation inhibition, while a near perfect inhibition was recorded when the inhibitor concentration was increased to achieve a 10:1 mole ratio relative to the hemoglobin. In scanning the inhibitors against HbS, Iqbal et al. employed an HbS concentration that is about three orders of magnitude smaller than the intracellular concentration of hemoglobin, which is 334 mg/mL assuming an RBC volume of 90 fl and mean corpuscular hemoglobin of 30 pg. At such higher cellular content of hemoglobin, a more efficient system of inhibition is probably needed. Thus, continuing searches for antisickling agents is warranted, independent of the successful progression of GBT440 through phase III clinical trial leading to its recent FDA approval.

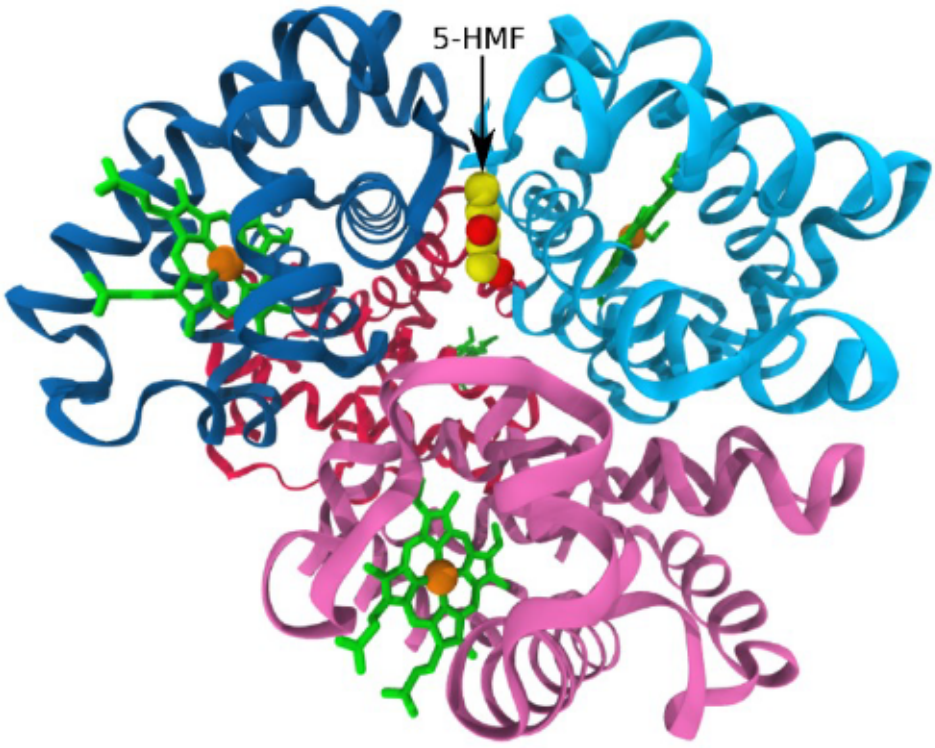

Figure 1.5: Binding of 5-hydroxymethylfurfural (5-HMF; yellow/red) in the α-cleft of HbS via hydrogen bonds and hydrophobic interactions formed with both α-globins, stabilizing the R-state conformation. The coloring scheme from Figure 1.2 is used for HbS. The figure was produced from PDB entry 5URC [104]

1.4.3 Interprotein Contacts during HbS Aggregation

In the quest to target HbS to directly disrupt polymerization as therapeutic approach, one should consider that this may be more challenging than it first seems because of the plethora of multiple binding sites that, when interfered with, may influence the conformational preferences of HbS that favor or disfavor polymerization. There exists a good number of data suggesting that both intra- and interpeptide contacts sponsor the polymerization process of HbS, which involves interactions at multiple sites on the hemoglobin molecule. Without doubt, the aberrant valine residue at position 6 of the β-globin is involved, believed to be in immediate contact with β'Phe85 and β'Leu88 (Figure 1.4). It is thought that concurrently, to this contact, a hydrogen bond between βThr4 and β'Asp73 is formed due to the spatial proximity between these residues. In addition to these primary contacts, secondary contacts, which involve hydrophobic and also a number of ionic interactions [90, 94, 117–120], have been identified and proposed to either influence directly the polymerization process, modulate the conformational equilibrium between the R and T state, or simply modify the solubility of deoxygenated HbS. For example, the αAsn78$\rightarrow$Lys mutation leads to an increase in the solubility of deoxy-HbS, alleviating the severity of SCD [120, 121]. Another challenge for the design of antisickling agents aimed at disrupting the aggregation process is a common problem when targeting proteinprotein interactions, because these interaction sites are typically flat and large, quite different from the "grooves" or pockets in which small molecules typically bind.

1.5 Aims and Objectives of the Study

The current understanding of SCD pathology suggests that effective management depends on the availability of therapeutic methods that can prevent and treat the hemorheological factors contributing to the development of crises that are common amongst SCD patients. Our present limited knowledge of the detailed mechanism of hemoglobin aggregation has resulted partly from the general reliance of national clinical guidelines on medicine focusing largely on the reduction of painful crises and improvement of hemotological functions. However, by developing a treatment that exploits the static factors and dynamic forces driving the protein-protein interactions in HbS, such treatment will be able to both prevent crises development and also reverse polymerization of HbS. Although, there are some herbal preparations that are capable of inducing the reversal of erythrocyte sickle formation at the cellular level, the recently approved voxelotor which is an orally administered medication is also capable of inhibiting HbS polymerization. Ultimately, the aim here is to understand the molecular mechanism underlying the aggregation of HbS. An understanding of the mechanism is very crucial to the development of cheap,readily available and effective drugs, especially for SCD sufferers living in developing countries. In summary, there is a need to understand the molecular factors that drive the aggregation of HbS and to identify compounds that have HbS aggregation inhibitory properties. Here in this thesis, using Molecular Dynamics (MD) simulations, we examine the structural and conformational basis for HbS aggregation, most importantly the role that the Glu6Val mutation plays in the aggregation process. Furthermore, using computer aided drug discovery methods that combine both high throughput virtual screening with MD simulations we aim to identify compounds that are capable of binding to HbS and also possess HbS aggregation inhibitory properties. The identified compounds were further validated by performing anti-sickling assays. The objectives of this thesis include:

1. Identifying protein-protein interaction hotspots to be prioritized in the search to develop HbS inhibitors that prevent from HbS aggregating into fibers for SCD therapy.

2. Screen libraries of small molecules e.g. organic molecules, natural products and D-enantiomeric peptides for their aggregation inhibitory properties with validation provided by *in vitro* experiments.

This thesis is presented in the following order: In chapter 2, the methods employed in this work are discussed in detail. In chapter 3, the results are presented. Finally, in chapter 4 the summary of all the results are provided and overall conclusions are drawn.

Chapter 2

Methods

2.1 Molecular Dynamics (MD) Simulations

In recent years, computational methods have become more relevant in the field of life sciences due to the intricacies involved with studying biological systems. With the invention of faster and more powerful computers, complex bio-molecular systems can be explored using computational simulations and modelling [122]. Dynamics simulation methods are widely used to obtain information about the conformations of biological macro-molecules e.g. proteins, nucleic acids, lipids etc. as well as the related kinetic and thermodynamic data [123–126]. They provide a bridge between experiments and theory, thus enhancing conventional experiments [127]. Through computer simulations, the motions of individual particles as a function of time can be precisely studied, additionally it may provide insights into mechanisms and processes that are not directly accessible through experiments [122, 127]. Computer simulations have a wide range of applications, which includes discovery and design of new drugs. Using a computer model to study the properties of a molecule is faster and less expensive than synthesizing and characterizing it in the laboratory. Computer aided drug discovery and design methods are commonly used in big pharmaceutical and biotech industries [127].

Molecular dynamics (MD) are a computational simulation technique that model complex systems at both atomic and molecular level. With the aid of MD simulations, biologically relevant macromolecules and their environment can be studied. MD simulations act as an interplay between the length and the timescales of the micro and macro-attributes [122]. In addition to providing insight into experiments, they reveal details which are difficult to discover in the laboratory due to complexity and cost.

2.1.1 The Molecular Dynamics Algorithm

A molecular dynamics simulation involves solving Newton's equation of motion for an atom i , with mass m_i and net force $\mathbf{F}_i$ that are within a system that consists of other interacting particles. The equation is given as:

$$\begin{aligned} \mathbf{F}_i &= m_i \mathbf{a}_i \\ &= m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} \end{aligned} \tag{2.1}$$

Using a set of initial system conditions,the molecular dynamics algorithm can provide accurate solutions for state properties, e.g. positions and velocities over time [128]. At time zero, an MD simulation typically starts with defining starting coordinates $\mathbf{R}(t = 0)$, which consist of coordinates $\mathbf{r}_{1,2,3,\ldots,N}$ for N atoms. Using experimental methods such as NMR, X-ray crystallography or homology modelling, if a structure has not been resolved yet by experimental means, we can derive the starting coordinates $\mathbf{R}(t = 0)$ for all atoms [129]. Most often, these initial structures obtained experimentally need to be refined structurally e.g. by adding missing atoms, before they can be used for MD simulations. There are several computational approaches available that can be used for these refinements [130].

After the initial atomic positions have been determined, it is necessary to define the initial velocities. Using a Maxwellian distribution, the initial velocities can be assigned using the initial temperature of the system that also needs to be specified:

$$P(v) = \sqrt{\frac{m}{2\pi kT}} e^{(-\frac{mv^2}{2kT})} \tag{2.2}$$

Here, $P(v)$ equals the probability, m is atomic mass, v is velocity and k represents the Boltzmann constant. There is, however, a tendency for the initial setup to be far from equilibrium, because the velocities are assigned randomly. Due to this, it might be necessary to further adjust these velocities [130]. To this end, the temperature is gradually raised until thermal equilibrium at the envisaged temperature is reached. The equipartition theorem (equation (2.4) can be used to link the thermal energy with the kinetic energy, based on which the temperature of the system can be adjusted by reassigning the velocities.

$$E_{\text{ther}} = \frac{1}{2} KT \tag{2.3}$$

$$T = (\frac{1}{k3N}) \sum_{i=1} \sum_{a=x,y,z} m_i {v_i}^2, a \tag{2.4}$$

In MD simulations, it is necessary to compute the forces acting on the atoms at each timestep. The type of system that is studied and the level of details being modelled determines how the force calculation is done. For example, in systems where full atomistic descriptions are needed, the numerical solutions of the Newton's equation of motions (see equation (2.1)) are required for each atom in the x,y,z coordinate respectively. By solving this equation we can find the atomic coordinates $\mathbf{r}_i$ for atom i at time t. A gradient of the atomic potential, $\nabla_i U$ is taken in respect of the position vector of atom i, which in turn is used to obtain the atomic forces from which positions and momenta are eventually computed: [130]

$$\begin{aligned} \mathbf{F}_i &= -\nabla_i U \\ &- \frac{\partial U(R)}{\partial r_i} \end{aligned} \tag{2.5}$$

For systems that do not require full atomic descriptions, other forms of equations of motions are solved to obtain the forces. An example is the Lagrangian equation used for studying whole domain motion [131].

Through time discretization and selection of an appropriate timestep for integrating the equations of motion, MD simulation involves sampling the phase space based on the time-dependent evolution of the system [130]. In this thesis, the MD simulations of the systems studied are based on the numerical integration of Newton's second equation of motion. After knowing the system coordinates, velocities, and forces at time t, one has to calculate subsequent positions of the interacting atoms. An MD trajectory is simply a sequence of positions over time. The main goal of the integration of Newton's equation of motion is to define position $\mathbf{r}(t+\Delta t)$ at time $t+\Delta t$ by reference to the already known positions at t and $t-\Delta t$ [122]. There are several algorithms that have been developed for the numerical integration of the equation of motions. The three most popular ones are: the Verlet algorithm, the Leapfrog algorithm and the Velocity-Verlet algorithm. The most commonly used algorithm in MD simulations is the Verlet algorithm, this is due to its simplicity and stability [122]. A disadvantage of the algorithm is that velocities at t can only be calculated if the positions $\mathbf{r}(t+\Delta t)$ are known. Using Taylor's expansion of the coordinate $\mathbf{r}_i$ of a particle at time $t+\Delta t$, we can derive this algorithm. The equation is written as:

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m_i}\Delta t^2 \tag{2.6}$$

where Δt is the MD time step, the estimated error in the new position is of the order of Δt^4. Essentially, the trajectories represent the limit of an immeasurably small integration step. It is often better i.e. more efficient to sample longer trajectories using larger time steps [122]. Fast motions in the system are what determine the Δt value; to ensure that the integration is stable, Δt has to be on the femtosecond scale because bonds that involve light atoms (e.g. O-H) vibrate with periods of several femtoseconds [122]. In the integration algorithm, even though the bond length can be constrained to eliminate the fastest and less important vibrations, achieving a time step higher than 5 fs in simulations of biomolecules is unlikely.

Integrating the Newton's second equation of motion, keeps the number of particles N, total volume of the simulation cell V and total energy of the system E constant. This means that the trajectory will be generated in the NVE or microcanonical ensemble. However, integration errors, force fluctuations, and lack of consistencies in the forces generated majorly by the cutoff can result slow drifts in the total energy [132]. The total energy contribution is constant, but not the kinetic and potential energy contributions, so a system that is not in equilibrium will exchange potential and kinetic energies whenever the temperature changes. Therefore, there is a need to control the temperature of the system. Most experiments performed in the laboratory are done under constant temperature or pressure, and in order to be able to compare experimental results with simulation results it will desirable to perform MD simulations under the same conditions. By applying a thermostat to the system, we can ensure that the temperature remains constant. The velocity rescaling thermostat, Nose-Hoover thermostat, Langevin thermostat, Andersen thermostat, and Berendsen thermostat are some of the methods commonly used to control the temperature in MD simulation. To keep the pressure stable and simultaneously the density of the system, a barostat is also used. The pressure can be controlled using the Berendsen barostat or Parrinello-Rahman barostat.

Various types of properties can be extracted from an MD trajectory, but the analysis performed ultimately depends on the question the simulation is intended to address. For the applicable length of the simulation T, the equations of motion are solved at every time step which results in a trajectory [130]. Using the coordinates, velocities, potential energy, pressure, etc., one can calculate the desired equilibrium properties over time [133]. An example is the Root Mean Square Deviation (RMSD) which can be used to monitor structural fluctuations by either averaging over selected coordinates or over time in which the relative fluctuations of different susbsets of

the simulation can be monitored. Analyzing MD-generated trajectories typically involves calculating the time average of the quantity A over the simulation period T [130]:

$$\langle A(\mathbf{r},\mathbf{p})\rangle_{\text{time}} = \frac{1}{T}\int_{t=0}^{T} A(r(t),p(t))dt \tag{2.7}$$

That is, $\langle A(\mathbf{r},\ \mathbf{p})\rangle_{\text{time}}$ which is dependent on the position $\mathbf{r}$ and momentum $\mathbf{p}$, this calculation is for a single point over the various time steps that constitute the simulation time.

2.2 Force Fields

A mathematical expression that describes the relationship between the energy and the coordinates of the system is referred to as a Force Field (FF). It is made up of interatomic potential energy $U(r_1, r_2, ..., r_{\text{N}})$, and a set of parameters. In most cases, the parameters are either obtained by fitting based experimental data such as neutron spectroscopy, NMR, X-ray and electron diffraction, infrared, spectroscopy etc. or from *ab initio*/semi empirical quantum calculations [132]. Molecules are defined as a series of atoms that are connected by bonds and by using a FF, the true potential is replaced with a simplified model relevant to the area being simulated [132]. A FF is used to describe the motions of bonds and bond angles as well as torsions, and also, non-bonding van der Waals and electrostatic interactions. There are many FFs available, with different complexities and each developed to treat different types of systems. Biomolecules such as proteins are most often simulated either using an atomistic or a coarse-grained FF. A typical FF is represented with the following equation:

$$\begin{aligned} U = &\sum_{bonds}\frac{1}{2}k_b(r-r_0)^2 + \sum_{angles}\frac{1}{2}k_a(\theta-\theta_0)^2 + \\ &\sum_{torsions}\frac{V_n}{2}[1+cos(n\phi-\delta)] + \\ &\sum_{impropers} V_{\text{imp}} + \\ &\sum_{LJ} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{elec}\frac{q_i q_j}{r_{ij}} \end{aligned} \tag{2.8}$$

In the equation above, the intramolecular contributions to the total energy in-

cluding bond stretching, angle bending, torsions and improper dihedrals are denoted by the first four terms, and the last two terms describe the non-bonded interactions; the repulsive and Van der Waals attractive dispersion in the form of the Lennard-Jones 12-6 potential and the Coulombic electrostatic potential respectively.

2.2.1 Intramolecular Terms

As presented in equation (2.8), the first term which is for bond stretching is most times represented by a harmonic function which is useful for correcting the length of covalent bonds. By using this harmonic form (with force constant k_{b}), the correct chemical structure will be produced but it prevents modelling chemical changes such as bond breaking which means certain chemical processes can not be studied. This is one of the main disadvantages FF-based MD simulations have over *ab initio* MD simulations [132]. Other functional forms (such as the Morse potential) may also be employed but they are computationally expensive. Due to this factor and the fact that the harmonic approximation is fairly good, existing potentials mostly use the simpler harmonic approximation [132].

The harmonic potential can also be used in representing angle bending which is often combined with a trigonometric function [132]:

$$U_{\mathrm{bending}} = \frac{1}{2} k_{\mathrm{a}} (cos\theta - cos\theta_0)^2 \tag{2.9}$$

In some cases, the Urey-Bradley potential is added to optimize the fitting to vibrational spectra [134]:

$$U_{\mathrm{UB}} = \sum_{angles} \frac{1}{2} k_{\mathrm{UB}} (S - S_0)^2 \tag{2.10}$$

Dihedral or torsional angles involve four atoms. Unlike in angle bending and specifically in bond stretching where the motions are of high frequency, torsional motions are hundred times less stiff, and therefore very important in ensuring the correct rigidity of the molecule and to replicate important changes in configuration of the molecule, that are as a result of rotation about bonds [132]. Thus, they play a vital role in determining the stability of different molecular conformations. As seen in equation (2.8), the torsional energy is represented by a cosine function, where ϕ is the torsional angle, the phase is δ , n represents the number of minima/maxima between 0 and 2π and V_{n} is energy barrier for the motion of the torsion. In some cases, two or more terms can be combined with different n in order to construct several minima with varying depths. The dihedral potentials can also be represented alternatively, for example in OPLS FF the dihedral potential is represented by this

equation.

$$U_{\text{tors}} = \sum_{torsions} k_0 + \frac{K_1}{2}(1 + cos\phi) + \frac{K_2}{2}(1 - cos2\phi) + \frac{K_3}{2}(1 + cos3\phi) \tag{2.11}$$

Lastly, in order to ensure desired planarity of some groups e.g. sp^2 hybridized carbons in carbonyl groups or in aromatic rings or in esters, it is necessary to define an additional term which is the improper torsion angles. They are used to ensure the correct geometry and chirality of certain conformations. The improper torsion terms are mostly represented by the equation below:

$$U_{\text{imp}} = \sum_{impropers} \frac{K_{\text{imp}}}{2}[1 + cos(2\omega - \pi)] \tag{2.12}$$

or

$$U_{\text{imp}} = \sum_{impropers} \frac{K_{\text{imp}}}{2}(\omega - \omega_0)^2 \tag{2.13}$$

where ω represents the deviation from planarity.

2.2.2 Intermolecular Terms

The fifth term in FF equation (2.8) is called the Lennard-Jones potential commonly referred to as 12-6-LJ potential. It describes the potential energy interaction between two atoms that are not bonded but separated by at least three covalent bonds. The LJ potential consists of two parts: the repulsive and attractive interactions between atoms. The repulsive interactions are for small distances and the weak attractive part are for longer distances. The interactions between particle i and j are described as:

$$V_{\text{LJ}} = \sum_{LJ} 4\epsilon_{\text{ij}} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.14}$$

Here, V is the intermolecular potential between the two molecules, ϵ_{ij} represents the energy minimum of this potential, σ_{ij} represents the distance at which the intermolecular potential is minimal, and r_{ij} is the distance of separation between both particles.

The last term in equation (2.8) describes electrostatic interactions. Using the Coulumbic function we can model the electrostatic interactions between charged atoms or group of atoms and the partial charges are assigned to the atoms involved:

$$U_{\text{coul}} = \frac{1}{4\pi\varepsilon_0} \frac{q_{\text{i}} q_{\text{j}}}{\epsilon_{\text{r}} r_{\text{ij}}} \tag{2.15}$$

From the equation, the partial charges of atom i and j are represented by q_i and q_j and ϵ_r is the relative dielectric constant. Although functions incorporating polarisation effects and higher multipoles have proved more accurate at modelling the electrostatic potential, they often computationally expensive, therefore making the Coulombic model the preferred choice [132].

2.3 Virtual Screening: A tool Used in Modern Drug Discovery

Developing and discovering novel drugs and therapeutics is typically a lengthy, laborious, and expensive process. In 2014, the average cost of developing a novel drug from scratch was estimated to be 2.5 billion dollars. This represents a 145% increase from the previous study made in 2003. This increase is primarily due to the drugs high failure rate recorded [135]. On the average, it could take 10–17 years before a new drug will hit the market [136], since it has to pass through all phases for new drug development, from target discovery to drug registration. Recently, the enthusiasm for rational approaches has been fueled by tremendous advances in computations and protein crystallography [137–140]. As a result, in-silico approaches have become increasingly popular and have become a crucial component of academic and industrial research, directing drug discovery and development.

In order to develop drug candidates against a particular disease, a potential biological target is identified that plays a critical role in the progression of the disease. Following the identification of a connection between the target and the disease, the next step is to identify potential candidates that would treat the disease [128]. The first step in this process is to identify molecules that demonstrate efficacy called "hits" through a simple screen. Screening involves using high-throughput assays to test and select a variety of compounds from databases based on their biological activity. This step is very crucial to the development of new drugs but it involves maintaining huge libraries of compounds and involves running several assays. This makes it very expensive for researchers and small pharmaceutical companies and it also increases cost for big pharmaceutical companies [141]. The next step after identifying the hits is modifying them chemically into structures with relevant biological activity, i.e. to improve their pharmaceutical properties; such compounds are referred to as "leads" [141–143]. In drug discovery, this strategy can be categorized into two; lead identification and lead optimization. It is imperative to start with a compound that exhibits some activity, against a particular receptor involved in a disease [144]. Using the screening procedure, the compound pharmacokinetics and toxicity can also be studied *in vivo*. Knowing other influencing factors is essential since it is not always easy to link the biological activity of a compound to a specific target [144]. The identification of a target protein and active site does not always lead to a rational conclusion in the process of drug discovery. Developing new drugs is hindered by many obstacles. The identification of structures for membrane proteins, for example, can be a difficult task, so modelled structures are used [144] Additionally, other factors such as water mediation at the interfaces of biomolecular complexes and protonation states play an important role in determining drug inter-

actions. Metal ion-binding sites, residues in the binding site, and changes caused by drug molecules binding are also crucial aspects to be accounted for when designing new drugs [145–147]. High throughput screening (HTS) allows us to identify hit and lead compounds by screening library containing several compounds for their biological activity. In HTS, many compounds are identified on a trial-and-error basis. HTS, however is very expensive, time intensive and therefore often not feasible. In order to overcome these challenges, innovative computational tools have been developed to screen and identify compounds with desired biological activity. This makes virtual screening a promising approach. Virtual screening has proven to be an alternative to HTS in recent years, it is used for screening several libraries of active compounds against the biological target in few days and identifying possible hits.

2.3.1 Virtual Screening

Several libraries of small compounds can be screened using computer-aided drug design (CADD) tools rather than the expensive experimental methods. Pharmaceutical companies have increasingly turned to virtual screening to search for new drug-like compounds or hits in recent years [148]. Virtual screening (VS) is a computational approach employed in drug discovery for screening libraries in search of molecules that will bind to a target i.e. identify potential hit candidates. Virtual screening aims to reduce the number of compounds that need to be tested experimentally *in vitro* by first screening them computationally [149, 150], thus reducing the time and costs associated with physical screening. In addition to identifying the most promising compounds that will most likely bind to the target protein or enzyme receptor, toxic compounds or those with unfavorable pharmacodynamic (for example, potency, affinity, selectivity) and pharmacokinetic (for example, absorption, metabolism, bioavailability) properties are identified by VS [151]. Therefore, VS techniques play an important role in the identification of new bioactive substances [152]. In recent years, VS has been widely used in drug development and has already contributed to the development of compounds currently on the market. Example of drugs that were discovered using VS include captopril (antihypertensive drug), saquinavir, ritonavir, and indinavir (three drugs for the treatment of human immunodeficiency virus), tirofiban (fibrinogen antagonist), dorzolamide (used to treat glaucoma), zanamivir (a selective antiviral for influenza virus), aliskiren (anti-hypertensive drug), boceprevir (protease inhibitor used for the treatment of hepatitis C), nolatrexed (in phase III clinical trial for the treatment of liver cancer) [153–156]. There are two commonly used VS approaches employed in CADD based on the knowledge of the target structure ; ligand based and structure based VS.

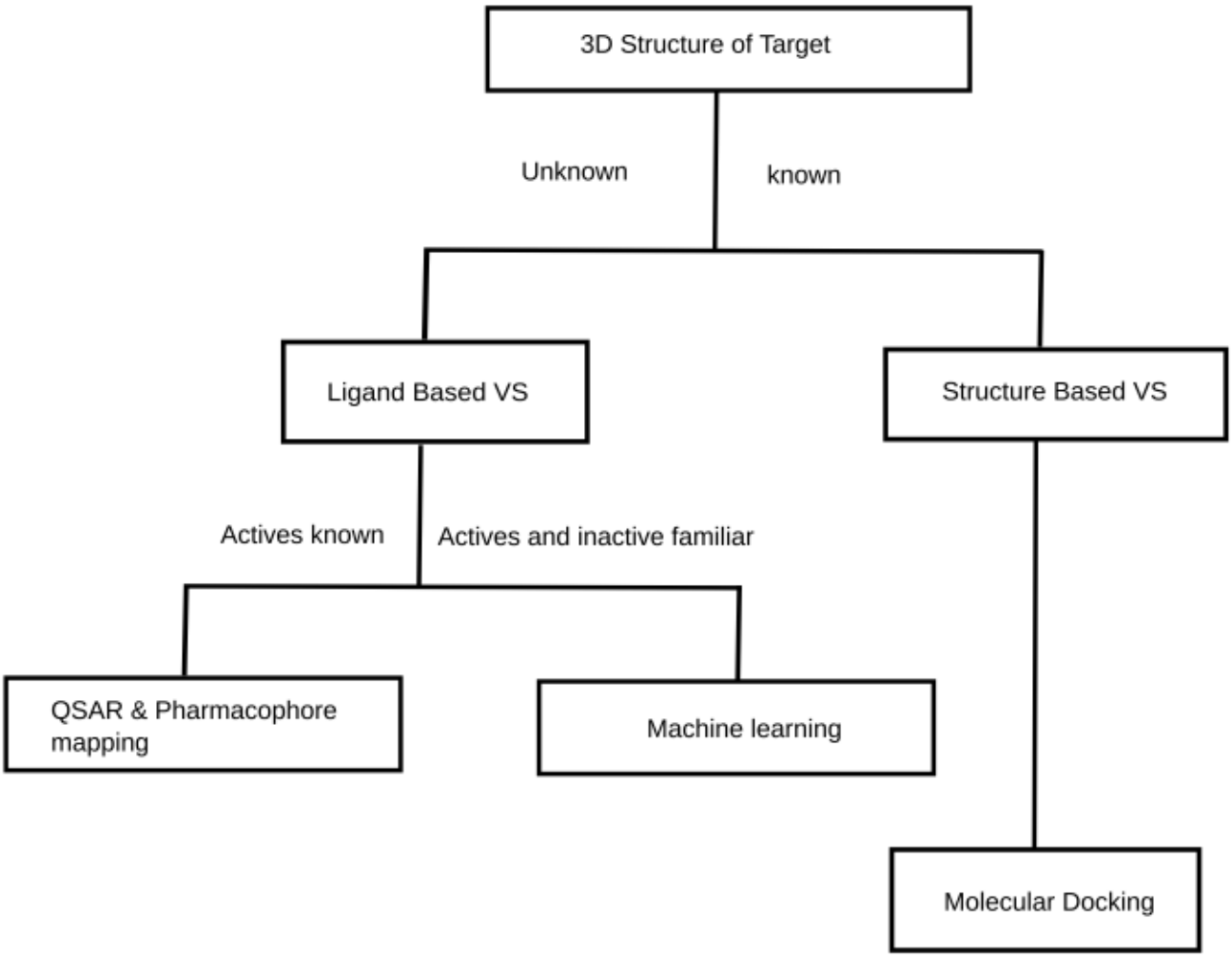

Figure 2.1: Types of virtual screening workflow

In ligand based virtual screening (LBVS) which is also called neighbourhood search, the protein or enzyme target structure is unknown, it relies on the information present in a known active ligand for lead identification and optimization. This is very common when dealing with G-protein-coupled receptor (GPCR) targets [157–163] or protein structures resolved in the apo form [164]. Ligand databases or libraries are searched to identify compounds that are similar to known active compounds (similarity searching) or have a pharmacophore in common with a known active ligand (pharmacophore substructure searching) [144]. It is based on the assumption that ligands similar to an active ligand are more likely to be active than random ligands. In LBVS, compounds that have specific pharmacokinetic, or toxicological properties based on their structure, and physicochemical properties derived from their ligand structure can be predicted by machine learning methods [165].

Unlike LBVS, in structure based VS (SBVS) the the three dimensional (3D) structure of the target protein is known. Typically, SBVS starts with identifying the potential active site where the ligand will bind to on the receptor [166]. Preferably, the active or target site is a pocket that contains hydrogen bond donors and acceptors, hydrophobic properties and a surface that a molecule can adhere to [167]. Molecular docking is the method used in SBVS, and due to the immense growth in 3D X-ray, NMR and cryo-EM structures, it has become very useful in the drug discovery process. It is a robust, cheap, useful and promising technique used in drug discovery. Therefore, in this thesis, structure based VS method "Docking" was employed for identifying potential aggregation inhibitors for sickle hemoglobin (HbS) aggregation. Details on this method will follow.

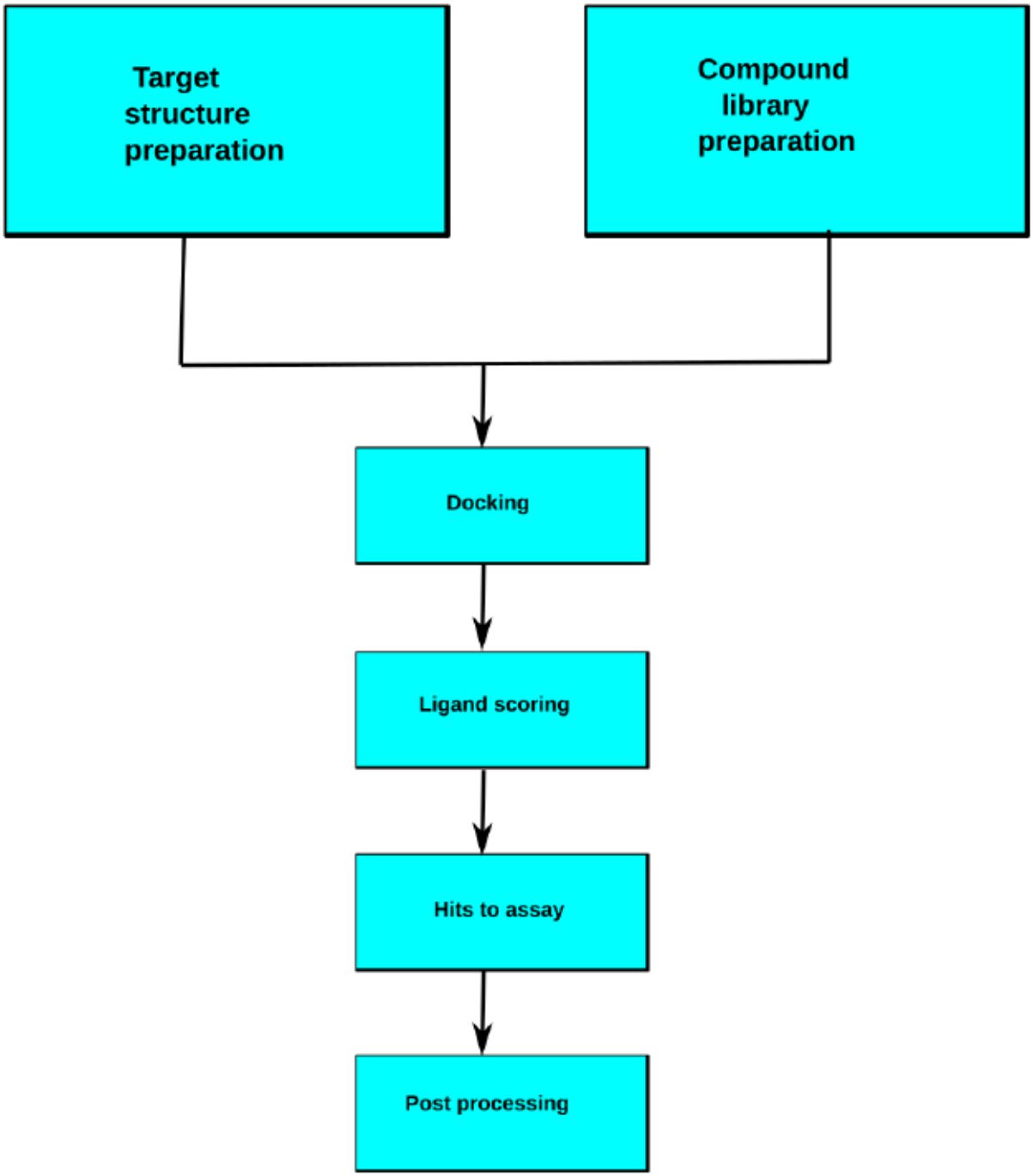

Figure 2.2: Typical workflow of a structure-based virtual screening (SBVS) [193].

2.4 Molecular Docking

Docking is the process that binds two molecules together computationally. Molecular docking helps in predicting the binding conformations of a small molecule i.e. the ligand, protein, to the appropriate target's (protein) binding site. To accurately carry out docking, one requires a three dimensional structure of the protein with a known or predicted binding site. These structures can be generated through experimental methods such as NMR and X-ray crystallography or via homology modelling [144]. Docking can be achieved in two steps: first, by sampling the conformation of the ligand in the active site of the protein, and then ranking the conformations based on a scoring function [168]. These two steps will be discussed further in the next subsections.

2.4.1 Sampling Algorithm

There are lots of binding modes possible between a ligand and a protein, generating all these binding modes is too computationally expensive and takes a lot of time. Therefore, several algorithms were developed to effectively sample the ligand in the binding site and present only the important conformations.

Matching algorithm (MA)[169–171]is the simplest of all the algorithms developed. and it is very fast which is useful when screening large libraries. This algorithm works by searching for molecules that are very similar in shape to the binding pocket of the receptor. Both protein and ligand are represented as pharmacophores, and the distance in the pharmacophore between the ligand and receptor is calculated for a match [152]. While matching, certain properties such as hydrogen-bond acceptors and donors are considered. Examples of software that use this algorithm include; DOCK [166], FLOG [172], LibDock [173] and SANDOCK [174].

Incremental construction (IC) [175–177] is another algorithm used in molecular docking. In IC, the ligands are placed in the binding pocket of the receptor in an incremental or fragmented way [168]. In order to dock into the active site, the ligand is fragmentented by breaking its rotatable bonds. After doing this, one of the fragments is selected to be docked into the active site first. The initial fragment selected is most times the largest or the one with a significant interaction with the receptor. The remaining fragments are now added into the binding site incrementally. Examples of software that use the IC method are DOCK 4.0 [178], FlexX [175], Hammerhead [179], SLIDE [180] and eHiTS [181].

Another algorithm used is the Monte Carlo (MC) [182, 183] technique and it is based on stochastic methods. The binding modes of the ligands are generated by modifying the ligand through bond rotation and translation or by rotation of the entire ligand. After this modification, the conformation generated is now tested using

energy calculation in the binding site. It is then either rejected or accepted based on the Boltzmann's distribution. The acceptance or rejection of the conformation depends largely on the change in energy with respect to temperature. This process of rejection or acceptance continues until the number of predefined conformation is collected. Examples of software applying the MC methods include an older version of AutoDock [184], ICM [185], QXP [186] and Affinity [187]. Its main disadvantage is the uncertainty of convergence, but this can be minimized by running multiple independent runs [141].

Genetic algorithm (GA) also uses the stochastic method and it is very similar to the MC method. GA is based on Darwin's theory of evolution. A set of parameters describing the rotation, translation and conformation of the ligand with respect to the protein defines the protein-ligand arrangement. These parameters are referred to as the "state variables". The total interaction energy between ligand and protein is known as the fitness function, this function stochastically evaluates the set parameters that are encoded in the chromosome. To generate a new chromosome (offspring), random pairs of chromosomes are merged. The new chromosome inherits genes from either parent based on the fitness value. Some of the new chromosomes created undergo random mutations in which one gene is modified. These mutations are only accepted if they result in a better fitness value. As a result, offsprings that adapt better in environment reproduce and the poorer ones die. GA is used in programs like AutoDock [188], GOLD [189], DIVALI [190] and DARWIN [191]. The limitation of the GA algorithm is also similar to that of MC algorithm, whereby there is convergence uncertainty.

Molecular dynamics [192–194] algorithm is also used in molecular docking. MD simulations allow for flexibility of both the ligands and receptors. A disadvantage of MD simulations is that they generally proceed in very small steps, so they are not able to overcome high-energy conformational barriers as a result, sampling might be inadequate [168]. However, it has an advantage of achieving local optimization efficiently. Therefore, it is advisable to generate the conformation of the ligand using a random search and follow with a short MD simulation for refinement after docking.

2.4.2 Scoring Functions

In molecular docking, scoring functions are used for separating the good binders from the bad ones. Scoring functions can be applied in different ways. In cases where we are interested in how a single ligand binds to a receptor, scoring function helps in accurately predicting the docked conformation which shows the 'true' structure of the receptor-ligand complex [141]. In cases where several ligands are evaluated, scoring functions help in evaluating and ranking these ligands after generating the

accurate poses [141]. Scoring functions simply estimate the binding energy of the complex using various assumptions and simplifications and not exactly calculate the binding affinity. Scoring functions can be divided into force-field-based, empirical and knowledge-based scoring functions [195].

Force field functions [196–198] simply estimate the binding affinities of the protein-ligand complex by calculating the sum of intermolecular / non-bonded interactions (van der Waals and electrostatic) and bonded interactions [199]. The electrostatic interactions are modeled by a Coulombic formulation while the van der Waals term is described by Lennard Jones potential functions. As an example, we provide the FF function which is used by DOCK [200].

$$E = \sum_{i} \sum_{j} \left(\frac{A_{ij}}{r^{12}{}_{ij}} - \frac{B_{ij}}{r^{6}{}_{ij}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \tag{2.16}$$

where r_{ij} is the distance between protein atom i and ligand atom j, A_{ij} and B_{ij} represent van der Waal parameters, q_i and q_j represent the atomic charges, and the distance-dependent dielectric constant is represented by $\varepsilon(r_{ij})$ The major limitation of the FF scoring function is that it takes into account only the protein and ligand interactions which is not robust enough. The Shoichet group [201] developed a more robust function by adding solvent effects using implicit water models to the existing protein-ligand ingteractions. The Poisson Boltzmann approach was also used to model the electrostatic potentials, the van der Waals interactions were calculated using the Lennard Jonnes potential and the electrostatic interactions between the receptor and the ligand was estimated using DelPhi [202].

Empirical scoring functions: The energy terms used in empirical scoring functions are simpler when compared to that of the FF functions. The number of types of interactions between two binding partners is counted in this model [203]. This can be done either by calculating the number of atoms of the protein and ligands interaction with each other. It can also be calculated based on the difference in the solvent accessible surface area (SASA) of the complex compared to the SASA of the uncomplexed ligand and protein. LUDI [204], PLP [205–207], ChemScore [208] are examples derived from empirical scoring functions. The coefficients of the scoring function are usually fit using multiple linear regression methods.

$$\Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{h-bonds} f(\Delta R, \Delta\alpha)$$
$$+\Delta G_{\text{ionic}} \sum_{ionicint.} f(\Delta R, \Delta\alpha) + \Delta G_{\text{lipo}} |A_{\text{lipo}}| + \Delta G_{\text{rot}} N_{\text{rot}}. \tag{2.17}$$

Here, ΔG_o is the binding energy without protein interactions, ΔG_{hb} is the hy-

drogen bonding contributions, ΔG_{ionic} is the binding energy from ionic interactions, ΔG_{lipo} is the binding energy contributions of liphophilic interactions, ΔG_{rot} describes the loss of binding energy due to freezing of internal degrees of freedom in the ligand while N_{rot} represents the number of rotatable bonds and $f(\Delta R, \Delta\alpha)$ is a penalty function that accounts for large deviations from ideal hydrogen bond and salt bridge geometry.

Knowledge based scoring functions: These are based on the the structural information that can be found in structures that have been solved experimentally and are present in databases e.g. the Protein Data Bank. The similarities in the information retrieved from the database is compared to that of the putative protein-ligand complexes. The functions analyze crystal structures of complexes so as to find interatomic contact frequencies between a protein and its ligand based on the hypothesis that the stronger the interaction, the greater its frequency of occurrence [141]. Using equation (2.18) the overall score can be calculated by accounting for both attractive and repulsive interactions between the atoms in the protein and ligand.

$$w(r) = kT1n[g(r)], \quad g(r) = (r)p(r)/p*(r) \tag{2.18}$$

where k_{B} is the Boltzmann constant, T is the absolute temperature, $p(r)$ is the number of density of the protein-ligand atom at distance r, and $p*(r)$ is the pair density in the reference state where the interatomic interactions are zero and $g(r)$ is the pair distribution function [141].

All the scoring function methods explained above have their major limitations, therefore to get better or improved results it is better to combine several scoring functions and this approach is referred to as the consensus scoring [209].

Another method that started gaining recognition recently is the machine learning method. The machine learning method captures binding effects that are too complex to be explicitly modeled [210]. This scoring function deduces the functional form directly from experimental data, rather than assuming [210]. In the prediction of binding affinity of protein-ligand complexes, machine-learning scoring functions have consistently transcended classical scoring functions. A 88.6% hit rate was recorded by a program that incorporated the machine learning methods [211].

Chapter 3

Results

This chapter is divided into three sub-chapters corresponding to three manuscripts, of which **Chapter 3.1** is accepted for publication in **Proteins: Structure, Function, and Bioinformatics**. **Chapter 3.2** and **Chapter 3.3** are manuscripts under preparation.

3.1 Multiscale MD Simulations of wild-type and sickle hemoglobin aggregation

ORIGINAL ARTICLE

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, 52425 Jülich, Germany

[2]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

[3]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Obafemi Awolowo University, Ile-Ife 220282, Nigeria

[4]Institute of Drug Research and Development, Afe Babalola University, Ado-Ekiti 361212, Nigeria

Correspondence
Email: b.strodel@fz-juelich.de

Sickle cell disease is a hemoglobinopathy resulting from a point mutation from glutamate to valine at position six of the β-globin chains of hemogblobin. 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 hydrohobic 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
sickle cell disease, Glu6Val mutation, MD simulation, protein aggregation, protein-protein interactions

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 give 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, damages to the RBC morphology, and RBC interaction with endothelial surfaces, premature damage to the erythrocytes[5;6], agonizing vaso-occlusive crisis, an overall poor health condition, and death[7;8;9;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 genetic level[13;14], only 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;18;19;20;21;22]. For instance, prevalence is concentrated in sub-Saharan Africa and parts of south-east Asia, with more than 75% of the cases are believed to be in Nigeria, Democratic Republic of 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;28;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;31;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 $\approx$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 ex-

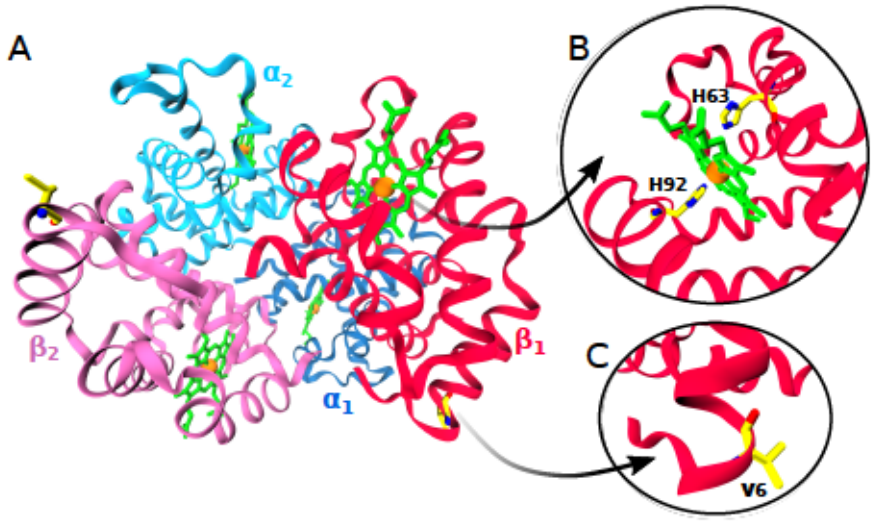

FIGURE 1 (A) The quaternary structure of HbS consisting of two α subunits (here denoted α_1 and α_2 for ease of distinction, shown in shades of blue) and two β (β_1 and β_1, shades of red) subunits. Each globin subunit carries one heme (green), including an Fe^{2+} ion (orange). (B) The hemes are linked 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.

hibited 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 [36;35]. 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 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 to 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 protein-protein interactions in the resulting aggregates are elucidated and their stability 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 quantify the relative strengths of molecular contacts and identify protein-protein interaction hot-spots 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 eqilibration 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 Nose-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

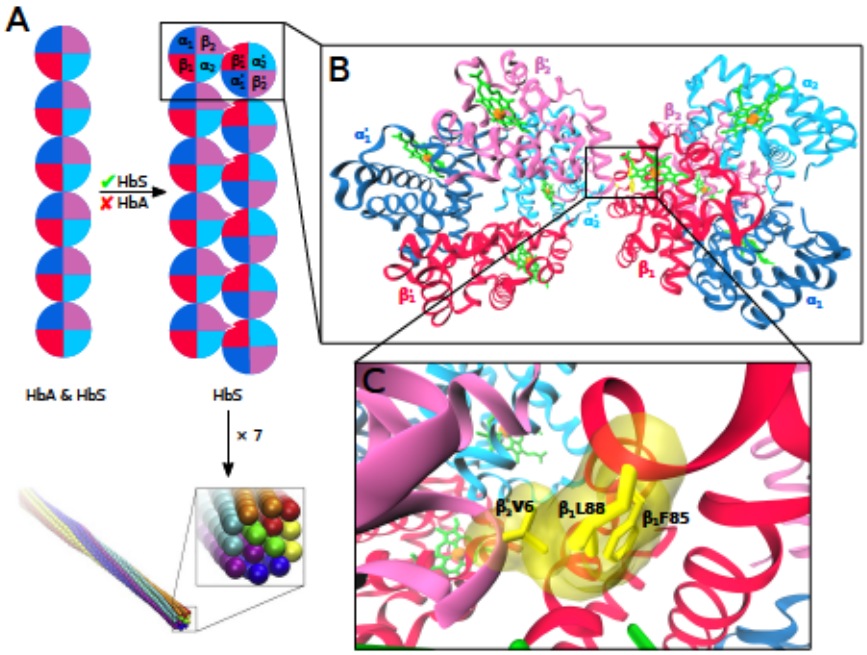

FIGURE 2 (A) Schematic representation of how the Glu6Val mutation modifies normal hemoglobin polymerization of HbS heterotetramers, involving linear Hb aggregates formed by both HbA and HbS (left) into double HbS filaments (right). The hemoglobin tetramer is represented as a circle, such that one quarter 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].

simulation box using the same box dimensions as in the corresponding AA-MD simulation. Energy minimization using the steepest decent algorithm in 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 decent algorithm was performed; afterwards, 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 ten 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 decent algorithm in 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 steep-

est 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 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 was 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]. 300 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

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

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 $D_{\text{solv}} = 80$ for the dielectric constant of the solvent (corresponding to water), $D_{\text{solute}} = 2$ 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 2,000 bootstrap steps.

3 | RESULTS AND DISCUSSION

3.1 | Conformational dynamics of monomeric hemoglobin

In order to asses 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 are 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], i.e., 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 structure 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 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 are 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 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 is caused by structural instabilities in either or both of the two HbS proteins composing the dimer (denoted as HbS-

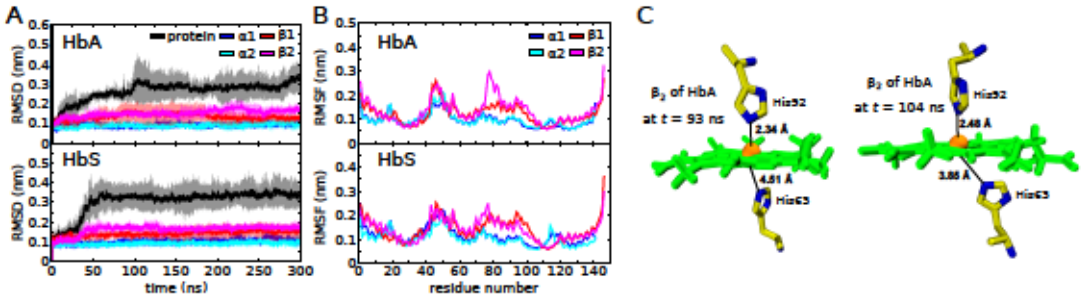

FIGURE 3 Results of the HbA and HbS monomer simulations. (A) The evolution of the Cα RMSD of HbA (top) and HbS (bottom) is shown (black lines) as well as of the corresponding individual chains (α_1: blue, α_2: cyan, β_1: red, β_2: magenta). For all RMSD calculations, the alignment was with respect to 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.

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 centres 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 changes 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 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.

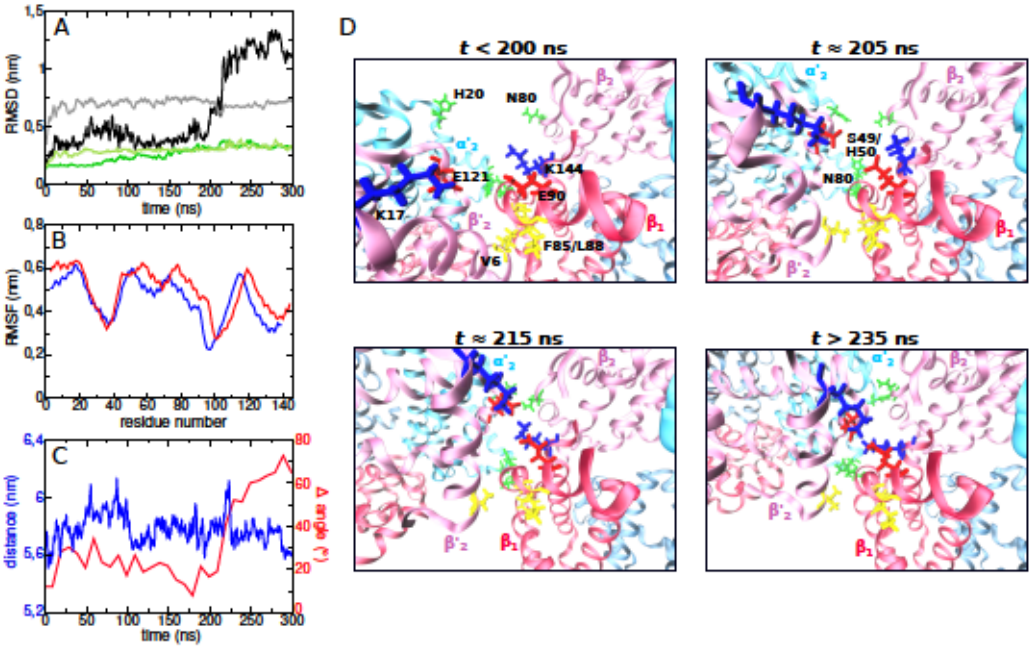

FIGURE 4 Results of the HbS dimer simulations. (A) The evolution of the RMSD of the whole HbS dimer (black) and of the composing HbS-1 and HbS-2 proteins (different shades of green) obtained from the AA-MD simulation is shown. The RSMD of the dimer in the CG-MD simulation is shown in gray. (B) The average RMSF values for the α and β chains (red and blue, respectively) support structural stability of the globins in the AA-MD simulation. (C) The evolution of the distance between the centres of mass of HbS-1 and HbS-2 in the AA-MD simulation (blue) confirms 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.

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 ro-

tate 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 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 absence of these interactions in the last third of the simulation. The electrostatic interactions, on the other hand, are only weakly present the the first 200 ns, but dominate in 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 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 simuations 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 Eqs. (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' are 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 ten HbS monomers to aggregate

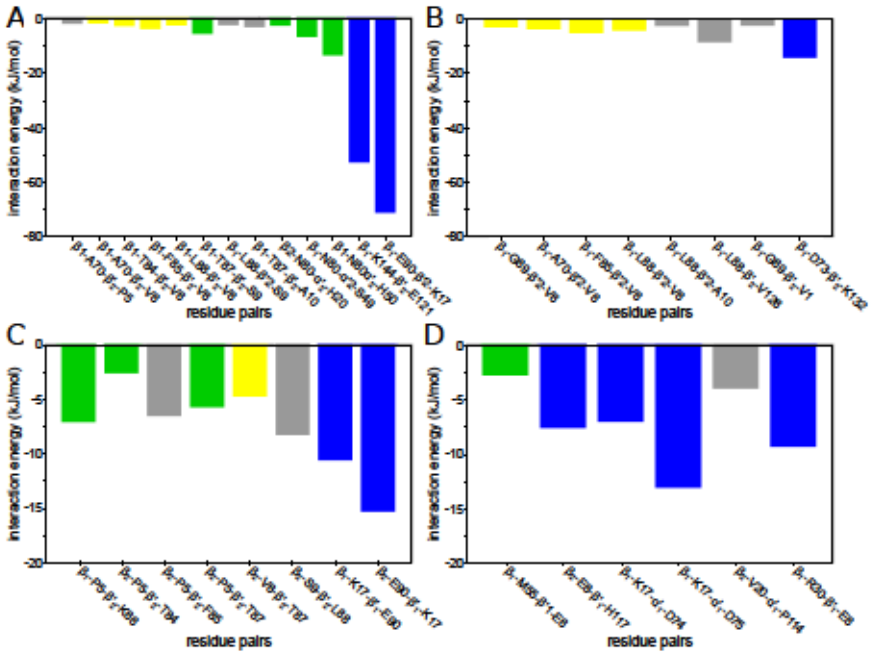

FIGURE 5 Time-averaged residue-residue interaction energies between hemoglobin molecules. In the upper panels, the energies between the two HbS proteins composing a dimer from (A) the AA-MD simulations and (B) the CG-MD simulations are shown. In the lower panels, the interaction energies between any 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.

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 cut-off 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 follow-

ing 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 ten such dimers were identified (Figure 6, left). To separate the weak and thus unimportant contacts from the strong ones, we calculated the interaction energies 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 are 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 interactionβ_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 ten 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 aggrgeation 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

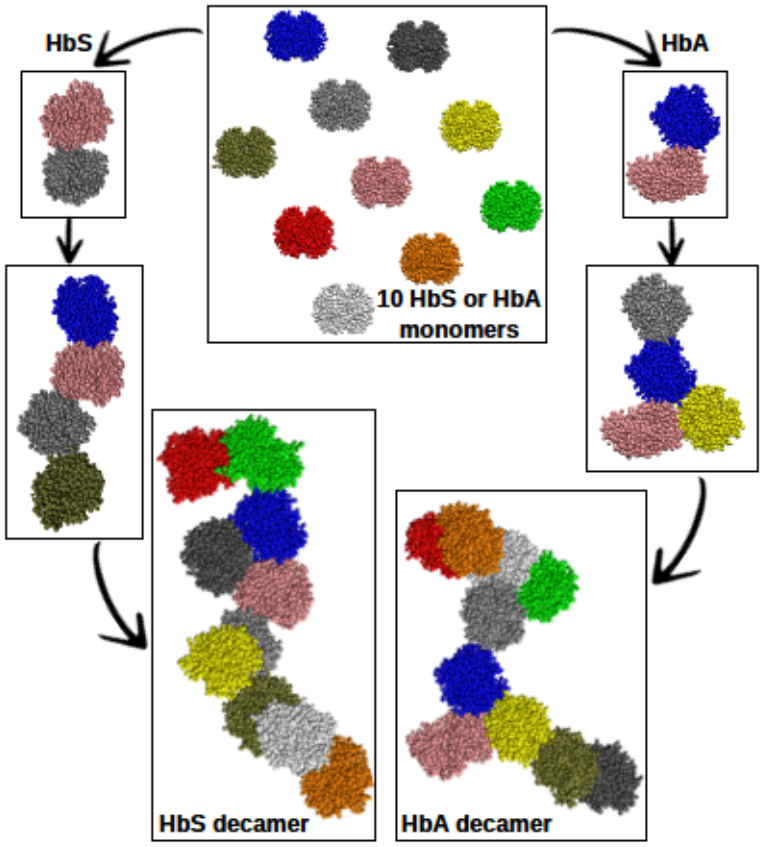

FIGURE 6 Hemoglobin aggregation pathways obtained from CG-MD simulations. The simulations were started from ten HbS or HbA monomers. The aggregation proceeded in a stepwise fashion until an HbS (left) or HbA (right) decamer formed.

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 ten HbS

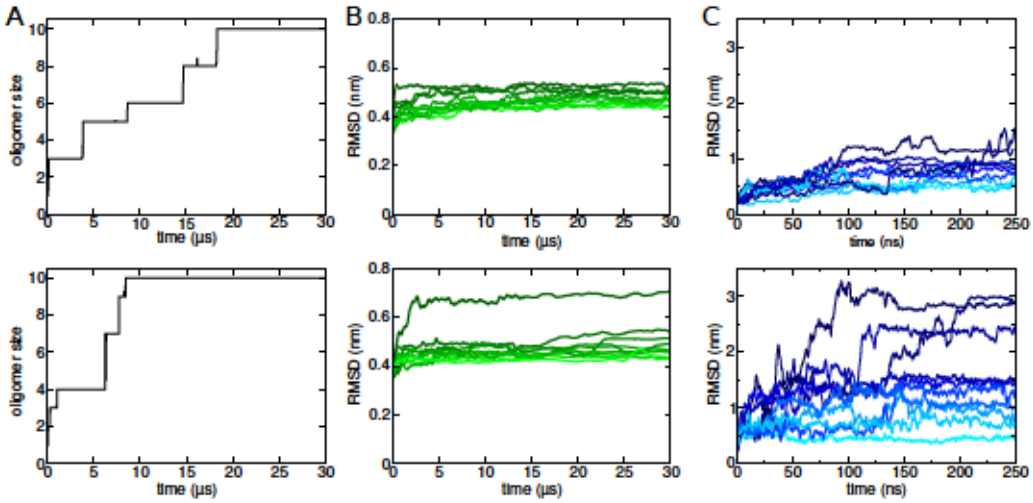

FIGURE 7 Results of the aggregation simulations of HbS (top) abd HbA (bottom). (A) The aggregation state in terms of number of hemoglobin proteins in the oligomer that formed is provided. (B) The evolution of RMSD of the individual proteins during the aggregation simulations confirms protein stability. (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.

and ten 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 ten 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 a 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, 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 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 to 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 hydrohobic 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 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[62;27;28;29], 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.

Acknowledgements

M.O.O. gratefully acknowledges 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.

Conflict of interest

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.

Supporting Information

Table S1: interprotein contacts in the HbS dimer in the 300 ns AA-MD simulation; Table S2: ΔG_{bind} values obtained from the 300 ns AA-MD simulation of the HbS dimer; Table S3: interprotein contacts in the HbS dimer

in the 300 ns CG-MD simulation; Figure S1: evolution of the His-Fe^{2+} distances in one of the 300 ns AA-MD simulations of the HbS and HbA monomer; Figure S2: components of the ΔG_{bind} values obtained from the 300 ns AA-MD simulation of the HbS dimer; Figure S3: analysis of the 250 ns AA-MD simulations of the back-mapped HbS and HbA dimers.

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. USA* 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 Jr WE. 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. USA* 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 Jr AD, 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.

44. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys* 2005; 7182(1981).

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.

Supporting Information

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, 52425 Jülich, Germany

[2] Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

[3] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Obafemi Awolowo University, Ile-Ife 220282, Nigeria

[4] Institute of Drug Research and Development, Afe Babalola University, Ado-Ekiti 361212, Nigeria

Correspondence
Email: b.strodel@fz-juelich.de

Table S1: Residue-residue contacts between the HbS proteins composing the HbS dimer obtained during the first 200 ns and last 100 ns of the all-atom MD simulation.

	First 200ns				Last 100ns			
Nr.	**Chain**	**Residue**	**Chain[a]**	**Residue**	**Chain**	**Residue**	**Chain[a]**	**Residue**
1	β1	Thr 87	β4	Val 6	β1	Thr 87	β4	Val 6
2	β1	Thr 87	β4	Ser 9	β1	Thr 87	β4	Ser 9
3	β1	Thr 84	β4	Val 6	β1	Thr 84	β4	Val 6
4	β1	Thr 87	β4	Ala 13	β1	Thr 87	β4	Ala 13
5	β1	Thr 87	β4	Ala 10	β1	Thr 87	β4	Ala 10
6	β1	Glu 90	β4	Lys 17	β1	Glu 90	β4	Lys 17
7	β1	Glu 90	β4	Ala 13	β1	Glu 90	β4	Ala 13
8	β1	Asp 79	α4	His 50	β1	Asp 79	α4	His 50
9	β1	Asn 80	α4	Ser 49	β1	Asn 80	α4	Ser 49
10	β1	Asp 79	α4	Ser 49	β1	Asp 79	α4	Ser 49
11	β1	Asn 80	α4	His 50	β1	Asn 80	α4	His 50
12	β1	Lys 144	β4	Glu 121	β1	Lys 144	β4	Glu 121
13	β1	Gly 69	β4	Pro 5	-	-	-	-
14	β1	Ala 70	β4	Pro 5	-	-	-	-
15	β1	Phe 85	β4	Pro 5	-	-	-	-
16	β1	Ala 70	β4	Val 6	-	-	-	-
17	β1	Phe 85	β4	Val 6	-	-	-	-
18	β1	Leu 88	β4	Val 6	-	-	-	-
19	β1	Gly 83	β4	Pro 125	-	-	-	-
20	β1	Leu 91	β4	Ala 13	-	-	-	-
21	β1	Pro 125	β4	Gly 83	-	-	-	-
22	β1	Leu 88	β4	Ser 9	-	-	-	-
23	β1	Val 67	β4	Pro 5	-	-	-	-
24	β1	Asn 80	α4	Leu 48	-	-	-	-
25	β1	His 77	β4	Glu 7	-	-	-	-
26	β1	Asp 73	β4	His 2	-	-	-	-
27	β1	Asp 94	β4	Lys 17	-	-	-	-
28	-	-	-	-	β2	Asn 80	α4	Tyr 24
29	-	-	-	-	β2	Asn 80	α4	His 20
30	-	-	-	-	β2	Thr 84	α4	His 112
31	-	-	-	-	β1	Lys 144	β4	Lys 120
32	-	-	-	-	β1	Glu 90	β4	Lys 120
33	-	-	-	-	β2	His 20	β4	Lys 120
34	-	-	-	-	β2	Asp 79	β4	Lys 120
35	-	-	-	-	β2	Val 1	β4	Lys 12

[a] β4 = β2'; α4 = α2'; - = contact not present

Table S2: Average binding free energies and decomposition into the contributions (including standard errors of the mean) calculated with the MM/PBSA method applied to the 300 ns all-atom MD simulation of the HbS dimer. Energies were averaged for the first 200 ns and last 100 ns of that simulation, using a time interval of 1 ns between the snapshots. All energies are in kJ/mol. In rows 3 and 4 the interaction energies within the binding site (BS) including all residues within 10 Å of the HbS-HbS interface are provided. Strongly attractive interactions are highlighted in bold.

System/Engery	ΔE_{LJ}	ΔE_{Coul}	ΔG_{polar}	$\Delta G_{nonpolar}$	ΔG_{bind}
HbS-1 to HbS-2; 0-200 ns	-178.3 ± 1.9	228.2 ± 10.5	216.7 ± 12.2	-22.5 ± 0.4	244.3 ± 12.5
HbS-1 to HbS-2; 200-300 ns	-141.6 ± 4.2	-317.6 ± 30.1	614.2 ± 25.4	-23.7 ± 0.8	131.0 ± 16.4
HbS-1:BS to HbS-2:BS; 0-200 ns	-156.2 ± 1.7	-80.8 ± 9.5	183.7 ± 6.3	-22.3 ± 0.3	-75.6 ± 6.8
HbS-1:BS to HbS-2:BS; 200-300 ns	-128.2 ± 4.0	-648.8 ± 30.4	513.3 ± 23.1	-23.5 ± 0.7	-284.9 ± 13.3
β1 to α2'; 0-200 ns	-41.0 ± 1.2	-75.6 ± 4.7	73.9 ± 6.0	-5.4 ± 0.2	-48.0 ± 6.0
β1 to α2'; 200-300 ns	-32.6 ± 1.8	-25.6 ± 6.5	69.3 ± 8.7	-4.6 ± 0.3	6.1 ± 8.5
β1 to β2'; 0-200 ns	-131.8 ± 1.6	-85.0 ± 9.7	171.7 ± 8.5	-18.0 ± 0.3	-62.9 ± 8.8
β1 to β2'; 200-300 ns	-43.7 ± 1.8	-470.9 ± 26.5	373.0 ± 18.9	-10.4 ± 0.3	-150.8 ± 11.5
β2 to α2'; 200-300 ns	-52.7 ± 3.4	-42.1 ± 6.0	37.9 ± 9.3	-7.4 ± 0.5	-64.2 ± 9.1
β2 to α2'; 200-300 ns	-10.7 ± 0.9	-110.0 ± 10.5	131.2 ± 10.1	-3.0 ± 0.2	7.6 ± 8.8

Table S3: Residue-residue contacts between the HbS proteins composing the HbS dimer obtained during the coarse-grained MD simulation.

Nr.	Chain[a]	Residue	Chain[a]	Residue
1	β1	Lys 65	β4	His 2
2	β1	Asp 73	β4	His 2
3	β1	Gly 69	β4	Val 1
4	β1	Gly 69	β4	Pro 5
5	β1	Gly 69	β4	Val 6
6	β1	Gly 69	β4	Val 6
7	β1	Ala 70	β4	Val 6
8	β1	Phe 85	β4	Val 6
9	β1	Leu 88	β4	Val 6
10	β1	Leu 88	β4	Ala 10
11	β1	Leu 88	β4	Val 11
12	β1	Leu 88	β4	Ala 13
13	β1	Leu 91	β4	Ala 10
14	β1	Leu 91	β4	Ala 13
15	β1	Leu 91	β4	Leu 14
16	β1	His 92	β4	Lys 17
17	β1	Phe 85	β4	Pro 125
18	β1	Thr 87	β4	Phe 118
19	β1	Thr 87	β4	Phe 121
20	β1	Thr 87	β4	Pro 125
21	β1	Thr 87	β4	Val 126
22	β1	Leu 88	β4	Pro 125
23	β1	Leu 88	β4	Val 126
24	β1	Leu 88	β4	Ala 129
25	β1	Leu 91	β4	Val 126
26	β1	Asp 73	β4	Lys 132
27	β1	Leu 91	β4	Val 126
28	β1	Leu 91	β4	Phe 118
29	β1	His 77	α4	Lys 40
30	β1	His 77	α4	Pro 37
31	β1	Asp 79	α4	Lys 40
32	β1	Asp 79	α4	Leu 34
33	β1	Asp 79	α4	Pro 37
34	β1	Asn 80	α4	Leu 34
35	β2	Asp 79	α4	Lys 56
36	β2	Asn 80	α4	Val 55
37	β2	Asn 80	α4	Ser 52
38	β2	Gly 83	α4	His 50

39	β2	Gly 83	α4	Ser 52
40	β2	Thr 4	α4	His 112
41	β2	Pro 5	α4	His 112
42	β2	Pro 5	α4	Leu 113
43	β2	Val 1	β4	Glu 121
44	β2	His 2	β4	Glu 121
45	β2	Leu 3	β4	Glu 121

[a] β4 = β2'; α4 = α2'

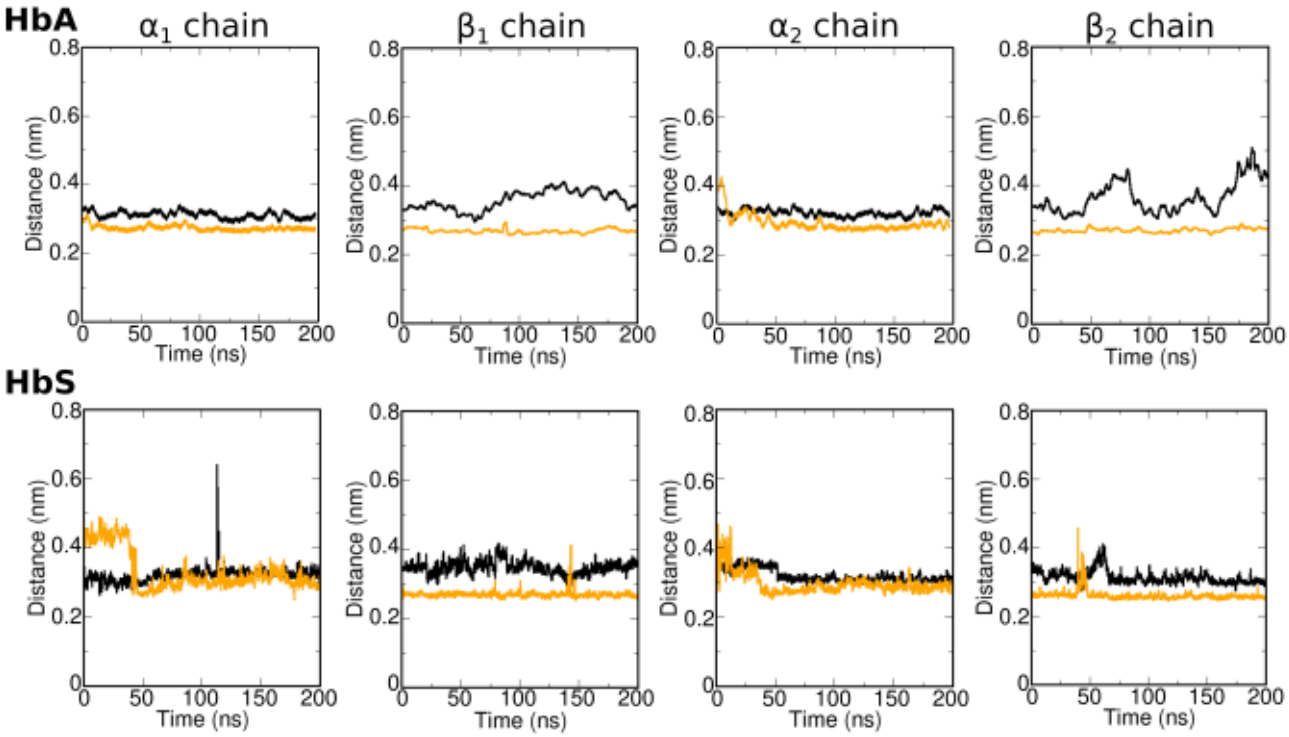

Figure S1: Evolution of the minimum distance between Fe^{2+} and the proximal proximal histidine residues (orange; His87 in the α chains and His92 in the β chains) and distal histidine residues (black; His58 in the α chains and His63 in the β chains) during all-atom MD simulations of the HbS and HbA monomer.

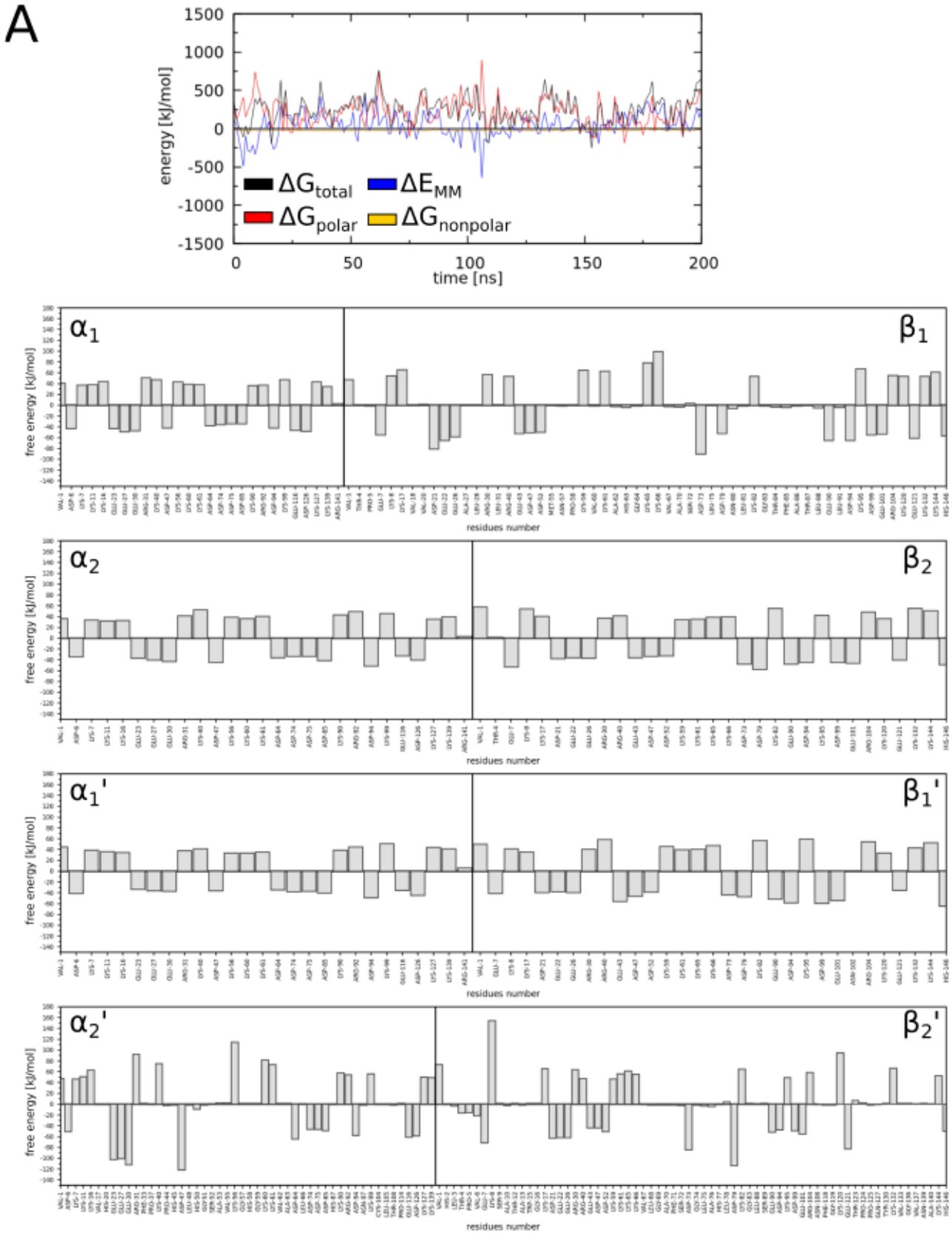

Figure S2: Results from the MM/PBSA analysis applied to the 300 ns AA-MD simulation of the HbS dimer. (A) Results for the first 200 ns and (B, see next page) results for the last 100 ns are shown. (Top) Evolution of the energetic contributions to ΔG_{bind} according to Eq. (1) and with E_{MM} = E_{bonded} + E_{Coul} + E_{LJ}. (Bottom) Decomposition of the time-averaged ΔG_{bind} into its per-residue contributions.

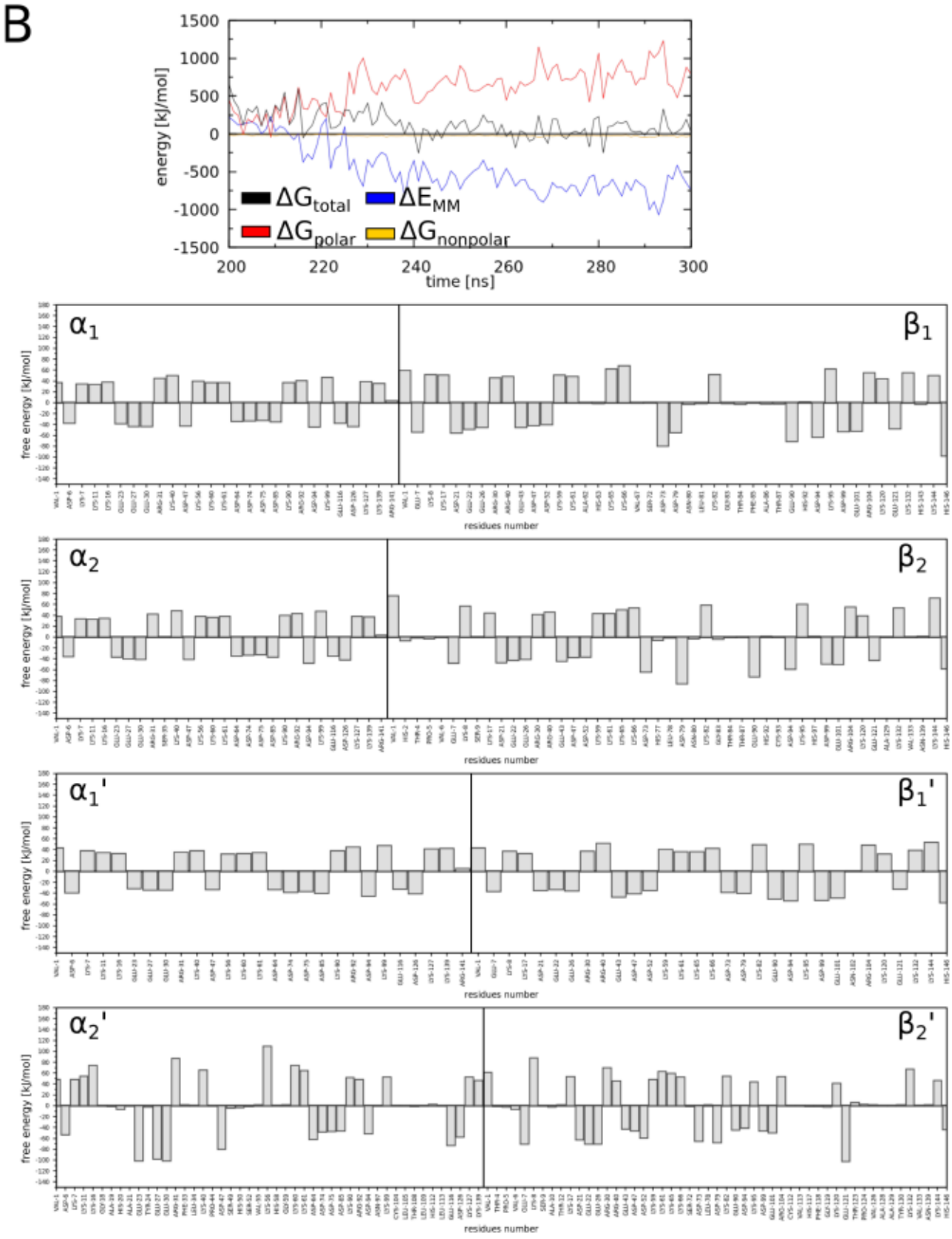

Figure S2 (continued): Results from the MM/PBSA analysis applied to the 300 ns AA-MD simulation of the HbS dimer. (A, see previous page) Results for the first 200 ns and (B) results for the last 100 ns are shown. (Top) Evolution of the energetic contributions to ΔG_{bind} according to Eq. (1) and with $E_{MM} = E_{bonded} + E_{Coul} + E_{LJ}$. (Bottom) Decomposition of the time-averaged ΔG_{bind} into its per-residue contributions.

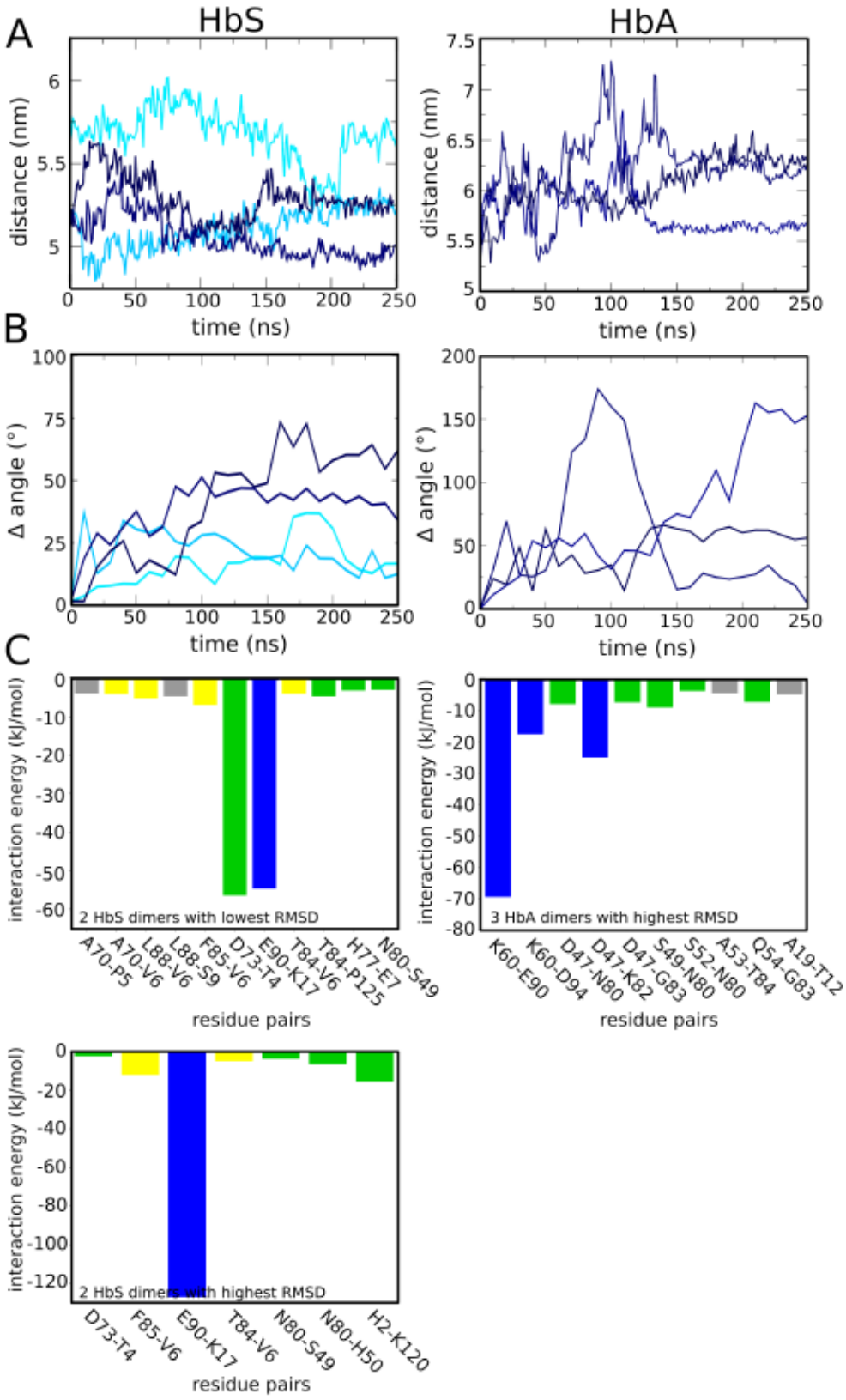

Figure S3: Analysis of the AA-MD simulations of the back-mapped dimers of HbS (left) and HbA (right). Results are shown for the most stable (light blue) and least stable dimers (dark blue). The colors in (A) and (B) correspond to the colors used for the different dimers in Fig. 7. (A) Evolution of the distance between the centres of mass of HbS-1 and HbS-2. (B) Change in the angle between the lines fitted through the atomic coordinates of HbS-1 and HbS-2. (C) Time- and dimer-averaged residue-residue interaction energies between hemoglobin molecules. Interaction energies involving Val6 of HbS are highligted in yellow, while electrostatic and polar interactions are colored in blue and green, respectively. Purely hydrophobic interactions as well as interactionsbetween hydrophobic and polar amino acids are shown by gray bars.

3.2 Computer-aided drug design-directed experimental identification of novel inhibitors of sickle hemoglobin polymerization

Computer-aided drug design-directed experimental identification of novel inhibitors of sickle hemoglobin polymerization.

Olagunju Maryam[1], Olujide Olubiyi[1,2,3], and Birgit Strodel[1,4,]

[1]Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany
[2]Department of Pharmaceutical and Medicinal Chemistry, College of Pharmacy, Afe Babalola University, Ado-Ekiti, 361212, Nigeria.
[3]Department of Pharmaceutical and Medicinal Chemistry, College of Pharmacy, Afe Babalola University, Ado-Ekiti, 361212, Nigeria.
[4]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
*e-mail: b.strodel@fz-juelich.de

ABSTRACT

Sickle cell disease is characterized by a single point mutation that involves the substitution of glutamic acid at position 6 to valine (E6V). This substitution in turn leads to the polymerization of sickle hemoglobin (HbS). The HbS aggregation causes the distortion of concave shaped red blood cells (RBC) into a crescent shape and anemia. In this present work, we aim to identify prospective HbS aggregation inhibitors using state of art computer aided drug design (CADD) techniques. To this end, we first screened a library of small molecule compounds including natural products, FDA approved drugs, non-FDA drugs and investigational drugs against the crystal structure of HbS. This was followed by ensemble docking that incorporates receptor dynamics. A total of 100 ligand-HbS complexes identified via docking were then selected for molecular dynamics simulations. By applying a range of parameters such as the ability of the ligands to form specific contacts with the Val6 binding site residues, fluctuation of the ligands in the binding site, proximity of the ligands to the binding site and lipophilicity, a total of 16 ligands were identified. The 16 best performing compounds will then be subjected to cell based assays and *in vitro* experiments to determine their HbS-aggregation inhibitory activities.

1 Introduction

One of the most common monogenic disorders that affect humans worldwide is sickle cell disease (SCD)[1]. SCD results from a single point mutation in the sixth position of the β hemoglobin chain which leads to the replacement of charged glutamic acid with hydrophobic valine (E6V)[2,3]. Under deoxygenated conditions, the mutated hemoglobin which is found in individuals with sickle cell diseases (HbS) polymerizes, thereby leading to the distortion of the flexible biconcave disk-like shaped red blood cells into a crescent shape. This, in turn, leads to clinical features observed in SCD patients such as several episodes of acute illness, hemolysis, organ and tissue damage, sore muscles, weakness, and most times premature deaths[2,4,5]. SCD is particularly common in many tropical countries, especially in sub-saharan Africa, India, and the Middle East, where malaria is or was common, as well as in any population with significant immigrant populations from these regions. This is due to their heterozygote advantage against various forms of malaria[6–8]. Worldwide, it is reported that about 4.4 million people suffer from SCD and about 43 million people carry the sickle cell trait[9]. According to a recent study, approximately 305,800 babies were born

with SCD in 2010, of which two-thirds were born in Africa, and that number could rise by 25% by 2050 to approximately 404,200[10,11]. About 40,000 children born with sickle cell are estimated to be born in India yearly, 10,000 in the Americas,10,000 in the Middle East, and 2,000 in Europe[12]. In Europe, the United States, and Jamaica, newborn screening programs have proven that the early detection of sickle cell anemia, especially during the first 5years of life, can help in the reduction of morbidity and mortality rates[13–16]. SCD is the first disease that was studied at a molecular level and one of the most studied diseases in the world, and SCD research of the disease is filled with several efforts that have helped in the understanding of the diseases better[17]. Recently, the disease is increasingly gaining recognition worldwide as a global health problem as a result of its high morbidity and mortality rates[18].

Researchers in both the academic field and pharmaceutical industries have revived efforts to provide a cure/ treatment for SCD[18,19]. The majority of the attempts at finding new treatment/therapy for the management of SCD have focused on using small molecules[20] and larger molecular weight peptides[21] to inhibit HbS polymerization because the HbS molecule is very important in the development of the disease. An example is the recently approved Voxelotor, whose mechanism of action involves the prevention of deformed red blood cells from polymerizing; studies have shown that it increases the production of normal hemoglobin in patients aged 12 and older[22,23]. In some parts of the world, e.g. the United States, stem cell transplant and gene therapy are being used for the treatment of SCD[24]. Another disease-modifying therapy used for SCD is red blood transfusion. The US Food and Drug Administration (FDA) approved the use of L-glutamine which reduces red blood cell oxidant injury to prevent acute pain experienced by SCD patients 5 years of age and older[18,25]. Since 1998, hydroxyurea has been approved for the treatment of SCD by the FDA; it has been shown to reduce the mortality in sickle cell disease. The mechanism of action involves the increase in the production of fetal hemoglobin (HbF) and a decrease in the production of HbS in patients[26–28]. Most of these treatments will not be available to SCD patients in developing countries for several years due to socio-economic and medical barriers. Therefore, there is an urgent need to develop cheap, yet effective medicines that are also readily available in developing countries[20]. Towards this end, various research efforts have focused on identifying potential leads and treatments from plant products[20]. The leaves of *Terminalia catappa, Carica papaya, Parquetina nigrescens, Citrus sinensis, Persia Americana, Zanthoxylum zanthoxyloides, and Cajanus Cajan* seeds are among the few plants that are commonly used in Nigeria and other African countries for the management of SCD[29–32].

In recent years attempts have been made to introduce rational methods of lead discovery to identify antisickling activities in plant products; these represent an interesting shift from a traditional method of investigation that relies exclusively on ethnomedicinal knowledge. Of note is the 2017 research investigation by Olubiyi et al[33]. As a proof of concept for this approach, the authors employed a combination of virtual screening and antisickling experiments to identify inhibitors of HbS polymerization from a list of FDA-approved drugs[33]. Drugs found to strongly inhibit sickle red blood cells (RBC) sickling in the work include glipizide, praziquantel, losartan, and ketoprofen. In a follow-up investigation, using a 3,000-strong library of natural products of Nigeria origin, the group again employed virtual screening as well as experimental antisickling and polymerization inhibitory assays in search of plants with the potential to inhibit sickle RBC sickling[9]. Top in the list of computationally identified plants were *Catharanthus roseus, Rauvolfia vomitoria, Hoslundia opposita, Lantana camara and Euphorbia hirta.* And after subjecting these to experimental assays that utilized hemoglobin polymerization inhibition and sickling reversal tests, sickling reversal levels of up to 68.50 % were obtained for *H. opposita.* Here, we present an experimental identification of novel antisickling compounds guided by a rigorous computational protocol that combines high throughput virtual screening with explicit solvent molecular dynamic (MD) simulations of multiple HbS-inhibitor binary complexes. The computer aided drug design (CADD)-identified best performing HbS binders were subjected to both cell and *in-vitro* HbS inhibition capabilities. The screened

compounds include drugs approved by the U.S. Food and Drug Administration (FDA) or other authorities (NFDA), investigational drugs, with possible drug repurposing for the treatment of SCD in mind. Previous MD simulations in our group revealed a number of secondary contacts as crucial for HbS aggregation with interactions involving βA70, βF85, and βL88 playing major roles (Figure 1)[34]. For this reason, we, therefore, limited the docking screening in the work reported here to an interaction cavity on multiple HbS structures (both crystal structure and MD-generated) defined by these three critical amino acids. The structure is a homotetramer with two interacting tetrameric units; each tetramer contains two α and βglobin subunits[35]. Two adjacent β-globin chains from two homotetramer form contacts, with β-globin subunit from one tetramer donating a convex Val6 knob which fits into a concave hydrophobic cleft from the β-globin of the other tetramer.

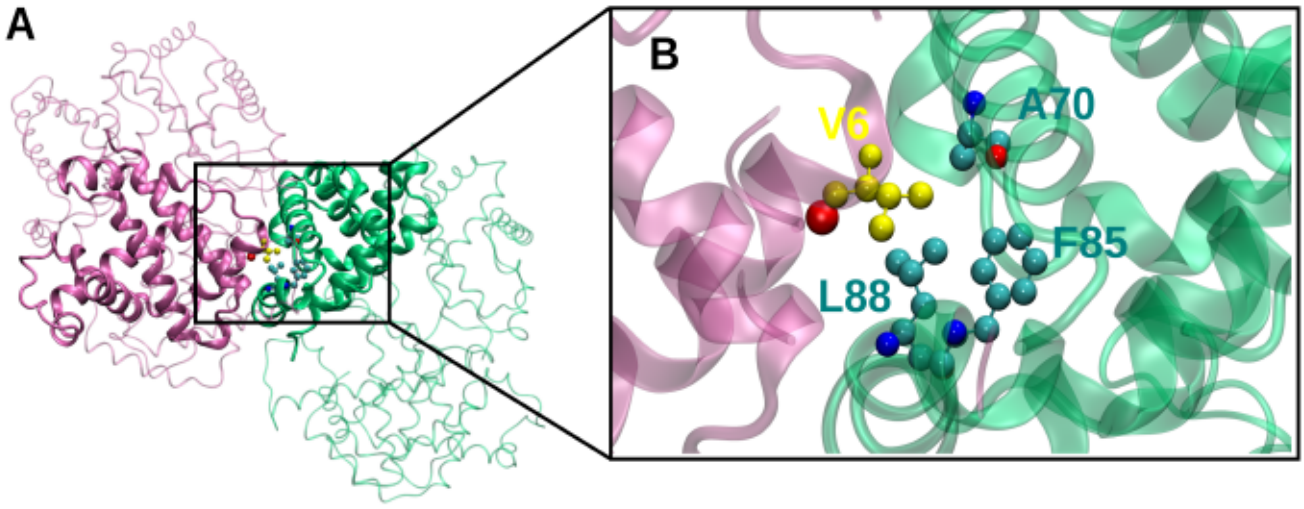

Figure 1. Crystal structure of HbS (PDB entry 2HBS) and the Val6 binding sites. (A) Cartoon representation of HbS consists of the eight subunits, with chain β1 in green (opaque) and chain β'_1 mauve (opaque). Val6 is shown in yellow ball and stick representation and A70, F85, and L88 are shown in cyan but using blue for N atoms and red for O atoms. (B) Zoom into the Val6 binding site of HbS[35]

2 Materials and Methods

2.1 Ligand library

The ligand library employed for the virtual screening contains small molecular compounds (molecular weight generally < 1000 g/mol) of natural origin, FDA drugs, investigational drug compounds, and other approved Non-FDA drugs (i.e. drugs approved by regulatory authorities of other countries). The major reason for including ligands from multiple sources in our library is to ensure decent coverage of

diverse chemical classes[36]. For the compounds of natural origin, a total number of 145,963 compounds downloaded from the ZINC database were included[37–39]. Using the DataWarrior software[40] we computed drug-likeness for all compounds and selected 65,038 druglike natural products for the virtual screening steps. We skipped the drug-likeness computation step for the approved drugs category composed of the FDA (2,101) and the non-FDA drugs (10,107). This produced a ligand library with 77,246 structural models mostly collated from the ZINC database[37–39].

2.2 Target HbS macromolecule

First, the X-ray crystallographic structure of dimerized human deoxygenated HbS (mutation E6V) was downloaded from the RCSB website[41], solved at 2.05 Å and with the Protein Databank Access code 2HBS[35]. The β-globin chain B with the hydrophobic cleft composed of βA70, βF85, and βL88 was extracted and employed in the present work as the receptor model. To account for receptor flexibility, we employed an RMSD-based geometric clustering protocol performed on a 300 ns explicit solvent molecular dynamics trajectory of the solvated sickle hemoglobin with the HbS crystal structure as the starting conformation. With an RMSD cut-off of 0.5 nm, 6 clusters were identified from which the representative coordinates of the four most populated clusters were selected for docking. Following the same protocol as that of the x-ray crystal structure, the β-globin chain B in each structure was extracted and employed for docking.

2.3 High throughput virtual screening protocol targeting HbS-aggregation

We carried out the screening in two phases. In the first phase, the ligand library was screened against the crystal structure of HbS and in the second phase, the four MD-generated HbS cluster structures were employed as the receptor. Using the AutoDock tool[42,43], a grid box was generated and centered at x, y, z coordinates of 20.007 Å, 44.453 Å, 31.458 Å corresponding to the hydrophobic cleft that the Val6-bearing globin chain binds to. This cleft is defined by some amino acid residues including βA70, βF85, and βL88. Polar hydrogen atoms and partial charges were added to the receptor structure after which they were saved as PDBQT files. PDBQT files were also generated for each ligand in the virtual library, after which the library was docked using Autodock Vina[44] against the prepared HbS receptor structure (the β-globin chain B extracted from the crystal HbS structure). The docking protocol treats the ligand structure as flexible while keeping the receptor molecule rigid. After the first round of screening against the crystal structure, using an energy cut-off of -9.0 kcal/mol, a total of 6,930 compounds of natural origin were identified. In the case of the FDA and non-FDA (including investigational drugs) drugs, we employed a -8.0kcal/mol energy cut-off to obtain 290 and 1,497 virtual hits, respectively. In total the first virtual screening round yielded 8,717 virtual hits which would later be employed in the second phase of screening. To incorporate protein dynamics, we employed an ensemble docking approach[45] where in addition to the crystal structure, the four cluster structures from HbS MD simulations were employed. We successfully utilized this approach in our recent work for identifying potent experimentally validated inhibitors of the SARS-CoV2 main protease enzyme[36]. A docking grid was generated for each of the four MD-generated HbS conformations (see Table S1) and using Autodock vina[44], the 8,717 virtual hits from the first screening were docked into the hydrophobic cleft in each of the four MD-generated HbS conformers. The computed affinity scores were then averaged for each ligand over the four receptor structures and the same energy cut-offs described above were used for the respective ligand classes. This yielded a total of 747 virtual hits(27 FDA drugs, 214 non-FDA, and investigational drugs, and 506 natural products). To further reduce the number of identified hits we computed the interaction distance of the ligand from the Val6 hydrophobic binding site; this was done by calculating the minimum distance between the ligands and the binding site amino acids. The top 100 compounds were selected for explicit solvent MD simulations to probe their

stability.

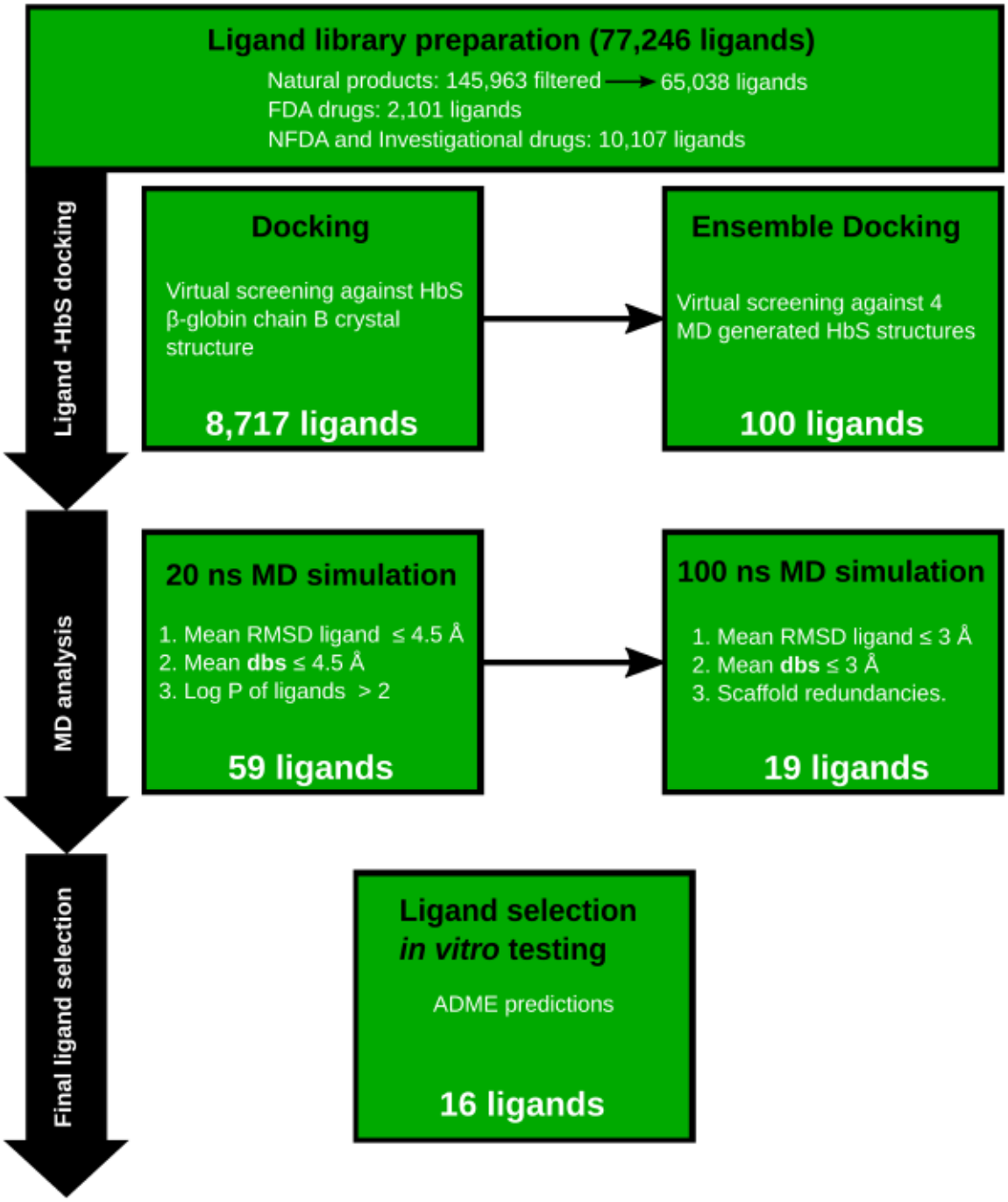

Figure 2. Flowchart of the approach employed in this work

2.4 Post-screening MD simulations

2.4.1 Simulation flow

To investigate their stability in the Val6 binding site we subjected the complexes formed between the 100 virtual hits and the HbS molecule to atomistic MD simulation. Generalized AMBER force field (GAFF)[46] parameters were generated for selected ligands in accordance with the protocol employed in our previous

work[47,48]. The MD sampling of the 100 HbS - ligand complexes was performed using GROMACS 2018[49] and AMBER14SB forcefield[50] and the TIP3P water model[51]. The simulations were initially carried out for 20 ns for each HbS - ligand complex. Using various quantities such as the root mean square deviation of the ligand RMSD $_{ligand}$, the distance between ligand and the binding site (d_{BS}), and logP values, we assessed the stability of the ligands in the binding sites. Of the 100 ligands, 59 were selected based on these criteria, and their trajectories extended to 100 ns. The best performing ligands based on the MD results were subsequently selected for *in vitro* HbS aggregation inhibitory and antisickling tests.

2.4.2 Analysis and Visualization software

Visualization of the protein-ligand structures was done using Visual Molecular Dynamics (VMD) software[52]. Figures analysing ligand properties were made with DataWarrior[40], the protein-ligand interactions were analyzed and plotted with LigPlot+[53,54]. The analysis of the simulations was also realized using various tools of the GROMACS package[55] while the resulting data was plotted using Xmgrace[56]. Figure 2 provides an overview of the approach used in this work as a flow chart.

3 Results and Discussion

In this work, we have employed both CADD and experimental approaches in search of compounds that are capable of inhibiting the aggregation of HbS, we focused our attention particularly on existing drugs to reduce the amount of time needed for the drugs to be ready clinically. Natural products, which exhibit a wide range of pharmacophores and a high degree of stereochemistry create a great source of possible hits. The protocol adopted here is similar to that followed by [37,48] in search of potential inhibitors for SARS-COV-2 main protease.

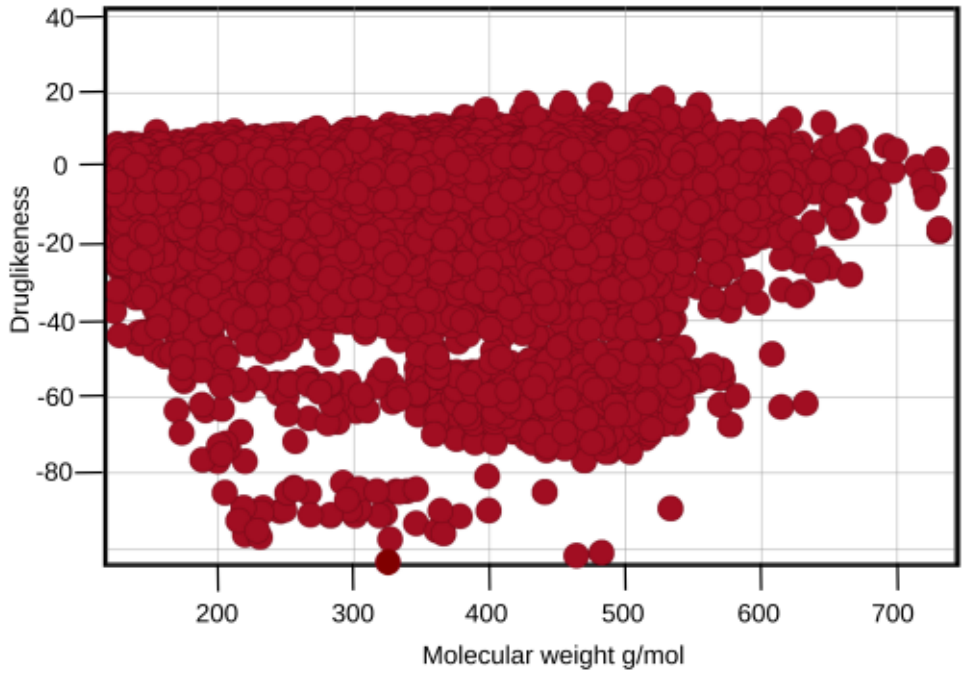

Figure 3. Distribution of ligands of natural origin based on drug-likeness.

3.1 Filtering of the Compounds of natural origin

Figure 3 shows the distribution of the compounds of natural origin based on drug-likeness, any compound that falls in the positive range (a total of 65,038) is said to be drug-like while those in the negative range are not drug-like. Selecting the drug-like compounds eliminates the need to spend computing resources investigating molecules with structurally inherent inability to cross biological membranes.

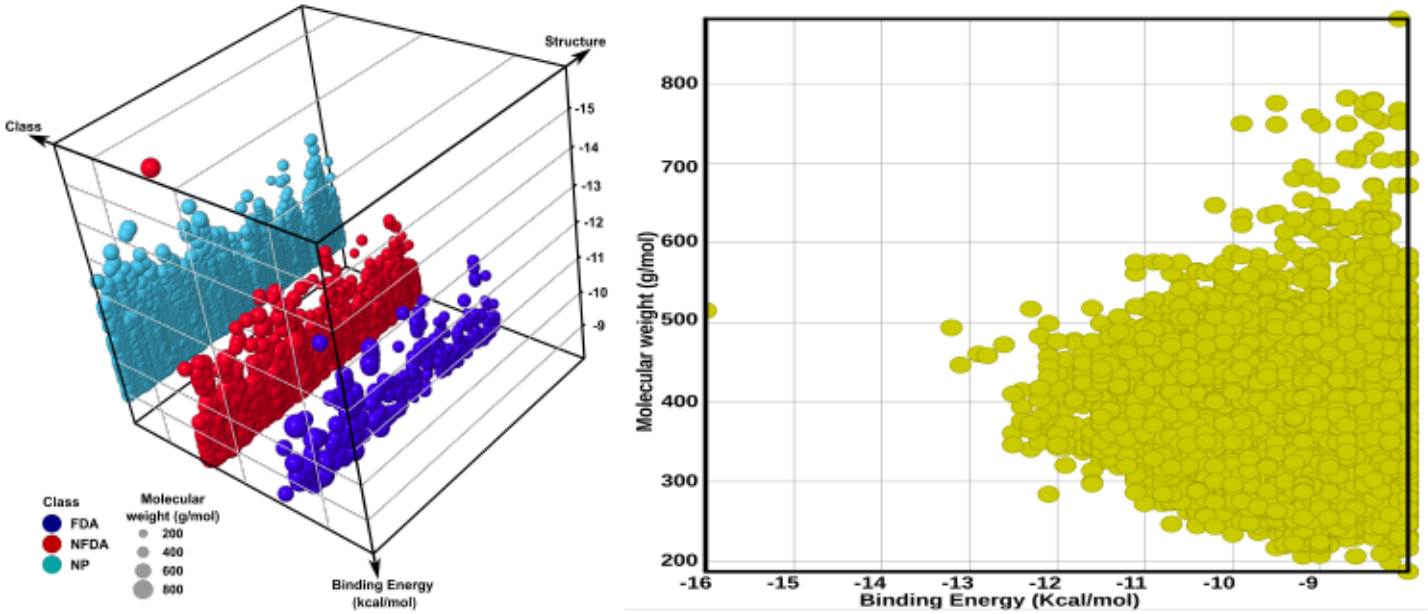

Figure 4. Distribution of the top 8,717 compounds selected from virtual screening against HbS. (A) A 3D plot of the compounds showing the distribution across ligand classes, in terms of the chemical structure of compounds, molecular weight (MW), and ΔG. The ligand library consists of FDA-approved drugs, non-FDA, investigational drugs, and natural products. (B) Distribution of the compounds in terms of binding free energies and molecular weight (g/mol).

3.2 High throughput virtual screening of the ligand library

As described in the methods section, the screening was carried out in two phases. Following the first screening, an energetic cutoff ($\Delta G \leq$ -9.0kcal/mol for compounds of natural origin and $\Delta G \leq$ -8kcal/mol for the existing drugs) was exerted to select the most promising ligands yielding a total of 8,717 compounds. The extent to which compounds from each database contribute to the top performing 8,717 compounds was analyzed (Figure 4A). From the analysis, the compounds from the natural products database dominate which is not surprising because the majority of the screened compounds are natural products. The physicochemical properties of the selected compounds were also analyzed. From Figure 4B it can be seen that compounds that fulfilled the energetic cut-off chosen have molecular weights varying from 200 to almost 900 g/mol with most having a molecular weight of 200 to 500 g/mol. From this figure, it is seen that there is no definite relationship between the molecular weight of the compounds and the binding free energy strength. Therefore, we decided not to prioritise our filtering based on physicochemical properties. The top hits were also analysed for the dominant chemical fragments, and most of the fragments found are present in commercially available drugs. The resulting compounds from the first screening were employed for the ensemble docking. After the screening, the ΔG value of each compound was obtained by averaging over the four HbS conformations. A total of 747 ligands that fulfilled the energetic cutoffS was selected.

These ligands were further streamlined to 100 based on their minimum distance to the binding site. The full list of the top 100 compounds chosen for *in-silico* validation is provided in Table S2.

3.3 Binding of the top three predicted compounds from each class

Here, we discuss the binding of the best three predictions from each class of the compounds employed in virtual screening, and further information about these compounds is provided. LigPlot+ [54,55] was used to analyze the protein-ligand interactions.

Compound 1: The compound is a natural product with ZINC ID (ZINC05433944) and MW 487.56g/mol. This ligand is buried within a side of the binding site of the receptor where it is able to form hydrophobic contacts with several residues, some of these contacts include Leu106, Gly107, Val134, Ala138, Val67, and Lys66. The pose also allows for hydrophobic contacts with Ala70, Leu88, and Phe85 as well as other amino acid residues in the Val6 binding site vicinity.

Compound 2: The ZINC database compound (ZINC01322039) is a natural product with MW 385.4g/mol. Unlike compound 1, this compound forms a hydrophobic contact with only one of the Val6 binding site residues, Ala70. It also establishes hydrogen bonding contacts with the His92 side chain.

Compound 3: This molecule is a natural product with ID (ZINC08952578, MW 508.578 g/mol). Similar to compound1, this compound is able to establish extensive contacts that are mostly hydrophobic in nature with the active site residues. Some of these contacts include Leu96, Leu106, Val98, Phe42, Leu28, Asp73,
Lys66, His92 and His63. The pose of this compound in the active site allows for hydrophobic contacts with only two out of the three Val6 binding site residues namely Ala70 and Leu88.

Compound 64: This compound with ZINC ID ZINC000098209140 and MW 356.389 g/mol) is an investigational drug. Similar to compound 3, this 5-ringed compound establishes hydrophobic contacts with Ala70 and Leu88. The compound relies solely on non-polar contacts without any indication of hydrogen bonds.

Compound 65: This compound with ZINC ID ZINC000034074273 is also an investigational drug. Similar to compound 64, this compound also contains 5 rings and it forms extensive contacts that are mostly hydrophobic in nature including a critical contact with Ala70. The three fluorine atoms form hydrophobic interactions with Phe103 and Leu106.

Compound 66: OnoRS–411, also known as Pranluksat is a cysteinyl leukotriene receptor-1 antagonist used for the treatment of allergic rhinitis and asthma symptoms. Pranluksat forms multiple hydrophobic contacts involving 13 residues, including the three Val6 binding site residues; Ala70, Leu88, and Phe85, and neighboring residues, e.g. Asp73 and Thr84.

Compound 92: Erivedge (MW: 421.3g/mol with generic name Vismodegib) is used to treat patients with locally advanced basal cell carcinoma. In interacting with the receptor (HbS), Erivedge forms hydrogen bond interaction with the side chain of His92 and several hydrophobic contacts, including critical contacts with Leu88 and Ala70 but not Phe85.

Compound 93: This compound is an FDA-approved drug, Nilotinib is a kinase inhibitor used for the treatment of Chronic Myeloid Leukemia (CML). Nilotinib just like the other compounds, also forms extensive hydrophobic contacts with all three hydrophobic residues in the Val6 binding site, as well as other hydrophobic residues that have been identified for other compounds. No hydrogen bond contact interaction is found here.

Compound 94: This is an approved drug called Lumacaftor, and it is sold under the brand name Orkambi; it is used in combination with Ivacaftor for the treatment/ management of cystic fibrosis. Lumacaftor has a characteristic 3-ring system that fits comfortably within the hydrophobic cavity. Similar to the contact found in Erivedge, Lumacaftor also forms a hydrogen bond with His63 side chain, while

also maintaining a hydrophobic contact with only one of the three of the Val6 binding site residues as well as most other contacts that have been identified for other representative compounds.

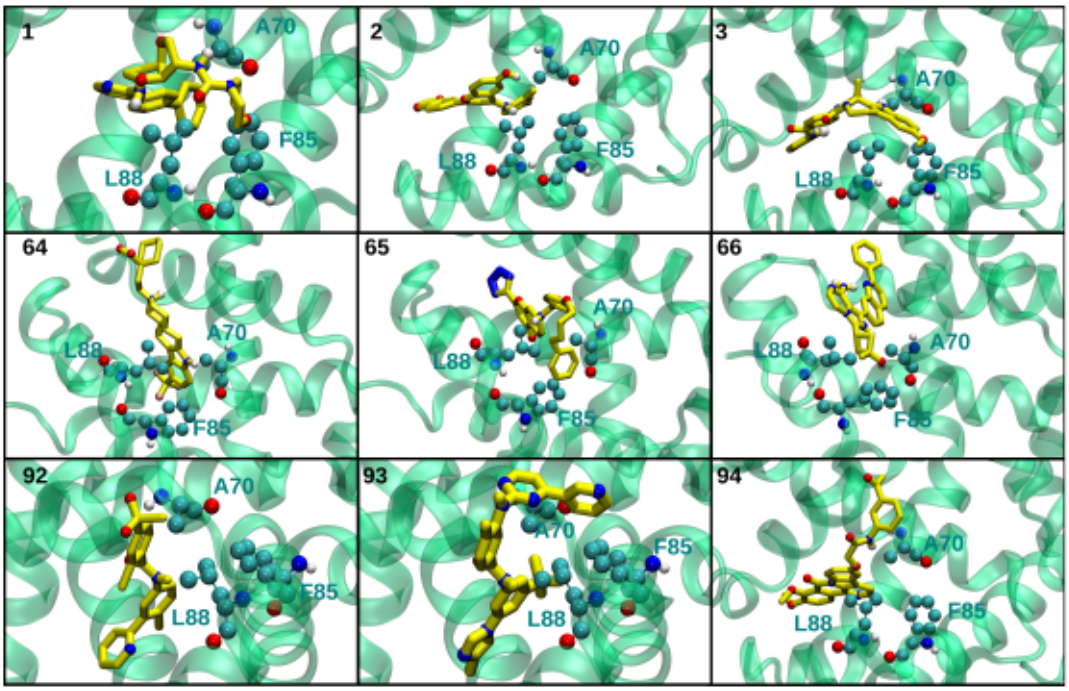

Figure 5. The binding poses of the top nine compounds 1, 2, 3, 64, 65, 66, 92, 93, 94. These compounds were chosen based on their binding free energy values ΔG and their proximity to the Val6 binding site residues as indicated by these plots. The same protein and ligand representation, as well as color scheme as in Figure 1 is used.

3.4 In-silico validation of selected ligands using MD-simulation

While molecular docking allows for the estimation of ligands that fit well into the HbS-Val6 binding site and might indicate that these compounds are good binders, MD simulation allows us to differentiate between the good and bad binders by taking into consideration the dynamics of the systems. The stability of the top-performing compounds identified via virtual screening was investigated by carrying out an atomistic MD simulation on the prospective HbS inhibitors. This study starts with 100 independent MD simulations for the ligand-HbS complexes for an initial 20 ns. Of this number, 63 are of compounds of natural origin, 9 FDA-approved drugs and 28 are either non-FDA/investigational (NFDA/INV) drugs. In order to select the compounds that will be extended to the next simulation phase, the criteria listed below were exerted:

1. The Root Mean square deviation of the ligand (mean RMSD $_{\text{ligand}} \leq 4.5$ Å).
2. The distance of the ligand to the Val 6 binding sites ($d_{BS} \leq 4.5$ Å).
3. LogP value of the ligands (logP > 2).

The first two parameters were computed for only the last 5 ns of the MD simulation, out of the 100 compounds, 61 ligands fulfilled both the distance and RMSD criteria. These resulting 62 ligands were further filtered based on their logP values (see Figure6). In total 59 ligands were selected, i.e. these ligands stayed bound within the last 5 ns of the 20 ns simulations, for which the MD simulations were then subsequently extended to 100 ns. For the complexes extended to 100 ns, the last 25 ns of the trajectories

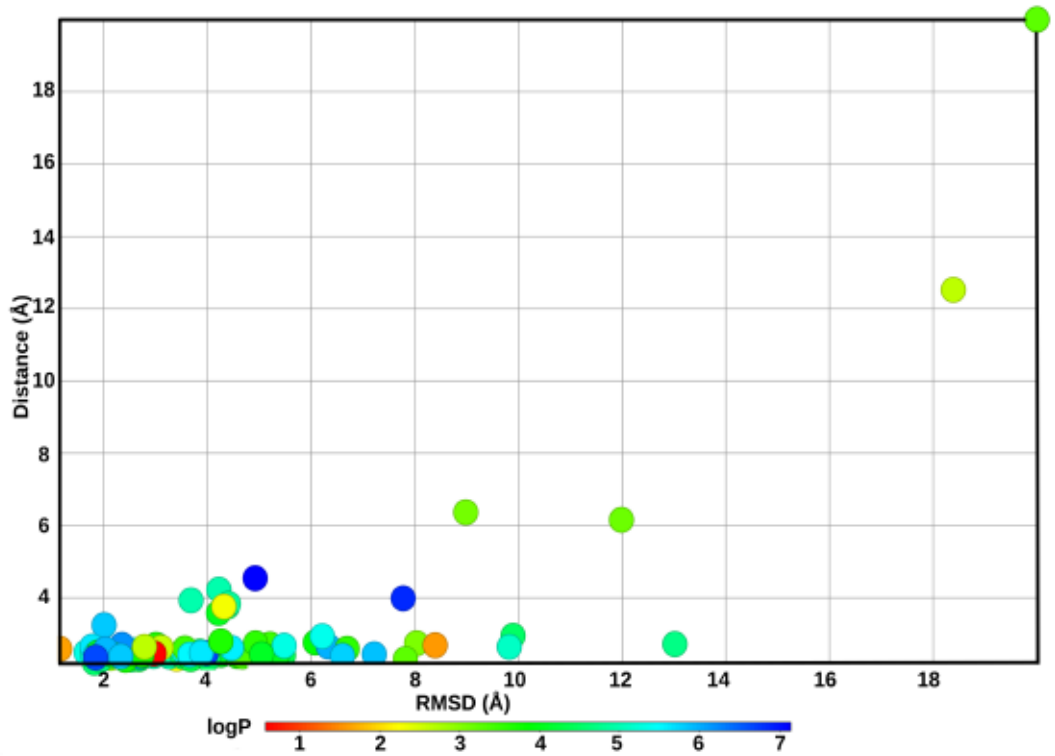

Figure 6. The distribution of ligands in terms of mean $RMSD_{ligand}$, (d_{BS} and LogP (color). RMSD and distance cut-off of ≤ 4.5 Å were chosen to determine the ligands for the extension of the MD simulations to 100 ns. 61 ligands met the chosen cutoff and were further filtered based on LogP. 2 out of the 61 ligands recorded logP value of less than 2.

were employed for analysis, applying criteria slightly different from the one above was used for the selection of the ligands that will be further tested *in-vitro*. The criteria are listed below:

1. RMSD of the ligand (mean $RMSD_{ligand} \leq 3$ Å cut-off)
2. Distance of the ligand to the binding site ($d_{BS} \leq 3$ Å cut-off)
3. Reducing scaffold/ structural redundancies.
4. ADME (absorption, distribution, metabolism, and elimination) predictions to identify compounds that are likely to efficiently penetrate the gastrointestinal (GI) epithelial barrier.

3.5 Filtering based on 20 ns MD simulation

3.5.1 Ligand flexibility (RMSD)

To quantify the flexibility and reorientation of the ligands that remained in the binding site within the simulation time, we computed the RMSD of the ligands for the last 5 ns of the 20 ns MD simulation. A high RMSD value usually indicates instability of the ligands in the binding site. Average $RMSD_{ligand}$ cut-off of ≤ 4.5 Å was chosen to differentiate between good and bad binders and also to streamline the number of compounds that will be extended to 100 ns. The values obtained for the selection library vary from 1.6 Å to 18.4 Å. Out of the 100 ligands, only 5 compounds recorded $RMSD_{ligand}$ mean values above 9 Å namely Idronoxyl (INV, 18.4 Å), Ono-rs 411 (NFDA, 13.0 Å), N-Desmethyleletriptan (NFDA, 12.0 Å), ZINC08792371 (NP, 9.9 Å), and Nilotinib (FDA, 9.8 Å). Some compounds that featured in the top binders list according to ΔG obtained from docking met the cut-off chosen here. Examples of these compounds include Erivedge (2.7Å), Linsitnib (1.9 Å), ZINC01322039 (1.67 Å), ZINC05433944 (3.4

Å), and ZINC08952578 (1.95 Å). Only one compound in the FDA class recorded a mean RMSD value above the catchment RMSD cut-off. Nilotinib, Ono-rs 41, and idronoxyl, which were identified to belong to the best binders via docking studies, display very high flexibility in the binding site, with an RMSD value way above the 4.5 Å cut-off chosen. In addition to these three compounds, there are also some other ligands with $RMSD_{ligand}$ value above the cut-off mark of 4.5 Å that showed great binding affinity energy values during the virtual screening. This highlights the importance of including structural dynamics when searching for prospective inhibitors. These include but are not limited to, ZINC4083487 (8.0 Å) which was part of the best 10 compounds in the natural products class, the investigational drug Equol (7.8 Å) and Adozelesin (5 Å).

3.5.2 Distance between the ligand and Val6 Binding site

To identify the best binding ligands based on 20 ns MD simulations, another criterion that was put into consideration is the distance between the ligand and any of the residues in the Val6 binding site. To identify these compounds that are in close proximity with the binding site a cut-off of 4.5 Å mean distance to the binding site was chosen. The majority of the compounds recorded mean distance values ranging between 2-4.5 Å. Some compounds such as Idronoxil recorded a very high mean distance value of 12.25 Å which is not surprising because it also showed high flexibility in the binding site and recorded the highest ligand RMSD value of 18.4 Å. Other compounds that recorded very high mean distance values include N-Desmethyleletriptan with 6.1 Å and investigational drug ZINC000068205 with a mean distance of 6.4 Å.

3.5.3 Lipophilicity (LogP) values of the ligands

Lipophilicity that is commonly referred to as LogP is a parameter employed in drug design due to its influence on absorption, distribution, permeability, as well as the routes of drug clearance. A total of 62 compounds that satisfied the RMSD and distance criteria described above were further filtered for their lipophilicity. Of the 61 compounds, only two of the compounds namely Genistein and Verubecestat recorded a logP value less than 2, meaning they have very low lipophilicity. In total 59 compounds were selected for the simulation extension, indicating that all these ligands stayed bound to the binding site and also have very high lipophilicity which makes them very good hit candidates.

3.6 Filtering based on 100 ns MD simulation

3.6.1 Ligand flexibility and Distance to the Val6 binding site

To identify the best binding ligands based on 100 ns simulations, the RMSD of the ligand ($RMSD_{ligand}$) and the distance (d_{BS}) between all the ligands simulated and the Val6 binding sites were computed using the last 25 ns of the 100 ns simulation. In addition to these, the ligands were also filtered based on scaffold redundancies after sorting-based d_{BS}. A cut-off of 3 Å was chosen for both the mean $RMSD_{ligand}$ and d_{BS}. The mean $RMSD_{ligand}$ ranges between 1.3 Å and 6.7 Å and the mean d_{bs} range between 1.5 Å and 5.4 Å. Out of the 59 ligands simulated, 45 ligands fulfilled the mean d_{BS} cut-off of 3.0 Å. These 45 ligands were now sorted based on their $RMSD_{ligand}$ and reducing the structural and scaffold redundancies. The criteria for the mean $RMSD_{ligand}$ cut-off, and scaffold /structural redundancies reduction were fulfilled by 19 out of the 45 ligands.

3.6.2 ADME predictions

After reducing the number of choices/ possibilities of potential HbS-aggregation inhibitors/ligands to a sufficiently small library of 19 ligands, the top binders were further analyzed to identify compounds that are likely to efficiently penetrate the GI epithelial barrier by performing ADME predictions using the swissADME webtool[57]. ADME predictions would give an indication of compounds that can cross the

plasma membrane. After performing the ADME predictions, the 19 ligands were further narrowed down to 16 ligands which will be further tested experimentally for their prospective HbS-aggregation inhibitory capabilities.

Conclusion

Using a rigorous computer aided drug design approach, we have collated and screened a ligand library of 77,246 compounds in search of compounds that are capable of binding to HbS and preventing its polymerization. The ligand library consists of compounds from natural origin, FDA drugs, non-FDA drugs and investigational drugs. To ensure the diversity of chemical classes, we decided to include ligands from different sources. After an initial molecular docking we identified a number of compounds with impressive energetic values but we did not limit the selection of our top compounds to their energetic values. The compounds were also evaluated for their ability to form contacts with any of the three residues (βA70, βL88 and βF85) that make up the Val6 binding site. The analysis of the top hits revealed that they belong to different chemical classes and most chemical fragments present in them are present in commercially available drugs.

The protein-ligand complexes of the top 100 compounds identified via molecular docking were further subjected to explicit solvent MD simulations to validate their stability in the Val6 binding site. In total, we identified 16 ligands that fulfill all the criteria chosen to differentiate the poor binders from the good binders. Out of the 16 ligands identified, 13 of them belong to the natural product class, 2 are FDA approved drugs and 1 is an investigational drug. The top 16 ligands identified via the computational approach will be tested experimentally to validate their inhibitory activities. We believe that the findings presented in this work will be of importance in the development of novel therapeutics for sickle cell disease.

Acknowledgements

The authors thank Jennifer Loschwitz for her contributions. We thank RWTH-Aachen University for providing the computing resources for all computational work under the project rwth0518.

Competing interests

The authors declare no competing interests

References

1. Kapoor, S., Little, J. A. & Pecker, L. H. Advances in the treatment of sickle cell disease. *Mayo Clin. Proc.* **93**, 1810–1824, DOI: https://doi.org/10.1016/j.mayocp.2018.08.001 (2018).
2. Pauling, L., Itano, H. A., Singer, S. J. & Wells, I. C. Sickle cell anemia, a molecular disease. *Science* **110**, 543–548 (1949).
3. Ingram, V. M. *et al.* Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* **180**, 326–328 (1957).
4. Creary, M., Williamson, D. & Kulkarni, R. Sickle cell disease: current activities, public health implications, and future directions. *J. women's health* **16**, 575–582 (2007).
5. Madigan, C. & Malik, P. Pathophysiology and therapy for haemoglobinopathies; part i: sickle cell disease. *Expert. reviews molecular medicine* **8**, 1–23 (2006).
6. Roberts, I. & de Montalembert, M. Sickle cell disease as a paradigm of immigration hematology: new challenges for hematologists in europe. *Haematologica* **92**, 865–871 (2007).
7. Ilesanmi, O. O. Pathological basis of symptoms and crises in sickle cell disorder: implications for counseling and psychotherapy. *Hematol. reports* **2** (2010).
8. Allison, A. C. Protection afforded by sickle-cell trait against subtertian malarial infection. *Br. medical journal* **1**, 290 (1954).
9. Olori, E. O., Olubiyi, O. O. & Babalola, C. P. Hoslundia opposita and other nigerian plants inhibit sickle hemoglobin polymerization and prevent erythrocyte sickling: a virtual screening-guided identification of bioactive plants. *bioRxiv* (2021).
10. Piel, F. B., Hay, S. I., Gupta, S., Weatherall, D. J. & Williams, T. N. Global burden of sickle cell anaemia in children under five, 2010–2050: modelling based on demographics, excess mortality, and interventions. *PLoS medicine* **10**, e1001484 (2013).
11. Piel, F. B., Steinberg, M. H. & Rees, D. C. Sickle cell disease. *New Engl. J. Medicine* **376**, 1561–1573 (2017).
12. Piel, F. B. *et al.* Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *The Lancet* **381**, 142–151 (2013).
13. Rogers, D. W. *et al.* Early deaths in jamaican children with sickle cell disease. *Br Med J* **1**, 1515–1516 (1978).
14. Vichinsky, E., Hurst, D., Earles, A., Kleman, K. & Lubin, B. Newborn screening for sickle cell disease: effect on mortality. *Pediatrics* **81**, 749–755 (1988).
15. Almeida, A. M., Henthorn, J. S. & Davies, S. C. Neonatal screening for haemoglobinopathies: the results of a 10-year programme in an english health region. *Br. journal haematology* **112**, 32–35 (2001).
16. Bardakdjian-Michau, J. *et al.* Neonatal screening for sickle cell disease in france. *J. Clin. Pathol.* **62**, 31–33 (2009).
17. Salinas Cisneros, G. & Thein, S. L. Recent advances in the treatment of sickle cell disease. *Front. Physiol.* **11**, 435 (2020).
18. Gardner, R. V. Sickle cell disease: advances in treatment. *Ochsner journal* **18**, 377–389 (2018).

19. Telen, M. J., Malik, P. & Vercellotti, G. M. Therapeutic strategies for sickle cell disease: towards a multi-agent approach. *Nat. reviews Drug discovery* **18**, 139–158 (2019).

20. Nurain, I. O., Bewaji, C. O., Johnson, J. S., Davenport, R. D. & Zhang, Y. Potential of three ethnomedicinal plants as antisickling agents. *Mol. pharmaceutics* **14**, 172–182 (2017).

21. Olubiyi, O. O., Olagunju, M. O. & Strodel, B. Rational drug design of peptide-based therapies for sickle cell disease. *Molecules* **24**, 4551 (2019).

22. Vichinsky, E. *et al.* A phase 3 randomized trial of voxelotor in sickle cell disease. *New Engl. J. Medicine* **381**, 509–519 (2019).

23. Henry, E. R. *et al.* Treatment of sickle cell disease by increasing oxygen affinity of hemoglobin. *Blood, The J. Am. Soc. Hematol.* **138**, 1172–1181 (2021).

24. Bank, A., Markowitz, D. & Lerner, N. Gene transfer. a potential approach to gene therapy for sickle cell disease. *Annals New York Acad. Sci.* **565**, 37–43 (1989).

25. Niihara, Y. *et al.* A phase 3 trial of l-glutamine in sickle cell disease. *New Engl. J. Medicine* **379**, 226–235 (2018).

26. Cokic, V. P. *et al.* Hydroxyurea induces fetal hemoglobin by the nitric oxide–dependent activation of soluble guanylyl cyclase. *The J. clinical investigation* **111**, 231–239 (2003).

27. Hassan, A., Awwalu, S., Okpetu, L., Waziri, A. D. *et al.* Effect of hydroxyurea on clinical and laboratory parameters of sickle cell anaemia patients in north–west nigeria. *The Egypt. J. Haematol.* **42**, 70 (2017).

28. Inusa, B. P. D. *et al.* Low-dose hydroxycarbamide therapy may offer similar benefit as maximum tolerated dose for children and young adults with sickle cell disease in low-middle-income settings. *F1000Research* **7** (2018).

29. Imaga, N. *et al.* Antisickling property of carica papaya leaf extract. *Afr. J. Biochem. Res.* **3**, 102–106 (2009).

30. Ekeke, G. & Shode, F. The reversion of sickled cells by cajanus cajan. *Planta medica* **51**, 504–507 (1985).

31. Moody, J. *et al.* Anti-sickling potential of a nigerian herbal formula (ajawaron hf) and the major plant component (cissus populnea l. cpk). *Phytother. Res. An Int. J. Devoted to Pharmacol. Toxicol. Eval. Nat. Prod. Deriv.* **17**, 1173–1176 (2003).

32. Nwosu, F., Dosumu, O. & Okocha, J. The potential of terminalia catappa (almond) and hyphaene thebaica (dum palm) fruits as raw materials for livestock feed. *Afr. J. Biotechnol.* **7** (2008).

33. Olubiyi, O. O., Olagunju, M. O., Oni, J. O. & Olubiyi, A. O. Structural basis of antisickling effects of selected fda approved drugs: A drug repurposing study. *Curr. computer-aided drug design* **14**, 106–116 (2018).

34. x. *Curr. computer-aided drug design* **x**, x–x (2022).

35. Harrington, D. J., Adachi, K. & Royer Jr, W. E. The high resolution crystal structure of deoxyhemoglobin s. *J. molecular biology* **272**, 398–407 (1997).

36. Olubiyi, O. O., Olagunju, M., Keutmann, M., Loschwitz, J. & Strodel, B. High throughput virtual screening to discover inhibitors of the main protease of the coronavirus sars-cov-2. *Molecules* **25**, 3193 (2020).

37. Irwin, J. J. & Shoichet, B. K. Zinc- a free database of commercially available compounds for virtual screening. *J. chemical information modeling* **45**, 177–182 (2005).

38. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. Zinc: a free tool to discover chemistry for biology. *J. chemical information modeling* **52**, 1757–1768 (2012).

39. Sterling, T. & Irwin, J. J. Zinc 15–ligand discovery for everyone. *J. chemical information modeling* **55**, 2324–2337 (2015).

40. Sander, T., Freyss, J., von Korff, M. & Rufener, C. Datawarrior: an open-source program for chemistry aware data visualization and analysis. *J. chemical information modeling* **55**, 460–473 (2015).

41. Berman, H. M. *et al.* The protein data bank. *Nucleic acids research* **28**, 235–242 (2000).

42. Goodsell, D. S., Morris, G. M. & Olson, A. J. Automated docking of flexible ligands: applications of autodock. *J. molecular recognition* **9**, 1–5 (1996).

43. Santos-Martins, D., Forli, S., Ramos, M. J. & Olson, A. J. Autodock4zn: an improved autodock force field for small-molecule docking to zinc metalloproteins. *J. chemical information modeling* **54**, 2371–2379 (2014).

44. Trott, O. & Olson, A. J. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. computational chemistry* **31**, 455–461 (2010).

45. Amaro, R. E. *et al.* Ensemble docking in drug discovery. *Biophys. journal* **114**, 2271–2278 (2018).

46. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. computational chemistry* **25**, 1157–1174 (2004).

47. Loschwitz, J. *et al.* Novel inhibitors of the main protease enzyme of sars-cov-2 identified via molecular dynamics simulation-guided in vitro assay. *Bioorganic Chem.* **111**, 104862 (2021).

48. Loschwitz, J. *et al.* Dataset of amber force field parameters of drugs, natural products and steroids for simulations using gromacs. *Data Brief* **35**, 106948 (2021).

49. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. chemical theory computation* **4**, 435–447 (2008).

50. Maier, J. A. *et al.* ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. chemical theory computation* **11**, 3696–3713 (2015).

51. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The J. chemical physics* **79**, 926–935 (1983).

52. Humphrey, W., Dalke, A. & Schulten, K. Vmd: visual molecular dynamics. *J. molecular graphics* **14**, 33–38 (1996).

53. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. Ligplot: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering, design selection* **8**, 127–134 (1995).

54. Laskowski, R. A. & Swindells, M. B. Ligplot+: multiple ligand–protein interaction diagrams for drug discovery (2011).

55. Van Der Spoel, D. *et al.* Gromacs: fast, flexible, and free. *J. computational chemistry* **26**, 1701–1718 (2005).

56. Turner, P. Xmgrace, version 5.1. 19. *Cent. for Coast. Land-Margin Res. Or. Graduate Inst. Sci. Technol. Beaverton, OR* (2005).

57. Daina, A., Michielin, O. & Zoete, V. Swissadme: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. reports* **7**, 1–13 (2017).

Supplementary Information

Computer-aided drug design-directed experimental identification of novel inhibitors of sickle hemoglobin polymerization

Olagunju Maryam[1], Olujide Olubiyi[1,2,3], and Birgit Strodel[1,4]

[1]Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany.

[2]Department of Pharmaceutical and Medicinal Chemistry, College of Pharmacy, Afe Babalola University, Ado-Ekiti,361212, Nigeria.

[3]Department of Pharmaceutical and Medicinal Chemistry, College of Pharmacy, Afe Babalola University, Ado-Ekiti,361212, Nigeria.

[3]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf,Germany.

Table S1. Showing the docking grid measurements generated for the X-ray crystallographic structure and each of the four MD-generated HbS conformations.

Structure	**Dimensions [Å]**		
	X	Y	Z
X-ray	42.00	34.00	20.00
Cluster 1	62.00	40.00	40.00
Cluster 2	46.00	38.00	18.00
Cluster 3	48.00	44.00	24.00
Cluster 4	60.00	36.00	40.00

Table S2. List of the top 100 compounds. These compounds include FDA-approved drugs, other drugs, natural products. The average DG values for the four representative structures of HbS obtained from MD simulation and the distance between the ligands and the val6 binding site residues are also presented are average values obtained. The compounds highlighted in bold are discussed in more detail in the text.

S/N	Accession ID	Compound name	ΔG [kcal/mol]	d_{bs} [Å]
Natural Products				
1	**ZINC05433944**	**—**	**-9.45**	**2.26**
2	**ZINC01322039**	**—**	**-9.25**	**2.31**
3	**ZINC08952578**	**—**	**-9.53**	**2.34**
4	ZINC08792157	—	-9.13	2.39
5	ZINC08918447	—	-9.18	2.42
6	ZINC03844548	—	-9.05	2.48
7	ZINC08918505	—	-9.08	2.52
8	ZINC08792251	—	-9.08	2.55
9	ZINC4083487	—	-9.08	2.55
10	ZINC08792168	—	-9.83	2.56
11	ZINC08792190	—	-9.20	2.57
12	ZINC08792170	—	-9.30	2.58
13	ZINC12604545	—	-9.05	2.58
14	ZINC04042527	—	-9.40	2.59
15	ZINC08792130	—	-9.40	2.62
16	ZINC08792274	—	-9.10	2.65
17	ZINC06167717	—	-9.38	2.68
18	ZINC13403038	—	-9.38	2.68
19	ZINC04235407	—	-9.18	2.69
20	ZINC12865606	—	-9.15	2.70
21	ZINC12865643	—	-9.12	2.70
22	ZINC08792371	—	-9.03	2.70
23	ZINC08918448	—	-9.25	2.75
24	ZINC02114520	—	-9.18	2.75
25	ZINC02117398	—	-9.23	2.76
26	ZINC08792280	—	-9.25	2.76
27	ZINC06090657	—	-9.13	2.78
28	ZINC02121309	—	-9.70	2.78
29	ZINC06197814	—	-9.25	2.79
30	ZINC08792234	—	-9.13	2.79

31	ZINC08918416	—	-9.40	2.80
32	ZINC09660050	—	-9.30	2.83
33	ZINC12661757	—	-9.18	2.84
34	ZINC13403731	—	-9.38	2.85
35	ZINC12865846	—	-9.43	2.88
36	ZINC02113346	—	-9.00	2.89
37	ZINC08792205	—	-9.15	2.90
38	ZINC08792140	—	-9.23	2.91
39	ZINC06111574	—	-9.08	2.91
40	ZINC02120287	—	-10.15	2.96
41	ZINC03846509	—	-9.48	2.97
42	ZINC08792253	—	-9.20	2.97
43	ZINC08918358	—	-9.13	2.97
44	ZINC06111661	—	-9.63	2.98
45	ZINC01900623	—	-9.70	2.98
46	ZINC12902247	—	-9.20	2.98
47	ZINC12900597	—	-9.13	2.98
48	ZINC02117462	—	-9.13	2.99
49	ZINC02126533	—	-9.38	2.99
50	ZINC11867167	—	-9.10	3.00
51	ZINC32124125	—	-9.03	3.01
52	ZINC08394882	—	-9.48	3.02
53	ZINC09660108	—	-9.13	3.03
54	ZINC08792129	—	-9.05	3.04
55	ZINC03846626	—	-9.20	3.05
56	ZINC00057871	—	-9.18	3.05
57	ZINC08792362	—	-9.05	3.05
58	ZINC06090653	—	-9.23	3.06
59	ZINC03190441	—	-9.58	3.06
60	ZINC09660168	—	-9.13	3.06
61	ZINC09660049	—	-9.53	3.06
62	ZINC01898162	—	-9.80	3.06
63	ZINC000000102175	—	-9.13	3.07
Non-FDA and investigational drugs				
64	**ZINC000098209140**	—	**-8.18**	**2.30**
65	**ZINC000034074273**	—	**-9.33**	**2.36**
66	**ZINC000015919406**	**Ono-Rs 411**	**-8.08**	**2.43**

67	ZINC000100071817	Linisitinib	-8.80	2.44
68	ZINC000001491943	Idronoxil	-8.00	2.48
69	ZINC000068205235	—	-8.05	2.56
70	ZINC000003806113	—	-8.35	2.63
71	ZINC000018710085	Chir-265	-8.10	2.69
72	ZINC000059749972	Radotinib	-8.95	2.70
73	ZINC000043204100	—	-8.75	2.71
74	ZINC000100037101	Clofazimine	-8.00	2.75
75	ZINC000022940637	Bafetinib	-8.38	2.76
76	ZINC000018825330	Genistein	-8.00	2.83
77	ZINC000006117750	Desmethylazelastine	-8.23	2.88
78	ZINC000003818809	—	-8.05	2.85
79	ZINC000004214704	Tariquidar	-8.70	2.85
80	ZINC000000388661	Equol	-8.03	2.91
81	ZINC000001482077	Gliquidone	-8.03	2.91
82	ZINC000096170454	—	-8.45	2.91
83	ZINC000003780340	Hypericin	-8.48	2.92
84	ZINC000000597434	Blonanserin	-8.03	2.94
85	ZINC000072316409	—	-8.25	2.96
86	ZINC000003922429	Adozelesin	-9.15	2.98
87	ZINC000002047214	—	-8.20	2.99
88	ZINC000077287124	N-Desmethyleletriptan	-8.13	3.01
89	ZINC000144542146	Verubecestat	-8.00	3.03
90	ZINC000013831791	Ptc124	-8.03	3.07
91	ZINC000043208634	Omipalisib	-8.13	3.08
FDA drugs				
92	**ZINC000040899447**	**Erivedge**	**-8.15**	**2.64**
93	**ZINC000006716957**	**Nilotinib**	**-8.63**	**2.61**
94	**ZINC000064033452**	**Lumacaftor**	**-9.18**	**2.75**
95	ZINC000068202099	Erismodegib	-8.45	2.77
96	ZINC000035328014	Ibrutinib	-8.25	
97	ZINC000011679756	Eltrombopag	-8.69	2.83
98	ZINC000005733652	Diosmetin	-8.18	2.84
99	ZINC000004175630	Orap (Pimozide)	-8.55	3.02
100	ZINC000011681534	Nebivolol	-8.38	3.07

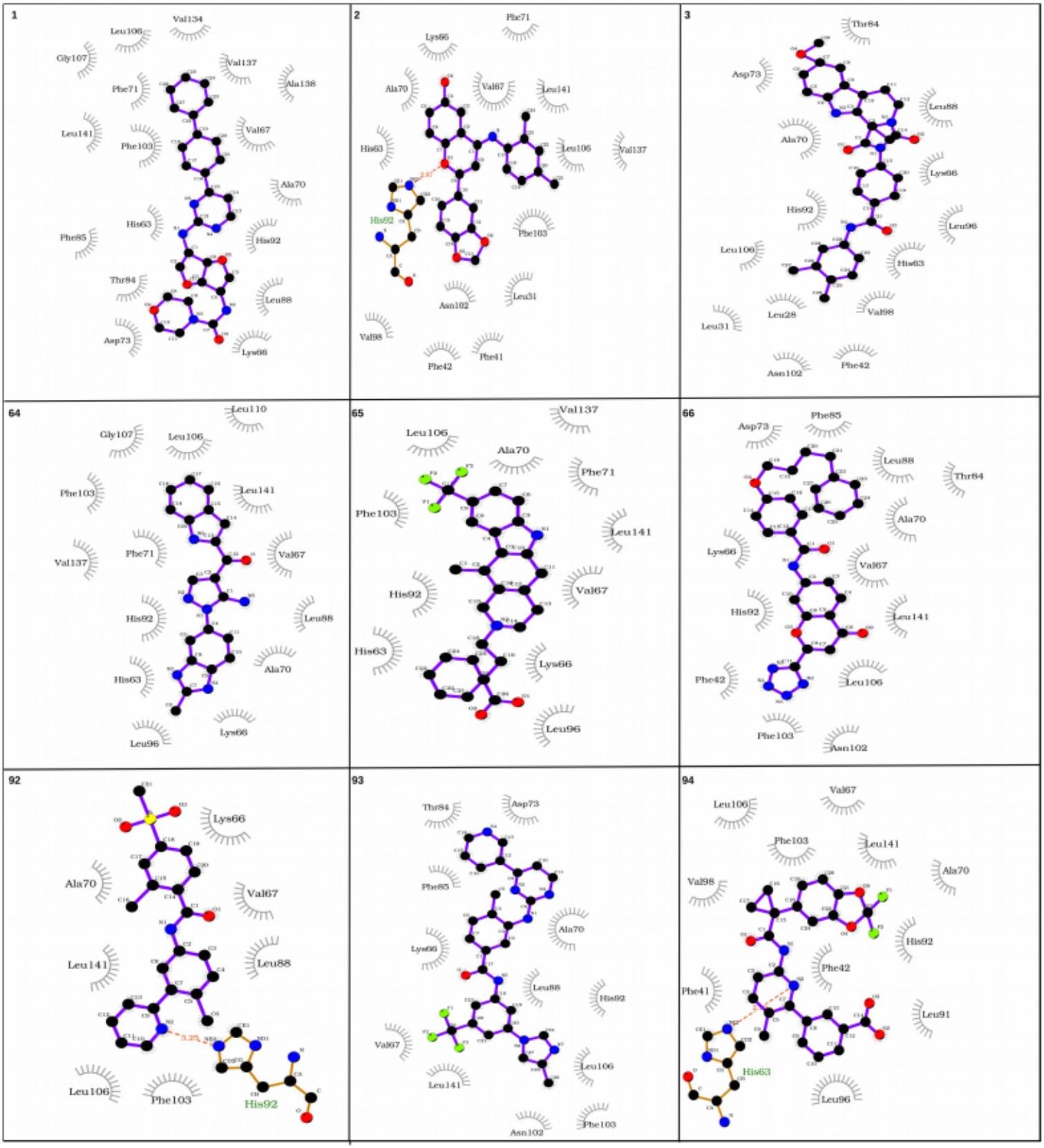

Figure S1. The interactions of the top 3 compounds from each class plotted with Ligplot+. Hydrogen bonds are indicated by orange dashed lines between the atoms involved and the

donor-acceptor distance is also written in orange and is given in Å, while the residue that forms hydrogen bond with the ligand is also shown in orange. Hydrophobic interactions are represented by gray arcs with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

3.3 *In silico* Identification of D-peptide Inhibitors of Sickle hemoglobin (HbS) Polymerization.

Declaration

This chapter was excerpted in parts from a review published in Molecules 24(24), 4551 (2019) by the authors Olujide O. Olubiyi, Maryam O. Olagunju, Birgit Strodel: Rational Drug Design of Peptide-Based therapies for Sickle Cell Disease.

Introduction

Peptide systems, short peptides in particular, have already been employed as potential inhibitors of protein aggregation in a number of pathological conditions involving protein aggregation [212–217]. The advantages associated with the use of short peptides include low overall toxicity resulting from the compatibility of peptide inhibitors with living tissues as opposed to small molecule inhibitors. Furthermore, the high chemical diversity, selectivity, and potency associated with peptide-based inhibitors are versatile, making them viable start-off points in drug discovery campaigns. With regard to protein aggregation in particular, peptide inhibitors, because of their chemical and structural composition, can offer good fits capable of interacting with protein surfaces sufficiently large to disrupt the process of protein aggregation [218]. In spite of these benefits associated with the use of peptides in therapeutics, it should be noted that they are often associated with poor pharmacokinetics relating to short half-life and low oral bioavailability [219, 220]. Because of the presence of peptidases, peptide drugs are rapidly degraded and cleared in different body compartments, leading to insufficient exposure of the target system to the administered drug. Available approaches for handling these challenges include the use of D-amino acids or non-natural residues, chemical modifications such as protecting the terminals with appropriate chemical groups (e.g., acetylating the N-terminal and amidating the C-terminal), cyclization, and incorporation of organic molecules in the peptide side chains [221–224]. Since these approaches alter the physicochemical attributes of the peptide, they can also be useful in improving the membrane partitioning of the peptide drugs. In practice, peptide penetration across cellular barriers has been accomplished via the incorporation of groups facilitating membrane crossing, like positively charged amino acids [225–227] or ligands (e.g., sugars), for recognition of membrane receptors [228]. The latter approach has been successfully employed to improve both the stability and the intestinal absorption of peptide drugs [229–231]. In the area of cancer drug delivery, where peptide-based chemotherapeutic agents are routinely required to be delivered to intracellular tar-

gets, increasing levels of success are being recorded with the development of innovative techniques like the use of cell-penetrating peptides, viral based-vectors, and nanoparticle-based systems [232–234]. It is expected that these new developments can be leveraged upon in delivering peptide-based HbS inhibitors into the intracellular compartment of RBCs.

One of the oldest ideas driving the design of HbS aggregation inhibitors relies on the acknowledgment of the causal role played by the Glu6Val β-globin mutation on disease development. Many of the earliest reported efforts sought to obtain compounds with the right combination of hydrophobicity, shape, and charge complementarity that, in principle, can bind within or in the immediate vicinity of the cavity formed by β'Phe85, β'Leu88 and β'Ala70, at the same time, possess charged groups oriented outwards. This outward projection is to prevent βVal6 of an incoming β-globin chain from binding as part of the lateral contact in HbS polymer. While the nature of βVal6 binding site would seem to place an upper limit on the molecular size of prospective inhibitors capable of binding to this site, in reality, conflicting reports have been published by different groups working on amino acid-derived inhibitors. In the late 1970s and early 1980s, Rich and co-workers examined short peptide inhibitors (up to pentapeptides) of HbS aggregation based on the belief that amphipathic nature was required to inhibit the polymerization of deoxygenated HbS [235, 236]. Out of the peptides examined, the lowest minimal inhibitor mole ratio (MIMR) of peptide to HbS necessary to prevent HbS polymerization was found for N-terminally succinylated (Phe)3, (Phe)3–Arg, and (Trp)2 (Table 4.1), where succinylation in each case served to enhance peptide solubility, or to modulate net charge, or both. It is, however, important to note that the concentrations of the peptides employed in these works were too high to be of any direct benefit in a clinical setting: The best inhibitory effects were achieved with peptide/HbS mole ratios of about 10. While structural data were lacking to categorically conclude on structure–activity relationship (SAR), the reported pattern of inhibition showed inhibitory activity increasing with peptide chain length. This could point to the fact that the nature of HbS–HbS interaction surface requires sufficiently large inhibitors to effectively disrupt crucial amino acid interactions. It is thus likely that more potent peptide inhibitors will be achieved with peptide lengths longer than those screened in these studies [235, 236]. Interestingly, a similar trend was observed with peptide inhibitors of amyloid-β aggregation, whereby highly potent aggregation inhibitors were achieved with 12-amino acid peptides, while shorter ones lacked this property [212–214]. In fact, a phage display work by Hanson et al. in 2013 successful identified a highly potent 12-residue peptide (Hb-B10, sequence CHNLLPTPWWCA) with a micromolar range (21 μmol/L) binding affinity for hemoglobin [237]. Even though the intention was not to target HbS polymerization but to aid

the clearance of circulating hemoglobin, the outcome of this research shows that indeed it is possible to obtain peptide-based systems with a HbS binding affinity required for clinical intervention.

Table 3.1: Short peptides with the best demonstrated inhibitory activity identified in [121,122], given as the minimal inhibitor mole ratio (MIMR) of peptide to HbS necessary to prevent HbS polymerization. The values are means ± standard error. "Suc" stands for succinyl: –OOC–$(CH2)_2$–CO–.

Peptide	MIMR
Suc-(L-Phe)-(L-Phe)-(L-Phe)	9.5 ± 0.5
Suc-(L-Phe)-(L-Phe)-(L-Phe)-(L-Arg)	10.0 ± 1.0
Suc-(Ln-Trp)-(L-Trp)	10.0 ± 0.5
Suc-(L-Trp)-(L-Phe)	12.5 ± 0.5

The work of Kubota and Yang was similarly founded on the special importance of the βVal6 residue during HbS polymerization by designing oligopeptides to mimic the N-terminal segments of the β-globin chain of Hb [238]. The idea behind this approach is that such peptides would interact with the βPhe85/βLeu88/βAla70 pocket (but only if the sequence was taken from HbS), or any other complementary binding site, and thus inhibit HbS polymerization. The tested peptides were indeed found to exhibit significant HbS aggregation inhibitory attributes, with the $\beta_{1\text{–}6}$ hexapeptides of the N-terminal end of both HbA (sequence VHLTPE) and HbS (sequence VHLTPV) molecules reported to increase the minimum gelling concentration (MGC) by about 75% [238]. The MGC is the concentration of HbS required to form a gel (or polymer), which is about 9.5 g/dL in the absence of peptide inhibitors, and an aggregation inhibitor is expected to increase this value. The highest inhibitory activities were obtained at peptide/heme mole ratios of between 2 and 2.5. Considering that there are four heme molecules per hemoglobin, this translates to a peptide/hemoglobin ratio of 8 to 10, which is in the MIMR range reported by Rich et al. [235, 236] and listed in Table 1. These concentrations, like those reported in [235, 236], are too high to have any clinical applicability. Truncating the length of the oligopeptides below six residues significantly reduced the inhibitory effect, which seems to suggest that the $\beta_{1\text{–}6}$ hexapeptides might indeed interact with the βVal6 binding site on the β-globin chain [238]. According to the authors, hexapeptides, but not shorter oligopeptides, are likely to preserve the secondary structure necessary to provide the complementary shape needed to interact with the βVal6 binding site. The lack of structural data, however, makes this interpretation of the experimental outcome, at best, speculative; it is possible that the peptides interacted at other sites of the HbS molecule. Hexapeptides mimicking both HbA and HbS N-terminal segments produced similar inhibitory effects, while increasing the peptide length beyond six did not improve activity, although shorter peptides were less

effective. Interestingly, in a separate work, it was observed that longer oligopeptide inhibitors involving sequences β_{1-12} (sequence VHLTPVEKSAVT), β_{3-13} (sequence LTPVEKSAVTA), β_{4-8} (sequence TPVEK), and β_{4-10} (sequence TPVEKSA) of HbS promote HbS polymerization [239], as they decrease the solubility of HbS [240]. The susceptible balance between peptide sequence, length, and structure for the capability to inhibit HbS polymerization is also demonstrated in a more recent work [241]. Akbar et al. studied the effects of 15-, 11-, 7-, and 3-mer peptides derived from one of the helices of the β-globin chain of hemoglobin. In the case of the 15-mer peptide, the sequence comprised the β-globin residues 6579 with sequence KKVLGAFS[**H**/**L**]GLAHLD, where, at position 73, the βHis73 and βLeu73 mutations were included instead of the native βAsp73, as, in HbS, these mutations were previously observed to inhibit HbS aggregation [242]. The shorter peptides with 3, 7, and 11 residues failed to inhibit polymerization, suggesting the importance of secondary structure and multiple contact points for the observed inhibitory activity. For the longer peptide, it was found that the βHis73 15-mer peptide more significantly inhibited polymerization compared with the βLeu73 15-mer peptide. The βHis73 15-mer peptide is believed to interact with βThr4 and thus disrupt the hydrogen bonding between βThr4 and β'Asp73, and also hydrophobic interactions involving βVal6 due to its spatial proximity. However, it should be mentioned that a peptide/HbS ratio of 3:1 was needed to obtain a noteworthy delay in HbS polymerization [242]. While it is likely that different hemoglobin binding sites were employed by these peptides, they represent about 70% improvement in potency over the peptides studied in earlier works [235, 236, 238]. The outcomes of the different experiments suggest that there is no simple relationship between peptide length and HbS polymerization inhibition.

In this present work, a rational approach that combines virtual screening with explicit solvent MD simulations, was used to identify prospective D-enantiomeric peptides that are capable of inhibiting HbS polymerization. We decided to focus on screening D-peptides in this study due to the advantages they offer over L-peptides in terms of high resistance to degradation by human proteases [243], stability, and good bioavalaibiliy [244]. D-peptides might therefore perform better than L-peptides as therapeutics. This approach is recently being used in developing new therapeutics for several diseases such as, but not limited to HIV/AIDs and Alzheimer's diseases [245, 246]. Here, we screened a library of 1,000 ten amino-acid long D-peptides against both the crystal structure of HbS and the MD-generated structures. The protocol employed here is the same as the protocol used in our previous study to identify prospective inhibitors against HbS aggregation from a library consisting of natural products, FDA approved drugs and non-FDA approved and investigational drugs (see chapter 3.2).

Materials and Methods

Peptide Library Generation A peptide library consisting of 1,000 ten amino-acid L-peptides was generated in total. To create the pdb files for the L-peptides, Python [247] was used to generate peptide sequences. Sequences were generated from the P10 to the P1 position. A temporary glutamine residue was used as a place holder at the beginning and end of each sequence. Using the Python library 'PeptideBuilder' [248] and the BioPython package [249] and Pymol [250] the sequences were then converted into pdb files. The glutamine used as a place holder was edited and replaced with capping groups. The N and C terminus were capped using acetyl group (ACE) and N-methyl amide (NME) group, respectively. After generating the peptides in a linear conformation, to obtain relaxed conformations for each peptide, MD simulation was carried out on each for 20 ns using GROMACS 2018 [251] and the CHARMM36 force field [252]. This was followed by cluster analysis using the GROMACS option 'gmx cluster', with a cutoff of 0.25 nm, and the center of the most populated cluster was then chosen for further analysis. After the MD simulation, the isomer configuration of each peptide was changed form L to D by flipping the signs of the x-coordinates for all atoms in a pdb file of an L-peptide to create the corresponding D-peptide [253]. The 1000 resulting D-peptides were assembled in a library.

Virtual Screening The docking screening was carried out in two phases. First, the peptides were docked against the hydrophobic cleft of β-globin chain B of the HbS crystal structure composed of βAla70, βPhe85 and βLeu88. This was followed by re-screening of the peptides against 4 MD generated HbS structures. After the first phase of docking, an energetic cut-off of -5.5 kcal/mol was exerted, i.e. all compounds that have a binding energy value ΔG of ≤ -5.5 kcal/mol were selected and moved to the second phase of the virtual screening. In the second phase of screening, an ensemble docking approach using the 4 MD structures was employed [254]. This was carried out in order to account for receptor flexibility because molecular docking treats the ligand as flexible while the receptor is kept rigid. This approach was utilized successfully in a recent work in our group for identifying inhibitors of the SARS-CoV-2 main protease enzyme [255, 256]. Using Autodock Vina, all the identified hits from the first screening were docked against the β globin chain B of the four MD generated HbS conformers. After the virtual screening, the computed energy values were averaged for each peptide over the four HbS structures. An energy cut-off of ≤ -4.5 kcal/mol was chosen. In order to further streamline the number of identified hits, the minimum distance between the peptides and the Val6 binding site was calculated and then averaged over all the four MD-generated structures. The top hits were then selected for explicit solvent MD simulations to

further probe their stability in the Val6 binding sites.

MD simulations To investigate the stability of the selected peptides obtained via virtual screening in the Val6 binding site, the peptide-HbS complexes were subjected to atomistic MD simulation. The MD sampling of the peptide-protein complexes were carried out using GROMACS 2018 [251], the CHARMM36 forcefield [252] and the TIP3P water model [257]. The protein-peptide complexes were inserted into a dodechaderon box, with a distance of at least 1.2 nm between the complex and the nearest box surfaces. The systems were solvated after which ions were added to both neutralize the system and achieve an NaCl concentration of 100 mM. Energy minimization was carried out on the systems using a steepest decent algorithm. This was followed by MD equilibration of the systems and, due to the size of the peptide-protein systems, the equilibration runs was carried out in two phases. In the first phase, the NVT equilibration was carried out for 100 ps, at a temperature T of 310 K using the velocity rescaling thermostat [258]. In the second phase, an NpT ensemble was realized by using the Nose-Hoover thermostat [259, 260] together with the Parrinello-Rahman barostat [261–263]. The pressure was adjusted to 1 bar while a temperature of 310 K was maintained during that 1 ns MD simulation. After the equilibration step, production runs were carried out for each system for an initial 20 ns. The LINCS algorithm [264] was used to constrain all bond lengths and the equations of motions were solved using the leapfrog algorithm with a time step of 2 fs. After the initial 20 ns, using two quantities, namely the $\mathrm{RMSD}_{\mathrm{peptide}}$ of the peptide and the distance between the peptide and the binding site (d_{bs}), we assessed the stability of the protein-peptide complexes. The peptides that fulfilled the cut-off criteria chosen were selected and their trajectories were extended to 100 ns.

Analysis and Visualization All complexes were visualized using Visual Molecular Dynamics (VMD) [265]. All the peptide-protein complex MD simulations were analysed using the GROMACS 2018 package [251] and the data plotted using XMGRACE [266].

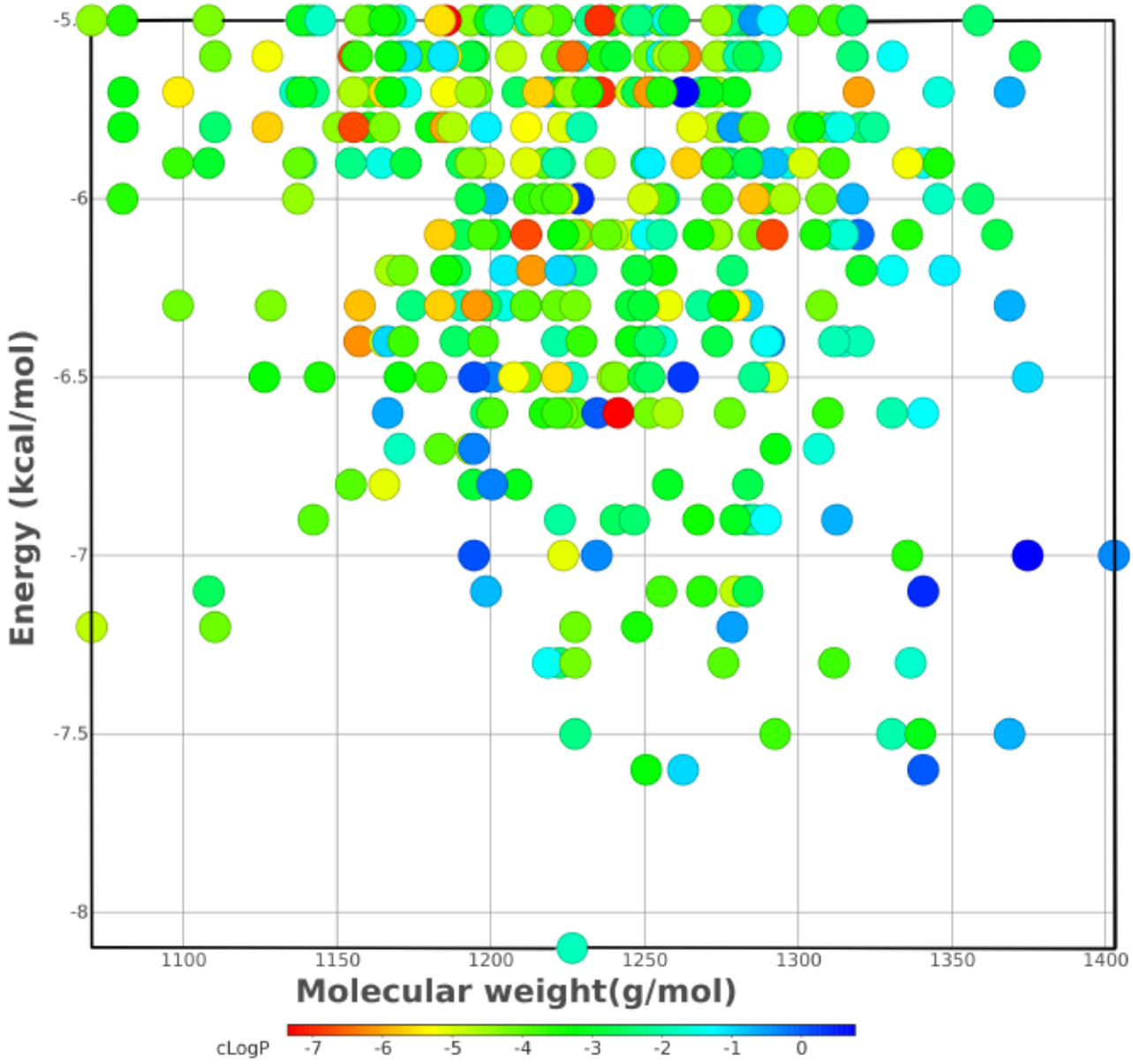

Figure 3.1: Distribution of the top 441 peptides in terms of ΔG, molecular weight and clogP value identified after virtual screening against the crystal structure of HbS.

3.3.1 Results and Discussion

A peptide library of 1000 D-peptides was generated and screened to identify possible inhibitors of HbS polymerization. To this end, we performed both docking against the β globin chain B extracted from the crystal structure of HbS, which was followed by ensemble docking. A subset of these initial peptides identified by molecular docking were then further studied by carrying out MD simulations, to account for dynamical changes and check for the stability of the identified peptides in the Val 6 binding site.

Virtual Screening of the D-peptide Library As explained above, the screening was carried out in two phases. Following the first screening, ΔG values were sorted to identify better and poorer binding peptides. An energetic cutoff of ΔG -5.5 kcal/mol was chosen to select the most promising peptides yielding a total of 441 compounds. We analyzed the physicochemical properties of the selected compounds using Data Warrior software [267]. Figure 3.1 shows the relationship between the molecular weight, binding strengths (ΔG) and clogP of all the top hits. From this figure, it can be seen that all the top hits have very large molecular weights within the range of 1070 to 1402 g/mol and the clogP values ranges from 0 to -7. Large molecular weights are present in all of the predicted peptides, which is not surprising because peptides are generally known for having high molecular weight in comparison to small molecules. Many of the compounds with logP values between -3 and -7 have ΔG values of -5.5 and -6.3 kcal/mol while the ones with smaller clogP values have better binding energy. It should be recalled that LogP is an important in the determination of drug-likeness of a compound. According to Lipinski's rule of 5, an oral drug should have a LogP value <5, ideally between 1.35 - 1.8 for good oral and intestinal absorption[268]. We decided not to prioritize filtering based on physicochemical properties such as the molecular weight, clogP values that are used when predicting oral bioavailability of drugs, since the focus of the present study is to identify peptides that are capable of binding to the hydrophobic amino acid residues in the Val6 binding site. From the figure, it is seen that there is no clearly defined relationship between the binding strengths and molecular weight.

To account for the influence of flexibility of the receptor in the peptide binding, the resulting 441 peptides selected from the initial screening against the crystal structure of HbS were employed for ensemble docking. The ΔG value obtained for each peptide were averaged over the four HbS conformations and and a total of 323 peptides that fulfilled the energetic cutoff were selected. To further streamline the number of peptides that will be tested further and to identify the peptides that interact intimately with the Val 6 binding site, we calculated the distances between the bound peptides and the Val6 binding site residues for the resulting 323

peptides and averaged over the four confirmations. The values per peptide were then employed in ranking the screened library and selecting the best peptides for further analysis. In total, 61 peptides were selected for *in silico* validation using MD simulation.

***In silico* Validation of Selected Peptide-protein Complexes using MD simulations** We carried out MD simulations to observe the changes in the dynamics of the protein-peptide complexes and investigate the stability of the top performing peptides identified via docking in the Val6 binding site over time. A similar approach was used in identifying small molecules capable of inhibiting HbS polymerization in a previous study (chapter 3.2). The MD study starts with 61 independent simulations for the protein-peptide complexes initially for 20 ns, several analyses were performed in order to differentiate the good binders from the poor binders, and all the peptide-protein complexes that fulfill the criteria chosen were then extended to the next phase of simulation. The following criteria were used for the selection compounds that were extended to 100 ns. (a) the Root Mean Square Deviation of the peptide (mean $RMSD_{peptide} \leq 7$Å), (b) the distance of the peptide to the Val6 binding site (d_{bs}) $\leq$ 5.5 Å). These analyses were performed for only the last 5 ns of the simulation. Out of the 61 peptides that were tested with the initial simulations, 21 peptides that fulfilled the chosen criteria were subsequently extended to 100 ns. After the 100 ns simulation of the peptide-protein complexes, the same analyses were performed to select suitable peptides (hits) that will be used for further analysis. Only the last 25 ns of the the 100 ns of the trajectories were employed for analysis.

Filtering based on 20 ns MD simulations

Peptide detachment

Molecular docking gives an idea of whether a compound fits well in the binding site of the potential target, and therefore might qualify as a good binder. In order to properly differentiate a good binder from a bad binder, it is therefore necessary to investigate the dynamics of such compounds using MD simulations [256] and not rely on only static analysis [269]. During the initial 20 ns MD simulation, some of the peptides did not remain in close proximity to the binding site by displaying very high distance values, which raises the question of whether the peptides remained bound to the binding site. In order to answer this question, we calculated the distance between the center of mass (COM) of the peptide and the COM of the binding site of HbS (d_{COM}). The binding site was defined by the three critical amino acid residues (βPhe85, βLeu88 and βPhe70) that are crucial for HbS aggregation identified in our previous MD simulation study. For a peptide to actually be considered detached, a peptide must have a d_{COM} greater than 18 Å for a total of 2 consecutive ns. This criterion is applied to the whole of the trajectory instead of just the last 5 ns of the simulation. As seen in Figure 4.2, where the d_{COM} is plotted for a few peptides, four peptides are revealed as detaching from the binding site. AMLNESFRVY is leaving the binding site already at about 5 ns, while for AMLNEFCKVY, AMLNQHLRVY and AMLNMHLRVY, the d_{COM} remained below 18 Å for the first 5 ns. For two of the four peptides, once detached they did not return to the binding site during the 20 ns of the simulation. For comparison, four peptides that recorded the lowest d_{COM} values are also shown. Interestingly, AMLNEELEAY was already identified via docking among the ligands with the smallest distances to the binding site residues.

Peptide flexibility

To determine the flexibility of the peptides within the binding site, the Root Mean Square Deviation ($RMSD_{peptide}$) was calculated for all the peptide-HbS complex conformations of the last 5 ns of the each trajectory by comparing them to the starting conformation of the MD simulation. Here, a cut-off of 7 Å was chosen to differentiate between good and bad binders and further reduce the number of peptides extended to 100 ns. The choice of a relatively higher cut-off is because of possible re-adjustement of the binding poses generated via docking to adapt to conformational changes in the Val6 binding site as a result of dynamical forces [256]. The mean RMSD values obtained for the peptide library ranges from 2 Å to 68.4 Å. The results obtained here for the mean $RMSD_{peptide}$ are not consistent with the results of the earlier study (chapter 3.2). In the previous study, values

ranging from 1.6 Å to 18.4 Å were obtained for FDA-approved drugs, investigational drugs and natural products. This shows that peptides are generally more flexible in the binding site in comparison to organic molecules. Of the 61 compounds, 26 peptides (42%) recorded a mean RMSD value over 10 Å and only 21 (34%) of the 61 compounds fulfilled the mean $RMSD_{peptide}$ cut-off chosen here. Interestingly, some of the peptides previously identified as the top binders in the initial virtual screening display poor dynamical characteristic, i.e. high flexibility in the binding site with mean RMSD values above the catchment RMSD cut-off of 7 Å, examples include AMLNKFFKAY(8.8 Å), AMLNQHFKAY (14.5 Å), AMLNRIFEAY (16.9 Å), AMLNRFIRAY (32.4 Å), AMLNTHCKAY (19.2 Å). These results indicate the significance of incorporating structural dynamics when searching for prospective inhibitors for HbS. Nonetheless, some of the top binders identified via molecular docking were also observed to have low flexibility in the binding site with RMSD values falling below the RMSD cut-off, examples are AMLNEGLEAY (2.5 Å) and AMLGTGCYVY (4.0 Å).

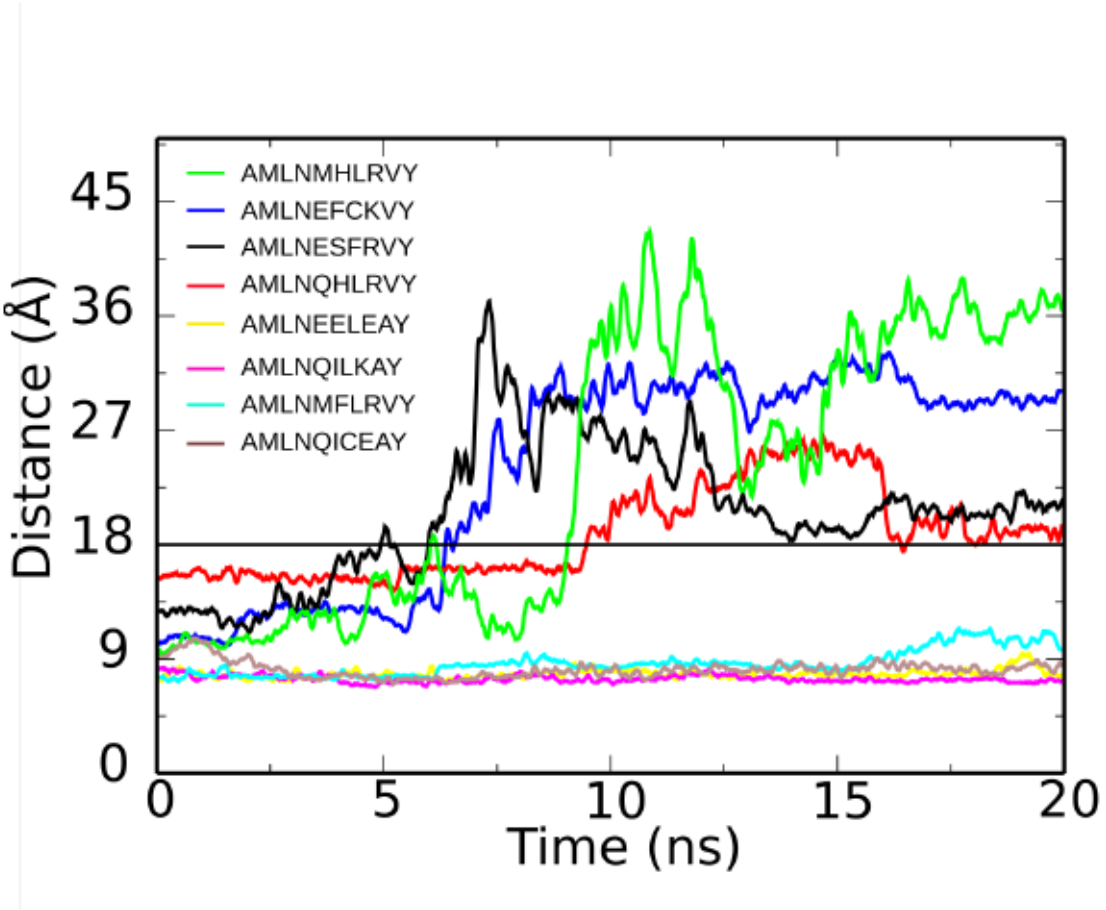

Figure 3.2: The distance between the COMs of selected peptides and the binding site residues. The distance is shown for the four detaching peptides. The cutoff distance at 18 Å is indicated by a black line. For comparison, the distances of the peptides displaying the smallest d_{COM} is also shown.

Distance between the peptide and binding site

To identify the best binding peptides based on 20 ns MD simulations, another criterion that was put into consideration is the distance between the peptide and any of the residues in the Val6 binding site (d_{bs}). The distances are defined as minimum distances between peptide and the group of residues in the Val6 binding site over the last 5 ns of the MD simulations. To identify these peptides that are in close proximity with the binding site, we set a cut-off of 5.5 Å mean distance. In the previous study, a cut-off of 4.5 was chosen to identify such compounds. The mean distance values obtained for the selection library ranges from 2.1 Å to 19.4 Å, with the majority of the peptides recording mean distance values between 2.1 Å to 4.5 Å. The results obtained here are very consistent with what was obtained from the previous study with the selection library also recording mean distance values ranging from 2 Å to 18.4 Å. Examples of compounds that recorded very high mean distance value above 10 Å include, but are not limited to AMLNEFCKVY (19.4 Å), AMLNQHCKVY (12.6 Å), AMLNKFIKAY (12.7 Å). These high mean distance values obtained are not surprising because these compounds also showed very high flexibility in the binding site and recorded very high RMSD values.

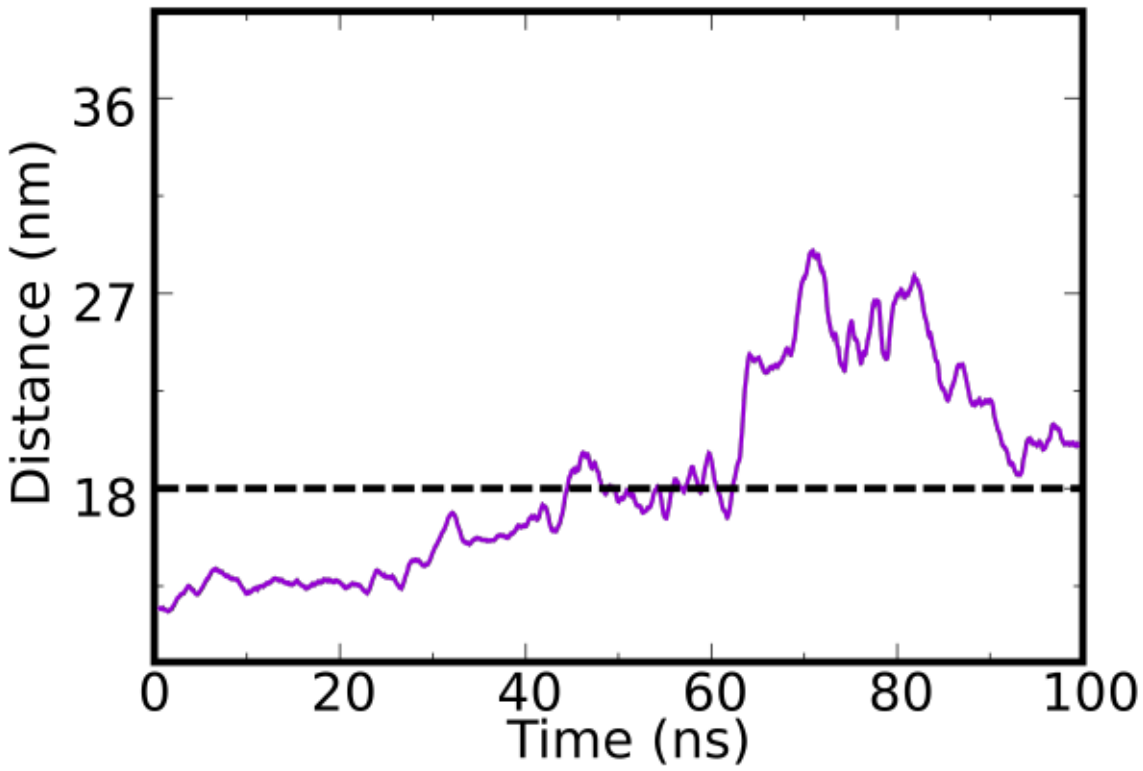

Figure 3.3: The distance between the COM of AMLNKHLRAY and the binding site residues. The cutoff distance af 18 Å is indicated by a dotted black line.

Filtering based on 100 ns MD-simulations

Peptide detachment

After applying the various cut-offs using the last 5 ns of the initial 20 ns MD simulations, the remaining 21 peptides that fulfilled the cut-offs were simulated for 100 ns to identify peptides that bind best to the binding site. The same cut-off chosen to identify peptides that detached from the binding site in the initial 20 ns was also applied here. During these 100 ns long simulations, three peptides detached from the Val6 binding site. One out of the three peptides started to visibly detach from the binding site at about 40 ns but returned to the binding site shortly after . It moved away from the binding site again at about 60 ns and never returned till the end of the simulation (see Figure 3.3). This observation of peptide still detaching even after 45 ns - 65 ns emphasizes the importance of a more detailed screening of compounds identified via docking and initial 20 ns MD simulation by extending the simulations.

Peptide flexibility and distance to the binding site

To identify the best binding peptides based on 100 ns simulations, the RMSD of all the 21 peptides ($RMSD_{peptide}$) simulated and their distances to the Val6 binding site were computed using the last 25 ns of the 100 ns simulation. The mean d_{bs} ranges between 2.2 Å and 11.2 Å and mean $RMSD_{peptide}$ between 3.9 Å and 28.2 Å. In comparison to our earlier study with organic molecules, the values obtained were smaller with the $RMSD_{ligand}$ values ranging between 1.3 Å and 6.7 Å and the mean d_{bs} was between between 1.5 Å and 5.4 Å. This shows that the compounds screened in the earlier study contained compounds with very low RMSD and distance values. This results suggest a higher affinity of the ligands to the binding site and are thus probably better binders than the peptides studied here. It further supports the assumption of a higher flexibility of the peptides compared to the ligands used in the earlier study. Of all the 21 peptides, 4 peptides did not stay in close proximity to the binding site (corresponding to 19% of the total peptides simulated) namely; AMLNWGIRAY (6.3 Å), AMLNKHLRAY (8.0 Å), AMLGMHFRVY (8.2 Å) and AMLGTGCKVY (11.2 Å). The criterion for the mean $RMSD_{peptide}$ cutoff is fulfilled by 11 (52% of the 21 peptides). In total 11 compounds met the two criteria chosen.

Interaction of peptides with the binding site residues

The interaction energy E_{int} composed both Coulomb and Lennard-Jones (LJ) energies of the peptides with the three Val6 binding site residues was calculated. The total interaction energy ($E_{int,bs}$) was calculated by summing the $E_{int,Ala70}$, $E_{int,Phe85}$ and $E_{int,Leu88}$. The interaction of the peptide to Ala70, Phe85 and Leu88 are determined based on a cutoff of -4 KJ/mol. From the 21 peptides simulated, 7 peptides formed contacts with only Leu88, 8 peptides with only Ala70, 3 peptides formed contacts with only Phe85 and only one peptide formed contacts with all three residues. However, some of the peptides did not interact with any of the these three critical residues and 5 out of these 7 peptides that did not form any contacts with the critical residues recorded very high mean $RMSD_{peptide}$ and d_{bs} values. Interestingly, 4 of the resulting 11 peptides identified as good binders earlier based on their RMSD and distance to the binding site values did not interact with any of the residues resulting in a total of 7 peptides in our library. The binding poses of these peptides in the Val6 binding site, along with their sequences is shown in Figure 3.4.

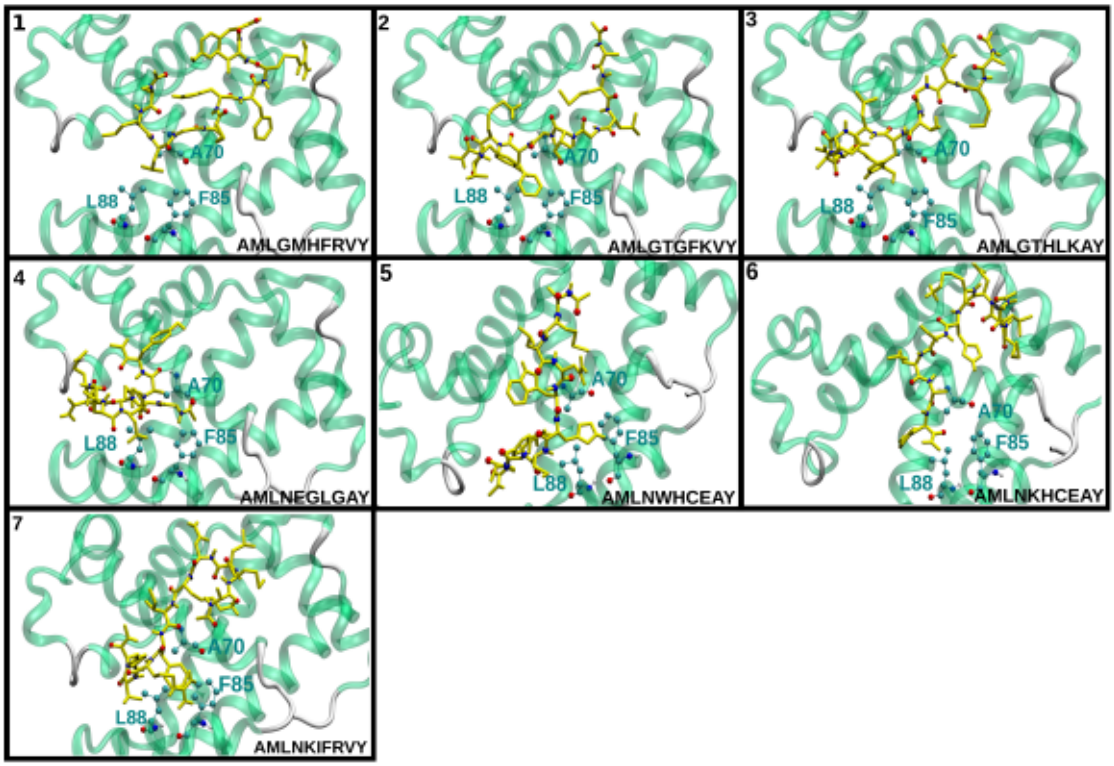

Figure 3.4: The binding poses of the top seven peptides with their respective sequences. The same protein and ligand representation, as well as color scheme as in Figure 1 in chapter 3.2 is used.

Conclusion

Over the years, computer aided drug design (CADD) has proven to be an effective way of identifying or developing new therapeutics for different diseases and thus, is very important in the drug discovery process. In this present work, CADD techniques were employed in identifying prospective D-enantiomeric peptides that are capable of inhibiting HbS polymerization. To this end, a virtual library of 1000 D-peptides was constructed and screened for their prospective HbS polymerization inhibitory abilities. The methodology adapted here was successfully used to identify inhibitors of the main protease enzyme, 3CLpro of the SARS-CoV-2 [256]. After an initial docking of the peptides against both the crystal structure of HbS and MD generated structures, a total of 61 peptides were identified. The selected peptides were then further employed in MD simulations to validate their stability in the binding site for an initial 20 ns. Subsequently, several analyses such as $RMSD_{peptide}$, d_{bs} and d_{COM} were performed to assess various binding properties of the peptides and streamline the number of peptides that would be extended to 100 ns MD simulation. In total, 21 peptides that met the criteria chosen after this initial MD simulation were simulated for 100 ns. A total of 11 peptides were identified that fulfilled the selection criteria chosen. The interaction energy of the selected peptides with any of the three critical residues present in the Val6 binding site was also analysed and we identified 7 peptides. In comparison to the results obtained in the earlier study, the values obtained here are significantly higher, i.e. the organic molecules screened significantly outperformed the peptides. It should be mentioned here that 3 orders of magnitude more small molecules were screened, i.e. the peptide screening done here is just a start of a larger initiative. 20^{10} ten amino acid long peptides would be possible. 10^7 times more than the library screened here.

In order to validate the results obtained from the CADD approach employed here, the top peptides identified should be subjected to both cell-based and *in vitro* assays to determine the inhibitory activities of these peptides experimentally. We expect that the findings presented here serves as a good starting point in identifying novel antisickling agents with potential of being applied clinically for treating sickle cell disease.

Chapter 4

Conclusions

Over the years, several diseases have been linked to protein aggregation, examples include Alzheimer's disease and Parkinson's diseases, amyotrophic lateral sclerosis, frontotemporal dementia, Huntington's disease as well as sickle cell disease (SCD) [270–272]. On a molecular level, aggregation of proteins is a result of electrostatic and hydrophobic interactions mediated by water. This balance is exemplified by sickle hemoglobin (HbS) which is a mutant of normal hemoglobin (HbA)[273] where a single substitution of glutamic acid to valine at the sixth position (E6V) of the HbA β chains causes the aggregation of deoxygenated HbS responsible for SCD. The Hydrophobic attraction between the substituted valine and hydrophobic residues βAla70, βLeu88 and βPhe85 encourages aggregation. This E6V mutation is believed to provide the shape and charge necessary to kick-start the aggregation process of hemoglobin.

Despite the fact that SCD is extensively studied and one of the most common genetic diseases worldwide, the treatment options available do not fully address the complex molecular manifestations of the disease. There is therefore an urgent need to develop effective, affordable and readily available drugs for the treatment of SCD. In order to develop new treatments for SCD, it is necessary to understand the structural and conformational basis of HbS aggregation and the role that the E6V mutation plays in the aggregation process. As a result of computational modeling and molecular dynamics (MD) simulations, we are able to study aggregation processes that are too rapid to be studied experimentally. Furthermore, they have been highly effective at simulating various motions within protein structures. In this thesis, using MD simulations we studied different aspects of both wild-type hemoglobin (HbA) and human sickle hemoglobin (HbS). These simulations were performed to probe the structural changes characterising the HbS aggregation paths, and also to identify protein-protein interaction hot-spots between HbS molecules, which were prioritised in search of HbS aggregation inhibitors. Using a rigorous computer guided approach that combines high throughput virtual screening with

explicit solvent MD simulations we identified a number of compounds whose ability to establish contacts with critical val6 binding site residues of HbS suggests them as good starting point for the design of inhibitors of HbS aggregation.

In chapter 3.1, several aspects of human sickle hemoglobin was studied along with the wild type hemoglobin as reference to reveal the structural effect of the Glu6Val mutation. We started with the conformational sampling of their monomeric units and ended with aggregation studies up to decamers and analysing the underlying protein-protein interactions. From our investigation, it was revealed that Glu6Val mutation starts with effects that may not be directly obvious. We found that the HbS both in monomeric and dimeric form is overall more rigid in comparison to the HbA and this is as a result of the mutation. Specifically, the β-globin chains of the HbS exhibited less flexibility than that of the HbA. Stabilization of the His 63-Fe^{2+} coordination which might play a role in the reduced binding of HbS to oxygen was also uncovered through our investigations. Through our analysis of the protein-protein contacts of each dimer extracted from the aggregated decamer, we identified both new and previously reported protein-protein contacts. Our investigations further reinforces the importance of the lateral contacts formed between hydrophobic βVal6 and the hydrophobic residues β'Phe85/Leu88 present in the neighbouring β'-globin chain. In addition to these hydrophobic interactions, we discovered that electrostatic interactions involving βLys17 and β'Glu90 also play a major role in aggregation. From our simulations comparing HbA with HbS we concluded that both the presence of Val6 and absence of Glu6 are important for the aggregation of HbS. In HbA, the closeness of the βGlu6 residue present to βLys17, prevents the latter from interacting with β'Glu90. In contrast, the presence of βVal6 in HbS permits the establishment of a network of contacts around the Val6 binding site. However, this observation is not fully consistent with an earlier study by Ghatge et al.,[69] in which they found that the presence of βVal6 was less important than the absence of βGlu6 for polymerization of HbS. For the wild type hemoglobin (HbA), it was concluded that it is also able to form aggregates but the aggregates formed are neither stable nor long-lived due to the presence of mainly axial contacts. From our study, we suggest that in the search for therapies that are capable of disrupting HbS aggregation, in addition to targeting hydrophobic interactions involving βVal6 and β'Ala70, β'Leu88 and β'Phe85, one could also aim at interrupting the electrostatic interactions involving βLys17 and β'Glu90 .

Based on our initial findings in chapter 3.1, we identified a number of secondary contacts crucial in the HbS aggregation process with interactions involving mutated βVal6 and β'Ala70, β'Leu88 and β'Phe85 playing a major roles. For this reason, we, therefore, limited the docking screening work reported in chapter 3.2 to an interaction cavity on multiple HbS structures (both crystal structure and MD-generated) defined by these three critical amino acids.

It is crucial to identify compounds that are suitable for binding to a drug target as a first step in drug discovery. Using a rigorous Computer aided drug design (CADD) method that combines high throughput virtual screening with explicit molecular dynamic (MD) simulations we identified prospective compounds capable of disrupting these interactions thereby inhibiting HbS polymerization. This method was successfully employed in a previous study in search of prospective inhibitors of the SARS-CoV-2 main protease enzyme [255, 256]. We started by building a virtual library that consists of 77,246 organic molecules from different sources such as natural origin, FDA approved drugs, non-FDA approved drugs and investigational drugs. We then performed an initial virtual screening of these compounds against the crystal structure of HbS, the top compounds were identified mainly based on their energetic values. The analysis of the physicochemical properties revealed that there is no definite relationship between the molecular weight of the top compounds and the binding free energy strength with most having a molecular weight between 200-900 g/mol. In order to incorporate protein dynamics, we screened the top 8,717 compounds identified against four MD-generated structures of HbS from previous MD simulations in chapter 3.1. At this stage, we decided to prioritize not only the energetic values but also the contact of the compound with at least one of the three residues present in the Val6 binding sites, namely; β'Ala70, β'Leu88 and β'Phe85. This yielded a total of 100 ligands for which their HbS-ligand complexes were then subjected to MD simulations to validate their stability in the HbS Val6 binding site. The MD simulations results revealed the importance of including structural dynamics when searching for prospective binders for a drug target and not simply relying on docking results. We first performed an initial 20 ns MD simulation of the complexes, after which we incorporated various parameters, such as, the distances between the ligand and the residues in the Val6 binding site, and the RMSD of the ligands. In total 61 ligands were identified. Even though, MD simulations allow identification of ligands that bind well to a receptor, they do not account for how the compounds reaches its drug target. It is necessary for an effective drug to cross the lipid bilayer. An important measure for determining the cell permeability is the log P. Therefore, we calculated the LogP values of these 61 ligands. A total of 59 compounds satisfying all the calculated parameters were identified, for which MD simulations were extended to 100 ns. Using the last 25 ns of the trajectories after the 100 ns MD

simulations, we applied two additional criteria in addition to RMSD and Distances; reduction of structural redundancy in the compounds identified and prediction of ADME (Adsorption, Distribution, Metabolism, and Elimination). ADME predictions help in identifying compounds that are likely to penetrate the gastrointestinal (GI) barrier efficiently. They are very important in predicting how a chemical compound is processed by a living organism. In total, 16 compounds met all the chosen criteria of which 13 of them are natural products, 2 are FDA-approved drugs and 1 is an investigational drug. These 16 compounds identified via CADD approach will be further probed experimentally (via *in vitro* and cell based assays) to validate their potential HbS- aggregation inhibitory properties.

The idea of using D-peptides to inhibit protein aggregation in diseases is not novel. D-peptides therapeutics are recently being developed against the treatment of several protein-aggregation disease such Alzheimer's disease. In chapter 3.3, following similar approach used in chapter 3.2, we built a virtual library consisting of 1000 D-peptides which were then screened for their prospective HbS polymerization inhibitory abilities. We decided to focus on D-peptides due to the advantages they offer over L-peptides in terms of resistance to degradation by human proteases. A comparison was made between the peptides studied here and the small molecules from the earlier study. After thorough analysis, we identified a total of 7 D-enantiomeric peptides with prospects of being able to bind to HbS and inhibit its aggregation. From our investigations, we observed that the peptides were relatively more flexible in the Val6 binding site recording considerably higher RMSD values in comparison to the organic molecules screened. We concluded that the organic molecules outperformed the peptides. It should however be noted that we screened a lot more organic molecules and the peptide screening here is merely a beginning of a larger initiative. As a final assessment of the results obtained here using various i*n silico* methods, the inhibitory activity of the best performing peptides can be experimentally determined by using *in vitro* methods. In conclusion, we expect that the findings presented in this thesis will be of importance in the development of novel therapeutics for managing SCD.

Bibliography

[1] Ingram, V. "Abnormal human haemoglobins: I. The comparison of normal human and sickle-cell haemoglobins by "fingerprinting"". *Biochimica et biophysica acta* 28 (1958), pp. 539–545.

[2] Poillon, W., Kim, B., Labotka, R., Hicks, C., Kark, J. "Antisickling effects of 2, 3-diphosphoglycerate depletion" (1995).

[3] Poillon, W. N., Kim, B. C. "2, 3-Diphosphoglycerate and intracellular pH as interdependent determinants of the physiologic solubility of deoxyhemoglobin S" (1990).

[4] Rees, D. C., Williams, T. N., Gladwin, M. T. "Sickle-cell disease". *The Lancet* 376.9757 (2010), pp. 2018–2031.

[5] Li, X., Dao, M., Lykotrafitis, G., Karniadakis, G. E. "Biomechanics and biorheology of red blood cells in sickle cell anemia". *Journal of biomechanics* 50 (2017), pp. 34–41.

[6] Conner, B. J., Reyes, A. A., Morin, C., Itakura, K., Teplitz, R., Wallace, R. B. "Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides". *Proceedings of the National Academy of Sciences* 80.1 (1983), pp. 278–282.

[7] Hahn, E. V., Gillespie, E. B. "Sickle cell anemia: report of a case greatly improved by splenectomy. Experimental study of sickle cell formation". *Archives of Internal Medicine* 39.2 (1927), pp. 233–254.

[8] Brugnara, C., de Franceschi, L., Alper, S. L., et al. "Inhibition of Ca (2+)-dependent K+ transport and cell dehydration in sickle erythrocytes by clotrimazole and other imidazole derivatives." *The Journal of clinical investigation* 92.1 (1993), pp. 520–526.

[9] Mohandas, N., Evans, E. "Adherence of sickle erythrocytes to vascular endothelial cells: requirement for both cell membrane changes and plasma factors" (1984).

[10] Chien, S., Usami, S., Bertles, J. F., et al. "Abnormal rheology of oxygenated blood in sickle cell anemia". *The Journal of clinical investigation* 49.4 (1970), pp. 623–634.

[11] Lopez, A. D., Williams, T. N., Levin, A., Tonelli, M., Singh, J. A., Burney, P. G., Rehm, J., Volkow, N. D., Koob, G., Ferri, C. P. "Remembering the forgotten non-communicable diseases". *BMC medicine* 12.1 (2014), pp. 1–19.

[12] Williams, T. N. "Sickle cell disease in sub-Saharan Africa". *Hematology/Oncology Clinics* 30.2 (2016), pp. 343–358.

[13] Piel, F. B., Hay, S. I., Gupta, S., Weatherall, D. J., Williams, T. N. "Global burden of sickle cell anaemia in children under five, 2010–2050: modelling based on demographics, excess mortality, and interventions". *PLoS medicine* 10.7 (2013), e1001484.

[14] Michael R DeBaun, M. A. G. "Sickle cell disease in sub-Saharan Africa". *UpToDate* (2022). URL: `https://www.uptodate.com/contents/sickle-cell-disease-in-sub-saharan-africa#H432682665` (visited on Mar. 22, 2022).

[15] Serjeant, G. "Geography and the clinical picture of sickle cell disease: An overview". *Annals of the New York Academy of Sciences* 565.1 (1989), pp. 109–119.

[16] Piel, F. B., Patil, A. P., Howes, R. E., Nyangiri, O. A., Gething, P. W., Williams, T. N., Weatherall, D. J., Hay, S. I. "Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis". *Nature communications* 1.1 (2010), pp. 1–7.

[17] McAuley, C. F., Webb, C., Makani, J., Macharia, A., Uyoga, S., Opi, D. H., Ndila, C., Ngatia, A., Scott, J. A. G., Marsh, K., et al. "High mortality from Plasmodium falciparum malaria in children living with sickle cell anemia on the coast of Kenya". *Blood, The Journal of the American Society of Hematology* 116.10 (2010), pp. 1663–1668.

[18] Makani, J., Komba, A. N., Cox, S. E., Oruo, J., Mwamtemi, K., Kitundu, J., Magesa, P., Rwezaula, S., Meda, E., Mgaya, J., et al. "Malaria in patients with sickle cell anemia: burden, risk factors, and outcome at the outpatient clinic and during hospitalization". *Blood, The Journal of the American Society of Hematology* 115.2 (2010), pp. 215–220.

[19] Tshilolo, L., Kafando, E., Sawadogo, M., Cotton, F., Vertongen, F., Ferster, A., Gulbis, B. "Neonatal screening and clinical care programmes for sickle cell disorders in sub-Saharan Africa: lessons from pilot studies". *Public health* 122.9 (2008), pp. 933–941.

[20] Ansong, D., Akoto, A. O., Ocloo, D., Ohene-Frempong, K. "Sickle cell disease: management options and challenges in developing countries". *Mediterranean journal of hematology and infectious diseases* 5.1 (2013).

[21] Gardner, R. V. "Sickle cell disease: advances in treatment". *Ochsner journal* 18.4 (2018), pp. 377–389.

[22] Steinberg, M. H., Barton, F., Castro, O., Pegelow, C. H., Ballas, S. K., Kutlar, A., Orringer, E., Bellevue, R., Olivieri, N., Eckman, J., et al. "Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment". *Jama* 289.13 (2003), pp. 1645–1651.

[23] Strouse, J. J., Lanzkron, S., Beach, M. C., Haywood, C., Park, H., Witkop, C., Wilson, R. F., Bass, E. B., Segal, J. B. "Hydroxyurea for sickle cell disease: a systematic review for efficacy and toxicity in children". *Pediatrics* 122.6 (2008), pp. 1332–1342.

[24] Inusa, B. P. D., Atoyebi, W., Hassan, A. A., Idhate, T., Dogara, L., Ijei, I., Qin, Y., Anie, K., Lawson, J. O., Hsu, L. "Low-dose hydroxycarbamide therapy may offer similar benefit as maximum tolerated dose for children and young adults with sickle cell disease in low-middle-income settings". *F1000Research* 7 (2018).

[25] Qureshi, A., Kaya, B., Pancham, S., Keenan, R., Anderson, J., Akanni, M., Howard, J., et al. "Guidelines for the use of hydroxycarbamide in children and adults with sickle cell disease: A British Society for Haematology Guideline." *British journal of haematology* 181.4 (2018), pp. 460–475.

[26] Hassan, A., Awwalu, S., Okpetu, L., Waziri, A. D., et al. "Effect of hydroxyurea on clinical and laboratory parameters of sickle cell anaemia patients in North–West Nigeria". *The Egyptian Journal of Haematology* 42.2 (2017), p. 70.

[27] Wong, T. E., Brandow, A. M., Lim, W., Lottenberg, R. "Update on the use of hydroxyurea therapy in sickle cell disease". *Blood, The Journal of the American Society of Hematology* 124.26 (2014), pp. 3850–3857.

[28] Murad, M., Hazem, A., Prokop, L., et al. "Hydroxyurea for Sickle Cell Disease: A Systematic Review of Benefits, Harms, and Barriers of Utilization, 2012 Prepared for the National Heart, Lung, and Blood Institute (NHLBI) Prepared by the Knowledge and Encounter Research Unit". *Mayo Clin.* Vol. 2012. 2012, pp. 1–116.

[29] Ware, R. E. "Optimizing hydroxyurea therapy for sickle cell anemia". *Hematology 2014, the American Society of Hematology Education Program Book* 2015.1 (2015), pp. 436–443.

[30] Agrawal, R. K., Patel, R. K., Nainiwal, L., Trivedi, B., et al. "Hydroxyurea in sickle cell disease: drug review". *Indian Journal of Hematology and Blood Transfusion* 30.2 (2014), pp. 91–96.

[31] Tripathi, A., Jerrell, J. M., Stallworth, J. R. "Clinical complications in severe pediatric sickle cell disease and the impact of hydroxyurea". *Pediatric blood & cancer* 56.1 (2011), pp. 90–94.

[32] Alvarez, O., Yovetich, N. A., Scott, J. P., Owen, W., Miller, S. T., Schultz, W., Lockhart, A., Aygun, B., Flanagan, J., Bonner, M., et al. "Pain and other non-neurological adverse events in children with sickle cell anemia and previous stroke who received hydroxyurea and phlebotomy or chronic transfusions and chelation: results from the SWiTCH clinical trial". *American journal of hematology* 88.11 (2013), pp. 932–938.

[33] Nzouakou, R., Bachir, D., Lavaud, A., Habibi, A., Lee, K., Lionnet, F., Hulin, A., Jouault, H., Préhu, C., Roudot-Thoraval, F., et al. "Clinical follow-up of hydroxyurea-treated adults with sickle cell disease". *Acta haematologica* 125.3 (2011), pp. 145–152.

[34] Wang, W. C., Ware, R. E., Miller, S. T., Iyer, R. V., Casella, J. F., Minniti, C. P., Rana, S., Thornburg, C. D., Rogers, Z. R., Kalpatthi, R. V., et al. "Hydroxycarbamide in very young children with sickle-cell anaemia: a multicentre, randomised, controlled trial (BABY HUG)". *The Lancet* 377.9778 (2011), pp. 1663–1672.

[35] Thornburg, C. D., Files, B. A., Luo, Z., Miller, S. T., Kalpatthi, R., Iyer, R., Seaman, P., Lebensburger, J., Alvarez, O., Thompson, B., et al. "Impact of hydroxyurea on clinical events in the BABY HUG trial". *Blood, The Journal of the American Society of Hematology* 120.22 (2012), pp. 4304–4310.

[36] Gilmore, A., Cho, G., Howard, J., Layton, M., Afif, M., Hughes, R. G., Philpott, N. J., Patankar, S., Davies, S. C., Haemoglobinopathy, N. L. "Feasibility and benefit of hydroxycarbamide as a long-term treatment for sickle cell disease patients: results from the North West London Sickle Cell Disease Registry". *American journal of hematology* 86.11 (2011), pp. 958–961.

[37] Lopes de Castro Lobo, C., Pinto, J. F., Nascimento, E. M., Moura, P. G., Cardoso, G. P., Hankins, J. S. "The effect of hydroxcarbamide therapy on survival of children with sickle cell disease". *British journal of haematology* 161.6 (2013), pp. 852–860.

[38] Sharef, S. W., Al-Hajri, M., Beshlawi, I., Al-Shahrabally, A., Elshinawy, M., Zachariah, M., Mevada, S. T., Bashir, W., Rawas, A., Taqi, A., et al. "Optimizing Hydroxyurea use in children with sickle cell disease: low dose regimen is effective". *European Journal of Haematology* 90.6 (2013), pp. 519–524.

[39] Voskaridou, E., Christoulas, D., Bilalis, A., Plata, E., Varvagiannis, K., Stamatopoulos, G., Sinopoulou, K., Balassopoulou, A., Loukopoulos, D., Terpos, E. "The effect of prolonged administration of hydroxyurea on morbidity and mortality in adult patients with sickle cell syndromes: results of a 17-year, single-center trial (LaSHS)". *Blood, The Journal of the American Society of Hematology* 115.12 (2010), pp. 2354–2363.

[40] Hansen, I. O., Sørensen, A. L., Hasselbalch, H. C. "Second malignancies in hydroxyurea and interferon-treated Philadelphia-negative myeloproliferative neoplasms". *European Journal of Haematology* 98.1 (2017), pp. 75–84.

[41] Steinberg, M. H., McCarthy, W. F., Castro, O., Ballas, S. K., Armstrong, F. D., Smith, W., Ataga, K., Swerdlow, P., Kutlar, A., DeCastro, L., et al.

"The risks and benefits of long-term use of hydroxyurea in sickle cell anemia: a 17.5 year follow-up". *American journal of hematology* 85.6 (2010), pp. 403–408.

[42] McGann, P. T., Flanagan, J. M., Howard, T. A., Dertinger, S. D., He, J., Kulharya, A. S., Thompson, B. W., Ware, R. E., Investigators, B. H. "Genotoxicity associated with hydroxyurea exposure in infants with sickle cell anemia: results from the BABY-HUG Phase III Clinical Trial". *Pediatric blood & cancer* 59.2 (2012), pp. 254–257.

[43] Niihara, Y., Miller, S. T., Kanter, J., Lanzkron, S., Smith, W. R., Hsu, L. L., Gordeuk, V. R., Viswanathan, K., Sarnaik, S., Osunkwo, I., et al. "A phase 3 trial of l-glutamine in sickle cell disease". *New England Journal of Medicine* 379.3 (2018), pp. 226–235.

[44] Rumen, N. "Inhibition of sickling in erythrocytes by amino acids". *Blood* 45.1 (1975), pp. 45–48.

[45] Mackenzie, R., Gleason, E., Schatzman, G., Cawein, M. "An in vitro method for screening compounds for the effect on the rate of sickling of erythrocytes". *Journal of International Medical Research* 4.6 (1976), pp. 375–381.

[46] Quinn, C. T. "l-Glutamine for sickle cell anemia: more questions than answers". *Blood, The Journal of the American Society of Hematology* 132.7 (2018), pp. 689–693.

[47] Zerez, C. R., Lachant, N. A., Lee, S. J., Tanaka, K. R. "Decreased erythrocyte nicotinamide adenine dinucleotide redox potential and abnormal pyridine nucleotide content in sickle cell disease". *Blood* 71.2 (1988), pp. 512–515.

[48] Niihara, Y., Zerez, C. R., Akiyama, D. S., Tanaka, K. R. "Increased red cell glutamine availability in sickle cell anemia: demonstration of increased active transport, affinity, and increased glutamate level in intact red cells". *Journal of Laboratory and Clinical Medicine* 130.1 (1997), pp. 83–90.

[49] Kiessling, K., Roberts, N., Gibson, J. S., Ellory, J. C. "A comparison in normal individuals and sickle cell patients of reduced glutathione precursors and their transport between plasma and red cells." *The Hematology Journal: the Official Journal of the European Haematology Association* 1.4 (2000), pp. 243–249.

[50] Morris, C. R., Suh, J. H., Hagar, W., Larkin, S., Bland, D. A., Steinberg, M. H., Vichinsky, E. P., Shigenaga, M., Ames, B., Kuypers, F. A., et al. "Erythrocyte glutamine depletion, altered redox environment, and pulmonary hypertension in sickle cell disease". *Blood, The Journal of the American Society of Hematology* 111.1 (2008), pp. 402–410.

[51] Bhatia, M., Sheth, S. "Hematopoietic stem cell transplantation in sickle cell disease: patient selection and special considerations". *Journal of blood medicine* 6 (2015), p. 229.

[52] Mentzer, W. C., Heller, S., Pearle, P. R., Hackney, E., Vichinsky, E. "Availability of related donors for bone marrow transplantation in sickle cell anemia." *The American journal of pediatric hematology/oncology* 16.1 (1994), pp. 27–29.

[53] Dallas, M. H., Triplett, B., Shook, D. R., Hartford, C., Srinivasan, A., Laver, J., Ware, R., Leung, W. "Long-term outcome and evaluation of organ function in pediatric patients undergoing haploidentical and matched related hematopoietic cell transplantation for sickle cell disease". *Biology of Blood and Marrow Transplantation* 19.5 (2013), pp. 820–830.

[54] Bernaudin, F., Socie, G., Kuentz, M., Chevret, S., Duval, M., Bertrand, Y., Vannier, J.-P., Yakouben, K., Thuret, I., Bordigoni, P., et al. "Long-term results of related myeloablative stem-cell transplantation to cure sickle cell disease". *Blood, The Journal of the American Society of Hematology* 110.7 (2007), pp. 2749–2756.

[55] Vermylen, C., Cornu, G., Ferster, A., Brichard, B., Ninane, J., Ferrant, A., Zenebergh, A., Maes, P., Dhooge, C., Benoit, Y., et al. "Haematopoietic stem cell transplantation for sickle cell anaemia: the first 50 patients transplanted in Belgium". *Bone marrow transplantation* 22.1 (1998), pp. 1–6.

[56] Gluckman, E., Cappelli, B., Bernaudin, F., Labopin, M., Volt, F., Carreras, J., Pinto Simões, B., Ferster, A., Dupont, S., De la Fuente, J., et al. "Eurocord, the Pediatric Working Party of the European Society for Blood and Marrow Transplantation, and the Center for International Blood and Marrow Transplant Research. Sickle cell disease: an international survey of results of HLA-identical sibling hematopoietic stem cell transplantation". *Blood* 129.11 (2017), pp. 1548–1556.

[57] Vichinsky, E., Hoppe, C. C., Ataga, K. I., Ware, R. E., Nduba, V., El-Beshlawy, A., Hassab, H., Achebe, M. M., Alkindi, S., Brown, R. C., et al. "A phase 3 randomized trial of voxelotor in sickle cell disease". *New England Journal of Medicine* 381.6 (2019), pp. 509–519.

[58] Abraham, D. J., Mehanna, A. S., Wireko, F. C., Whitney, J., Thomas, R. P., Orringer, E. P. "Vanillin, a potential agent for the treatment of sickle cell anemia" (1991).

[59] Nakagawa, A., Lui, F. E., Wassaf, D., Yefidoff-Freedman, R., Casalena, D., Palmer, M. A., Meadows, J., Mozzarelli, A., Ronda, L., Abdulmalik, O., et al. "Identification of a small molecule that increases hemoglobin oxygen affinity and reduces SS erythrocyte sickling". *ACS chemical biology* 9.10 (2014), pp. 2318–2325.

[60] Blair, H. A. "Voxelotor: first approval". *Drugs* 80.2 (2020), pp. 209–215.

[61] Howard, J., Hemmaway, C. J., Telfer, P., Layton, D. M., Porter, J., Awogbade, M., Mant, T., Gretler, D. D., Dufu, K., Hutchaleelaha, A., et al. "A phase 1/2 ascending dose study and open-label extension study of voxelotor in patients with sickle cell disease". *Blood, The Journal of the American Society of Hematology* 133.17 (2019), pp. 1865–1875.

[62] Stocker, J. W., De Franceschi, L., McNaughton-Smith, G. A., Corrocher, R., Beuzard, Y., Brugnara, C. "ICA-17043, a novel Gardos channel blocker, prevents sickled red blood cell dehydration in vitro and in vivo in SAD mice". *Blood, The Journal of the American Society of Hematology* 101.6 (2003), pp. 2412–2418.

[63] Ataga, K. I., Reid, M., Ballas, S. K., Yasin, Z., Bigelow, C., James, L. S., Smith, W. R., Galacteros, F., Kutlar, A., Hull, J. H., et al. "Improvements in haemolysis and indicators of erythrocyte survival do not correlate with acute vaso-occlusive crises in patients with sickle cell disease: a phase III randomized, placebo-controlled, double-blind study of the Gardos channel blocker senicapoc (ICA-17043)". *British journal of haematology* 153.1 (2011), pp. 92–104.

[64] McArthur, J. G., Svenstrup, N., Chen, C., Fricot, A., Carvalho, C., Nguyen, J., Nguyen, P., Parachikova, A., Abdulla, F., Vercellotti, G. M., et al. "A

novel, highly potent and selective phosphodiesterase-9 inhibitor for the treatment of sickle cell disease". *Haematologica* 105.3 (2020), p. 623.

[65] Iyamu, E. W., Turner, E. A., Asakura, T. "In vitro effects of NIPRISAN (Nix-0699): a naturally occurring, potent antisickling agent". *British journal of haematology* 118.1 (2002), pp. 337–343.

[66] Swift, R., Abdulmalik, O., Chen, Q., Asakura, T., Gustafson, K., Simon, J. E., Zaman, V., Quiusky, K. A., Hassell, K. L., Shapira, I., et al. "SCD-101: a new anti-sickling drug reduces pain and fatigue and improves red blood cell shape in peripheral blood of patients with sickle cell disease". *Blood* 128.22 (2016), p. 121.

[67] Eaton, W. A., Bunn, H. F. "Treating sickle cell disease by targeting HbS polymerization". *Blood, The Journal of the American Society of Hematology* 129.20 (2017), pp. 2719–2726.

[68] Kassa, T., Wood, F., Strader, M. B., Alayash, A. I. "Antisickling drugs targeting βCys93 reduce iron oxidation and oxidative changes in sickle cell hemoglobin". *Frontiers in Physiology* 10 (2019), p. 931.

[69] Ghatge, M. S., Ahmed, M. H., Omar, A. S. M., Pagare, P. P., Rosef, S., Kellogg, G. E., Abdulmalik, O., Safo, M. K. "Crystal structure of carbonmonoxy sickle hemoglobin in R-state conformation". *Journal of structural biology* 194.3 (2016), pp. 446–450.

[70] Vitagliano, L., Mazzarella, L., Merlino, A., Vergara, A. "Fine Sampling of the R→ T Quaternary-Structure Transition of a Tetrameric Hemoglobin". *Chemistry–A European Journal* 23.3 (2017), pp. 605–613.

[71] Mehanna, A. "Sickle cell anemia and antisickling agents then and now". *Current medicinal chemistry* 8.2 (2001), pp. 79–88.

[72] Chikezie, P. C. "Sodium metabisulfite–induced polymerization of sickle cell hemoglobin incubated in the extracts of three medicinal plants (Anacardium occidentale, Psidium guajava, and Terminalia catappa)". *Pharmacognosy magazine* 7.26 (2011), p. 126.

[73] Chen, K., Ballas, S. K., Hantgan, R. R., Kim-Shapiro, D. B. "Aggregation of normal and sickle hemoglobin in high concentration phosphate buffer". *Biophysical Journal* 87.6 (2004), pp. 4113–4121.

[74] Bridges, K., Barabino, G., Brugnara, C., Cho, M., Christoph, G., Dover, G., Ewenstein, B., Golan, D., Guttmann, C., Hofrichter, J., et al. "A multiparameter analysis of sickle erythrocytes in patients undergoing hydroxyurea therapy" (1996).

[75] Delalic, Z., Takashima, S., Adachi, K., Asakura, T., Gilbert, G. "Dielectric constant of sickle cell hemoglobin: Dielectric properties of sickle cell hemoglobin in solution and gel". *Journal of molecular biology* 168.3 (1983), pp. 659–671.

[76] Asakura, T., Ohnishi, S. T., Adachi, K., Ozguc, M., Hashimoto, K., Devlin, M. T., Schwartz, E. "Effect of piracetam on sickle erythrocytes and sickle hemoglobin". *Biochimica et Biophysica Acta (BBA)-Protein Structure* 668.3 (1981), pp. 397–405.

[77] Adachi, K., Asakura, T. "Gelation of deoxyhemoglobin A in concentrated phosphate buffer. Exhibition of delay time prior to aggregation and crystallization of deoxyhemoglobin A". *The Journal of biological chemistry* 254.24 (1979), pp. 12273–12276.

[78] Adachi, K., Asakura, T. "Hemoglobin gelation". *Texas reports on biology and medicine* 40 (1980), pp. 251–260.

[79] Adachi, K., Matarasso, S. L., Asakura, T. "Nucleation-controlled aggregation of deoxyhemoglobin S effect of organic phosphates on the kinetics of aggregation of deoxyhemoglobin S in concentrated phosphate buffer". *Biochimica et Biophysica Acta (BBA)-Protein Structure* 624.2 (1980), pp. 372–377.

[80] Adachi, K., Asakura, T. "Polymerization of deoxyhemoglobin CHarlem (β6 Glu→ Val, β73 Asp→ Asn): The effect of β73 asparagine on the gelation and crystallization of hemoglobin". *Journal of Molecular Biology* 144.4 (1980), pp. 467–480.

[81] Adachi, K., Asakura, T. "Aggregation and crystallization of hemoglobins A, S, and C. Probable formation of different nuclei for gelation and crystallization." *Journal of Biological Chemistry* 256.4 (1981), pp. 1824–1830.

[82] Adachi, K., Asakura, T. "Gelation and crystallization of sickle (Hb S and Hb C Harlem) and non-sickle hemoglobins (Hb A and Hb C) in concentrated

phosphate buffer". *Progress in clinical and biological research* 55 (1981), pp. 123–144.

[83] Adachi, K., Asakura, T. "Kinetics of the polymerization of hemoglobin in high and low phosphate buffers." *Blood cells* 8.2 (1982), pp. 213–224.

[84] Adachi, K., Kim, J., Travitz, R., Harano, T., Asakura, T. "Effect of amino acid at the beta 6 position on surface hydrophobicity, stability, solubility, and the kinetics of polymerization of hemoglobin. Comparisons among Hb A (Glu beta 6), Hb C (Lys beta 6), Hb Machida (Gln beta 6), and Hb S (Val beta 6)." *Journal of Biological Chemistry* 262.27 (1987), pp. 12920–12925.

[85] Perutz, M. F., Fermi, G., Abraham, D. J., Poyart, C., Bursaux, E. "Hemoglobin as a receptor of drugs and peptides: X-ray studies of the stereochemistry of binding". *Journal of the American Chemical Society* 108.5 (1986), pp. 1064–1078.

[86] Abraham, D. "Correlation of partition coefficients with antisickling activity of simple alcohols, amides and ureas." *Blood Cells* 8.2 (1982), pp. 345–355.

[87] Abraham, E., Stallings, M., Abraham, A., Garbutt, G. "Modification of sickle hemoglobin by acetaldehyde and its effect on oxygenation, gelation and sickling". *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 705.1 (1982), pp. 76–81.

[88] Abraham, D. J., Perutz, M. F., Phillips, S. "Physiological and x-ray studies of potential antisickling agents". *Proceedings of the National Academy of Sciences* 80.2 (1983), pp. 324–328.

[89] Lu, L., Li, X., Vekilov, P. G., Karniadakis, G. E. "Probing the twisted structure of sickle hemoglobin fibers via particle simulations". *Biophysical journal* 110.9 (2016), pp. 2085–2093.

[90] Harrington, D. J., Adachi, K., Royer Jr, W. E. "The high resolution crystal structure of deoxyhemoglobin S". *Journal of molecular biology* 272.3 (1997), pp. 398–407.

[91] Oder, E., Safo, M. K., Abdulmalik, O., Kato, G. J. "New developments in anti-sickling agents: can drugs directly prevent the polymerization of sickle haemoglobin in vivo?" *British journal of haematology* 175.1 (2016), pp. 24–30.

[92] Iqbal, Z., Li, M., McKendry, R., Horton, M., Caruana, D. J. "Investigation of Sickle-Cell Haemoglobin Polymerisation under Electrochemical Control". *Chemphyschem* 14.10 (2013), pp. 2143–2148.

[93] Ferrone, F. A., Hofrichter, J., Eaton, W. A. "Kinetics of sickle hemoglobin polymerization: II. A double nucleation mechanism". *Journal of molecular biology* 183.4 (1985), pp. 611–631.

[94] Eaton, W. A., Hofrichter, J. "Sickle cell hemoglobin polymerization". *Advances in protein chemistry* 40 (1990), pp. 63–279.

[95] Lu, L., Li, Z., Li, H., Li, X., Vekilov, P. G., Karniadakis, G. E. "Quantitative prediction of erythrocyte sickling for the development of advanced sickle cell therapies". *Science advances* 5.8 (2019), eaax3905.

[96] Castle, B. T., Odde, D. J., Wood, D. K. "Rapid and inefficient kinetics of sickle hemoglobin fiber growth". *Science advances* 5.3 (2019), eaau1086.

[97] Oosawa, F., Asakura, S., et al. *Thermodynamics of the Polymerization of Protein*. Academic Press, 1975.

[98] Gardner, M. K., Charlebois, B. D., Jánosi, I. M., Howard, J., Hunt, A. J., Odde, D. J. "Rapid microtubule self-assembly kinetics". *Cell* 146.4 (2011), pp. 582–592.

[99] Wang, Y., Ferrone, F. A. "Dissecting the energies that stabilize sickle hemoglobin polymers". *Biophysical journal* 105.9 (2013), pp. 2149–2156.

[100] Cao, Z., Ferrone, F. A. "Homogeneous nucleation in sickle hemoglobin: stochastic measurements with a parallel method". *Biophysical journal* 72.1 (1997), pp. 343–352.

[101] Otto, J. M., Plumb, J. O., Clissold, E., Kumar, S. B., Wakeham, D. J., Schmidt, W., Grocott, M. P., Richards, T., Montgomery, H. E. "Hemoglobin concentration, total hemoglobin mass and plasma volume in patients: implications for anemia". *Haematologica* 102.9 (2017), p. 1477.

[102] Akinbami, A., Dosunmu, A., Adediran, A., Oshinaike, O., Adebola, P., Arogundade, O. "Haematological values in homozygous sickle cell disease in steady state and haemoglobin phenotypes AA controls in Lagos, Nigeria". *BMC research notes* 5.1 (2012), pp. 1–6.

[103] Abere, T. A., Okoye, C. J., Agoreyo, F. O., Eze, G. I., Jesuorobo, R. I., Egharevba, C. O., Aimator, P. O. "Antisickling and toxicological evaluation of the leaves of Scoparia dulcis Linn (Scrophulariaceae)". *BMC Complementary and Alternative Medicine* 15.1 (2015), pp. 1–7.

[104] Xu, G. G., Pagare, P. P., Ghatge, M. S., Safo, R. P., Gazi, A., Chen, Q., David, T., Alabbas, A. B., Musayev, F. N., Venitz, J., et al. "Design, synthesis, and biological evaluation of ester and ether derivatives of antisickling agent 5-HMF for the treatment of sickle cell disease". *Molecular pharmaceutics* 14.10 (2017), pp. 3499–3511.

[105] Abdulmalik, O., Ghatge, M. S., Musayev, F. N., Parikh, A., Chen, Q., Yang, J., 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". *Acta Crystallographica Section D: Biological Crystallography* 67.11 (2011), pp. 920–928.

[106] Dash, B. P., Archana, Y., Satapathy, N., Naik, S. K. "Search for antisickling agents from plants". *Pharmacognosy reviews* 7.13 (2013), p. 53.

[107] Li, Q., Henry, E. R., Hofrichter, J., Smith, J. F., Cellmer, T., Dunkelberger, E. B., Metaferia, B. B., Jones-Straehle, S., Boutom, S., Christoph, G. W., et al. "Kinetic assay shows that increasing red cell volume could be a treatment for sickle cell disease". *Proceedings of the National Academy of Sciences* 114.5 (2017), E689–E696.

[108] Padmos, M., Sackey, K., Roberts, G., Kulozik, A., Bail, S., Morris, J., Serjeant, B., Serjeant, G. "Two different forms of homozygous sickle cell disease occur in Saudi Arabia". *British journal of haematology* 79.1 (1991), pp. 93–98.

[109] El-Hazmi, M. A. "Heterogeneity and variation of clinical and haematological expression of haemoglobin S in Saudi Arabs". *Acta haematologica* 88.2-3 (1992), pp. 67–71.

[110] Ivanova, M., Jasuja, R., Kwong, S., Briehl, R. W., Ferrone, F. A. "Nonideality and the nucleation of sickle hemoglobin". *Biophysical journal* 79.2 (2000), pp. 1016–1022.

[111] Safo, M. K., Ko, T.-P., Schreiter, E. R., Russell, J. E. "Structural basis for the antipolymer activity of Hb $\zeta 2\beta s2$ trapped in a tense conformation". *Journal of molecular structure* 1099 (2015), pp. 99–107.

[112] Metcalf, B., Chuang, C., Dufu, K., Patel, M. P., Silva-Garcia, A., Johnson, C., Lu, Q., Partridge, J. R., Patskovska, L., Patskovsky, Y., et al. "Discovery of GBT440, an orally bioavailable R-state stabilizer of sickle cell hemoglobin". *ACS medicinal chemistry letters* 8.3 (2017), pp. 321–326.

[113] Pagare, P. P., Ghatge, M. S., Musayev, F. N., Deshpande, T. M., Chen, Q., Braxton, C., Kim, S., Venitz, J., Zhang, Y., Abdulmalik, O., et al. "Rational design of pyridyl derivatives of vanillin for the treatment of sickle cell disease". *Bioorganic & medicinal chemistry* 26.9 (2018), pp. 2530–2538.

[114] Oksenberg, D., Dufu, K., Patel, M. P., Chuang, C., Li, Z., Xu, Q., Silva-Garcia, A., Zhou, C., Hutchaleelaha, A., Patskovska, L., et al. "GBT 440 increases haemoglobin oxygen affinity, reduces sickling and prolongs RBC half-life in a murine model of sickle cell disease". *British Journal of Haematology* 175.1 (2016), pp. 141–153.

[115] Dufu, K., Patel, M., Oksenberg, D., Cabrales, P. "GBT440 improves red blood cell deformability and reduces viscosity of sickle cell blood under deoxygenated conditions". *Clinical hemorheology and microcirculation* 70.1 (2018), pp. 95–105.

[116] Nnamani, I. N., Joshi, G. S., Danso-Danquah, R., Abdulmalik, O., Asakura, T., Abraham, D. J., Safo, M. K. "Pyridyl derivatives of benzaldehyde as potential antisickling agents". *Chemistry & biodiversity* 5.9 (2008), pp. 1762–1769.

[117] Manno, M., San Biagio, P., Palma, M. "The role of pH on instability and aggregation of sickle hemoglobin solutions". *Proteins: Structure, Function, and Bioinformatics* 55.1 (2004), pp. 169–176.

[118] Ferrone, F. A. "Polymerization and sickle cell disease: a molecular view". *Microcirculation* 11.2 (2004), pp. 115–128.

[119] Nagel, R., Johnson, J., Bookchin, R., Garel, M., Rosa, J., Schiliro, G., Wajcman, H., Labie, D., Moo-Penn, W., Castro, O. "β-Chain contact sites in the haemoglobin S polymer". *Nature* 283.5750 (1980), pp. 832–834.

[120] Rhoda, M.-D., Martin, J., Blouquit, Y., Garel, M.-C., Edelstein, S. J., Rosa, J. "Sickle cell hemoglobin fiber formation strongly inhibited by the stanleyville II mutation (α78 Asn$\rightarrow$ Lys)". *Biochemical and Biophysical Research Communications* 111.1 (1983), pp. 8–13.

[121] Benesch, R. E., Kwong, S., Edalji, R., Benesch, R. "alpha Chain mutations with opposite effects on the gelation of hemoglobin S." *Journal of Biological Chemistry* 254.17 (1979), pp. 8169–8172.

[122] Jarosaw, M. "Molecular Dynamics". *ENCYCLOPEDIA OF LIFE SCIENCES* (2001), pp. 1–17.

[123] Cheatham III, T. E., Kollman, P. A. "Molecular dynamics simulation of nucleic acids". *Annual review of physical chemistry* 51.1 (2000), pp. 435–471.

[124] Karplus, M., McCammon, J. A. "Molecular dynamics simulations of biomolecules". *Nature structural biology* 9.9 (2002), pp. 646–652.

[125] Norberg, J., Nilsson, L. "Molecular dynamics applied to nucleic acids". *Accounts of chemical research* 35.6 (2002), pp. 465–472.

[126] Moraitakis, G., Purkiss, A. G., Goodfellow, J. M. "Simulated dynamics and biological macromolecules". *Reports on Progress in Physics* 66.3 (2003), p. 383.

[127] Allen, M. P. et al. "Introduction to molecular dynamics simulation". *Computational soft matter: from synthetic polymers to proteins* 23.1 (2004), pp. 1–28.

[128] Ippoliti, E. "Parallel Algorithms for Short-Range Molecular Dynamics: What is Molecular Dynamics". *Biophysics Application Project* (2011).

[129] Becker, O. M., MacKerell, A. D., Roux, B., Watanabe, M. *Computational biochemistry and biophysics*. Marcel Dekker New York, 2001.

[130] Olubiyi Olujide, O. *Investigation of the interaction between Alzheimer's abeta peptide and aggregation inhibitors using molecular simulations*. 2013.

[131] Scheraga, H. A., Khalili, M., Liwo, A. "Protein-folding dynamics: overview of molecular simulation techniques". *Annu. Rev. Phys. Chem.* 58 (2007), pp. 57–83.

[132] González, M. "Force fields and molecular dynamics simulations". *École thématique de la Société Française de la Neutronique* 12 (2011), pp. 169–200.

[133] Tsai, C. S. *An introduction to computational biochemistry.* John Wiley & Sons, 2003.

[134] Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids". *Journal of the American Chemical Society* 118.45 (1996), pp. 11225–11236.

[135] Tufts, C. "Cost to Develop and Win Marketing Approval for a New Drug Is 2.6*Billion*". *Tufts CSDD* 19 (2014), pp. 1–7.

[136] Leelananda, S. P., Lindert, S. "Computational methods in drug discovery". *Beilstein journal of organic chemistry* 12.1 (2016), pp. 2694–2718.

[137] Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Galpin, S. A., Handa, B. K., Kay, J., Kröhn, A., et al. "Rational design of peptide-based HIV proteinase inhibitors". *Science* 248.4953 (1990), pp. 358–361.

[138] Erickson, J., Neidhart, D., VanDrie, J., Kempf, D., Wang, X., Norbeck, D. "Plattner n, Rittenhouse J, Turon M, Wideburg N, Kohlbrenner WE, Simmer R, Helfrich R, Paul D and Knigg M, Design, activity and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-I protease". *Science* 249 (1990), pp. 527–533.

[139] BD, D. "Levin RB. McDaniel SL. Vacca JP. Guare JP. Darke PL. Zugay JA. Emini EA. Schleif WA. Quintero JC. Lin JH. Chen I.-W. Holloway MK. Fitzgerald PMD. Axel MG. Ostovic D. Anderson PS". *Huff JR. J. Med. Chem* 37 (1994), p. 3443.

[140] Schapira, M., Raaka, B. M., Samuels, H. H., Abagyan, R. "Rational discovery of novel nuclear hormone receptor antagonists". *Proceedings of the National Academy of Sciences* 97.3 (2000), pp. 1008–1013.

[141] Sethi, A., Joshi, K., Sasikala, K., Alvala, M. "Molecular docking in modern drug discovery: Principles and recent applications". *Drug discovery and development-new advances* 2 (2019), pp. 1–21.

[142] Pickett, S. D., McLay, I. M., Clark, D. E. "Enhancing the hit-to-lead properties of lead optimization libraries". *Journal of chemical information and computer sciences* 40.2 (2000), pp. 263–272.

[143] Kenakin, T. "Predicting therapeutic value in the lead optimization phase of drug discovery". *Nature Reviews Drug Discovery* 2.6 (2003), pp. 429–438.

[144] Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H., Sastry, G. N. "Virtual screening in drug discovery-a computational perspective". *Current Protein and Peptide Science* 8.4 (2007), pp. 329–351.

[145] Chourasia, M., Sastry, G. M., Sastry, G. N. "Proton binding sites and conformational analysis of H+ K+-ATPase". *Biochemical and biophysical research communications* 336.3 (2005), pp. 961–966.

[146] Bindu, P. H., Sastry, G. M., Sastry, G. N. "Characterization of calcium and magnesium binding domains of human 5-lipoxygenase". *Biochemical and biophysical research communications* 320.2 (2004), pp. 461–467.

[147] Murthy, J. N., Nagaraju, M., Sastry, G. M., Rao, A. R., Sastry, G., et al. "Active site acidic residues and structural analysis of modelled human aromatase: A potential drug target for breast cancer". *Journal of computer-aided molecular design* 19.12 (2005), pp. 857–870.

[148] Shoichet, B. K. "Virtual screening of chemical libraries". *Nature* 432.7019 (2004), pp. 862–865.

[149] F Sousa, S., MFSA Cerqueira, N., A Fernandes, P., Joao Ramos, M. "Virtual screening in drug design and development". *Combinatorial chemistry & high throughput screening* 13.5 (2010), pp. 442–453.

[150] Klebe, G. "Virtual ligand screening: strategies, perspectives and limitations". *Drug discovery today* 11.13-14 (2006), pp. 580–594.

[151] Maia, E. H. B., Assis, L. C., De Oliveira, T. A., Da Silva, A. M., Taranto, A. G. "Structure-based virtual screening: from classical to artificial intelligence". *Frontiers in chemistry* 8 (2020), p. 343.

[152] Berman, H. M., Kleywegt, G. J., Nakamura, H., Markley, J. L. "The future of the protein data bank". *Biopolymers* 99.3 (2013), pp. 218–222.

[153] Talele, T. T., Khedkar, S. A., Rigby, A. C. "Successful applications of computer aided drug discovery: moving drugs from concept to the clinic". *Current topics in medicinal chemistry* 10.1 (2010), pp. 127–141.

[154] Sliwoski, G., Kothiwale, S., Meiler, J., Lowe, E. W. "Computational methods in drug discovery". *Pharmacological reviews* 66.1 (2014), pp. 334–395.

[155] Devi, R. V., Sathya, S. S., Coumar, M. S. "Evolutionary algorithms for de novo drug design–A survey". *Applied Soft Computing* 27 (2015), pp. 543–552.

[156] Nunes, R. R., Fonseca, A. L. d., Pinto, A. C. d. S., Maia, E. H. B., Silva, A. M. d., Varotti, F. d. P., Taranto, A. G. "Brazilian malaria molecular targets (BraMMT): selected receptors for virtual high-throughput screening experiments". *Memórias do Instituto Oswaldo Cruz* 114 (2019).

[157] Rai, B. K., Tawa, G. J., Katz, A. H., Humblet, C. "Modeling G protein-coupled receptors for structure-based drug discovery using low-frequency normal modes for refinement of homology models: Application to H3 antagonists". *Proteins: Structure, Function, and Bioinformatics* 78.2 (2010), pp. 457–473.

[158] Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. L., Teller, D. C., Okada, T., Stenkamp, R. E., et al. "Crystal structure of rhodopsin: AG protein-coupled receptor". *science* 289.5480 (2000), pp. 739–745.

[159] Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K., et al. "High-resolution crystal structure of an engineered human β2-adrenergic G protein–coupled receptor". *science* 318.5854 (2007), pp. 1258–1265.

[160] Jaakola, V.-P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P., Stevens, R. C. "The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist". *Science* 322.5905 (2008), pp. 1211–1217.

[161] Kim, D., Xu, D., Guo, J.-t., Ellrott, K., Xu, Y. "PROSPECT II: protein structure prediction program for genome-scale applications". *Protein engineering* 16.9 (2003), pp. 641–650.

[162] Petrey, D., Xiang, Z., Tang, C. L., Xie, L., Gimpelev, M., Mitros, T., Soto, C. S., Goldsmith-Fischman, S., Kernytsky, A., Schlessinger, A., et al. "Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling". *Proteins: Structure, Function, and Bioinformatics* 53.S6 (2003), pp. 430–435.

[163] Simons, K. T., Kooperberg, C., Huang, E., Baker, D. "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions". *Journal of molecular biology* 268.1 (1997), pp. 209–225.

[164] Jegerschöld, C., Pawelzik, S.-C., Purhonen, P., Bhakat, P., Gheorghe, K. R., Gyobu, N., Mitsuoka, K., Morgenstern, R., Jakobsson, P.-J., Hebert, H. "Structural basis for induced formation of the inflammatory mediator prostaglandin E2". *Proceedings of the National Academy of Sciences* 105.32 (2008), pp. 11110–11115.

[165] Ma, X. H., Jia, J., Zhu, F., Xue, Y., Li, Z. R., Chen, Y. Z. "Comparative analysis of machine learning methods in ligand-based virtual screening of large compound libraries". *Combinatorial chemistry & high throughput screening* 12.4 (2009), pp. 344–357.

[166] Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., Ferrin, T. E. "A geometric approach to macromolecule-ligand interactions". *Journal of molecular biology* 161.2 (1982), pp. 269–288.

[167] Anderson, A. C. "The process of structure-based drug design". *Chemistry & biology* 10.9 (2003), pp. 787–797.

[168] Meng, X.-Y., Zhang, H.-X., Mezei, M., Cui, M. "Molecular docking: a powerful approach for structure-based drug discovery". *Current computer-aided drug design* 7.2 (2011), pp. 146–157.

[169] Fischer, D., Norel, R., Wolfson, H., Nussinov, R. "Surface motifs by a computer vision technique: searches, detection, and implications for protein–ligand recognition". *Proteins: Structure, Function, and Bioinformatics* 16.3 (1993), pp. 278–292.

[170] Subramanian, J., Sharma, S., B-Rao, C. "Modeling and Selection of Flexible Proteins for Structure-Based Drug Design: Backbone and Side Chain Move-

ments in p38 MAPK". *ChemMedChem: Chemistry Enabling Drug Discovery* 3.2 (2008), pp. 336–344.

[171] Norel, R., Fischer, D., Wolfson, H. J., Nussinov, R. "Molecular surface recognition by a computer vision-based technique". *Protein Engineering, Design and Selection* 7.1 (1994), pp. 39–46.

[172] Miller, M. D., Kearsley, S. K., Underwood, D. J., Sheridan, R. P. "FLOG: a system to select 'quasi-flexible'ligands complementary to a receptor of known three-dimensional structure". *Journal of computer-aided molecular design* 8.2 (1994), pp. 153–174.

[173] Diller, D. J., Merz Jr, K. M. "High throughput docking for library design and library prioritization". *Proteins: Structure, Function, and Bioinformatics* 43.2 (2001), pp. 113–124.

[174] Burkhard, P., Taylor, P., Walkinshaw, M. "An example of a protein ligand found by database mining: description of the docking method and its verification by a 2.3 Å X-ray structure of a Thrombin-Ligand complex". *Journal of molecular biology* 277.2 (1998), pp. 449–466.

[175] Rarey, M., Kramer, B., Lengauer, T., Klebe, G. "A fast flexible docking method using an incremental construction algorithm". *Journal of molecular biology* 261.3 (1996), pp. 470–489.

[176] DesJarlais, R. L., Sheridan, R. P., Dixon, J. S., Kuntz, I. D., Venkataraghavan, R. "Docking flexible ligands to macromolecular receptors by molecular shape". *Journal of medicinal chemistry* 29.11 (1986), pp. 2149–2153.

[177] Leach, A. R., Kuntz, I. D. "Conformational analysis of flexible ligands in macromolecular receptor sites". *Journal of Computational Chemistry* 13.6 (1992), pp. 730–748.

[178] Ewing, T. J., Makino, S., Skillman, A. G., Kuntz, I. D. "DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases". *Journal of computer-aided molecular design* 15.5 (2001), pp. 411–428.

[179] Welch, W., Ruppert, J., Jain, A. N. "Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites". *Chemistry & biology* 3.6 (1996), pp. 449–462.

[180] Schnecke, V., Kuhn, L. A. "Virtual screening with solvation and ligand-induced complementarity". *Virtual Screening: An Alternative or Complement to High Throughput Screening?* Springer, 2000, pp. 171–190.

[181] Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S., Peter Johnson, A. "eHiTS: an innovative approach to the docking and scoring function problems". *Current Protein and Peptide Science* 7.5 (2006), pp. 421–435.

[182] Goodsell, D. S., Lauble, H., Stout, C. D., Olson, A. J. "Automated docking in crystallography: analysis of the substrates of aconitase". *Proteins: Structure, Function, and Bioinformatics* 17.1 (1993), pp. 1–10.

[183] Hart, T. N., Read, R. J. "A multiple-start Monte Carlo docking method". *Proteins: Structure, Function, and Bioinformatics* 13.3 (1992), pp. 206–222.

[184] Goodsell, D. S., Olson, A. J. "Automated docking of substrates to proteins by simulated annealing". *Proteins: Structure, Function, and Bioinformatics* 8.3 (1990), pp. 195–202.

[185] Abagyan, R., Totrov, M., Kuznetsov, D. "ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation". *Journal of computational chemistry* 15.5 (1994), pp. 488–506.

[186] McMartin, C., Bohacek, R. S. "QXP: powerful, rapid computer algorithms for structure-based drug design". *Journal of computer-aided molecular design* 11.4 (1997), pp. 333–344.

[187] Sutter, J., Li, J., J Maynard, A., Goupil, A., Luu, T., Nadassy, K. "New features that improve the pharmacophore tools from Accelrys". *Current computer-aided drug design* 7.3 (2011), pp. 173–180.

[188] Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., Olson, A. J. "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function". *Journal of computational chemistry* 19.14 (1998), pp. 1639–1662.

[189] Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., Taylor, R. D. "Improved protein–ligand docking using GOLD". *Proteins: Structure, Function, and Bioinformatics* 52.4 (2003), pp. 609–623.

[190] Clark, K. P. "Flexible ligand docking without parameter adjustment across four ligand–receptor complexes". *Journal of Computational Chemistry* 16.10 (1995), pp. 1210–1226.

[191] Taylor, J. S., Burnett, R. M. "DARWIN: a program for docking flexible molecules". *Proteins: Structure, Function, and Bioinformatics* 41.2 (2000), pp. 173–191.

[192] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A. "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules". *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197.

[193] Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Weiner, P. "A new force field for molecular mechanical simulation of nucleic acids and proteins". *Journal of the American Chemical Society* 106.3 (1984), pp. 765–784.

[194] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. a., Karplus, M. "CHARMM: a program for macromolecular energy, minimization, and dynamics calculations". *Journal of computational chemistry* 4.2 (1983), pp. 187–217.

[195] Liao, C., Sitzmann, M., Pugliese, A., Nicklaus, M. C. "Software and resources for computational medicinal chemistry". *Future medicinal chemistry* 3.8 (2011), pp. 1057–1085.

[196] Kollman, P. "Free energy calculations: applications to chemical and biochemical phenomena". *Chemical reviews* 93.7 (1993), pp. 2395–2417.

[197] Åqvist, J., Luzhkov, V. B., Brandsdal, B. O. "Ligand binding affinities from MD simulations". *Accounts of chemical research* 35.6 (2002), pp. 358–365.

[198] Carlson, H. A., Jorgensen, W. L. "An extended linear response method for determining free energies of hydration". *The Journal of Physical Chemistry* 99.26 (1995), pp. 10667–10673.

[199] Huang, N., Kalyanaraman, C., Bernacki, K., Jacobson, M. P. "Molecular mechanics methods for predicting protein–ligand binding". *Physical Chemistry Chemical Physics* 8.44 (2006), pp. 5166–5177.

[200] Meng, E. C., Shoichet, B. K., Kuntz, I. D. "Automated docking with grid-based energy evaluation". *Journal of computational chemistry* 13.4 (1992), pp. 505–524.

[201] Shoichet, B. K., Leach, A. R., Kuntz, I. D. "Ligand solvation in molecular docking". *Proteins: Structure, Function, and Bioinformatics* 34.1 (1999), pp. 4–16.

[202] Nicholls, A., Honig, B. "A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation". *Journal of computational chemistry* 12.4 (1991), pp. 435–445.

[203] Böhm, H.-J. "Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs". *Journal of computer-aided molecular design* 12.4 (1998), pp. 309–309.

[204] Böhm, H.-J. "LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads". *Journal of Computer-Aided Molecular Design* 6.6 (1992), pp. 593–606.

[205] Gehlhaar, D. K., Verkhivker, G. M., Rejto, P. A., Sherman, C. J., Fogel, D. R., Fogel, L. J., Freer, S. T. "Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming". *Chemistry & biology* 2.5 (1995), pp. 317–324.

[206] Gehlhaar, D. K., Moerder, K. E., Zichi, D., Sherman, C. J., Ogden, R. C., Freer, S. T. "De novo design of enzyme inhibitors by Monte Carlo ligand generation". *Journal of medicinal chemistry* 38.3 (1995), pp. 466–472.

[207] Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., Freer, S. T., Larson, V., Luty, B. A., Marrone, T., et al. "Deciphering common failures in molecular docking of ligand-protein complexes". *Journal of computer-aided molecular design* 14.8 (2000), pp. 731–751.

[208] Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P. "Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes". *Journal of computer-aided molecular design* 11.5 (1997), pp. 425–445.

[209] Charifson, P. S., Corkery, J. J., Murcko, M. A., Walters, W. P. "Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins". *Journal of medicinal chemistry* 42.25 (1999), pp. 5100–5109.

[210] Ballester, P. J., Mitchell, J. B. "A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking". *Bioinformatics* 26.9 (2010), pp. 1169–1175.

[211] Wójcikowski, M., Ballester, P. J., Siedlecki, P. "Performance of machine-learning scoring functions in structure-based virtual screening". *Scientific Reports* 7.1 (2017), pp. 1–10.

[212] Olubiyi, O., Frenzel, D., Bartnik, D., Gluck, J., Brener, O., Nagel-Steger, L., Funke, S., Willbold, D., Strodel, B. "Amyloid aggregation inhibitory mechanism of arginine-rich D-peptides". *Current medicinal chemistry* 21.12 (2014), pp. 1448–1457.

[213] Wiesehan, K., Buder, K., Linke, R. P., Patt, S., Stoldt, M., Unger, E., Schmitt, B., Bucci, E., Willbold, D. "Selection of D-amino-acid peptides that bind to Alzheimer's disease amyloid peptide Aβ1–42 by mirror image phage display". *Chembiochem* 4.8 (2003), pp. 748–753.

[214] Aileen Funke, S., Willbold, D. "Peptides for therapy and diagnosis of Alzheimer's disease". *Current pharmaceutical design* 18.6 (2012), pp. 755–767.

[215] Jiang, Y., Jiang, X., Shi, X., Yang, F., Cao, Y., Qin, X., Hou, Z., Xie, M., Liu, N., Fang, Q., et al. "α-Helical Motif as Inhibitors of Toxic Amyloid-β Oligomer Generation via Highly Specific Recognition of Amyloid Surface". *Iscience* 17 (2019), pp. 87–100.

[216] Frenkel-Pinter, M., Tal, S., Scherzer-Attali, R., Abu-Hussien, M., Alyagor, I., Eisenbaum, T., Gazit, E., Segal, D. "Naphthoquinone-tryptophan hybrid inhibits aggregation of the tau-derived peptide PHF6 and reduces neurotoxicity". *Journal of Alzheimer's Disease* 51.1 (2016), pp. 165–178.

[217] Saelices, L., Nguyen, B. A., Chung, K., Wang, Y., Ortega, A., Lee, J. H., Coelho, T., Bijzet, J., Benson, M. D., Eisenberg, D. S. "A pair of peptides inhibits seeding of the hormone transporter transthyretin into amyloid fibrils". *Journal of Biological Chemistry* 294.15 (2019), pp. 6130–6141.

[218] Cunningham, A. D., Qvit, N., Mochly-Rosen, D. "Peptides and peptidomimetics as regulators of protein–protein interactions". *Current opinion in structural biology* 44 (2017), pp. 59–66.

[219] Marqus, S., Pirogova, E., Piva, T. J. "Evaluation of the use of therapeutic peptides for cancer treatment". *Journal of biomedical science* 24.1 (2017), pp. 1–15.

[220] Leader, B., Baca, Q. J., Golan, D. E. "Protein therapeutics: a summary and pharmacological classification". *Nature reviews Drug discovery* 7.1 (2008), pp. 21–39.

[221] Sato, A. K., Viswanathan, M., Kent, R. B., Wood, C. R. "Therapeutic peptides: technological advances driving peptides into development". *Current opinion in biotechnology* 17.6 (2006), pp. 638–642.

[222] Vlieghe, P., Lisowski, V., Martinez, J., Khrestchatisky, M. "Synthetic therapeutic peptides: science and market". *Drug discovery today* 15.1-2 (2010), pp. 40–56.

[223] Thayer, A. M. "Improving peptides". *Chemical & Engineering News* 89.22 (2011), pp. 13–+.

[224] Werle, M., Bernkop-Schnürch, A. "Strategies to improve plasma half life time of peptide and protein drugs". *Amino acids* 30.4 (2006), pp. 351–367.

[225] Li, Z. J., Cho, C. H. "Peptides as targeting probes against tumor vasculature for diagnosis and drug delivery". *Journal of translational medicine* 10.1 (2012), pp. 1–9.

[226] Fasano, A. "Innovative strategies for the oral delivery of drugs and peptides". *Trends in biotechnology* 16.4 (1998), pp. 152–157.

[227] Teesalu, T., Sugahara, K. N., Kotamraju, V. R., Ruoslahti, E. "C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration". *Proceedings of the National Academy of Sciences* 106.38 (2009), pp. 16157–16162.

[228] Albert, R., Marbach, P., Bauer, W., Briner, U., Fricker, G., Brums, C., Pless, J. "SDZ CO 611: a highly potent glycated analog of somatostatin with improved oral activity". *Life sciences* 53.6 (1993), pp. 517–525.

[229] Kihlberg, J., Aahman, J., Walse, B., Drakenberg, T., Nilsson, A., Soederberg-Ahlm, C., Bengtsson, B., Olsson, H. "Glycosylated peptide hormones: pharmacological properties and conformational studies of analogs of [1-Desamino, 8-d-arginine] vasopressin". *Journal of medicinal chemistry* 38.1 (1995), pp. 161–169.

[230] Varamini, P., Mansfeld, F. M., Blanchfield, J. T., Wyse, B. D., Smith, M. T., Toth, I. "Synthesis and biological evaluation of an orally active glycosylated endomorphin-1". *Journal of medicinal chemistry* 55.12 (2012), pp. 5859–5867.

[231] Kovalszky, I., Surmacz, E., Scolaro, L., Cassone, M., Ferla, R., Sztodola, A., Olah, J., Hatfield, M., Lovas, S., Otvos Jr, L. "Leptin-based glycopeptide induces weight loss and simultaneously restores fertility in animal models". *Diabetes, Obesity and Metabolism* 12.5 (2010), pp. 393–402.

[232] Habault, J., Poyet, J.-L. "Recent advances in cell penetrating peptide-based anticancer therapies". *Molecules* 24.5 (2019), p. 927.

[233] Singh, T., Murthy, A. S., Yang, H.-J., Im, J. "Versatility of cell-penetrating peptides for intracellular delivery of siRNA". *Drug delivery* 25.1 (2018), pp. 1996–2006.

[234] Tesauro, D., Accardo, A., Diaferia, C., Milano, V., Guillon, J., Ronga, L., Rossi, F. "Peptide-based drug-delivery systems in biotechnological applications: recent advances and perspectives". *Molecules* 24.2 (2019), p. 351.

[235] Gorecki, M., Votano, J., Rich, A. "Peptide inhibitors of sickle hemoglobin aggregation: Effect of hydrophobicity". *Biochemistry* 19.8 (1980), pp. 1564–1568.

[236] Votano, J. R., Gorecki, M., Rich, A. "Sickle hemoglobin aggregation: a new class of inhibitors". *Science* 196.4295 (1977), pp. 1216–1219.

[237] Hanson, M. S., Xu, H., Flewelen, T. C., Holzhauer, S. L., Retherford, D., Jones, D. W., Frei, A. C., Pritchard Jr, K. A., Hillery, C. A., Hogg, N., et al. "A novel hemoglobin-binding peptide reduces cell-free hemoglobin in murine hemolytic anemia". *American Journal of Physiology-Heart and Circulatory Physiology* 304.2 (2013), H328–H336.

[238] Kubota, S., Yang, J. T. "Oligopeptides as potential antiaggregation agents for deoxyhemoglobin S". *Proceedings of the National Academy of Sciences* 74.12 (1977), pp. 5431–5434.

[239] Noguchi, C. T., Schechter, A. N. "Inhibition of sickle hemoglobin gelation by amino acids and related compounds". *Biochemistry* 17.25 (1978), pp. 5455–5459.

[240] Schechter, A. N. "Stereospecific inhibitors of the gelation of sickle hemoglobin". *Hemoglobin* 4.3-4 (1980), pp. 335–345.

[241] Adachi, K., Ding, M., Wehrli, S., Reddy, K. S., Surrey, S., Horiuchi, K. "Effects of different β73 amino acids on formation of 14-stranded fibers of Hb S versus double-stranded crystals of Hb C-Harlem". *Biochemistry* 42.15 (2003), pp. 4476–4484.

[242] Akbar, M. G., Tamura, Y., Ding, M., Ding, H., Rosenblatt, M. M., Reddy, K. S., Surrey, S., Adachi, K. "Inhibition of hemoglobin S polymerization in vitro by a novel 15-mer EF-helix β73 histidine-containing peptide". *Biochemistry* 45.27 (2006), pp. 8358–8367.

[243] Bruno, B. J., Miller, G. D., Lim, C. S. "Basics and recent advances in peptide and protein drug delivery". *Therapeutic delivery* 4.11 (2013), pp. 1443–1467.

[244] Miller, S. M., Simon, R. J., Ng, S., Zuckermann, R. N., Kerr, J. M., Moos, W. H. "Comparison of the proteolytic susceptibilities of homologous L-amino acid, D-amino acid, and N-substituted glycine peptide and peptoid oligomers". *Drug Development Research* 35.1 (1995), pp. 20–32.

[245] Welch, B. D., Francis, J. N., Redman, J. S., Paul, S., Weinstock, M. T., Reeves, J. D., Lie, Y. S., Whitby, F. G., Eckert, D. M., Hill, C. P., et al. "Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance". *Journal of virology* 84.21 (2010), pp. 11235–11244.

[246] Funke, S. A., Willbold, D. "Mirror image phage display—a method to generate D-peptide ligands for use in diagnostic or therapeutical applications". *Molecular BioSystems* 5.8 (2009), pp. 783–786.

[247] Van Rossum, G., Drake, F. "Python 3 reference manual createspace". *Scotts Valley, CA* (2009).

[248] Tien, M. Z., Sydykova, D. K., Meyer, A. G., Wilke, C. O. "PeptideBuilder: A simple Python library to generate model peptides". *PeerJ* 1 (2013), e80.

[249] Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. "Biopython: freely available Python tools for computational molecular biology and bioinformatics". *Bioinformatics* 25.11 (2009), pp. 1422–1423.

[250] Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A., Kmiecik, S. "CABSdock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site". *Nucleic acids research* 43.W1 (2015), W419–W424.

[251] Abraham, M. J., van der Spoel, D., Lindahl, E., Hess, B., Tgd, T. "GROMACS user manual version 2018". *Royal Institue of Technology and Uppsala University* (2018).

[252] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., et al. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields". *Journal of computational chemistry* 31.4 (2010), pp. 671–690.

[253] Garton, M., Nim, S., Stone, T. A., Wang, K. E., Deber, C. M., Kim, P. M. "Method to generate highly stable D-amino acid analogs of bioactive helical peptides using a mirror image of the entire PDB". *Proceedings of the National Academy of Sciences* 115.7 (2018), pp. 1505–1510.

[254] Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., Smith, J. C. "Ensemble docking in drug discovery". *Biophysical journal* 114.10 (2018), pp. 2271–2278.

[255] Olubiyi, O. O., Olagunju, M., Keutmann, M., Loschwitz, J., Strodel, B. "High throughput virtual screening to discover inhibitors of the main protease of the coronavirus SARS-CoV-2". *Molecules* 25.14 (2020), p. 3193.

[256] Loschwitz, J., Jackering, A., Keutmann, M., Olagunju, M., Eberle, R. J., Coronado, M. A., Olubiyi, O. O., Strodel, B. "Novel inhibitors of the main protease enzyme of SARS-CoV-2 identified via molecular dynamics simulation-guided in vitro assay". *Bioorganic Chemistry* 111 (2021), p. 104862.

[257] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. "Comparison of simple potential functions for simulating liquid water". *The Journal of chemical physics* 79.2 (1983), pp. 926–935.

[258] Bussi, G., Donadio, D., Parrinello, M. "Canonical sampling through velocity rescaling". *The Journal of chemical physics* 126.1 (2007), p. 014101.

[259] Evans, D. J., Holian, B. L. "The nose–hoover thermostat". *The Journal of chemical physics* 83.8 (1985), pp. 4069–4074.

[260] Braga, C., Travis, K. P. "A configurational temperature Nosé-Hoover thermostat". *The Journal of chemical physics* 123.13 (2005), p. 134101.

[261] Parrinello, M., Rahman, A. "Crystal structure and pair potentials: A molecular-dynamics study". *Physical review letters* 45.14 (1980), p. 1196.

[262] Parrinello, M., Rahman, A. "Polymorphic transitions in single crystals: A new molecular dynamics method". *Journal of Applied physics* 52.12 (1981), pp. 7182–7190.

[263] Parrinello, M., Rahman, A. "Strain fluctuations and elastic constants". *The Journal of Chemical Physics* 76.5 (1982), pp. 2662–2666.

[264] Hess, B., Bekker, H., Berendsen, H. J., Fraaije, J. G. "LINCS: a linear constraint solver for molecular simulations". *Journal of computational chemistry* 18.12 (1997), pp. 1463–1472.

[265] Humphrey, W., Dalke, A., Schulten, K. "VMD: visual molecular dynamics". *Journal of molecular graphics* 14.1 (1996), pp. 33–38.

[266] Turner, P. "XMGRACE, Version 5.1. 19". *Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR* (2005).

[267] Sander, T., Freyss, J., von Korff, M., Rufener, C. "DataWarrior: an open-source program for chemistry aware data visualization and analysis". *Journal of chemical information and modeling* 55.2 (2015), pp. 460–473.

[268] Bhal, S. K. "LogP—making sense of the value". *Advanced Chemistry Development: Toronto, ON, Canada* (2007), pp. 1–4.

[269] Salmaso, V., Moro, S. "Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview". *Frontiers in pharmacology* (2018), p. 923.

[270] Chiti, F., Dobson, C. M. "Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade". *Annual review of biochemistry* 86 (2017), pp. 27–68.

[271] Ross, C. A., Poirier, M. A. "Protein aggregation and neurodegenerative disease". *Nature medicine* 10.7 (2004), S10–S17.

[272] Noguchi, C. T., Schechter, A. N. "Sickle hemoglobin polymerization in solution and in cells". *Annual review of biophysics and biophysical chemistry* 14.1 (1985), pp. 239–263.

[273] Galamba, N., Pipolo, S. "On the binding free energy and molecular origin of sickle cell hemoglobin aggregation". *The Journal of Physical Chemistry B* 122.30 (2018), pp. 7475–7483.