Chapter 4
Conclusions and Future Outlook

HbS fiber formation is a fairly complex process in which specific contacts are formed between several residues from both α- and β-chains of different HbS tetramers. The identity of these residues are determined by two experimental methods. Based on solution polymerization studies, a residue is defined as a contact residue if introducing a mutation on that site results in altered polymerization behavior compared to native HbS. In the fiber models, a distance based cutoff is used to define contact residues; two residues from two different tetramers are considered to form a contact site only if the distance between any pairs of atoms from those two residues are 5 Å or less. Here in this study we showed the first definition does not always give correct results because mutation of residues outside the contact regions are also capable of altering polymerization behavior of HbS. Molecular dynamics simulations showed that such mutations exert their effect up to distant sites by altering the dynamic behavior of the protein.

The effect of simultaneous mutations of the contact residues have been experimentally demonstrated for some sites but the mechanistic basis of additive or non-additive coupling between such mutated sites is not clear. Based on the static structure of the fiber one might expect that in the absence of direct contact between the sites, the effect of simultaneous mutations must be simply additive but this is not found to be always the case and the effects can be non-additive also. Results from our molecular dynamics simulations suggest non-additive effects of multiple mutation are due to communication between the mutated sites through a dynamically linked network of residues between the sites of the mutations.
However a limitation of the analyses reported in this thesis is that the results obtained are of a qualitative nature. The protocols used by us did not allow for a quantitative estimate of the mutation induced change in the free energy of binding between adjacent hemoglobin tetramers nor could an accurate prediction of the experimental $C_{\text{sat}}$ values be obtained. Estimation of changes in binding free energy due to the mutations could have been made by sophisticated molecular dynamic based protocol like free energy perturbation methods. Such studies should be carried out in the future.

Delineation of intrinsic strength of each site and interaction-linkage between two or more contact and non-contact sites are of considerable importance in the context of designing new mutants with novel properties. Prediction and designing such mutations is best possible only if the dynamic behavior of the system is taken into the account. Considering the huge number of single and multiple mutations and time consuming nature of molecular dynamics simulations, we developed a coarse-grained method for modeling dynamic behavior of the system. This method was developed on the basis of the Gaussian Network Model (GNM). In traditional GNM, a simplified representation of the molecules is used where each residue is represented by a dimensionless point. Interactions between pairs of residues are represented by a harmonic spring connecting the points. Residues are thought to interact with each other if the distance separating them is less than a specified cutoff and not otherwise. Furthermore all types of interactions are treated equally by specifying a single harmonic force constant to all the springs. Therefore it is impossible to study the role of mutations on the dynamic behavior of the system. In our method, the diverse interactions between different residues within the protein was modeled by considering non-uniform radii of interaction for different residues which were weighted according to the strength of the interaction between two residues. The weights were optimized until the predicted dynamic behavior of the system best matched with that obtained from explicit MD simulations. The optimized GNM model could very well reproduce dynamic properties like root
mean squared fluctuations obtained from explicit MD simulations. Optimizing the GNM against mutant \(\alpha\)-chain simulations of HbS could recover the essence of long rage interactions between different sites in the molecule.

In an already optimized GNM model against a native simulation, one can change the interaction radius for a particular residue. Changing the radius of interaction for one residue results in changes in interactions between that residue with other residues within the system. So such alteration can be thought of as equivalent to introducing mutations in that site. This step can be followed up by re-optimizing the GNM against the native MD simulation data by letting other radii of interactions to vary. The residues with maximum change in their radius of interactions are most likely those residues that dynamically interact with the mutated residue and mutating these are likely to nullify the effect of the first mutation.

So in practice one only needs to perform one all-atom explicit MD simulation for the native protein and optimize the GNM against that simulation. This allows for a rapid method of \textit{in silico} mutagenesis that takes into account roles played by the overall dynamics of the molecules as well as the effect of interactions between different mutation sites.

By implementing this method on the HbS \(\alpha\)-chain we identified dynamically coupled residues which we had earlier pin pointed by explicit MD simulations.

The overall results, although satisfactory for identification of residue positions mutating which could have desired effects fell short of identifying the exact residue substitution required. Further research in this area should attempt not only to find the residue positions for mutagenesis but also should attempt to identify the exact residue substitutions required. One way to do this would be to construct GNMs from all heavy atoms of the molecule so that the chemical identities of individual residues are now explicit. Training such all atom GNMs against explicit MD simulations can
possibly identify the exact residue substitutions necessary to obtain a desired phenotype.