FOR THE RECORD

A model for the sickle hemoglobin fiber using both mutation sites

ADAM ROUFBERG AND FRANK A. FERRONE

Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104

(RECEIVED January 3, 2000; FINAL REVISION March 27, 2000; ACCEPTED March 30, 2000)

Abstract: The standard molecular model of the fiber of the sickle hemoglobin (HbS: $\beta 6$ Glu $\rightarrow$ Val) has been revised to allow both $\beta 6$ mutation sites to participate in intermolecular contacts, rather than only one $\beta 6$ site as previously thought, for four molecules per 14-molecule fiber cross section. This structure accurately predicts the copolymerization of hybridized mixtures of HbS with HbA or HbC ($\beta 6$ Glu $\rightarrow$ Lys), which could not be reconciled with prior models in which only half the $\beta 6$ sites were required for assembly. This model suggests new contacts within the fiber and raises the question of whether these cross-linked double strands could possess added stability important in such processes as nucleation.

Keywords: copolymerization; hemoglobin; nucleation; sickle; structure

Sickle cell disease arises because a negatively charged amino acid ($\beta 6$ Glu) has been mutated to a hydrophobic Val on the surface of the hemoglobin molecule. The consequence of this mutation is the formation of long polymers that rigidify the cell and compromise circulation. Electron microscopy has revealed that these polymers contain 14 individual strands twisted in a rope-like fashion (Dykes et al., 1979). Though the importance of the atomic level structure of the polymers is evident, experimental difficulties have thwarted attempts to determine fiber structure by X-ray diffraction. The crystals that do form from concentrated deoxygenated hemoglobin S solutions contain linear double strands rather than twisted fibers (Wishner et al., 1975). The structure of such double strands has been determined (Wishner et al., 1975; Padlan & Love, 1985) and refined to 2 Å (Harrington et al., 1997). Work on natural mutants has established that the same contacts between molecules seen in the double strands occur in the polymers (cf. Eaton & Hofrichter, 1990). This has led to the hypothesis that the double strands form a basic element of the polymer structure, which should be accordingly viewed as seven double-strand pairs. Using this motif and a template deduced from electron microscopy, models of the fiber

have been constructed, permitting hypotheses to be formulated and tested regarding the pairing of contacts that do not appear in the double strands themselves (Watowich et al., 1989, 1993; Cretegny & Edelstein, 1993). Of these models, the most complete is that of Watowich et al. (1993), which we label here as WJ. Clearly, by standards of modern crystallography, these models are far from refined.

The double strands form two classes of contacts—axial and lateral. The axial contacts comprise hydrophobic interactions occurring across a broad group of amino acids between successive molecules in either of the members of the double-strand pair. The lateral contacts are between a given molecule on one strand and both neighboring half-staggered molecules on the adjacent strand. The lateral contact contains the $\beta 6$ Val (donor) interacting with a hydrophobic pocket comprised of $\beta 88$ Leu and $\beta 85$ Phe and several heme atoms. Each molecule thus interacts with one lateral neighbor as a donor, and the other lateral neighbor as a receptor.

A common feature in the fiber models is the participation in polymer contacts of only one of the two $\beta 6$ mutation sites, which directly reflects the findings of double-strand crystals in which only one contact is observed. The position of the remaining 14 $\beta 6$ amino acids is clearly of interest. Four of these amino acid positions appear on the surface of the fiber, where they are hypothesized to participate in the heterogeneous nucleation reaction that gives sickle cell polymerization its characteristic abruptness (and delay time) (Mirchev & Ferrone, 1997). The remaining 10 have had no particular assigned contacts, and from the model of Watowich et al. (1989), the unaccounted $\beta 6$ groups appear in positions accessible to pockets of solvent within the polymer.

Such models can be tested by examining the behavior of sickle hemoglobin mixed with nonpolymerizing hemoglobin A. Tetrameric hemoglobin will naturally dissociate into its constituent dimers, and thus a mixture of HbA and HbS will reconstitute hybrid molecules with one $\beta 6$ Val and one $\beta 6$ Glu (as well as molecules of pure HbA and pure HbS). The copolymerization of such hybrids can be deduced by measuring the composition or solubility of gels made from such solutions. The WJ model predicts a copolymerization probability of 0.5, but data collected over 10 years by seven different laboratories produces mutually consistent and significantly lower copolymerization probability

Reprint requests to: Frank Ferrone, Department of Physics, Drexel University, 32nd and Chestnut Streets, Philadelphia, Pennsylvania 19104; e-mail: fferrone@einstein.drexel.edu.

Fig. 1. Copolymerization probability for various mixtures of Hb with HbS. Because hemoglobin naturally dissociates into $\alpha\beta$ dimer pairs, a mixture of hemoglobins reproporionates, creating hybrids whose copolymerization can be determined. The top two bars (diagonal shading) show copolymerization of HbC ($\beta 6$ Glu $\rightarrow$ Lys) with HbS; the solid bars show copolymerization of HbA with HbS. In the latter group, the darker bars show copolymerization determined by assaying the composition of the supernatant following centrifugation of gels, while the lighter bars represent copolymerization probability determined from solubility measured of the supernatant. Copolymerization probabilities were computed and tabulated by Eaton and Hofrichter (1990). The dashed vertical line at 0.5 represents the expected value if only a single $\beta 6$ is involved in contacts for each molecule and clearly is not representative of the data. The vertical line at 0.357 represents the calculated value for the model presented here.

(see Fig. 1). Such a result suggests that more than half the $\beta 6$ sites are found in contacts.

Results and discussion: Upon inspection of the model of WJ, we observed two locations in which the internal $\beta 6$ Val were relatively close to their normal crystal contact partners. We hypothesized that it might be possible to dock a second $\beta 6$ into a corresponding double-strand receptor pocket for two of the inner 10 unassigned $\beta 6$ valines (Fig. 2). The energetic motivation for this is based on the assumption that a significant amount of the stabilization of the fiber arises from the hydrophobic interaction that buries $\beta 6$ and simultaneously occludes the hydrophobic pocket. Such motivation is justified, based on solvent-accessible areas, and standard values for the energy associated with hydrophobicity by which an energy of about 3 kcal/mol per contact site was deduced out of 6–7 kcal/mol total contact energy (Cao & Ferrone, 1997). Manual docking was thus performed, guided by the requirement that the new site has a geometry similar to that found in the crystalline contact.

A cross section of the fiber is shown in Figure 2. The four outer strands on each side of the fiber (drawn as darker circles) unite two double strands in a geometry we shall refer to as a “tetrad.” (There are two tetrads per fiber cross section.) The pairing is tighter than in the WJ structure, as can be seen comparing Figure 2A (WJ) with Figure 2B (RF), but the internal structure of the double-strand pairs (e.g., contact between 1 and 2 and between 10 and 11) is not changed. The second $\beta 6$ contact is designated in Figure 2C by the filled circles between molecules 1 and 11 or 6 and 13. A detailed view of the contact region from the viewpoint of the val $\beta 6$ is

Fig. 2. Cross section of the fiber. **A:** Pairing as modeled by Watowich and Josephs (Watowich et al., 1989, 1993), and denoted the WJ model. The Hb molecules have had a central core of amino acids computationally removed for ease of display. **B:** Pairing as proposed here, denoted RF. Note that molecules 1 and 11 are closer than in **A**. Placement of 1 relative to 2 and 10 relative to 11 are unchanged. **C:** Cross section of the fiber, showing the new contacts. The tetrad of **B** is drawn with a dark outline. Each $\beta 6$ contact is shown as a circle. Such a circle actually denotes two contacts: one as donor up the fiber, and one as a receptor down the fiber. This is due to the half stagger nature of the double strands. Thus, if 10 is donor to 11 at one layer, then 11 donates to the next upper 10. The $\beta 6$ contacts between double strand partners are drawn with an open circle; the new contact proposed here is drawn with a filled circle. Note that in the WJ prescription, molecule 10 does not maintain double-strand contacts as found in the crystal. The molecule 10 used in our investigations was generated by using the crystal structure as a template starting from molecule 11, which is the double-strand partner to molecule 10 in the fiber.

shown in Figure 3. The contact created here cannot be directly compared with the crystallographic double-strand contact because only one $\beta 6$ is in contact within the crystal double strand, and thus the $\beta 6$ we are viewing has no contact partner in the crystal. For

Fig. 3. Comparison of the proposed contact region with a crystal-like contact. The Val side chain, from the α carbon to the end of the side chain, is shown for the donor; the receptor region is drawn in a space-filling representation. The $\beta 88$ Leu in the receptor pocket is the unlabeled space-filling group in the center, colored in purple. The crystal geometry would give the red Val (left), while our structure produces the green Val on the right. Because only one $\beta 6$ is in contact within the crystal double strand, the contact created here cannot be directly compared with the crystallographic double-strand contact. Thus, the comparison is made for a hypothetical molecule in an orientation that would have given a correct contact at the primary $\beta 6$ site. The overall displacement of the α carbons is about 3.4 Å.

comparison, therefore, we placed a hypothetical molecule in an orientation that would have given a correct contact at the primary $\beta 6$ site. The position of that hypothetical $\beta 6$ is then shown in Figure 3 (the left Val, shown in red) to compare with the structure obtained by the procedure described above (right Val, in green). As can be seen, the overall displacement is not great, being about 3.4 Å for the α carbons, and the Val remains generally within the hydrophobic pocket.

The resulting structure was systematically surveyed for drastically unfavorable contacts, such as similar charges in close proximity. All contacts <5 Å between like charges occur within the double strands themselves and not between the double strands after we have repositioned them. Hence, charge pairing creates no problems for the tetrad geometry.

This model is capable of rationalizing the copolymerization experiments that have been extensively performed on mixtures of HbS with HbA and HbC, a naturally occurring mutation in which the $\beta 6$ site becomes positive, viz. $\beta 6$ Glu $\rightarrow$ Lys. The copolymerization probabilities are the same for HbS with HbA or HbC within experimental uncertainty, implying that a similarly large penalty appears for the burial of a charged residue in HbA and HbC. Analysis of the mixture experiments yields a copolymerization probability, i.e., the likelihood that a given molecule will be found in the polymer. Such a probability depends on the free energy of discrimination, but in the case of a highly unfavored energy, the free energy penalty is great enough that the probability becomes an all or none measure, and thus an effective determination of the fraction of accessible contacts.

In our case, within the tetrad there are two molecules with both $\beta 6$ sites participating in contacts. Such a site entails a significant penalty for incorporating an S/A or S/C hybrid. Therefore, 4 of the 14 molecules in a cross section would have effectively zero copolymerization probability of incorporating an SA or SC hybrid. The remaining 10 molecules require only one "active" $\beta 6$, and thus each have 0.5 copolymerization probability. When the weighted average is computed, the copolymerization probability becomes $4/14 \times 0 + 10/14 \times 0.5 = 0.35$, which is drawn as the vertical line in Figure 1. This very close agreement with the copolymerization data thus supports this double $\beta 6$ model.

This geometry was then surveyed for new contacts that may be of significance for the stabilization of the polymer, specifically for any possible salt bridges. We found a 4.9 Å separation between the groups $\beta 2$ Lys 144 and $\beta 2$ Asp 21 in molecules 1 and 11, respectively. The "reversed" pairing, i.e., between molecule 11 and next higher molecule 1, has similar partners at somewhat farther separation, viz. 6.7 Å. Clearly, an energy minimization would be required to determine if any of the potential salt bridges suggested by these alignments can be realized.

Historically, intermolecular contact sites have been validated by the effects of specific mutations on the polymerization of sickle hemoglobin. Such a strategy works very effectively when all molecules have the same contacts, but is less effective when the contact sites are of more than one type. Because most molecules retain the original WJ contact geometry, all mutants found to have an effect will be similarly identified in the RF model. The only way the mutant data can be used is if there are mutation sites whose effects have not been ascribed and can be rationalized in this model. Four sites with effects but no contacts have been identified (Watowich et al., 1993), of which three are plausibly explained by their proximity to known contact partners. The one site without a structural correlation is $\alpha 6$, which for which the replacement (Asp

$\rightarrow$ Ala) is said to lose hydrogen bonds that stabilize the T structure. Nonetheless, this mutant favors polymerization. The tetrad picture provides a way to explain this phenomenon, by considering that the weakened T state can more flexibly contact both $\beta 6$ sites (recall the displacement in Fig. 3).

Finally, we raise the tantalizing question of whether this tetrad geometry is an enhanced stability, or compensates for the stretch of the outer fibers that is required to wrap them around the core. It has been suggested that such stretch limits the fiber diameter (Watowich et al., 1989). If the tetrad represents a net gain in stability relative to the core of the fiber, it may play a role in the kinetic processes that form the fiber and play a major role in the pathophysiology of the disease. In addition, this new geometry suggests a set of contact sites that could be explored in attempts to develop therapeutic drugs to treat this disease.

Methods: Two docking procedures were employed. The first used as its starting point the WJ model. The separation of the lateral contacts was minimized, even though axial overlaps occurred. Then axial overlaps were relaxed by axial motion. The specific details are as follows: the two-molecule element of the double strand was taken as the elementary unit. Consequently, the intermolecular double-strand contact (primary $\beta 6$) is not perturbed during the repositioning. A two-dimensional translation of the double-strand molecules 1 and 2 (see Fig. 2 for notation) of net distance 11.43 Å was performed relative to molecules 10 and 11 in a given plane, i.e., the elementary pair was translated within the cross section perpendicular to the fiber axis so as to minimize the $\beta 6$ – $\beta 88$ distance between molecules 1 and 11 without overlap in this plane of investigation. The minimized pair must then be propagated along the fiber axis to build the fiber. Because the axis of the fiber is not the same as the axis of the twisting constituent double strands, the previous minimization creates small overlaps above and below the minimized plane of investigation once we attempted to propagate the new geometry. Thus, one of the double-strand pairs (molecules 1 and 2) of the contact region was displaced along the fiber axis (normal to the plane of investigation) to relieve the overlap and again laterally repositioned to closest contact. This procedure was iterated several times in steps of <3 Å with a twist of $<4^\circ$. In a second procedure, we attempted to retain the same contact in both $\beta 6$ regions. This led to a significant overlap of the van der Waals radii. Beginning with this overlapped geometry, we relaxed the contact by translation and rotation, keeping a minimum distance without overlap. This procedure was not as successful as the first protocol, although the final structures were quite similar.

Acknowledgments: The authors thank Prof. H.-X. Zhou and Dr. Vijayakumar for helpful discussions. This work was supported by NIH Grants HL 57549 and HL58512.

References

- Behe MJ, Englander SW. 1979. Mixed gelation theory. Kinetics, equilibrium and gel incorporation in sickle hemoglobin mixtures. *J Mol Biol* 133:137–160.
- Benesch RE, Edalji R, Benesch R, Kwong S. 1980. Solubilization of hemoglobin S by other hemoglobins. *Proc Natl Acad Sci USA* 77:5130–5134.
- Bookchin RM, Balazs T. 1986. Ionic strength dependence of the polymer solubilities of deoxyhemoglobin S + C and S + A mixtures. *Blood* 67:887–892.
- Bunn HF, Noguchi CT, Hofrichter J, Schechter GP, Schechter AN, Eaton WA. 1982. Molecular and cellular pathogenesis of hemoglobin SC disease. *Proc Natl Acad Sci USA* 79:7527–7531.

- Cao Z, Ferrone FA. 1997. Homogeneous nucleation in sickle hemoglobin. Stochastic measurements with a parallel method. *Biophys J* 72:343–372.
- Cretegny I, Edelstein SJ. 1993. Double strand packing in hemoglobin S fibers. *J Mol Biol* 230:733–738.
- Dykes GW, Crepeau RH, Edelstein SJ. 1979. Three dimensional reconstruction of 14-filament fibers of hemoglobin S. *J Mol Biol* 130:451–472.
- Eaton WA, Hofrichter J. 1990. Sickle cell hemoglobin polymerization. *Adv Protein Chem* 40:63–280.
- Goldberg MA, Husson MA, Bunn HF. 1977. Participation of hemoglobins A and F in the polymerization of sickle hemoglobin. *J Biol Chem* 252:3414–3421.
- Harrington DL, Adachi K, Royer WE Jr. 1997. The high resolution crystal structure of deoxyhemoglobin S. *J Mol Biol* 272:398–407.
- Mirchev R, Ferrone FA. 1997. The structural origin of heterogeneous nucleation and polymer cross linking in sickle hemoglobin. *J Mol Biol* 265:475–479.
- Padlan EA, Love WE. 1985. Refined crystal structure of deoxyhemoglobin S I. Restrained least squares refinement at 3.0 Å resolution. *J Biol Chem* 260:8272–8279.
- Sunshine HR, Hofrichter J, Eaton WA. 1979. Gelation of sickle cell hemoglobin in mixtures with normal adult and fetal hemoglobins. *J Mol Biol* 133:435–467.
- Watowich SJ, Gross LJ, Josephs R. 1993. Analysis of the intermolecular contacts within sickle hemoglobin fibers: Effect of site-specific substitutions, fiber pitch and double-strand disorder. *J Struct Biol* 111:161–179.
- Watowich SJ, Gross LJ, Josephs RJ. 1989. Intramolecular contacts within sickle hemoglobin fibers. *J Mol Biol* 209:821–828.
- Wishner BC, Ward KB, Lattman EE, Love WE. 1975. Crystal structure of sickle-cell deoxyhemoglobin at 5 Å resolution. *J Mol Biol* 98:179–194.