Expression of Recombinant Monomer Hemoglobins (Component IV) from the Marine Annelid *Glycera dibranchiata*: Evidence for Primary Sequence Positional Regulation of Heme Rotational Disorder†

Steve L. Alam,‡ David P. Dutton,§ and James D. Satterlee*,§

Departments of Biochemistry and Biophysics and of Chemistry, Washington State University, Pullman, Washington 99164

Received March 9, 1994; Revised Manuscript Received May 23, 1994®

ABSTRACT: A description of the efficient high-level expression of the monomer hemoglobin (GMG4) from *Glycera dibranchiata* is presented. The cDNA described by Simons and Satterlee [Simons, P. C., & Satterlee, J. D. (1989) *Biochemistry 28*, 8525–8530] was subcloned into an expression system, and conditions were found that led to the production of large amounts of soluble apoprotein (rec-gmg). These conditions included lowering the temperature during the induction period and growth in a rich medium with a higher ionic strength. Characterization of this reconstituted recombinant protein showed that it was not identical to the native GMH4 protein. Both UV–visible and ^{1}H NMR data indicated differences within the holoprotein (rec-gmh) heme pocket compared to the native protein, the major difference being that two nonidentical heme orientations are significantly populated in rec-gmh. This phenomenon has been seen previously in other heme proteins, where these heme orientational isomers are described by a 180-deg rotation about the heme α–γ meso axis. This work prompted the production of a complete chemical sequence for the native GMH4 [Alam S. L., Satterlee, J. D., & Edmonds, C. G. (1994) *J. Protein Chem. 13*, 151–164], which showed that the expressed rec-gmg protein differed at three primary sequence positions (41, 95, and 123) from the native component IV globin (GMG4). Subsequently, we have produced the triple-revertant mutations required to express the recombinant wild-type protein (recGMG4). The physical characteristics of the active site in the holoprotein (recGMH4) are identical to those of the native protein. In addition, we have evidence, based on the model of GMH4, that position 41 may be crucial in determining the orientation of the heme within the active site.

The monomeric hemoglobin component IV (GMH4[1]) is one of three major monomeric hemoglobins contained within the nucleated erythrocytes of *Glycera dibranchiata* (Vinogradov et al., 1970; Kandler & Satterlee, 1983; Kandler et al., 1984). Interest in these proteins has arisen mainly due to the observation that the conserved distal histidine, seen in most hemoglobins and myoglobins, is replaced by a leucine in all three naturally occuring *Glycera dibranchiata* monomer hemoglobins. This unusual amino acid substitution was confirmed in a crystal structure of a monomeric hemoglobin (Arents & Love, 1989) by NMR data on all three proteins (Cooke & Wright, 1985a–c, 1987; Cooke et al., 1987) and by amino acid sequences obtained for all three major monomeric hemoglobins (Li & Riggs, 1971; Imamura et al., 1972; Alam et al., 1994; unpublished data from this laboratory). The kinetic implications of such an amino acid replacement are immediately recognizable and are emphasized by the finding that cyanide binding kinetics for each of the three monomeric hemoglobins showed on-rate constants that were several orders of magnitude smaller than those of sperm whale myoglobin and several other ferriheme proteins (Mintorovitch & Satterlee, 1988; Mintorovitch et al., 1989). This, in combination with the observation of an increased K_d for O_2(g) in all three monomers (Parkhurst et al., 1980), emphasizes the importance a peculiar distal residue may have in influencing ligand binding dynamics in these monomeric hemoglobins.

Previously, it was reported that a cDNA obtained from *Glycera dibranchiata* corresponded to the first 25 amino acids of the N-terminally sequenced native component IV protein (Simons & Satterlee, 1989). On this basis, the subsequently expressed protein originally was expected to correspond to wild-type component IV; however, the characterization of this protein (now termed rec-gmg) described herein ultimately led to the complete chemical sequencing of native component IV (GMH4; Alam et al., 1994) and has initiated primary sequencing efforts for the remaining two major monomeric components. The chemical sequencing described elsewhere (Alam et al., 1994) showed three amino acid differences between the native protein and the translated cDNA insert previously thought to encode for GMG4. This has not only led us to an understanding of the physical differences between the two proteins but has also enabled us to produce the revertant mutations required to express a recombinant form of the native protein (recGMG4).

One of the more interesting features of the three major *Glycera dibranchiata* monomer hemoglobins is that two forms

† J.D.S. acknowledges support of his work from the National Institutes of Health (GM47645).

* Author to whom correspondence should be addressed: Department of Chemistry, Washington State University, Pullman, WA 99164-4630. Telephone: (509)335-8620. Fax: (509)335-8867. E-mail: Hemeteam@cosy.chem.wsu.edu.

‡ Department of Biochemistry and Biophysics.

§ Department of Chemistry.

® Abstract published in *Advance ACS Abstracts,* August 1, 1994.

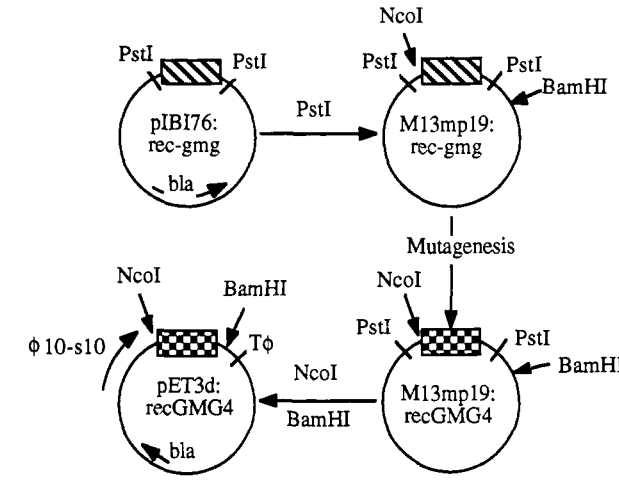
[1] Abbreviations: 2XYT, 2× yeast extract/tryptone medium; BSA, bovine serum albumin; CM, carboxymethyl; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate; GMG4, native glycera monomer globin component IV (apoprotein); GMH4, native glycera monomer hemoglobin component IV (holoprotein); rec-gmg, recombinantly expressed globin from initial cDNA insert (Simons & Satterlee, 1989); rec-gmh, holoprotein of rec-gmg; IPTG, isopropyl β-D-thiogalactopyranoside; recGMG4, recombinantly expressed glycera monomer globin component IV; recGMH4, recombinantly expressed glycera monomer hemoglobin component IV; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

0006-2960/94/0433-10337$04.50/0

are simultaneously present in solution. The major form of the native proteins constitutes ~85–90% of the total protein in any given solution. Proton NMR studies have suggested that the source of the minor form is due to heme reversal, so that the major and minor forms differ only in the orientation of the heme in the heme pocket (Constantinidis et al., 1988; Constantinidis & Satterlee, 1987). Moreover, several studies have shown that the heme orientation in each major form of the *G. dibranchiata* monomer hemoglobins is reversed compared to the predominant heme orientation of the myoglobins (Arents & Love, 1989; Cooke et al., 1987; Cooke & Wright, 1985a–c, 1987; Constantinidis et al., 1988). This type of heme orientational isomerism has been documented within other hemoglobins and myoglobins by many physical techniques, including proton NMR (Ahmad & Kincaid, 1983; La Mar et al., 1984; Lecomte et al., 1985; Yee & Peyton, 1991), UV–visible spectroscopy (Deeb & Peyton, 1992), and circular dichroism (Santucci et al., 1988; Aojula et al., 1986). For most other heme proteins, the appearance of a heme reversed orientation is transient. Generally, both heme orientations are present in nearly equal populations initially following heme reconstitution, but with time the populations relax to one predominant form (>90%). The relevance of these observations to the work presented here is that we have expressed a *G. dibranchiata* triple-mutant monomer hemoglobin (compared to the native GMG4 sequence), called recgmh, which presents proton NMR spectra consistent with the presence of nearly equal amounts of two protein forms. These forms are stable and persist in solution for long periods of time. We present evidence in this work that the two forms are due to heme rotational disorder. We also identify the most logical amino acid position in the primary sequence that functions as a heme orientation regulatory site using primary sequence comparisons, characterization of the expressed proteins, and molecular modeling. This regulatory site is a previously unrecognized heme contact.

MATERIALS AND METHODS

The cDNA coding for rec-gmg (Simons & Satterlee, 1989) was removed from pIBI76:GMG4 using *Pst*I-cut sites that flank the insert. The 745-base-pair fragment was purified using standard agarose gel protocols [standard protocols were followed as in Sambrook et al. (1989), unless otherwise stated] and recovered using a solid matrix purification kit (Prepagene) supplied by Bio-Rad. This fragment was ligated into *Pst*I-cut M13mp19 phage, so that the 3′-end of the insert was contiguous with the *Bam*HI site within the multiple cloning site of M13mp19 (Figure 1A). Since two orientations were possible, the desired orientation was confirmed using Sanger's dideoxy-sequencing methods following the protocols from the kit supplied by U.S. Biochemicals. The three mutations required to create a cDNA that encodes recGMG4 (Figure 1B: I41M–ATC→ATG:Y95N–TAT→AAC:T123N–ACT → AAC) were created using single-stranded site-directed mutagenesis (Kunkel et al., 1987). Briefly, all three mutagenic oligonucleotides (of about 25 bases in length) were annealed onto a uracil-containing template simultaneously. The complementary strand was generated with T4 DNA polymerase. This reaction mixture was transformed into JM101-(f^+), and the secreted phage was purified. Mutations at all three positions were confirmed by dideoxy-DNA sequencing. The complete 745-base pair insert was sequenced for all triple-mutant candidates to ensure the lack of nonspecific mutations. Once confirmed, the replicating form of the phage M13mp19:recGMG4 was isolated and directionally subcloned into the T7 expression plasmid, pET3d (Studier & Moffatt, 1986; Rosenburg et al., 1987; Studier et al., 1990), using the restriction enzymes *Nco*I and *Bam*HI to create pET3d:recGMG4 (Figure 1A). Expression of the recombinant monomeric globin (recGMG4) was achieved within the BL21 (DE3) pLysS strain of *Escherichia coli* (Studier & Moffatt, 1986; Studier et al., 1990).




```
                    40
                    |
1) 1--------LSAHH DIAAVFG----------------

2) 1--------LSAHH DMAAVFG----------------

     90                  120
     |                   |
1) H KGYGYKHIKA-----IGG KMTAAAKD---147

2) H KGYGNKHIKA-----IGG KMNAAAKD---147
```

FIGURE 1: (A, top) Schematic diagram depicting subcloning of the rec-gmg insert (striped box) from pIBI76:rec-gmg (Simons & Satterlee, 1989) into M13mp19 using the endonuclease *Pst*I. Single-stranded mutagenesis was performed as outlined in Materials and Methods using M13mp19:rec-gmg to construct M13mp19:recGMG4 (checkered box). This insert was subcloned into pET3d using the enzymes *Nco*I and *Bam*HI to produce the expression plasmid pET3d:recGMG4. (B, bottom) Comparison of regions within the complete sequences of (1) rec-gmg and (2) recGMG4 where the three amino acid replacements are found. The differences are highlighted in boldface print (positions 41, 95, and 123).

Induction experiments were performed to optimize the final concentration of IPTG added (0.1–1 mM), the length of the induction period (0.5–20 h), the temperature at which the culture was maintained during induction (21–37°C), and the growth medium used: (1) 2XYT: 1.6% Bactotryptone, 1% yeast extract, and 0.5% NaCl; or (2) tryptone phosphate: 2% Bactotryptone, 0.2% Na_2HPO_4, 0.1% KH_2PO_4, 0.8% NaCl, 1.5% yeast extract, and 0.2% glucose. All cultures were maintained with final concentrations of 100 μg ampicillin/mL and 30 μg chloramphenicol/mL. A typical induction experiment consisted of inoculating a culture by diluting an overnight culture 1/100 into the fresh medium and monitoring the growth progress at 37 °C by observing the optical density at 600 nm (Perkin-Elmer 559A spectrometer). Once the A_{600} reached 0.8, the induction temperature was set and IPTG was added to the desired concentration. The cells were aerated by shaking for the desired induction period. Cells were harvested by centrifugation, resuspended in one-tenth the culture volume of lysis buffer [50 mM Tris Cl, 1 mM EDTA, and 100 mM NaCl (pH 8.0)], and frozen at −80 °C. The

frozen cells were allowed to thaw and were gently shaken at room temperature to initiate the action of the internally produced lysozyme.

Cell lysis was accomplished by one of two methods, depending on the scale of induction. For small scale inductions (1–3 mL), cell lysis was achieved by vigorous vortexing of the thawed cells in the presence of acid-washed glass beads (0.45–0.5-mm diameter), a method commonly used for the mechanical lysis of yeast (Sambrook et al., 1989). For large scale inductions (1–4 L at an A_{600} of ~8.0), the resuspended cells (2 vol of 200 mL) were disrupted using 3 × 20 s sonication bursts on ice. This crude extract was called the whole cell fraction. A soluble fraction was obtained by clarifying the whole cell extract with additional centrifugation (12 000g) for 20 min at 4 °C. Induction efficiency and localization of the expressed recGMG4 (and rec-gmg) were determined by electrophoresis of the samples using a 15% SDS–polyacrylamide gel system (Laemmli, 1970). Proteins were visualized using Coomassie or silver stains.

Once induction conditions were optimized, estimates of the amount of recGMG4 were gathered using a heme titration technique (Ascoli et al., 1981) modified to account for nonspecific heme binding to other soluble proteins. Cultures (25 mL) of induced and noninduced bacteria were grown side-by-side in identical conditions, less the IPTG in the noninduced case. Initially, both cultures were grown to an A_{600} of 0.8 at 37 °C. Both cultures were incubated further at 21 °C for 5 h (±IPTG). The cells were harvested and lysed as described for the large scale inductions above. For both cultures, soluble fractions were diluted about 200 times in the lysis buffer to give a final A_{280} reading between 1 and 2 units. The noninduced sample was placed in the reference cell, while the induced supernatant was placed in the sample cell of a Perkin-Elmer 559A UV–visible spectrophotometer. Difference spectra were scanned from 700 to 190 nm. Equal aliquots (1–10 μL) of a heme stock solution were added to both the sample and reference cells, mixed by repeated pipeting, and allowed to stand at 23 °C for 5 min prior to taking the next difference spectrum in the series. The heme stock solution was prepared by dissolving 10–15 mg of hemin chloride (Porphyrin Products) in 250 μL of 1 N NaOH, diluting to 10 mL in lysis buffer, and followed by a 1/10 dilution in the same buffer to give a final hemin concentration of 1–1.5 mg of hemin/10 mL. Difference spectra were obtained until the absorbance at 392 nm (the Soret band of the native GMH4) remained constant. This end-point absorbance was used to calculate the amount of recGMH4 produced by using the absorptivity value of 129 000 M^{-1} cm^{-1} (Seamonds et al., 1971). Estimates of the total protein concentration within the soluble fraction were obtained from a colorimetric BCA assay (Pierce) using BSA as the standard.

Large scale preparations of recGMG4 (4 L) and rec-gmg (for UV–visible and NMR experiments) were further purified from the soluble fraction using ion exchange chromatography. The crude lysate (400 mL) was dialyzed against a 10× volume of 10 mM potassium phosphate at pH 6.0. At this point the sample was reconstituted with hemin. First, a hemin solution was prepared by dissolving a 10× molar excess of hemin chloride (relative to the expected recGMG4 concentration) in a minimal amount of 1 N KOH, followed by dillution to a total volume of 10 mL with dialysis buffer. This was followed by the addition of an equimolar amount of potassium ferricyanide to ensure that the free heme was in the oxidized state during reconstitution. The hemin solution was slowly stirred into the dialyzed lysate over a 1-h period, and the resulting solution was left to stand overnight in the dark at 4 °C. The reconstituted lysate (~410 mL) was pumped directly onto a CM-Sepharose column (6 × 90 cm) equilibrated with the dialysis buffer (pH 6.0). After the column had been washed with 3 vol of dialysis buffer (~2 L), the pH of the buffer was changed to 6.80 (10 mM potassium phosphate), and the column was again washed with a minimum of 3 vol of this buffer (>2 L). Column eluent was monitored at 280 nm (ISCO UA-5). Finally, the red/brown band of recGMH4 was slowly eluted using a shallow gradient of potassium chloride (0–100 mM), which required over 5 h at a flow rate of 5 mL/min. The red/brown fractions containing recGMH4 were pooled and brought to a final concentration of 1.2 M ammonium sulfate by the addition of solid ammonium sulfate. After stirring overnight, the solution was clarified by centrifugation at 12 000g for 60 min and desalted using a Sephadex G-50 gel filtration column equilibrated with 100 mM potassium phosphate and 100 mM potassium chloride, pH 6.80. Protein purity was assayed using 15% SDS–polyacrylamide gels. Native GMH4 was purified as previously described (Kandler et al., 1984; Constantinidis & Satterlee, 1987).

UV–vis spectra were obtained on an HP-8452A. Ferric forms of the three proteins (native GMH4, rec-gmg, and recGMH4) were obtained by diluting frozen stocks of the ferric proteins in 100 mM potassium phosphate and 100 mM potassium chloride (pH 6.80) until the absorbance of the Soret band (392 nm) was about 1.0 absorbance unit for all three proteins. Cyanide-ligated forms of the proteins were produced from the ferric samples by adding a small amount of 6 μM KCN buffered in 100 mM potassium phosphate and 100 mM KCl (pH 6.8). The completion of the reaction was monitored by the red shift in the Soret region. Ferrous-oxy forms of the proteins were obtained by adding a 10% molar excess of dithionite to a small amount of each of the ferric stock solutions, followed by quick passage down a G-25 desalting column (DiIorio, 1981). Each sample was diluted to give a Soret maximum of about 1 absorbance unit.

One-dimensional NMR spectra of the high-spin form (metaquo) of the proteins were gathered on a Varian VXR500s spectrometer operating at the nominal proton frequency of 500 MHz. All samples were 2–4 mM in concentration and were buffered in 100 mM potassium phosphate and 100 mM potassium chloride (pH′ 6.80) in 99.9% ^2H_2O (ISOTEC). pH′ is the direct pH meter reading in ^2H_2O without correction for the deuterium isotope effect. An 80 000-Hz spectral width was acquired using 32K data points at 25 °C. Typical experiments used 90-deg pulse widths of 5 μs and repetition rates of 10 s^{-1}. Residual water was presaturated with the decoupler during the relaxation period. All spectra were processed with 10-Hz line-broadening using the Varian VNMR software. The spectra were referenced to external DSS through the residual water peak (4.70 ppm at 25 °C).

All model structures were visualized using the Biosym INSIGHTII package. Construction of the model structure for GMH4 was described previously (Alam et al., 1994), while a model for the Met41Ile single mutant was constructed in a similar manner. The reversed heme orientation was modeled into both of the above structures by superimposing the meso protons (i.e., α on α, γ on γ, β on δ, and δ on β) of a reversed heme onto those of the heme of normal orientation within the models and only displaying the reversed form.

N-Terminal sequencing of the expressed protein (recGMG4) was carried out in the Center for Biotechnology at Washington State University on reverse-phase-purified apo-

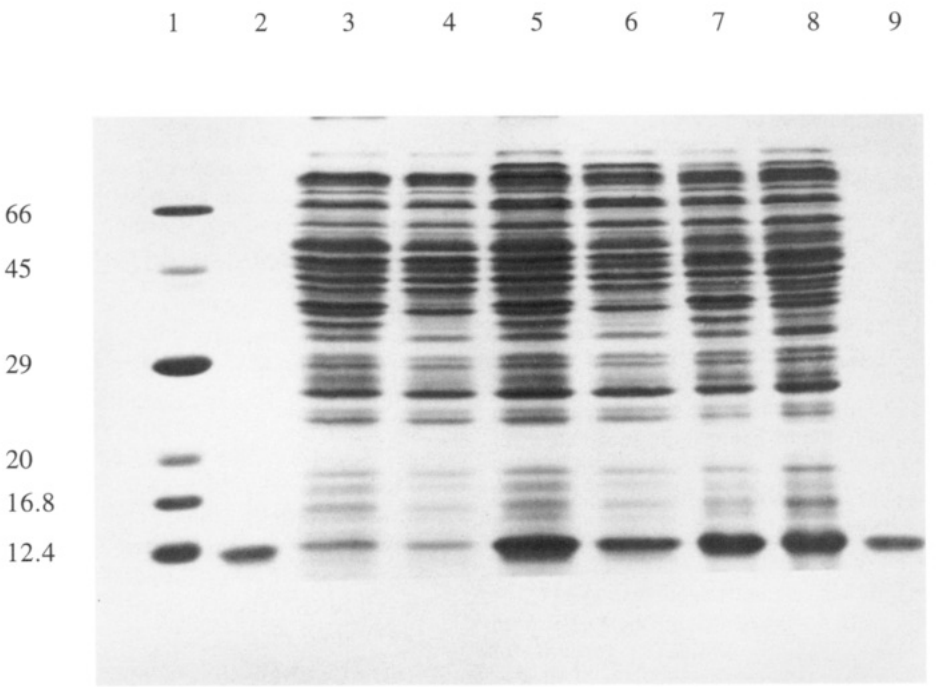


FIGURE 2: SDS–polyacrylamide gel electrophoresis (15%) showing a comparison of the level of soluble expressed recGMG4 under different induction conditions: lanes 1, protein molecular weight marker with molecular weights ($\times 10^3$) labeled to the left; lane 2, native GMH4 protein standard; lane 3, whole cell fraction of uninduced cells after 5 h of growth at 37 °C; lane 4, soluble fraction of lane 3; lane 5, 37 °C induced whole cell fraction; lane 6, 37 °C induced soluble fraction; lane 7, 21 °C induced whole cell fraction; lane 8, 21 °C induced soluble fraction; lane 9, native GMH4 protein standard.

protein samples, as described previously for the sequencing of native GMH4 (Alam et al., 1994).

RESULTS AND DISCUSSION

Complete amino acid sequencing of the native *Glycera dibranchiata* monomer hemoglobin component IV protein (GMH4; Alam et al., 1994) has shown that the cDNA thought to encode for GMG4 (Simons & Satterlee, 1989) differs at three amino acid positions (41, 95, and 123; Figure 1B) from the native GMH4 primary sequence. Physical evidence, presented here, indicates that the expressed protein translated from the original cDNA insert (rec-gmg) dramatically differs from the native protein, and these results stimulated our chemical sequencing of the native protein. This paper describes the revertant mutations required to express true native recombinant GMG4 (recGMG4) and the physical characterization of the three heme proteins (native GMH4, rec-gmh, and recGMH4). In addition, evidence is presented to indicate that an amino acid in close proximity to the heme, M41, is a factor in directing the orientation of the heme prosthetic group within the active site.

Subcloning the rec-gmg cDNA insert into M13mp19 for the purpose of single-stranded mutagenesis was accomplished using *Pst*I restriction sites that flank the cDNA insert in pIBI76:rec-gmg and inserting this fragment into the multiple cloning region of M13mp19 (Figure 1A). The three single amino acid changes (I41M, Y95N, and T123N; Figure 1B) were made simultaneously by annealing the three mutagenic oligonucleotides to the target and completing the elongation reaction. All three changes were confirmed using single-stranded dideoxy-DNA sequencing. Figure 1A also illustrates how the recGMG4 insert was subcloned into the pET3d expression plasmid (Studier & Moffatt, 1986; Studier et al., 1990).

Expression of recGMG4 in *E. coli* strain BL21(DE3) pLysS was monitored using SDS–PAGE as illustrated in Figure 2. It was discovered that during induction (5 h at 37 °C) a majority of the recGMG4 (~16 kDa) protein aggregated into insoluble inclusion bodies, with less of the expressed protein being found in the soluble fraction of the cell lysate (lane 5 versus lane 6). However, when the induction temperature was lowered to 21 °C (lanes 7 and 8), the expressed recGMG4 appeared primarily in the soluble fraction.

These qualitative assessments of expression results were derived from experiments like those shown in Figure 2. The amount of soluble protein, that is, expressed protein not found in inclusion bodies, was estimated using a uniform protocol whereby cells were induced at the two temperatures (5 h at 37 and 21 °C) and then lysed identically. For both temperatures this lysis mixture was divided into two volumes. One volume of lysed cells was treated directly with SDS–loading buffer and boiled to solubilize all proteins (termed the whole cell fraction). This sample contained proteins found in inclusion bodies and membrane proteins, as well as the cytosolic proteins. The second volume was clarified of large aggregates by centrifugation (termed soluble fraction) before being treated with the SDS–loading buffer. This sample contained only proteins soluble in the cytosol. Equal volumes of each extract were then subjected to SDS gel electrophoresis and the results are shown in Figure 2.

Lanes 5 and 6 of Figure 2 compare results for a 37 °C induction, while lanes 7 and 8 contain results from the 21 °C induction. Lane 5 corresponds to the whole cell fraction of the 37 °C induction. It is clear from this figure that the GMG4 band intensity is higher for this lane than for lane 6, the soluble fraction. We conclude from our results that at 37 °C a significant amount of expressed protein is found in insoluble inclusion bodies. This difference in GMG4 band intensity is not found in lanes 7 and 8, which compare the identically produced fractions for the induction experiments carried out at 21 °C. Lane 7 presents results for the whole cell fraction, while lane 8 presents results for the soluble

fraction. In contrast to the higher temperature incubation, there is no detectable difference in the GMG4 band intensities in lanes 7 and 8, leading us to conclude that most, if not all, of the expressed protein remains soluble in the cell cytosol during lower induction temperatures. Thus, inclusion body formation is minimized by induction at 21 °C.

Additional experiments showed that optimal conditions for induction include maintaining the 21 °C temperature during induction (Van Heeke & Schuster, 1989) and using a richer growth medium (tryptone phosphate; Moore et al., 1993) over longer induction periods (5–10 h). Heme titration experiments conducted on the expressed protein, as described in the Materials and Methods section, estimate that 48.8 ± 11.7 mg of recGMH4 is produced per liter of induced cells at 21 °C, representing approximately 5–8% of the total cellular protein. Several attempts at solubilizing and renaturing the inclusion bodies produced during 37 °C inductions were made, but they failed to produce soluble hemoglobin (data not shown).

A three-step purification generated a pure form of rec-GMH4, as illustrated by a single band after SDS–PAGE (data not shown). The protein is estimated to be >95% pure, while 40–50 mg of *pure* protein is obtained from every 4 L of cells (10–15 mg/L). The protein referred to as rec-gmh was purified in an identical manner with similar expression results. Both proteins migrate identically to the native GMH4 standard.

UV–visible spectra of the native protein and the two recombinant proteins are shown in Figure 3. Figure 3A displays the ferric high-spin metaquo forms of the three proteins, while Figure 3B,C presents the metcyano and ferrous-oxy forms, respectively. Qualitative comparisons show that for each form the spectra of the three proteins are very similar and possess spectral features characteristic of heme proteins (Seamonds et al., 1971; Eaton & Horfrichter, 1981). Closer inspection, particularly of the metaquo forms of the proteins (Figure 3A), shows that there is a distinct 4-nm red shift in the Soret band of the rec-gmh protein, while the recGMH4 and native proteins appear to be nearly identical. This difference is not apparent in the other two ligation states of the protein. This was the first evidence that suggested that the rec-gmh and native protein active sites differ enough to perturb the heme electronic structure.

One-dimensional ^1H NMR traces of the hyperfine-shifted high-frequency region of the high-spin metaquo forms of all three proteins in ^2H_2O are presented in Figure 4. Previous assignments made for the native protein (GMH4) are shown above Figure 4A, including assignments for all four of the heme methyl groups in both the major and minor heme isomeric forms (Constanidis et al., 1988). All three spectra were recorded under identical temperature and pH to simplify comparison. As is clearly seen, the expressed rec-gmh protein's high-frequency resonance pattern (Figure 4B) does not match that of the native protein (Figure 4A) in either chemical shift or relative intensity. There seems to be a doubling of nearly all of the hyperfine-shifted resonances, suggesting that two forms of the expressed protein are simultaneously present in solution. This can be explained by earlier proposals that the heme prosthetic group can adopt two orientions within the heme pocket, described by a 180-deg rotation of the heme about the α–γ meso axis (Constantinidis et al., 1988; Santucci et al., 1988; Mintorovitch et al., 1990). The major heme orientation constitutes about 85% of the native protein population, whereas the reversed minor form accounts for the remaining 15%. With the rec-gmh protein, the spectral comparisons shown in Figure 4B indicate that the minor form

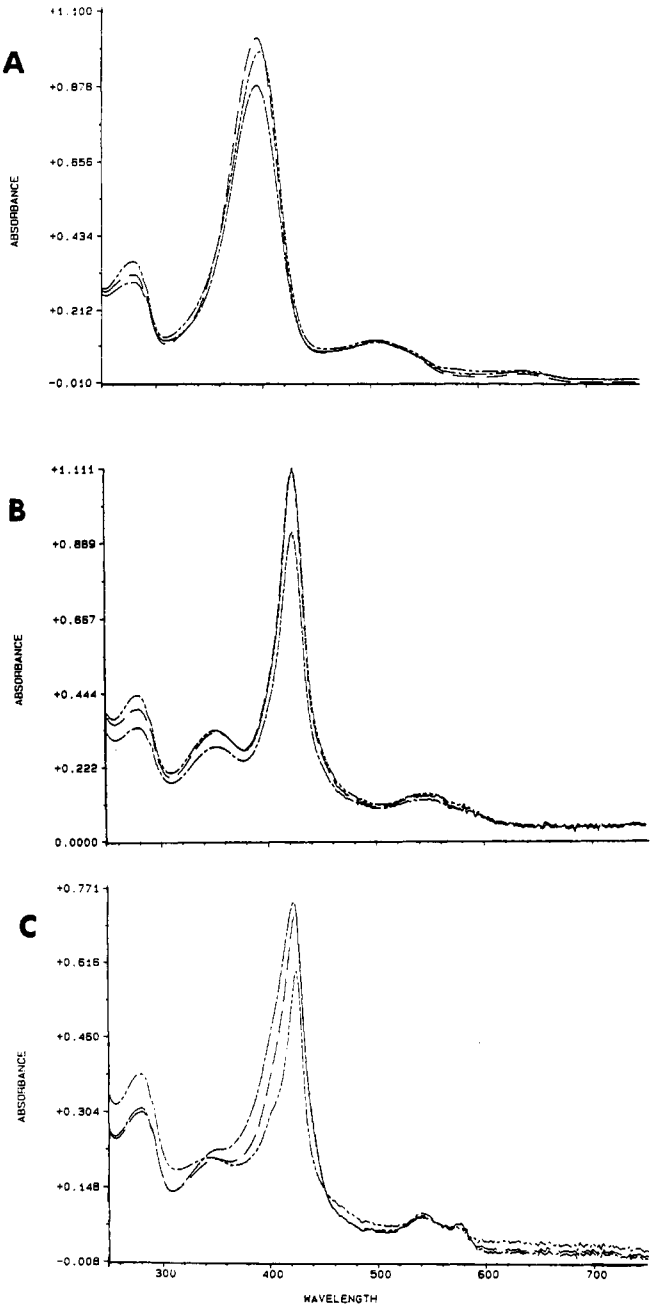


FIGURE 3: UV–visible spectra of three ligation forms for the three holoproteins: native GMH4, – – –; rec-gmg, – · · –; recGMH4, —. (A) Metaquo form of all three proteins (Fe^{3+}–H_2O); (B) ferric cyanide-ligated form of all three proteins; (C) ferrous–O_2 form of all three proteins.

has become more populated (~60%). As seen in Figure 4C, the recombinant wild-type (recGMH4) protein's high-frequency proton hyperfine-shift region is identical to that of the native protein in both chemical shifts and relative resonance intensities. This suggests that one of the three amino acid differences (or a combination of the three) between rec-gmh and both the native and recGMH4 proteins influences the populations of the two heme orientations at equilibrium within the heme pocket.

Using the newly constructed model of GMH4 (Alam et al., 1994), it becomes apparent that only one of the three amino acid side chains lies in a position to directly influence the heme environment (i.e., position 41). The methionine present at position 41 in the native protein has been replaced by isoleucine in rec-gmg. By modeling both heme orientations

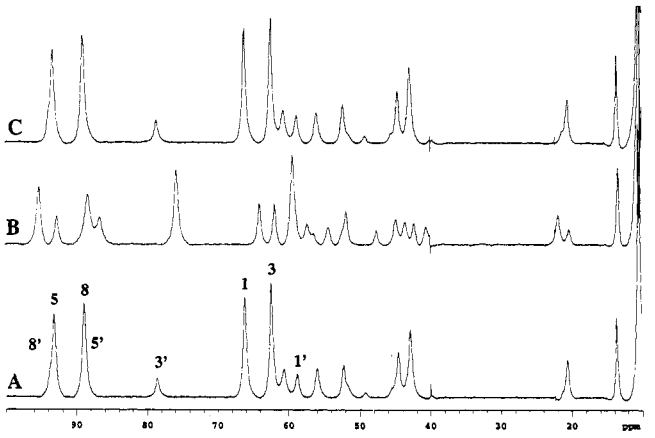


FIGURE 4: Downfield region of the 1D ^1H NMR spectra for the high-spin metaquo forms of all three proteins in 99.9% ^2H_2O, 100 mM potassium phosphate, and 100 mM sodium chloride (pH′ 6.80) at 20 °C. (A) Native GMH4 with previously assigned heme methyl groups of the major (1, 3, 5, 8) and minor (1′, 3′, 5′, 8′) heme orientations labeled above the spectrum. The same expanded spectral region for the ferric high-spin forms of the recombinant hemoglobins (B) rec-gmh and (C) recGMH4.

within the active site of the native model structure, a severe steric contact is detected when the minor form of the heme isomer is inserted. Figure 5A shows an edited stereoview of the major heme orientation within the active site (heme and position 41) of GMH4. The side chain of Met41 and the heme 1-CH_3 group are displayed with van der Waals surfaces. For the major heme orientation there is little contact between the heme and Met41. Figure 5B shows a stereoview of the native GMH4 model structure, with the minor heme orientation (i.e., heme reversed) placed within the active site of GMH4 and with van der Waals surfaces drawn for both Met41 and the heme 4-vinyl group. An obvious steric contact between Met41 and the heme 4-vinyl is created when the heme in this orientation is placed within the active site. It is believed that this van der Waals contact lowers the stability of the minor heme orientation during heme insertion. Figure 6A,B shows the active site of the single-site-modeled mutation M41I with both the major and minor heme orientations, respectively (with respect to native GMH4). This model may represent the active site of rec-gmh in which Ile occupies position 41. With the shorter side chain of isoleucine at position 41, the steric interaction observed with the minor heme orientation in native GMH4 is reduced in rec-gmh (Figure 6A,B) and may allow a higher percentage of the minor heme form. This is evident in the ^1H NMR spectrum of rec-gmh (Figure 4B). Thus, upon first inspection there seems to be an apparent correlation between side chain length at position 41 and relative populations of heme isomers within this monomeric hemoglobin.

Earlier studies on heme orientation within several different proteins illustrate the complexity of forces that combine to regulate heme orientation. These include steric interactions with the buried 2- and 4-vinyl groups from nearby amino acids (Ishimori & Morishima, 1988; La Mar et al., 1981, 1985; Haukson et al., 1990; Lee et al., 1990), H-bonding to either of the propionic acid side chains (Santucci et al., 1993; Haukson et al., 1990), and dependence on the oxidation state and ligation state of the heme iron (Ishimori & Morishima, 1988). All of these factors may play roles at various stages along the complicated, multistep process of heme insertion and reorientation. One could conclude from that work that the actual heme orientation displayed by a particular protein is due to a balance of many heme–globin interactions. Our experimental observation is simply interpreted as identifying primary sequence position 41 as a hitherto unrecognized heme contact which, under appropriate conditions, may help regulate heme orientation. The apparent correlation of heme orientation with the size of the position 41 amino acid side chain is an observation that suggests a hypothesis that is now being tested. For this monomeric hemoglobin (GMH4), when position 41 contains long amino acid side chains (i.e., Met), its steric interaction with the heme contributes to steering the heme into the reversed orientation (relative to Mb). However, when position 41 contains a smaller amino acid side chain (i.e., Ile), that additional steric interaction is relieved, allowing the heme to adopt an orientation similar to that in myoglobin. Of course, we recognize that the situation is probably more complicated and perhaps involves different folding states as the apoprotein is assembled with heme to form the holoprotein. Probing the reconstitution process (particularly heme reori-

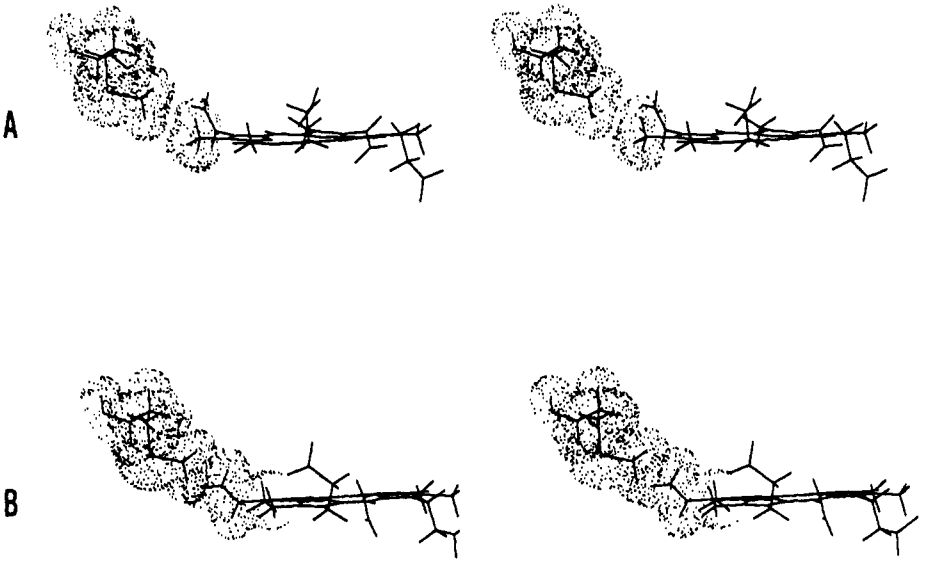


FIGURE 5: Stereoviews of an edited version of the heme active site within the model GMH4 structure showing only the methionine that occupies position 41 and the heme group. (A) van der Waals surfaces are shown around the Met41 side chain and the heme 1-CH_3 as seen with the major heme orientation of GMH4. (B) This stereoview shows the minor heme orientation modeled into the GMH4 active site with van der Waals surfaces on both the Met41 side chain and the heme 4-vinyl group. Note the direct steric contact between Met41 and the heme 4-vinyl group.

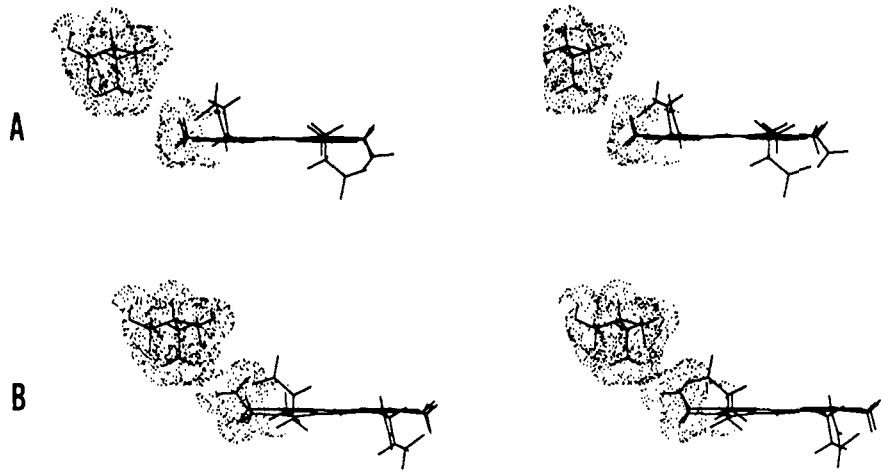


FIGURE 6: Model structure stereoviews for an edited version of the heme active site around position 41 for the single mutant M41I GMH4. This is thought to represent the heme area for the rec-gmh hemoglobin, in which Ile occupies position 41. (A) The heme orientation, corresponding to the major form of native GMH4, modeled into the M41I structure showing van der Waals surfaces on the Ile41 side chain and the heme 1-CH_3. (B) The heme site for the M41I GMH4 model structure constructed with the heme orientation corresponding to the minor form of native GMH4 with Ile41 and the heme 4-vinyl group represented as van der Waals surfaces.

entation) in this monomeric hemoglobin is currently the focus of an extended study involving five additional GMG4 recombinant proteins.

CONCLUSIONS

We have described conditions for the high-level expression of the recombinant forms of GMG4. The cDNA sequence originally described (now called rec-gmg) has been previously shown to differ from the native protein at three positions (Alam et al., 1994). It was the physical characterization of the rec-gmg protein described in this work that initiated the complete chemical sequencing of the native GMH4 protein. Both the UV–visible spectra (especially the ferric form) and the proton hyperfine NMR resonance pattern of the met-rec-gmh gave evidence that the expressed recombinant protein differed at the heme active site compared to the native hemoglobin. Proton NMR spectra of the native hemoglobin (met-GMH4) previously have revealed a minor protein form that is present to the extent of 10–15% and that this heterogeneity is due to heme orientational reversal, or rotational isomerism within the heme pocket. For met-recGMH4, a proton NMR spectrum virtually identical to the native protein NMR spectrum is observed (Figure 4C), indicating that it also displays only 10–15% of the heme reversed form. However, the proton NMR spectrum of the met-rec-gmh displays two nearly equally populated protein forms, which we conclude are due to a nearly equal occurrence of heme rotational isomers. This is unusual because, although this type of heme rotational isomerism has been observed in many hemoglobins and myoglobins, at equilibrium one isomer predominates in most cases. Only a few examples are available of hemoglobins and myoglobins that exhibit nearly equal populations of both heme orientations at equilibrium (La Mar et al., 1980; Levy et al, 1985). This extent of heterogeneity is absent in the expressed revertant recombinant product (recGMH4). Of the three site-specific changes required to make recGMG4 from rec-gmg, position 41 is the only site in close contact with the heme within the holoprotein (Alam et al., 1994). When Ile occupies this position (as in rec-gmh), the shorter side chain removes a steric contact with the heme 4-vinyl (of the minor heme isomer) that is observed when methionine resides there (as in both native GMH4 and recGMH4). We believe that there is a direct correlation between the side chain length at position 41 and the amount of rotational disorder in this monomeric hemoglobin.

It is interesting to note that position 41 in sperm whale myoglobin (Phillips, 1981) is occupied by the even shorter side chain of threonine. Since sperm whale myoglobin has the major heme orientation corresponding to the minor form heme orientation of GMH4 (Lecomte et al., 1985; Phillips, 1981), this observation leads to the intriguing possibility that side chain size at this position may play a role in allowing this specific heme orientation to be stably incorporated into the myoglobin active site.

ACKNOWLEDGMENT

We gratefully acknowledge the use of the Protein Data Bank at Brookhaven National Laboratory.

REFERENCES

Ahmad, M. B., & Kincaid, J. R. (1983) *Biochem. J. 215*, 117–122.
Alam, S. L., Satterlee, J. D., & Edmonds, C. G. (1994) *J. Prot. Chem. 13*, 151–164.
Aojula, H. S., Wilson, M. T., & Drake, A. (1986) *Biochem. J. 237*, 613–616.
Arents, G., & Love, W. E. (1989) *J. Mol. Biol. 210*, 149–161.
Ascoli, F., Rossi Fanelli, M. R., & Antonini, E. (1981) *Methods Enzymol. 76*, 72–88.
Constantinidis, I., & Satterlee, J. D. (1987) *Biochemistry 26*, 7779–7786.
Constantinidis, I., Satterlee, J. D., Pandey, R. K., Leung, H.-K., & Smith, K. M. (1988) *Biochemistry 27*, 3069–3076.
Cooke, R. M., & Wright, P. E. (1985a) *Biochim. Biophys. Acta 832*, 357–364.
Cooke, R. M., & Wright, P. E. (1985b) *Biochim. Biophys. Acta 832*, 365–372.
Cooke, R. M., & Wright, P. E. (1985c) *FEBS Lett. 187*, 219–223.
Cooke, R. M., & Wright, P. E. (1987) *Eur. J. Biochem. 166*, 409–414.
Cooke, R. M., Dalvit, C., Narula, S. S., & Wright, P. E. (1987) *Eur. J. Biochem. 166*, 399–408.
Deeb, R. S., & Peyton, D. H. (1992) *Biochemistry 31*, 468–474.
DiIorio, E. E. (1981) *Methods Enzymol. 76*, 57–72.
Eaton, W. A., & Hofrichter, J. (1981) *Methods Enzymol. 76*, 175–262.
Hauksson, J. B., La Mar, G. N., Pande, U., Pandey, R. K., Parish, D. W., Singh, J. P., & Smith, K. M. (1990) *Biochim. Biophys.*

Acta 1041, 186–194.
Imamura, T., Baldwin, T. O., & Riggs, A. (1972) *J. Biol. Chem. 247*, 2785–2797.
Ishimori, K., & Morishima, I. (1988) *Biochemistry 27*, 4747–4753.
Kandler, R. L., & Satterlee, J. D. (1983) *Comp. Biochem. Physiol. 75b*, 499–503.
Kandler, R. L., Constantinidis, I., & Satterlee, J. D. (1984) *Biochem. J. 226*, 131–138.
Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367–382.
Laemmli, U. K. (1970) *Nature 227*, 680–685.
La Mar, G. N., Smith, K. M., Gersonde, K., Sick, H., & Overkamp, M. (1980) *J. Biol. Chem. 255*, 66–70.
La Mar, G. N., Burns, P. D., Jackson, J. T., Smith, K. M., & Langry, K. C. (1981) *J. Biol. Chem. 256*, 6075–6079.
La Mar, G. N., Toi, H., & Krishnamoorthi, R. (1984) *J. Am. Chem. Soc. 106*, 6395–6401.
La Mar, G. N., Yamamoto, Y., Jue, T., Smith, K. M., & Pandey R. K. (1985) *Biochemistry 24*, 3826–3831.
Lecomte, J. T. J., Johnson, R. D., & La Mar, G. N. (1985) *Biochim. Biophys. Acta 829*, 268–274.
Lee, K.-B., La Mar, G. N., Kehres, L. A., Fujinari, E. M., & Smith, K. M. (1990) *Biochemistry 29*, 9623–9631.
Levy, M. J., La Mar, G. N., Jue, T., Smith, K. M., Pandey, R. K., Smith, W. S., Livingston, D. J., & Brown, W. D. (1985) *J. Biol. Chem. 25*, 13694–13698.
Li, S. L., & Riggs, A. F. (1971) *Biochim. Biophys. Acta 236*, 208–210.
Mintorovitch, J., & Satterlee, J. D. (1988) *Biochemistry. 27*, 8045–8050.
Mintorovitch, J., van Pelt, D., & Satterlee, J. D. (1989) *Biochemistry 28*, 6099–6104.
Mintorovitch, J., Satterlee, J. D., Pandey, R. K., Leung, H.-K., & Smith, K. M. (1990) *Inorg. Chim. Acta 170*, 157–159.
Moore, J. T., Uppal, A., Maley, F., & Maley, G. F. (1993) *Protein Expression Purif. 4*, 160–163.
Parkhurst, L. J., Sima, P., & Goss D. J. (1980) *Biochemistry 19*, 2688–2692.
Phillips, S. E. V. (1981) Brookhaven Protein Data Bank.
Rosenburg, A. H., Lade, B. N., Chui, D.-S., Lin, S.-W., Dunn, J. J., & Studier, F. W. (1987) *Gene 56*, 125–135.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Santucci, R., Mintorovitch, J., Constantinidis, I., Satterlee, J. D., & Ascoli, F. (1988) *Biochim. Biophys. Acta 953*, 201–204.
Santucci, R., Ascoli, F., La Mar, G. N., Parish, D. W., & Smith, K. M. (1990) *Biophys. Chem. 37*, 251–255.
Seamonds, B., Forster, R. E., & George, P. (1971) *J. Biol. Chem. 246*, 5391–5397.
Simons, P. C., & Satterlee, J. D. (1989) *Biochemistry 28*, 8525–8530.
Studier, F. W., & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–89.
Van Heeke, G., & Schuster, S. M. (1989) *J. Biol. Chem. 10*, 5503–5509.
Vinogradov, S. H., Machlik, C. A., & Chao, L. L. (1970) *J. Biol. Chem. 245*, 6533–6538.
Yee, S., & Peyton, D. H. (1991) *FEBS Lett. 290*, 119–122.