Owellen, R. J., Owens, A. H., & Donigian, D. W. (1972) *Biochem. Biophys. Res. Commun. 47*, 686–691.
Patzett, C., Singh, A., LeMarchand, Y., Orci, L., & Jeanrenaud, B. (1975) *J. Cell. Biol. 66*, 609–620.
Ray, K., Bhattacharyya, B., & Biswas, B. B. (1981) *J. Biol. Chem. 256*, 6241–6244.
Ray, K., Bhattacharyya, B., & Biswas, B. B. (1984) *Eur. J. Biochem. 142*, 577–581.
Raybin, D., & Flavin, M. (1977) *J. Cell Biol. 73*, 492–504.
Ringel, I., & Sternlicht, H. (1984) *Biochemistry 23*, 5644–5653.
Roach, M. C., Bane, S., & Luduena, R. F. (1985) *J. Biol. Chem. 260*, 3015–3023.
Rosner, M., Capraro, H. G., Jacobson, A. E., Atwell, L., Brossi, A., Iorio, M. A., Williams, T. H., Sik, R. H., & Chignell, C. F. (1981) *J. Med. Chem. 24*, 257–261.
Scatchard, G. (1949) *Ann. N.Y. Acad. Sci. 51*, 660–672.
Schmitt, H., & Atlas, D. (1976) *J. Mol. Biol. 102*, 743–758.
Serrano, L., Avila, J., & Maccioni, R. B. (1984) *J. Biol. Chem. 259*, 6607–6611.
Studier, F. W. (1983) *J. Mol. Biol. 79*, 237–248.
Williams, R. F., Aivaliotis, M. J., Robinson, A. K., & Barnes, L. D. (1983) *J. Chromatogr. 266*, 141–150.
Williams, R. F., Mumford, C. L., Williams, G. A., Floyd, L. J., Aivaliotis, M. J., Martinez, R. A., Robinson, A. K., & Barnes, L. D. (1985) *J. Biol. Chem. 260*, 13794–13802.
Wilson, L., Geswell, K. M., & Chin, D. (1975) *Biochemistry 14*, 5586–5592.
Yamaoka, K., Tanigawara, U., Nakagawa, T., & Uno, T. (1981) *J. Pharmacobiodyn. 4*, 879–885.
Yeh, J. C., Chrzanowska, M., & Brossi, A. (1988) *FEBS Lett. 229*, 82–86.
Zweig, N. H., & Chignell, C. F. (1973) *Biochem. Pharmacol. 22*, 2141–2150.

cDNA Cloning and Predicted Amino Acid Sequence of *Glycera dibranchiata* Monomer Hemoglobin IV[†,‡]

Peter C. Simons and James D. Satterlee*

Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131, and Department of Chemistry, Washington State University, Pullman, Washington 99164

Received April 14, 1989; Revised Manuscript Received June 16, 1989

ABSTRACT: The three major monomer hemoglobins from *Glycera dibranchiata* erythrocytes isolated in this laboratory were sequenced from their N-termini. A stretch of amino acid sequence identity was used to determine the sequence of a mixed oligodeoxynucleotide that would be complementary to all 12 possible mRNA sequences coding for the amino acids. A cDNA library was constructed by using poly(A+) RNA from *G. dibranchiata* erythrocytes, the library was probed with the oligonucleotide, and the longest positive inserts found were subcloned into a sequencing plasmid and then sequenced. The first one was 745 bases long, containing 85 bases of 5′-untranslated RNA, an open reading frame of 444 bases coding for 148 amino acids, and a 3′-untranslated region of 216 bases. The predicted amino acid sequence matches the first 25 amino acids of *G. dibranchiata* monomer globin component IV. The sequence contains an N-terminal methionine plus 18 other mostly conservative sequence changes compared to the published sequence of Imamura et al. (1972), which appears from our partial sequencing to be monomer globin component II. We confirm the presence of leucine in the E7 position, which is histidine in most myoglobins and hemoglobins.

The marine bloodworm *Glycera dibranchiata* has nucleated erythrocytes that contain hemoglobins separable on Sephadex chromatography into a polymeric fraction and a monomeric fraction (Vinogradov et al., 1970). Crystals obtained from this monomeric fraction led to partial (Li & Riggs, 1971) and complete amino acid sequences (Imamura et al., 1972) and a 2.8-Å crystal structure (Padlan & Love, 1974) that showed an unusual substitution at position E7 in the ligand binding pocket. The almost universally found histidine was replaced by leucine. Subsequently, we used CM-cellulose chromatography to reproducibly separate the monomer hemoglobin fraction into two minor components (Ia and Ib) and three major components (II–IV), all with molecular weights of approximately 16 kDa (Kandler & Satterlee, 1983; Kandler et al., 1984; Constantinidis & Satterlee, 1987; Constantinidis et al., 1989). Isoelectric focusing confirms the high purity of our preparations of these proteins (Constantinidis & Satterlee, 1987; Constantinidis et al., 1989).

Recent progress in characterizing the three major monomeric hemoglobins has focused on ligand binding kinetics since that is the function of these proteins. With available instrumentation, we have found that each of the monomer methemoglobins binds cyanide at a rate several orders of magnitude slower than those of metmyoglobin and other monomeric ferriheme proteins (Mintorovitch & Satterlee, 1985, 1987, 1988; Mintorovitch et al., 1989). Such slow rates have led to significantly revised estimates of the equilibrium constant for cyanide binding. These anomalous dynamics, in conjunction with limited structural and spectroscopic data (Satterlee, 1984; Cooke & Wright, 1985, Cooke et al., 1987), imply that all three major monomer hemoglobin components

[†] This work was supported by National Institutes of Health Grants AM30912 and HL01758 (Research Career Development Award) to J.D.S.

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02873.

* Address correspondence to this author at the Department of Chemistry, Washington State University, Pullman, WA 99164-4630.

0006-2960/89/0428-8525$01.50/0

lack histidine E7 and that the E7 amino acid plays a pivotal role in ligand binding dynamics.

A point that is frequently raised is whether or not we actually isolated three different major hemoglobins. Their functional similarity, indistinguishability by electrophoresis, and very similar behavior upon isoelectric focusing contribute to the idea that we could possibly be isolating chemically modified variants of the same precursor protein. However, the sequence results presented here, in conjunction with published sequence information (Li & Riggs, 1971; Imamura et al., 1972), provide strong evidence for the view that the *G. dibranchiata* monomer hemoglobin fraction contains three major individual, but related, proteins.

In order to determine the basis of the heterogeneity of the combined monomer fraction and to be able to use site-directed mutagenesis to explore the unusual ligand binding properties of *G. dibranchiata* monomer hemoglobins, we have constructed a cDNA library from these erythrocytes. Our success proves that, at this phylogenetic level, the blood cells are transcriptionally active. We then performed N-terminal amino acid sequencing on the three major monomer globins in order to compare their limited sequences and to find regions of identity among the three that had a minimal number of possible mRNA codings. A mixed oligodeoxynucleotide was then synthesized to be complementary to the mRNAs, and clones that hybridized to the probe were isolated from the cDNA library. We report here the isolation and sequencing of the first clone, which corresponds to *G. dibranchiata* monomer hemoglobin IV.

Materials and Methods

Protocols not described here can be found in Maniatis et al. (1982), which remains the indispensable background for cloning methods.

Protein Sequencing. The monomeric hemoglobins were isolated as previously described (Kandler et al., 1984; Constantinidis & Satterlee, 1987). The globins were separated from their heme groups by fully oxidizing the iron with a dilute solution of potassium ferricyanide, reducing the pH to 4.6, and extracting the heme with an equal volume of cold 2-butanone (Teale, 1959). The globins were sequenced on an Applied Biosystems protein sequenator at the UNM Protein Laboratory using 300–500 pmol per run.

RNA Preparation. Fifty bloodworms (Maine Bait Co., Portland, ME) were minced to 1-in. lengths over a Büchner funnel, allowing the blood to flow down into 50 mL of ice-cold 0.5 M KCl, 0.13 mM EDTA,[1] 50 mM potassium phosphate, pH 7.2. The cells were pelleted by spinning at 500*g* for 10 min, rinsed in the above buffer, and then resuspended in the lysis buffer for isolated cells described by Maniatis et al. (1982), using vanadyl–ribonucleoside complexes (BRL, Bethesda Research Laboratories, Gaithersburg, MD) to inhibit RNases. The preparation of RNA was continued as described by two passages through oligo(dT)–cellulose (Pharmacia, Piscataway, NJ), except that elution of poly(A+) RNA used diethyl pyrocarbonate treated water with no salt.

In Vitro Translation of RNA. The mRNA [poly(A+) RNA] obtained was translated into protein by using the wheat germ system and protocol supplied by BRL, and [^{35}S]-methionine (Du Pont/NEN). The incorporation of label was the same as that obtained by using rabbit globin mRNA supplied with the kit.

Oligonucleotide Probe. N-Terminal amino acid sequencing of globins II–IV each showed a region of sequence homology that had a minimal number of possible mRNA codings (Figure 1H). A 12-fold-degenerate 14-mer cDNA probe was synthesized (OCS Labs, Denton, TX). The crude probe was deblocked with 200 μL of concentrated ammonium hydroxide at 60 °C overnight, acetic acid was added to attain pH 7–8, and the DNA was precipitated with 2 volumes of ethanol. The precipitate was resuspended in TE, and a sample was purified by gel electrophoresis in 8 M urea/20% acrylamide/1% bisacrylamide/TBE, with a gel size of 0.15 × 14 × 16 cm, for 3.5 h at 400 V, using 2 A_{260} units/cm of well. The bands were visualized by ultraviolet shadowing on a fluorescent thin-layer plate covered with plastic wrap. The darkest band [see Maniatis et al. (1982) for approximate migration rates] was excised and cut into 2 mm wide lengths, and the two absorbance units were eluted into 250 μL of water in a microfuge tube by shaking overnight. The DNA was separated from low molecular weight contaminants by chromatography using water as eluent through a 6-ml Sephadex column. Fractions of 0.4 mL were analyzed at 260 nm; the highest fraction was frozen for future use.

cDNA Library Preparation and Screening. Double-stranded cDNA was prepared with 10 μg of poly(A+) RNA as described in the BRL cDNA Cloning Manual. This was dC-tailed according to standard methods (Deng & Wu, 1981). The dC-tailed insert was mixed with an optimized amount of *Pst*I-cut, dG-tailed pBR322 (BRL), in our case, 2.5 μg, and annealed in 5 mL of 0.1 M Tris, pH 7.5, 1 mM EDTA, and 0.1 M NaCl from 90 to 35 °C over 30 min. This solution was used to transform *Escherichia coli* DH5 which had been treated for high-efficiency uptake of DNA (Hanahan, 1985). The transformed cells were spun down, resuspended in 10 mL of growth medium, plated on L + tetracycline plates, and grown overnight. The colonies were rinsed off the plates with 3 mL of L broth per plate, diluted 10-fold in 20% glycerol in L broth, frozen in aliquots in liquid nitrogen, and stored at −70 °C. A total of 1.2 million independent transformants were obtained. A 50-μL aliquot scraped from the top of a frozen mass of cells was diluted 100-fold in L broth for screening as described (Berent et al., 1985). The 12-fold-degenerate oligonucleotide probe was labeled as described (Maniatis et al., 1982).

Plasmid Purification, Subcloning, and Sequencing. Plasmid minipreparations utilized the lysis by boiling method (Maniatis et al., 1982) through precipitation with 0.6 volume of 2-propanol and redissolution in 50 μL of TE. High molecular weight RNA and protein were eliminated with an ammonium acetate precipitation followed by an ethanol precipitation of the DNA (Crouse & Amorese, 1987), and the final precipitate was redissolved in 25 μL of TE. Positive colonies from the screening were grown, plasmid minipreparations were made, the DNA from 5 μL was digested with *Pst*I, and the digests were run on 1% agarose TAE gels (Maniatis et al., 1982). The inserts from the digests having the four largest inserts were recovered from the gel with Geneclean, following the protocol of the supplier (Bio 101, La Jolla, CA). These were ligated to *Pst*I-cut, bacterial alkaline phosphatase treated pIBI76 (International Biotechnologies, Inc., New Haven, CT) according to standard methods (Maniatis et al., 1982). The ligation mixtures were used to transform *E. coli* NM522 which had been treated for high-efficiency uptake of DNA (Hanahan, 1985). Single-stranded template for sequencing was

[1] Abbreviations: Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; TE, 10 mM Tris-HCl, pH 8, 1 mM EDTA; TBE, 89 mM Tris–borate, pH 8, 1 mM EDTA; TAE, 40 mM Tris–acetate, pH 8, 2 mM EDTA; SDS, sodium dodecyl sulfate.

```
Sequence Number:        1        10        20        30        40
                        ▼        ▼         ▼         ▼         ▼
A. Component II:        GLSAAQRQVI AATWKDIAGN DNGAGV__D_
B. Component III:       GLSAAQRQVV ASTWKDIAGA DNGAGVGKEC LIKFIHA_P_ M
C. Component IV:        GLSAAQRQVV ASTWKDIAGS DNGAGV
D. Nucleotide-1:      M GLSAAQRQVV ASTWKDIAGS DNGAGVGKEC FTKFLSAHHD M
E. Nucleotide-2:      M GLSAAQRQVV ASTWKDIAGS DNGAGVGKEC FTKFLSAHHD M
F. Imamura et al.:      GLSAAQRQVI AATWKDIAGN DNGAGVGKDC LIKHLSAHPQ M
G. Li & Riggs:          GLSAAQRQVV AS_WKDIAGA D_GAGVGKE_ LIKFISAAPE M

H. Probe Sequence:                    WKDIA

I. Oligonucleotide Probe:  5'-GC {A/G/T}AT {A/G}TC {T/C}TT CCA-3'
```

FIGURE 1: Amino-terminal sequence comparison of *G. dibranchiata* monomer hemoglobin components II–IV isolated in this laboratory (A–C); two protein sequences inferred from cDNA sequences (D and E); a partial protein sequence from a crystalline sample (G) (Li & Riggs, 1971); and a complete protein sequence from a crystalline sample (F) (Imamura et al., 1972). (H) Probe sequence; (I) oligonucleotide probe used.

obtained by superinfection of plasmid-containing bacteria with M13K07 phage as described (Vieira & Messing, 1987), followed by their purification. Sequencing by Sanger's dideoxy chain-termination method was done by using the kit and protocol of U.S. Biochemicals (Cleveland, OH) and was confirmed by sequencing the opposite strand, which was generated by inserting the double-stranded cDNA into the vector in the opposite orientation. The reactions were labeled by using [α-^{35}S]dATP (500 mCi/mmol, Du Pont/NEN) and separated on 6% polyacrylamide/8 M urea gels.

Results and Discussion

In Vitro Translation of mRNA. The mRNA [poly(A+) RNA] obtained from *G. dibranchiata* erythrocytes was used as a template for in vitro translation, using a wheat germ in vitro translation kit. Incorporation of [^{35}S]methionine was as high as incorporation using the rabbit globin mRNA supplied with the kit. Separation of reaction products on a 15% polyacrylamide/SDS gel showed the major band of isotope incorporation (about 50%) in the *G. dibranchiata* translation comigrating with authentic *G. dibranchiata* globins II–IV, and with the major band of incorporation of rabbit globin message (data not shown).

N-Terminal Amino Acid Sequencing. A substantial body of chemical (Constantinidis et al., 1989; Constantinidis & Satterlee, 1987), spectroscopic (Kandler & Satterlee, 1983; Kandler et al., 1984; Constantinidis et al., 1988; Cooke & Wright, 1985a,b; Cooke et al., 1987), and kinetic (Mintorovitch & Satterlee, 1985, 1987, 1988) knowledge is accumulating for our consistent preparations of these three major monomeric hemoglobin components from *G. dibranchiata.* Collectively, these results strongly support the idea that each component is a true individual protein and that each is not simply a chemically modified derivative or degradation product. In order to confirm this, *G. dibranchiata* monomer globin components II–IV isolated in this laboratory were partially sequenced from their N-termini, and the results are shown in Figure 1A–C.

The conclusions that can be drawn from these partial sequences are as follows. First, the globins differ in amino acid composition and are not merely different states of degradation of a common precursor. This can be seen by comparing components II and III at sequence positions 10, 12, 20, and 29. Components II and IV also differ at these same positions. Components III and IV differ at position 20.

Second, prior to this work, partial (Li & Riggs, 1971) and complete (Imamura et al., 1972) sequences of *G. dibranchiata* monomer globins were published. Insofar as these two sequences can be compared (Figure 1F,G), they differ significantly despite having the common origin of both being crystalline samples isolated in the same laboratory. The partial sequence of Li and Riggs (1971) (Figure 1G) corresponds to the initial 42 residues of our component III hemoglobin. The first 25 amino acids (plus residue 29) of the complete sequence of Imamura et al. (1972) are identical with our component II hemoglobin.

Third, a sequence of identity was found that had only a small number of possible mRNAs (Figure 1H,I). A mixed oligonucleotide complementary to these mRNAs was synthesized to serve as a probe (Figure 1I, vide infra). The results of cDNA cloning and sequencing (next section) gave two identical inferred amino acid sequences whose initial 42 residues are also shown in Figure 1D,E. This inferred sequence is identical in the first 26 amino acids with that of component IV monomer globin (Figure 1C).

These partial sequences allow unambiguous assignment of the two published protein sequences to hemoglobin components III (Li & Riggs, 1971) and II (Imamura et al., 1972). This strengthens the conclusion (Constantinidis & Satterlee, 1987; Kandler & Satterlee, 1983; Kandler et al., 1984, 1989; Cooke & Wright, 1985a) that three similar, but not identical, monomer hemoglobin components are major constituents of the *G. dibranchiata* monomer hemoglobin fraction. It is also interesting to note that the two crystalline samples from Warner Love's laboratory that were sequenced (Li & Riggs, 1971; Imamura et al., 1972) were different proteins. Li and Riggs (1971) noted some contamination of their sample by other protein components, and it is likely that the contamination was from one of the other protein components in the monomer fraction. This apparent cocrystallization of multiple monomer components may be a manifestation of the extremely similar three-dimensional structures of each of the three monomer components. Certainly, the very conservative primary sequence substitutions that are obvious when the complete sequences of components II and IV are compared (next section) make a good argument for very similar three-dimensional structures. The published crystal structure (Padlan & Love, 1974) thus corresponds to monomer component II.

cDNA Cloning and Sequencing. A cDNA library was generated at the *Pst*I cut site in the plasmid pBR322 as described under Materials and Methods, giving 1.2 million independent transformants from 10 μg of mRNA. An aliquot from the frozen mass of cells was diluted to give about 500 colonies on a nitrocellulose filter. Probing a replica filter with

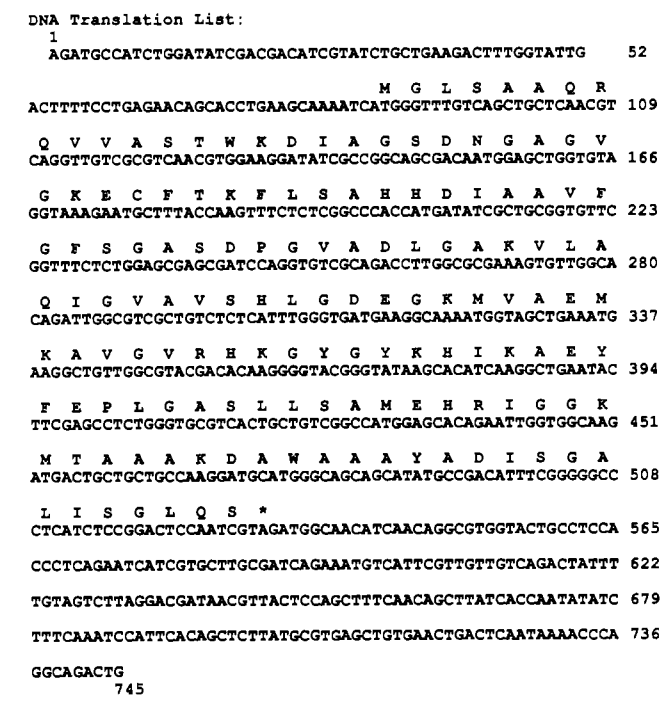

FIGURE 2: Nucleotide sequence (lower) and inferred protein sequence (upper) from the first sequenced insert, corresponding to globin component IV.

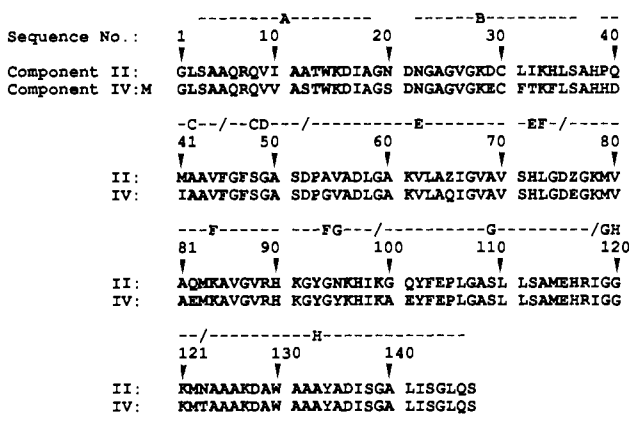

FIGURE 3: Primary sequence comparison. Component IV: Monomer globin component IV complete sequence inferred from the nucleotide sequence determined in this work. Component II: Globin sequence from Imamura et al. (1972). Shown above the sequence numbers are the helical segments derived by Imamura et al. (1972) from alignment with myoglobin.

the mixed oligonucleotide gave a dozen strong signals. Plasmid minipreparations from these dozen showed various length inserts on agarose gel electrophoresis. The four largest single inserts were recovered from TAE gels with Geneclean and ligated into the sequencing vector, pIBI76. The first two inserts to be sequenced gave identical cDNA sequences and the derived amino acid sequence shown in Figure 2. This corresponds to *G. dibranchiata* monomer globin IV on the basis of the predicted sequence's identity with the first amino acids in the N-terminal amino acid sequence of authentic monomer globin IV (Figure 1C–E). As noted above, the inferred protein sequence is similar to (but not identical with) that given by Imamura et al. (1972) (Figure 3), having 18 mostly conservative substitutions out of 147 amino acid residues (12% difference) and an N-terminal methionine as shown in the Figure 3 comparison.

Sequence Comparisons. The close sequence homology of components II and IV (Figure 3), identical electrophoresis behavior (Constantinidis & Satterlee, 1987; Kandler et al., 1989), and NMR studies (Cooke & Wright, 1985a,b; Cooke et al., 1987; Kandler et al., 1984; Satterlee, 1984; Constantinidis et al., 1987) suggest highly similar three-dimensional structures for components II–IV. We are, therefore, in a position to compare heme–protein contacts and other specific sequence questions by referring to the published crystal structure of component II (Padlan & Love, 1974). Table I compares the heme–globin contacts identified from the crystal structure (Padlan & Love, 1974) of component II. According to Padlan and Love's criteria, identical amino acids occupy these same sequence positions for component IV, and from this we may conclude that the heme–globin contacts in both monomer hemoglobins are identical.

Table I: Amino Acids Implicated in Heme–Globin Contacts in the *G. dibranchiata* Monomer Hemoglobin Components II (Imamura et al., 1972) and IV (This Work) in Comparison to Corresponding Positions in Sperm Whale Myoglobin

Glycera	amino acid		sperm whale myoglobin	
position	CII	CIV	position	amino acid
44 C6	Val	Val	C7	Lys
45 CD1	Phe	Phe	CD1	Phe
58 E7	Leu	Leu	E7	His
62 E11	Val	Val	E11	Val
65 E14	Glx	Gln	E14	Ala
66 E15	Ile	Ile	E15	Leu
83 F9	Met	Met	F1	Leu
86 F12	Val	Val	F4	Leu
90 F16	His	His	F8	His
93 FG3	Tyr	Tyr	FG2	His
98 FG8	Ile	Ile	FG5	Ile
102 G4	Tyr	Tyr	G4	Tyr
103 G5	Phe	Phe	G5	Leu
106 G8	Leu	Leu	G8	Ile
134 H12	Tyr	Tyr	H11	Leu
137 H15	Ile	Ile	H14	Phe
141 H19	Leu	Leu	H18	Ile

Furthermore, residue E7 is confirmed to be Leu in component IV, just as in component II. Comparing the E helix sequence (Figure 3) of the two monomer hemoglobins shows that only position 54 (Ala/Gly) and possibly position 65 (Glx/Gln) differ, and these are conservative changes. Consequently, the distal heme environment seems essentially identical for the two proteins. This is consistent with the concept of a uniform "bound ligand" environment developed from NMR and infrared spectroscopic studies of heme-bound $^{13}CO/^{12}CO$ in the monomer CO–hemoglobin fraction (Satterlee, 1984). That work revealed no resolvable differences in either chemical shift or infrared stretching vibration for heme-bound CO among the monomer components.

The F helix displays similar high homology for components II and IV, with possible conservative changes at positions 76 (Glx/Gln) and 82 (Gln/Glu). The other regions of the heme environment, CD and FG, are identical in components II and IV. The concept that emerges from the analysis is that components II and IV have essentially identical heme environments.

Table I also presents the amino acids responsible for heme–globin contacts in sperm whale myoglobin. The position designations are sometimes different than those for the *G. dibranchiata* monomer hemoglobins because the overall primary sequence homology in the proteins from the two species is low (Padlan & Love, 1974). Despite this, their overall gross structural features are remakably similar (Padlan & Love, 1974). However, their kinetic behavior is radically different. We have recently completed cyanide ligand binding studies for the met forms of the three monomer hemoglobin compo-

Table II: Comparison of Experimental Amino Acid Analyses for Monomer Hemoglobins A–C Isolated by Cooke and Wright (1985) with Those Calculated from the Sequences of Component II (Imamura et al., 1972) and Component IV

	CIV	CII	HbA	HbB	HbC
Ala (A)	27	28	27.00	27.12	28.13
Arg (R)	3	3	2.92	2.43	2.56
Asn (N)	1	4	*a*	*a*	*a*
Asp (D)	8	8	*a*	*a*	*a*
Cys (C)	1	1			
Gln (Q)	4	6	*b*	*b*	*b*
Glu (E)	6	2	*b*	*b*	*b*
Gly (G)	20	20	20.30	20.85	20.29
His (H)	6	6	5.99	4.93	6.00
Ile (I)	7	8	6.58	6.41	6.84
Leu (L)	10	11	10.26	9.89	10.99
Lys (K)	11	11	10.74	11.21	10.94
Met (M)	4	5	3.66	4.56	4.61
Phe (F)	5	3	5.36	4.22	3.07
Pro (P)	2	3	1.96	3.12	2.81
Ser (S)	12	10	10.27	11.71	10.14
Thr (T)	3	1	1.67	1.09	1.17
Trp (W)	2	2			
Tyr (Y)	4	3	2.69	3.18	2.85
Val (V)	11	10	10.76	8.98	8.79
Glx (Z)		2	*b*	*b*	*b*
Asx (N + D)	9	12	12.28	10.33	12.93
Glx (Q + E + Z)	10	10	11.55	13.97	11.89

[a] See values below for Asn (N + D). [b] See values below for Glx (Q + E + Z).

nents isolated in this lab (Mintorovitch & Satterlee, 1988; Mintorovitch et al., 1989). Component IV methemoglobin displays an apparent bimolecular rate constant, k_1^{app} (M^{-1} s^{-1}), that is approximately 220 times lower than that reported for metmyoglobin (Mintorovitch et al., 1989). The E7 substitution of His(Mb) for Leu (component IV, II) undoubtedly plays an important role in mediating this kinetic reactivity due to its proximity to the heme ligand binding site. However, there are kinetic differences in k_1^{app} for cyanide binding even among the three *G. dibranchiata* monomer methemoglobins. All are very much slower than k_1^{app} for metmyoglobin (approximately 400 M^{-1} s^{-1}; Awad & Badro, 1967), but component IV exhibits the largest k_1^{app} (1.82 M^{-1} s^{-1}) followed by component II (0.49 M^{-1} s^{-1}) and component III (0.30 M^{-1} s^{-1}). Component IV has a rate constant more than 3 times larger than that of component II, despite having essentially identical heme contacts and primary sequences in the CD, FG, E helix, and F helix regions. Compared to the large difference with metmyoglobin, this difference is small, but it serves to emphasize that more subtle variations in three-dimensional structure or heme pocket polarity may also play roles in mediating precise functional characteristics.

The data in Table I reveal that the heme environment in components II and IV display opposite polarity to that in myoglobin, at least as judged from the heme contacts. For example (neglecting the proximal His F16), component IV possesses one neutral polar residue (Gln E14), three acidic polar amino acids (Tyr FG3, G4, H12), and 12 nonpolar residues that are heme contacts. Similarly (neglecting the proximal His F8), myoglobin possesses three basic amino acids (Lys C7, His E7, His FG2), one acidic amino acid (Tyr G4), and 12 nonpolar amino acids. Heme environment polarity may play a role in mediating kinetic properties, and the precise nature of this effect should be revealed by kinetic studies on *G. dibranchiata* site-directed mutant monomer hemoglobins. These studies are currently in progress.

Other Comparisons. Other isolations and kinetic data have been reported for the *G. dibranchiata* monomer hemoglobins. However, different methods of isolation were used, and high-resolution analytical techniques were not employed for characterization, making a correlation of those reported properties with these sequences pointless until ambiguities in the various preparations are resolved.

In one case, however, Cooke and Wright (1985a) used a procedure very similar to ours (Kandler et al., 1984; Constantinidis & Satterlee, 1987) for isolating three monomer hemoglobins and reported amino acid analyses of each. These data are collected in Table II along with the predicted analyses from the component IV sequence (this work) and from the sequence of Imamura et al. (1972) (which corresponds to our component II). Comparing these, we agree with the assessment by Cooke and Wright (1985a) that their HbC corresponds to the sequence of Imamura et al. (1972), and it may be relabeled as component II. In addition, we have shown elsewhere that their HbC corresponds to our component II on the basis of isoelectric focusing and NMR spectroscopy (Kandler & Satterlee, 1983; Kandler et al., 1984; Constantinidis & Satterlee, 1987; Constantinidis et al., 1988).

It also appears from Table II that the best correlation for component IV is Cooke and Wright's (1985a) HbA, although there are significant differences in the content of Ser, Thr, Tyr, Asx, and Glx reported for HbA compared to our sequence of component IV. These errors are probably due to inaccuracies in the amino acid analysis technique or to impure HbA preparation. A limited comparison of the NMR spectra of HbA (Cooke & Wright, 1985a) and component IV (Kandler et al., 1984; Constantinidis et al., 1988) shows that they have nearly identical heme methyl resonance patterns, further indicating that HbA is component IV.

Using their preparations of the monomer hemoglobins, combined with elegant NMR technique, Cooke et al. (1987) were able to draw several conclusions concerning the amino acid sequences of all three monomer hemoglobins. For component IV, these can now be confirmed from the sequencing. (1) Component IV contains (Figure 3) A12 = W, CD1 = CD3 = F, E7 = L, E11 = V, F8 = H, G5 = F, G8 = L, and H8 = W, as predicted by NMR. (2) In component IV, position B10 is occupied by Phe (predicted by NMR), not Leu as it is in component II. (3) The fifth Phe spin system, which was identified by NMR but could not be assigned a sequence position, is seen to be B13 (residue 34, Figure 3). Thus, in all cases elucidated, the NMR experiments allowed correct, albeit limited, sequencing for component IV.

References

Asher, S. A. (1981) *Methods Enzymol. 76*, 371–413.

Berent, S. L., Mahmoudi, M., Torczynski, R. M., Bragg, P. W., & Bollon, A. P. (1985) *Biotechniques 3*, 208–220.

Constantinidis, I. C., & Satterlee, J. D. (1987) *Biochemistry 26*, 7779–7786.

Constantinidis, I. C., Satterlee, J. D., Pandey, R. K., Leung, H.-K., & Smith, K. M. (1988) *Biochemistry 27*, 3069–3076.

Constantinidis, I. C., Kandler, R., & Satterlee, J. D. (1989) *Comp. Biochem. Physiol. 92B*, 619–622.

Cooke, R. M., & Wright, P. E. (1985a) *Biochim. Biophys. Acta 832*, 357–364.

Cooke, R. M., & Wright, P. E. (1985b) *FEBS Lett. 187*, 219–223.

Cooke, R. M., Dalvit, C., Narula, S. S., & Wright, P. E. (1987) *Eur. J. Biochem. 166*, 399–408.

Crouse, J., & Amorese, D. (1987) *Focus 9*, 3–5.

Deng, G., & Wu, R. (1981) *Nucleic Acids Res. 9*, 4173–4188.

Hanahan, D. (1985) in *DNA Cloning* (Glover, D. M., Ed.) Vol. 1, pp 109–135, IRL Press, Washington, D.C.

Imamura, T., Baldwin, T. O., & Riggs, A. (1972) *J. Biol. Chem. 247*, 2785–2797.
Kandler, R. L., & Satterlee, J. D. (1983) *Comp. Biochem. Physiol., B: Comp. Biochem. B75*, 449–503.
Kandler, R. L., Constantinidis, I. C., & Satterlee, J. D. (1984) *Biochem. J. 225*, 131–138.
Li, S. L., & Riggs, A. F. (1971) *Biochim. Biophys. Acta 236*, 208–210.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, NY.
Mintorovitch, J., & Satterlee, J. D. (1985) *Rev. Port. Quim. 27*, 199–200.
Mintorovitch, J., & Satterlee, J. D. (1987) *Recl.: J. R. Neth. Chem. Soc. 106*, 271.
Mintorovitch, J., & Satterlee, J. D. (1988) *Biochemistry 27*, 8045–8050.
Mintorovitch, J., van Pelt, D., & Satterlee, J. D. (1989) *Biochemistry 28*, 6099–6104.
Padlan, E. A., & Love, W. E. (1974) *J. Biol. Chem. 249*, 4067–4078.
Satterlee, J. D. (1984) *Biochim. Biophys. Acta 791*, 384–394.
Teale, F. W. J. (1959) *Biochim. Biophys. Acta 35*, 543–546.
Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3–11.
Vinogradov, S. H., Machlik, C. A., & Chao, L. L. (1970) *J. Biol. Chem. 245*, 6533–6537.

Affinity Chromatography Using Protein Immobilized via Arginine Residues: Purification of Ubiquitin Carboxyl-Terminal Hydrolases†

Penelope J. Duerksen-Hughes,‡ Michael M. Williamson,§ and Keith D. Wilkinson*
Departments of Biochemistry and Chemistry, Emory University, Atlanta, Georgia 30322
Received February 27, 1989; Revised Manuscript Received June 8, 1989

ABSTRACT: 4-(Oxoacetyl)phenoxyacetic acid (OAPA) forms a stable, covalent bond between its glyoxal group and the guanidino group of arginine and arginine derivatives [Duerksen, P. J., & Wilkinson, K. D. (1987) *Anal. Biochem. 160*, 444–454]. Studies were carried out to determine the chemical nature of this linkage, and the structure of the stable adduct between OAPA and methylguanidine was elucidated. The stable product results from an internal oxidation–reduction of the Schiff base adduct to form a cyclic α-aminoamide, 4-[4-(carboxymethoxy)phenyl]-2-(methylimino)-5-oxoimidazolidine. OAPA coupled to polyacrylamide beads was used to immobilize ubiquitin via its arginine residues, and the resulting affinity support was shown to specifically and reversibly bind a previously described enzyme, ubiquitin carboxyl-terminal hydrolase [Pickart, C. M., & Rose, I. A. (1985) *J. Biol. Chem. 260*, 7903–7910]. The resin was then used to isolate three newly identified ubiquitin carboxyl-terminal hydrolytic activities, which did not bind to ubiquitin immobilized via lysine residues. Significant purification was achieved in each case, and one isozyme was further purified to homogeneity.

One of the most useful tools in modern protein purification has been affinity chromatography, in which the biological ligand-binding properties of a desired protein or group of proteins are exploited to separate it from other macromolecules. The most frequently used methods of preparing affinity resins focus on covalent attachment of the protein to the bead via nucleophilic side chains, such as those of lysine or cysteine residues [reviewed by Srere and Uyeda (1976)]. Useful as such resins are, they are limited in application, in large part because a number of alternative sites of attachment are available, which results in many possible binding configurations. Even when the number of possible linkages are small, the location of the reactive residue will determine what portion of the ligand is exposed to the media. Frequently, a protein of interest will be unable to recognize and bind to the portion of the ligand exposed.

We wished to develop the chemistry to present alternative regions or conformations of ligands coupled to solid matrices for use as affinity supports. Immobilization via arginine residues would be a logical alternative to currently available supports since there are often fewer of these residues than lysine residues in proteins. We have previously synthesized and characterized a linker molecule, 4-(oxoacetyl)phenoxyacetic acid (OAPA),[1] which was capable of binding to arginine residues of a ligand protein via a glyoxal functional group, and also of being linked to a solid matrix via a carboxyl group. Our studies indicated that many proteins immobilized in this manner were irreversibly bound (Duerksen & Wilkinson, 1987). Since this characteristic is highly desirable for affinity chromatographic work, we sought to chemically define the reaction through X-ray crystallography and nuclear magnetic

† This work was supported by grants awarded to K.D.W. from the NIH (GM30308) and the NSF (DMB-8703828). K.D.W. is the recipient of an NIH Research Career Development Award (HL01929).
* Correspondence should be addressed to this author at 255 Woodruff Memorial Building, Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322.
‡ Present address: Department of Immunology, Emory University School of Medicine, Atlanta, GA 30322.
§ Present address: Department of Biology, University of California at San Diego, LaJolla, CA 92093.

[1] Abbreviations: OAPA, 4-(oxoacetyl)phenoxyacetic acid; UCH, ubiquitin carboxyl-terminal hydrolase; NMR, nuclear magnetic resonance; HPLC, high-performance liquid chromatography, Tris·HCl, tris(hydroxymethyl)aminomethane hydrochloride; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; Ub-OAPA-R, the affinity resin obtained by coupling ubiquitin to OAPA immobilized on polyacrylamide beads as described by Duerksen and Wilkinson (1987).

0006-2960/89/0428-8530$01.50/0