Kannan, K. K., Notstrand, B., Fridborg, K., Lövgren, S., Ohlsson, A., & Petef, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72,* 51.
Khalifah, R. G. (1977) *Biochemistry 16,* 2236.
Khalifah, R. G., Strader, D. J., Bryant, S. H., & Gibson, S. M. (1977) *Biochemistry 16,* 2241.
Liljas, A. (1971) *Crystal Structure Studies of Human Erythrocyte Carbonic Anhydrase C at High Resolution,* Thesis, Acta Universitatis Upsaliensis.
Liljas, A., Kannan, K. K., Bergsten, P.-C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, U., Järup, L., Lövgren, S., & Petef, M. (1972) *Nature (London), New Biol. 235,* 131.
Lindskog, S. (1966) *Biochemistry 6,* 2641.
Lindskog, S., & Coleman, J. E. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70,* 2505.
Lindskog, S., Henderson, L. E., Kannan, K. K., Liljas, A., Nyman, P. O., & Strandberg, B. (1971) *Enzymes, 3rd Ed. 5,* 587.
Maren, T. M., Rayburn, C. S., & Liddell, N. E. (1976) *Science 191,* 469.
Notstrand, B. (1974) *The High Resolution Structure of Human Carbonic Anhydrase, Form B,* Ph.D. Thesis, University of Uppsala.
Pocker, Y., & Watamori, N. (1971) *Biochemistry 10,* 4843.
Prestegard, J. H., & Grant, D. M. (1978) *J. Am. Chem. Soc. 100,* 4664.
Strader, D. J., & Khalifah, R. G. (1976) *J. Am. Chem. Soc. 98,* 5043.
Whitney, P. L. (1970) *Eur. J. Biochem. 16,* 126.
Whitney, P. L., Nyman, P. O., & Malmstrom, B. G. (1967) *J. Biol. Chem. 242,* 4212.

# Complete Amino Acid Sequence of the Major Component Myoglobin from the Humpback Whale, *Megaptera novaeangliae*†

Lee D. Lehman, Francis E. Dwulet,‡ Barry N. Jones, Richard A. Bogardt, Jr., Steven T. Krueckeberg,§ Ronald B. Visscher,# and Frank R. N. Gurd*

ABSTRACT: The complete primary structure of the major component myoglobin from the humpback whale, *Megaptera novaeangliae,* was determined by specific cleavage of the protein to obtain large peptides which are readily degraded by the automatic sequencer. Over 80% of the amino acid sequence was established from the three peptides resulting from the cleavage of the apomyoglobin at the two methionine residues with cyanogen bromide along with the four peptides resulting from the cleavage of the acetimidated apomyoglobin at the three arginine residues with trypsin. The further digestion of the central cyanogen bromide peptide with trypsin and *S. aureus* strain V8 protease enabled the determination of the remainder of the covalent structure. This myoglobin differs from that of sperm whale, *Physeter catodon,* at 12 positions, and dwarf sperm whale, *Kogia simus,* at 14 positions, finback whale *Balaenoptera physalus* at 3 positions, minke whale, *Balaenoptera acutorostrata* at 2 positions, and California gray whale *Eschrichtius gibbosus,* at 1 position. All of the substitutions observed in this sequence fit readily into the three-dimensional structure of sperm whale myoglobin.

The complete sequence reported here for the myoglobin from the humpback whale, *Megaptera novaeangliae,* is in total agreement with that found for the first 60 residues by Edman & Begg (1967) in their classical introduction of automated sequencing methodology. The complete amino acid sequence of the myoglobin from Amazon River dolphin (Dwulet et al., 1975), California gray whale (Bogardt et al., 1976), Atlantic bottlenosed dolphin (Jones et al., 1976), arctic minke whale (Lehman et al., 1977), dwarf sperm whale (Dwulet et al., 1977), Pacific common dolphin (Wang et al., 1977), finback whale (DiMarchi et al., 1978a), pilot whale (Jones et al., 1978), and Dall porpoise (Meuth et al., 1978) have been reported. All of these sequences of cetacean myoglobin were determined by automated Edman degradation. This paper reports the application of the peptide fragmentation and analytical procedures that were used in these papers in determining the complete amino acid sequence of the major component myoglobin from the humpback whale. Completion of this sequence extends the number of known cetacean myoglobin sequences to 14. In addition to the above-mentioned proteins, the primary structures of the myoglobins from the Black Sea dolphin (Kluh & Bakardjieva, 1971), common porpoise (Bradshaw & Gurd, 1969; Meuth et al., 1978), sperm whale (Edmundson, 1965; Romero-Herrera & Lehmann, 1974), and killer whale (Castillo et al., 1977) have also been reported.

## Experimental Section

### *Materials*

The principal component of humpback whale myoglobin was isolated from muscle tissue as described by Hapner et al. (1968). Phosphate buffer (pH 6.4, 0.1 ionic strength) was used to effect the purification of the crude homogenate on Sephadex CM-50. The homogeneity of the purified myoglobin was shown

† From the Department of Chemistry, Indiana University, Bloomington, Indiana 47401. *Received March 21, 1978.* This is the 97th paper in a series dealing with coordination complexes and catalytic properties of proteins and related substances. For the preceding paper, see DiMarchi et al. (1978b). This work was supported by U.S. Public Health Service Research Grant HL-05556. L.D.L., F.E.D., and R.A.B. were supported by U.S. Public Health Service Grant T-1 GM 1046-14.

‡ Present address: Department of Biology, Indiana University, Bloomington, Ind. 47401.
§ Present address: Scripps Clinic and Research Foundation, La Jolla, Calif. 92037.
# Present address: Proctor and Gamble, Cincinnati, Ohio 45202.

0006-2960/78/0417-3736$01.00/0 

```
                5                  10                  15
 1  Val Leu Ser Asp Ala Glu Trp Gln Leu Val Leu Asn Ile Trp Ala
16  Lys Val Glu Ala Asp Val Ala Gly His Gly Gln Asp Ile Leu Ile
31  Arg Leu Phe Lys Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys
46  Phe Lys His Leu Lys Thr Glu Ala Glu Met Lys Ala Ser Glu Asp
61  Leu Lys Lys His Gly Asn Thr Val Leu Thr Ala Leu Gly Gly Ile
76  Leu Lys Lys Lys Gly His His Glu Ala Glu Leu Lys Pro Leu Ala
91  Gln Ser His Ala Thr Lys His Lys Ile Pro Ile Lys Tyr Leu Glu
106 Phe Ile Ser Asp Ala Ile Ile His Val Leu His Ser Arg His Pro
121 Ala Asp Phe Gly Ala Asp Ala Gln Ala Ala Met Asn Lys Ala Leu
136 Glu Leu Phe Arg Lys Asp Ile Ala Ala Lys Tyr Lys Glu Leu Gly
151 Phe Gln Gly
```

FIGURE 1: The amino acid sequence of humpback whale myoglobin. The hyphens between the amino acid residues have been omitted for clarity.

TABLE I: Amino Acid Composition of *Megaptera novaeangliae* Myoglobin.

| amino acid | no. of residues from acid hydrolysates[a] | no. of residues from the sequence |
|---|---|---|
| Asp | 12.2 | 12 |
| Thr | 5.1 | 5 |
| Ser | 5.1 | 5 |
| Glu | 17.2 | 17 |
| Pro | 3.8 | 4 |
| Gly | 10.1 | 10 |
| Ala | 19.2 | 19 |
| Val | 6.1 | 6 |
| Met | 1.9 | 2 |
| Ile | 9.7 | 10 |
| Leu | 18.1 | 18 |
| Tyr | 1.9 | 2 |
| Phe | 6.9 | 7 |
| Lys | 19.9 | 20 |
| His | 10.9 | 11 |
| Arg | 3.1 | 3 |
| Trp[b] | 1.8 | 2 |

[a] Ferrimyoglobin samples were hydrolyzed for 24, 48, and 72 h and duplicate analyses were performed on each hydrolysate. The values obtained for each residue were averaged except for serine, threonine, valine, isoleucine, and leucine. The values of threonine and serine were obtained by extrapolation to zero time. The values of valine, isoleucine, and leucine were the maximum values (72 h). [b] Tryptophan was determined by the method of Liu & Chang (1971).

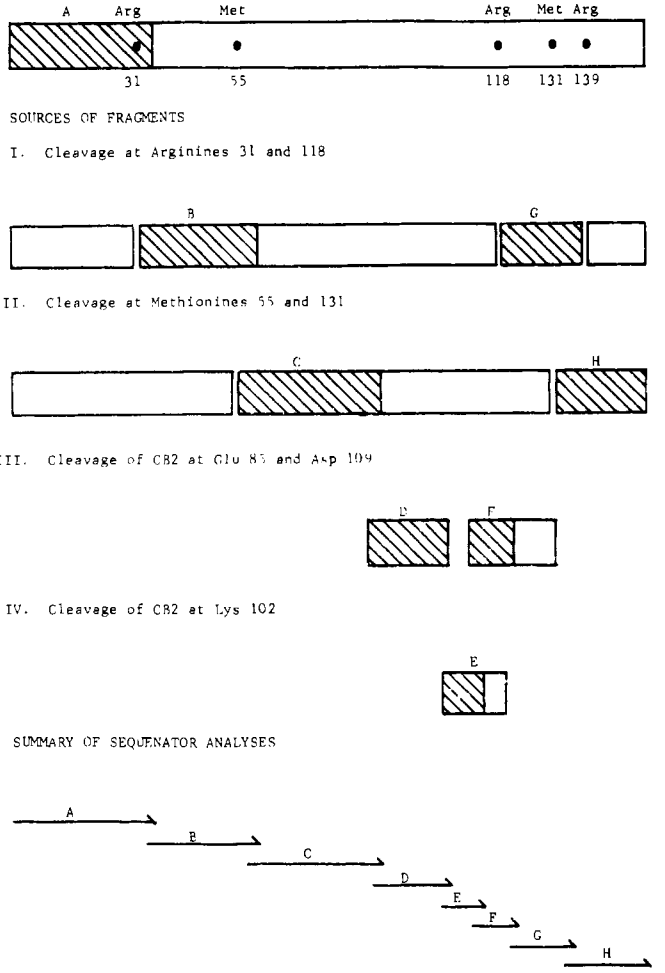


FIGURE 2: Diagrammatic summary of fragments generated from the humpback whale myoglobin for sequenator analysis. The top bar represents the whole myoglobin and the residues that are important for its fragmentation. The capital letters A–H identify the sequenator analyses in the order in which they are described in the text. A hatched section in each horizontal bar indicates the segment of the sequence determined by that analysis. A summary of overlaps is shown in the lower portion by the labeled arrows. Thus the sequenced portions are (A) 1–34; (B) 32–59; (C) 56–89; (D) 86–105; (E) 103–113; (F) 110–121; (G) 119–135; and (H) 132–152.

by cellulose-acetate electrophoresis at pH 9.2 and 5.2. Apomyoglobin was prepared by the method of Yonetani (1967). All other materials were as described in Dwulet et al. (1977).

### *Methods*

*Peptide Nomenclature.* For all cleavage methods the resulting peptides are numbered from the amino terminus to the carboxyl terminus of the complete sequence. The cyanogen bromide fragments are designated by the symbol CB, peptides resulting from the cleavage with trypsin at the arginine residues in the acetimidated protein by the symbol MT, and the tryptic and staphylococcal protease peptides obtained by the digestion of the middle cyanogen bromide peptide, CB2, by the symbols TCB2 and PCB2, respectively.

*Amino Acid Analysis.* Acid hydrolysis was performed with constant-boiling HCl at 110 °C unless otherwise specified. The amino acids were analyzed by the method of Spackman et al. (1958) on a Model 121 Beckman amino acid analyzer interfaced with a Texas Instruments 980A minicomputer which performed the identification and integration of the amino acid chromatograms (Bogardt, 1978; Jones et al., 1978). Tryptophan was determined by the method of Liu & Chang (1971).

All chemical modifications, enzymatic and chemical cleavages, and peptide fractionation procedures were as described previously (Dwulet et al., 1975; Bogardt et al., 1976).

## Results

*Amino Acid Composition.* The amino acid composition of the principal component myoglobin from the humpback whale was obtained from 24-, 48-, and 72-h hydrolysates of the ferrimyoglobin. The results are summarized in Table I.

*Peptide Separation.* The peptides resulting from the cyanogen bromide cleavage, tryptic digestion of the acetimidated apomyoglobin, tryptic digestion of CB2 (56–131), and staphylococcal protease digestion of CB2 (56–131) were purified as described previously (Dwulet et al., 1975; Bogardt et al., 1976; Jones et al., 1976). The amino acid analyses of all peptides obtained were found to be in good agreement with the

| Residue Number | 3 | 4 | 5 | 8 | 12 | 13 | 21 | 31 | 35 | 45 | 51 | 65 | 74 | 109 | 121 | 122 | 129 | 132 | 151 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Species | | | | | | | | | | | | | | | | | | | |
| Humpback Whale | Ser | Asp | Ala | Gln | Asn | Ile | Val | Arg | Gly | Lys | Thr | Asn | Gly | Asp | Ala | Asp | Ala | Asn | Phe |
| Gray Whale | Ser | Asp | Ala | Gln | Asn | Ile | Val | Arg | Gly | Lys | Thr | Asn | Gly | Asp | Gly | Asp | Ala | Asn | Phe |
| Minke Whale | Ser | Asp | Ala | His | Asn | Ile | Val | Arg | Gly | Lys | Thr | Asn | Gly | Asp | Ala | Glu | Ala | Asn | Phe |
| Finback Whale | Thr | Asp | Ala | His | Asn | Ile | Val | Ser | Gly | Lys | Thr | Asn | Gly | Asp | Ala | Glu | Ala | Asn | Phe |
| Sperm Whale | Ser | Glu | Gly | Gln | His | Val | Val | Arg | Ser | Arg | Thr | Val | Ala | Glu | Gly | Glu | Gly | Asn | Tyr |
| Dwarf Sperm Whale | Ser | Glu | Gly | Gln | His | Val | Ile | Arg | His | Arg | Ser | Val | Ala | Glu | Ala | Glu | Gly | Ser | Tyr |

FIGURE 3: Comparison of the amino acid sequences of other cetacean myoglobins. Only those positions in which differences occur are reported. All other positions are conserved.

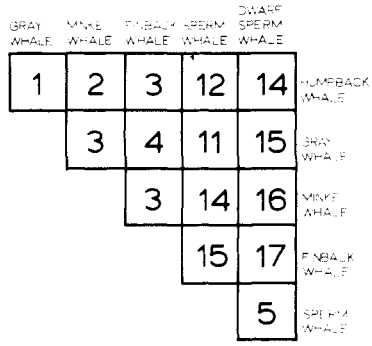


FIGURE 4: Difference matrix for cetacean myoglobins obtained by summing the number of different amino acids between pairs of proteins.

expected values and the results can be found in the supplementary material (see paragraph concerning supplementary material at the end of this paper).

*Sequence Investigation.* Only the sequence data necessary to establish the entire primary structure are reported here.

*Sequencer Results.* The complete primary structure of humpback whale myoglobin is shown in Figure 1. The sequence strategy used to obtain the primary structure is outlined diagrammatically in Figure 2. Sequencer analysis A (Figure 2) represents the first 34 amino terminal residues obtained by the automatic Edman degradation of the intact apomyoglobin. Peptide MT2 (32–118) in sequencer analysis B provided a 3-residue overlap with the intact protein analysis and extended the sequence 25 residues to position 59. Sequencer analysis C of peptide CB2 (56–131) provided an overlap of sequencer analysis B of 4 residues and extended the sequence 30 residues to position 89. Analysis D yielded the entire sequence of peptide PCB2-4 (86–105), which overlapped analysis C by 4 residues and extended the sequence 18 residues to position 105. Peptide TCB2-6 (103–118) in sequence analysis E overlapped analysis D with the only tyrosine in CB2 (56–131) and extended the sequence to position 113. Analysis F of PCB2-6 (110–131) extended the sequence to position 121, overlapping arginine 118. Peptide MT3 (119–139) was used in analysis G to overlap peptide PCB2-6 (110–131) and extended the sequence to position 135. The final sequencer analysis H of peptide CB3 (132–153) overlapped analysis G starting at position 132 and extended the sequence to the carboxyl terminus of the protein at position 153.

In all sequencer runs the yields for phenylthiohydantoins were similar to those previously discussed (Dwulet et al., 1975).

## Discussion

The present report is the 10th in a series[1,2] of complete cetacean myoglobin sequences determined by automated Edman degradation. The information derived from these sequence investigations has been used in the study of electrostatic interactions within the myoglobin molecule (Shire et al., 1975; Matthew et al., 1978), in the comparison of oxygen equilibria of myoglobin from different species (Shire 1974), in comparison of the acid denaturation of different myoglobins (Friend et al., 1977), in the interpretation of proton NMR results in which p$K_a$ values of individual histidine residues were assigned (Botelho, 1975), and in the development of a computer model of cetacean phylogenetics (Bogardt, 1978).

The sequence of humpback whale myoglobin is compared in Figure 3 with those of other baleen whales, the gray, minke, and finback whales, and with the toothed sperm and dwarf sperm whales. As seen in the difference matrix in Figure 4, humpback has close similarity to the other baleen whales with one amino acid difference from gray, two from minke, and three from finback whale. The one amino acid difference from gray whale is at position 121, which is a glycine in the gray whale and an alanine in the humpback whale. Both glycine and alanine are common at this position in other cetacean myoglobins. This is a conservative change and the analogous residue of sperm whale myoglobin is found on the surface of the molecule (Watson, 1969; Takano, 1977). The above sequence is compatible with the three-dimensional structure of sperm whale myoglobin and no significant change in the backbone conformation between these proteins is to be anticipated.

## Acknowledgments

The authors are grateful to Mrs. K. Verhey, Mr. J. Gutfleisch, and Mr. D. Stiff for their excellent technical help and to Mrs. D. Embry for typing this manuscript. We are also grateful to Dr. G. I. H. Hanania and Dr. E. Mitchell for helpful advice.

## Supplementary Material Available

Experimental material including elution profiles, sequencer

[1] B. N. Jones, work in progress, *Stenella attenuata graffmani*.

[2] J. A. Dwulet, work in progress, *Mesophodon carlhubbsi*.

repetitive yield plots, and amino acid composition tables are provided (18 pages). Ordering information is given on any current masthead page.

## References


Bogardt, R. A. (1978) Ph.D. Thesis, Indiana University.

Bogardt, R. A., Dwulet, F. E., Lehman, L. D., Jones, B. N., & Gurd, F. R. N. (1976) *Biochemistry 15,* 2497.

Botelho, L. H. (1975) Ph.D. Thesis, Indiana University.

Bradshaw, R. A., & Gurd, F. R. N. (1969) *J. Biol. Chem. 244,* 267.

Castillo, O., Lehmann, H., & Jones, L. T. (1977) *Biochim. Biophys. Acta 491,* 23.

DiMarchi, R. D., Wang, C.-C., Hemenway, J. B., & Gurd, F. R. N. (1978a) *Biochemistry 17,* 1968.

DiMarchi, R. D., Garner, W. H., Wang, C.-C., Hanania, G. I. H., & Gurd, F. R. N. (1978b) *Biochemistry 17,* 2822.

Dwulet, F. E., Bogardt, R. A., Jones, B. N., Lehman, L. D., & Gurd, F. R. N. (1975) *Biochemistry 14,* 5336.

Dwulet, F. E., Jones, B. N., Lehman, L. D., & Gurd, F. R. N. (1977) *Biochemistry 16,* 873.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1,* 80.

Edmundson, A. B. (1965) *Nature (London) 205,* 883.

Friend, S. H., Hanania, G. I. H., Gurd, R. S., & Gurd, F. R. N. (1977) *Fed. Proc., Fed. Am. Soc. Exp. Biol., 36,* 890.

Hapner, K. D., Bradshaw, R. A., Hartzell, C. R., & Gurd, F. R. N. (1968) *J. Biol. Chem. 243,* 683.

Jones, B. N., Vigna, R. A., Dwulet, F. E., Bogardt, R. A., Lehman, L. D., & Gurd, F. R. N. (1976) *Biochemistry 15,* 4418.

Jones, B. N., Dwulet, F. E., Lehman, L. D., Garner, M. H., Bogardt, R. A., Garner, W. H., & Gurd, F. R. N. (1978) *Biochemistry 17,* 1971.

Kluh, J., & Bakardjieva, A. (1971) *FEBS Lett. 17,* 31.

Lehman, L. D., Dwulet, F. E., Bogardt, R. A., Jones, B. N., & Gurd, F. R. N. (1977) *Biochemistry 16,* 706.

Liu, T. Y., & Chang, Y. T. (1971) *J. Biol. Chem. 246,* 2842.

Matthew, J. B., Friend, S. H., Botelho, L. H., Lehman, L. D., Hanania, G. I. H., & Gurd, F. R. N. (1978) *Biochem. Biophys. Res. Commun. 81,* 416.

Meuth, J. L., Jones, B. N., Garner, W. H., & Gurd, F. R. N. (1978) *Biochemistry 17* (in press).

Romero-Herrera, A. E., & Lehmann, H. (1974) *Biochim. Biophys. Acta 366,* 318.

Shire, S. J. (1974) Ph.D. Thesis, Indiana University.

Shire, S. J., Hanania, G. I. H., & Gurd, F. R. N. (1975) *Biochemistry 14,* 1352.

Spackman, D. H., Moore, S., & Stein, W. H. (1958) *Anal. Chem. 30,* 1190.

Takano, T. (1977) *J. Mol. Biol. 110,* 537.

Wang, C.-C., Avila, R., Jones, B. N., & Gurd, F. R. N. (1977) *Biochemistry 16,* 4978.

Watson, H. C. (1969) *Prog. Stereochem. 4,* 299.

Yonetani, T. (1967) *J. Biol. Chem. 242,* 5008.