Complete Amino Acid Sequence of Myoglobin from the Pilot Whale, *Globicephala melaena*†

Barry N. Jones, Francis E. Dwulet,‡ Lee D. Lehman, Margaret H. Garner,§ Richard A. Bogardt, Jr.,¶ William H. Garner,§ and Frank R. N. Gurd*

ABSTRACT: The complete amino acid sequence of the major component myoglobin from the pilot whale, *Globicephala melaena,* was determined by specific cleavage of the protein to obtain large peptides which are readily degraded by the automatic sequencer. The apomyoglobin was selectively cleaved at the two methionyl residues with cyanogen bromide and the acetimidated apomyoglobin was cleaved at the three arginyl residues by trypsin. From the sequence analysis of four of these peptides and the apoprotein, over 90% of the covalent structure of the protein was obtained. The remainder of the primary structure was determined by sequence analysis of three of the tryptic peptides isolated from the central cyanogen bromide fragment after modification of its single arginyl residue with 1,2-cyclohexanedione. This myoglobin differs from that of the Black Sea dolphin at four positions and from the myoglobin of the killer whale, Pacific common dolphin, and Atlantic bottlenosed dolphin at two positions. The above differences reflect the close taxonomic relationship of these five species of Cetacea. This sequence determination was aided by the use of a Texas Instruments 980A minicomputer system which performed peak integrations for all samples subjected to amino acid analysis.

In preceding papers, 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) and finback whale (DiMarchi et al., 1978) was reported. All of these sequences of catacean myoglobins 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 myoglobin from pilot whale, *Globicephala melaena.* Completion of this sequence extends the number of complete cetacean myoglobin sequences to 12. In addition to the above-mentioned proteins, the primary structures of the myoglobins from the Black Sea dolphin (Kluh and Bakardjieva, 1971), common porpoise (Bradshaw and Gurd, 1969), sperm whale (Edmundson, 1965; Romero-Herrera and Lehmann, 1974) and killer whale (Castillo et al., 1977) have also been reported.

Experimental Section

Materials

The principal component of pilot whale myoglobin was isolated from the frozen muscle tissue by the procedure of Hapner et al. (1968). Phosphate buffer (pH 6.6), ionic strength 0.1, was used to effect purification of the crude homogenate on C-50 CM[1]-Sephadex. The homogeneity of the purified myoglobin was verified by cellulose acetate electrophoresis at pH 9.2 and 5.2. Apomyoglobin was prepared essentially by the procedure of Teale (1959).

Methyl acetimidate hydrochloride was prepared according to the procedure of Hunter and Ludwig (1962) and 3-sulfophenyl isothiocyanate sodium salt was prepared according to the method of Dwulet and Gurd (1976). Tos-PheCH_2Cl[1]-treated trypsin was purchased from Worthington Biochemical Laboratories and 1,2-cyclohexanedione was a product of Aldrich Chemical Co. Sequencer reagents of "Sequencer" grade were obtained from Beckman Instruments. All other chemicals were the highest grade available.

Methods

Amino Acid Analysis. Peptides were hydrolyzed with constant-boiling HCl in evacuated tubes for 24 h at 110 °C unless otherwise specified. Amino acids were analyzed on either a Model 120B or 120C Beckman amino acid analyzer by the method of Spackman et al. (1958). Tryptophan was determined by the method of Liu and Chang (1971).

Computer Integration of Amino Acid Chromatograms. A Texas Instruments 980A minicomputer with 12K memory was used to perform all calculations involved in the integration of chromatographic peaks obtained from amino acid analyses. The minicomputer was interfaced on line to three amino acid analyzers so as to integrate chromatographic peaks as they are eluted. In order to reduce memory requirements, only the peak elution times and areas are stored. At the end of each analysis, the chromatographic peaks are identified according to their elution times, the areas are converted to nanomoles, and the results are printed out. Time sharing allows the minicomputer to perform other desired tasks without interference to its amino acid analysis integration duties. A diagram of the integration system and a typical printout of an analysis are presented in the Supplementary Material. Details for the construction of the system are reported by Bogardt (1978).

† From the Department of Chemistry, Indiana University, Bloomington, Indiana 47401. *Received December 20, 1977.* This is the 94th paper in a series dealing with coordination complexes and catalytic properties of proteins and related substances. For the preceding paper see DiMarchi et al. (1978). This work was supported by Public Health Service Research Grant HL-05556. F.E.D., L.D.L., and R.A.B. were supported by Public Health Service Grant No. T01 GM 1046-14.

‡ Present address: Department of Biology, Indiana University, Bloomington, Ind. 47401.

§ Present address: Department of Ophthalmology, Columbia University, New York, N.Y. 10032.

¶ Present address: Scripps Clinic and Research Foundation, La Jolla, Calif. 92037.

[1] Abbreviations used: Tos-PheCH_2Cl, L-1-tosylamido-2-phenylethyl chloromethyl ketone; CM, carboxymethyl.

0006-2960/78/0417-1971$01.00/0

TABLE I: Amino Acid Composition of *Globicephala melaena* Myoglobin.

Amino acid	No. of residues from acid hydrolysates[a]	No. of residues from the sequence
Asp	11.1	11
Thr	5.1	5
Ser	5.1	5
Glu	17.2	17
Pro	4.0	4
Gly	13.2	13
Ala	17.2	17
Val	4.8	5
Met	2.0	2
Ile	8.9	9
Leu	18.9	19
Tyr	2.0	2
Phe	7.0	7
Lys	19.8	20
His	11.9	12
Arg	3.0	3
Trp[b]	2.0	2

[a] Samples were hydrolyzed for 24, 48, and 72 h, and duplicate analyses were performed on each hydrolysate. The results obtained for each residue were averaged, except for threonine, serine, and isoleucine. The values for threonine and serine were obtained by extrapolation to zero time. The value for isoleucine was obtained by extrapolation to 96 h. The number of amino acid residues was calculated on the basis of 153 amino acids in the protein. [b] Tryptophan was determined by the method of Liu and Chang (1971).

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

FIGURE 1: The amino acid sequence of pilot whale myoglobin. The three-letter amino acid code has been used for simplicity and the hyphens between the amino acid residues have been omitted for clarity.

Peptide Nomenclature. The peptides isolated from the tryptic cleavage of the acetimidated apomyoglobin are given the symbol MT. The cyanogen bromide fragments from the apomyoglobin are designated by the symbol CB and those from the acetimidated apomyoglobin are labeled MCB. The tryptic peptides obtained by digestion of the central cyanogen bromide fragment, CB2, after modification of its single arginyl residue with cyclohexanedione are given the symbol CCB2-T. All peptides are numbered from the amino terminus to the carboxyl terminus of the completed sequence.

Specific Cleavage and Sequencing Techniques of Peptides. Preparation of the acetimidated apomyoglobin and its subsequent digestion with trypsin were accomplished as previously described by Dwulet et al. (1975). Cleavage of the apomyoglobin and the acetimidated apomyoglobin with cyanogen bromide was performed according to the procedure of Dwulet et al. (1975). The modification of the single arginyl residue in the central cyanogen bromide fragment, CB2 (56-131), with 1,2-cyclohexanedione and the subsequent digestion of this fragment with trypsin were accomplished according to the method of Dwulet et al. (1977).

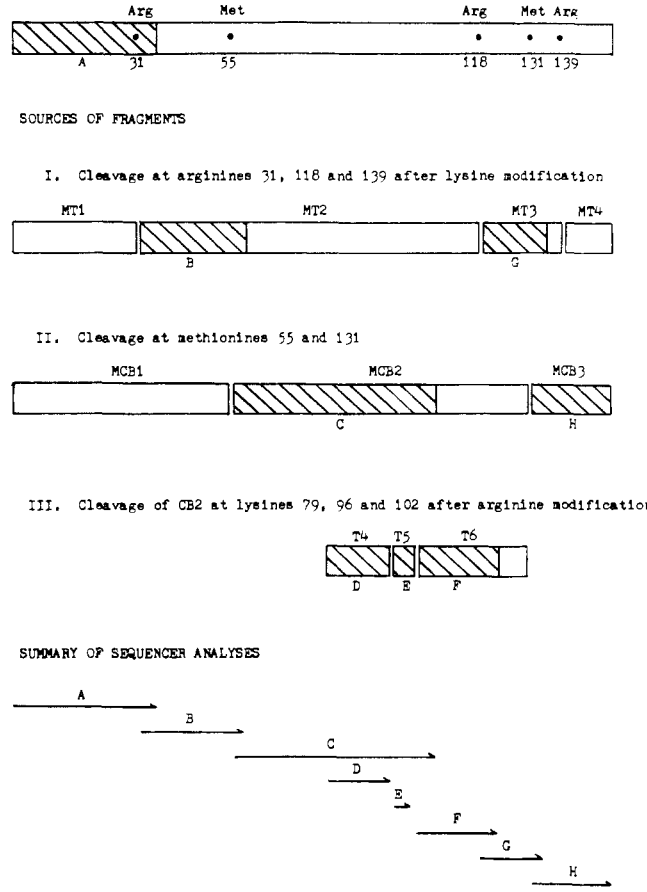

FIGURE 2: Diagrammatic summary of fragments generated from pilot whale myoglobin for sequencer analysis. The top bar represents the whole protein and the residues that are important for its fragmentation. A hatched segment in each horizontal bar indicates the segment of sequence determined by sequence analysis on that fragment. A summary of overlaps is shown in the lower portion by the labeled arrows. Thus, the sequenced portions are: A, 1-36; B, 32-59; C, 56-107; D, 80-96; E, 97-102; F, 103-123; G, 119-135; and H, 132-153.

Automatic phenyl isothiocyanate degradations of peptides were performed on a Beckman 980C sequencer. The sequencing techniques and the methods for the identification of phenylthiohydantoin amino acids used in this sequence determination are identical to those previously reported by Dwulet et al. (1975).

Results

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

Peptide Isolation. All the techniques used in the previous papers in this series to isolate the needed peptides were also used here. The elution profiles and peptide compositions can be found in the Supplementary Material.[2]

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

Sequencer Results. The complete primary structure of pilot

[2] Results of established procedures can be found in the Supplementary Material.

whale myoglobin is presented in Figure 1. The sequence strategy used to establish the structure is outlined diagrammatically in Figure 2. The numbers of residues represented in Figure 2 by each hatched section are: A, 36; B, 28; C, 52; D, 17; E, 6; F, 21; G, 17; and H, 22.

Discussion

Comparison with the Myoglobin Sequences of Other Delphinids. The present report is the eighth in a series of complete cetacean myoglobin sequences[3,4] determined by automated Edman degradation. The information derived from the amino acid sequence has aided previous studies of pilot whale myoglobin in which the pK_a values were determined for the individual histidine residues (Botelho, 1975) as well as the α-amino group (Garner et al., 1973) of the protein. In addition, the sequence analysis of pilot whale myoglobin has given a better definition to the phylogenetic relationships among the members of the cetacean family Delphinidae (dolphins) (Jones, 1977; Bogardt, 1978).

The sequence of pilot whale myoglobin is compared in Figure 3 with the known myoglobin sequences of other delphinids. As can be seen from this figure, the pilot whale myoglobin differs from that of the Black Sea dolphin at not less than four residue positions[5] and from each of the myoglobins from the killer whale, Atlantic bottlenosed dolphin, and Pacific common dolphin at only two positions. Examination of the substitution pattern at positions 21 and 28 indicates that these positions may be examples of Pitch's "covarions" (Fitch and Markowitz, 1970). Covarions are concomitantly variable codons represented by a limited set of amino acids within a protein structure. In the majority of the known myoglobin sequences, an increase in the side-chain volume of residue 21 from a valine to a leucine is accompanied by a decrease in the side-chain volume at position 28 from an isoleucine to a valine. Both of these positions are located within the interior of the myoglobin molecule, but they are not in direct contact with each other (Watson, 1969; Takano, 1977). However, some form of long-range stereochemistry may favor these exchanges in order to maintain internal molecular structure. The myoglobins of the pilot whale and killer whale are two of the very few exceptions to this phenomenon.

Position 66 has been established as a hypervariable locus in the myoglobin sequence and has been found to be occupied by asparagine in most cases, but there are also examples of mammalian myoglobins that have either histidine,[6] lysine (Vigna et al., 1974), threonine (Darbre et al., 1975), valine (Edmundson, 1965; Dwulet et al., 1977), glycine (Nauman, 1973), or alanine (Romero-Herrera and Lehmann, 1973) at this position. The occurrence of aspartic acid at this position has only been observed in the myoglobin from the Black Sea dolphin.

Glutamic acid at position 83 is the usual residue at this site in the myoglobin sequence. Aspartic acid at this position has been found only in the myoglobins from four of the delphinids in Figure 3 and the California sea lion (Vigna et al., 1974).

The final site of difference between these five delphinid myoglobins occurs at position 122. The usual residue at this position is either aspartic acid or glutamic acid. The only reported cases in which an amide residue occurs at this position are in the myoglobins from the killer whale and Black Sea dolphin. The large degree of similarity between these five proteins is consistent with the close taxonomic relationships between these species of Cetacea (Jones, 1977; Jones et al., 1977).

Species \ Residue Number	21	28	66	83	122
Pilot Whale	Leu	Ile	Asn	Glu	Glu
Killer Whale	Leu	Ile	Asn	Asp	Gln
Atlantic Bottlenosed Dolphin	Leu	Val	Asn	Asp	Glu
Pacific Common Dolphin	Leu	Val	Asn	Asp	Glu
Black Sea Dolphin	Val	Ile	Asp	Asp	Gln

FIGURE 3: Comparison of the amino acid sequences of delphinid myoglobins whose sequences have been completed to date. Only those positions in which differences occur are reported. All other positions are conserved and are the same as in the pilot whale myoglobin sequence.

Computer Integration System for Amino Acid Analyses. The sequence determination of pilot whale myoglobin was greatly aided by a computer system which was designed to integrate chromatographic peaks obtained from amino acid analyses and thereby eliminate the time-consuming process of manual integrations. This system is relatively inexpensive to assemble and operate as well as being flexible for general use where versatility is required. A comparison of the computer integration with hand integrations indicates that the computer integration is more sensitive to baseline variations than manual computations by an experienced person. However, the overall mean error of the computer, 1.09 ± 0.61%, is not significantly different from the hand integration, 0.89 ± 0.01%, and both are less than the 2% error considered inherent in the amino acid analyzer.[7]

Acknowledgments

The authors are grateful to Mrs. K. Verhey, Mrs. D. Embry, and Mr. J. Gutfleisch for their excellent technical help, Dr. J. C. Kendrew for his help in the procurement of the animal material, Mr. R. Sporleder for the development and maintenance of the computer hardware, Mr. N. Loy for building the amplifiers, Messrs. J. Russ and J. Schneider for programming assistance, and Drs. G. I. H. Hanania and Edward Mitchell for advice.

Supplementary Material Available

Supplementary data are available including elution profiles, peptide compositions, repetitive yield plots, and computer-related materials (20 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., and Gurd, F. R. N. (1976), *Biochemistry 15,* 2597.
Botelho, L. H. (1975), Ph.D. Thesis, Indiana University.
Bradshaw, R. A., and Gurd, F. R. N. (1969), *J. Biol. Chem. 244,* 267.
Castillo, O., Lehmann, H., and Jones, L. T. (1977), *Biochim. Biophys. Acta 491,* 23.

[3] J. L. Meuth, work in progress.
[4] J. A. Dwulet, work in progress.
[5] A fifth point of difference is uncertain. Kluh and Bakardjieva (1971) first assigned position 26 as glutamine and on subsequent repetition were only able to detect glutamic acid. It was thought possible that prolonged exposure to an alkaline medium might have caused deamidation.
[6] L. D. Lehman, unpublished observation.
[7] A comparison of the reproducibility of this integration system with the hand-calculation method is presented in the Supplementary Material.

Darbre, P. D., Romero-Herrera, A. E., and Lehmann, H. (1975), *Biochim. Biophys. Acta 393,* 201.
DiMarchi, R. D., Wang, C.-C., Hemenway, J. B., and Gurd, F. R. N. (1978), *Biochemistry* (preceding paper in this issue).
Dwulet, F. E., Bogardt, R. A., Jones, B. N., Lehman, L. D., and Gurd, F. R. N. (1975), *Biochemistry 14,* 5336.
Dwulet, F. E., and Gurd, F. R. N. (1976), *Anal. Biochem. 76,* 530.
Dwulet, F. E., Jones, B. N., Lehman, L. D., and Gurd, F. R. N. (1977), *Biochemistry 16,* 873.
Edmundson, A. B. (1965), *Nature (London) 205,* 883.
Fitch, W. M., and Markowitz, E. (1970), *Biochem. Genet. 4,* 579.
Garner, M. H., Garner, W. H., and Gurd, F. R. N. (1973), *J. Biol. Chem. 248,* 5451.
Hapner, K. D., Bradshaw, R. A., Hartzell, C. R., and Gurd, F. R. N. (1968), *J. Biol. Chem. 243,* 683.
Hunter, M. J., and Ludwig, M. L. (1962), *J. Am. Chem. Soc. 84,* 3491.
Jones, B. N. (1977), Ph.D. Thesis, Indiana University.
Jones, B. N., Dwulet, F. E., Lehman, L. D., DiMarchi, R. D., Wang, C.-C., Meuth, J. L., and Bogardt, R. A. (1977), *Fed. Proc. Fed. Am. Soc. Exp. Biol. 36,* 754.
Jones, B. N., Vigna, R. A., Dwulet, F. E., Bogardt, R. A., Lehman, L. D., and Gurd, F. R. N. (1976), *Biochemistry 15,* 4418.
Kluh, J., and Bakardjieva, A. (1971), *FEBS Lett. 17,* 31.
Lehman, L. D., Dwulet, F. E., Bogardt, R. A., Jones, B. N., and Gurd, F. R. N. (1977), *Biochemistry 16,* 706.
Liu, T. Y., and Chang, Y. T. (1971), *J. Biol. Chem. 246,* 2842.
Nauman, L. W. (1973), Ph.D. Thesis, University of Alaska.
Romero-Herrera, A. E., and Lehmann, H. (1973), *Proc. R. Soc. London 186,* 249.
Romero-Herrera, A. E., and Lehmann, H. (1974), *Biochim. Biophys. Acta 366,* 249.
Spackman, D. H., Moore, S., and Stein, W. H. (1958), *Anal. Chem. 30,* 1190.
Takano, T. (1977), *J. Mol. Biol. 110,* 537.
Teale, F. W. J. (1959), *Biochim. Biophys. Acta 35,* 543.
Vigna, R. A., Gurd, L. J., and Gurd, F. R. N. (1974), *J. Biol. Chem. 249,* 4144.
Wang, C.-C., Avila, R., Jones, B. N., and Gurd, F. R. N. (1977), *Biochemistry 16,* 4978.
Watson, H. C. (1969), *Prog. Stereochem. 4,* 299.

Characterization of a Long-Wavelength Feature in the Absorption and Circular Dichroism Spectra of β-Nicotinamide Adenine Dinucleotide. Evidence for a Charge Transfer Transition†

Richard R. Reisbig and Robert W. Woody*

ABSTRACT: The absorption spectrum of β-NAD^+ has a weak, broad feature extending to wavelengths well above 300 nm. A corresponding positive band is observed in the circular dichroism (CD) spectrum. α-NAD^+ has a somewhat stronger long-wavelength band in absorption and a negative CD band at 312 nm, opposite in sign to the main 260-nm band. These long-wavelength features are markedly diminished by lowering the pH, raising the temperature, or adding organic solvents. Enzymatic cleavage of both β- and α-NAD^+ abolishes the long-wavelength absorption and CD bands. We have also studied pyridine-3-aldehyde adenine dinucleotide, 3-acetylpyridine adenine dinucleotide, deamino-NAD^+, and $NADP^+$. All of these analogues exhibit long-wavelength absorption and CD bands analogous to those of NAD^+. The breadth of these features and their dependence on external conditions and on the integrity of the dinucleotide all support the assignment of this band to an intramolecular charge transfer transition, with the adenine moiety as the donor and the nicotinamide as the acceptor. The existence of such a charge-transfer transition in β-NAD^+ and related molecules provides conclusive evidence for a significant population of conformations in which the adenine and nicotinamide rings are stacked. Comparison of the extinction coefficient of N^1-methylnicotinamide-adenosine complex with that of β-NAD^+ at 320 nm provides an estimate of 40% stacking in β-NAD^+. In addition to characterizing the long-wavelength band of β-NAD^+ and various analogues, higher energy transitions have been characterized and some assignments made. In particular, the $n\pi^*$ transitions characteristic of the 3-substituent of the pyridinium ring are assigned by analogy to the corresponding substituted benzene derivative.

The pyridine nucleotide coenzyme,[1] β-NAD^+, is required as an electron carrier in metabolic processes in all known forms of life. The conformation of the coenzyme when bound to several enzymes has been determined by x-ray diffraction studies (Brändèn et al., 1975; Moras et al., 1975; Rossmann et al., 1975; Holbrook et al., 1975; Liljas & Rossmann, 1974;

† From the Department of Biochemistry, Colorado State University, Fort Collins, Colorado 80523. *Received October 12, 1977; revised manuscript received February 7, 1978.* This work was supported by Grants GM 17850 and GM 22994 from the U.S. Public Health Service.

[1] Abbreviations used: β-NAD^+, β-nicotinamide adenine dinucleotide; α-NAD^+, α-nicotinamide adenine dinucleotide (in α-NAD^+ the anomeric carbon of the ribose bound to the nicotinamide is in the α conformation); β-NMN^+, β-nicotinamide mononucleotide; α-NMN^+, α-nicotinamide mononucleotide; AMP, 5′-adenosine monophosphate; ε-NAD^+, nicotinamide 1,N^6-ethenoadenine dinucleotide; paNAD$^+$, 3-pyridinealdehyde adenine dinucleotide; acNAD$^+$, 3-acetylpyridine adenine dinucleotide; deamino-NAD^+, nicotinamide hypoxanthine dinucleotide; CD, circular dichroism; NMR, nuclear magnetic resonance; ^{1}H NMR, proton magnetic resonance; UV, ultraviolet.

0006-2960/78/0417-1974$01.00/0