Structure Elucidation of a Purple Peptide Found During the Purification of a Recombinant Protein from *Escherichia coli*

Andrew W. Guzzetta[a,*], Pamela Clark[b], Robert Tappe[c], Herbert Budzikiewicz[c]

[a] Department of Analytical Chemistry
[b] Department of Protein Chemistry, Scios Inc., 2450 Bayshore Parkway, Mountain View CA, USA
[c] Institute of Organic Chemistry, University at Köln, Greinstr. 4, D-50939 Köln, Germany

Z. Naturforsch. **54c**, 175–180 (1999); received December 9/December 30, 1998

E. coli, Purple Peptide, 3-Carboxy-3,4-dihydro-2*H*-naphtho[2,3-b][1,4]thiazine-5,10-dione

A purple substance (**4**) partially co-purified with a recombinant human B-type natriuretic peptide (hBNP), following an *E. coli* fermentation. The structure of the compound was elucidated by NMR, electrospray and FAB mass spectrometry. The chromophore is a 1,4-naphthoquinone condensed with the N-terminal cysteine of a heptapeptide by its NH_2- and SH-groups to form a dihydro-thiazine ring. The peptide sequence was determined as Cys-Lys-Val-Leu-Arg-Arg-His by mass spectrometric techniques. CID and data base matching identified it as the C-terminus of the 32-amino-acid recombinant peptide hBNP. This modification of an N-terminal Cys may be a more general phenomenon with implications for the production of heterologous proteins by microorganisms.

Introduction

Colored compounds are routinely found in the fermentation broth of microorganisms and frequently contain a peptidic portion. They may be siderophores such as the pyoverdins from fluorescent *Pseudomonas* spp. (Budzikiewicz, 1997) or photoactive molecules such as bacteriorhodopsin from *Halobacterium halobium* (Ovchinnikov *et al.*, 1978). In the production of heterologous proteins they are usually ignored as they can be easily removed by purification from the protein of interest. We now report the isolation and characterization of a purple compound noted during the purification of a recombinant protein, hBNP, from *Escherichia coli.*

Abbreviations: Amino acids, 3-letter code; hBNP, human B-type natriuretic peptide; LC, liquid chromatograph(y); HPLC, high performance liquid chromatography; EI, electron ionization; ESI, electrospray ionization; FAB, fast atom bombardment; LC-MS, liquid chromatograph coupled with a mass spectrometer; CID, collision induced decomposition; MS^n, CID of a selected ion in an ion trap mass spectrometer at a selected stage of decomposition; NMR, nuclear magnetic resonance; COSY, correlation spectroscopy; HMBC, heteronuclear multiple bond correlation; HMQC: heteronuclear multiple quantum coherence.

Reprint requests to Dr. A. W. Guzzetta.
Fax: +1-650-462-5900
E-mail: <guzzetta@sciosinc.com>

Materials and Methods

Instruments

Chromatography: Liquid chromatograph HP 1090 Series II (Hewlet-Packard, Palo Alto CA, USA) with a Vydac C18 0.32x250 mm 3 mμ 500 A capillary column (Micotech. Inc., Sunnyvale CA, USA) maintained at 40 °C. Flow rate 6 μl/min for LC/MS after splitting the flow rate of 200 μl/min emerging from the LC. Solvents 0.05% trifluoroacetic acid in H_2O (A) and in acetonitrile (v/v) (B), gradient 2 to 60% solvent B in 40 min.

Mass spectrometer: (a) For LC/MS a Finnigan (San Jose CA, USA) LCQ ion trap equipped with an ESI source was used. Scan parameters were the basic "triple play" where the instrument cycles through a full scan, a zoom scan for the base peak and an MS/MS analysis of the base peak. In addition, selected ions were further fragmented by CID (MS^n). Capillary temperature 250 °C, ion injection time 200 msec, source voltage 4.9 kV, 35% relative collision energy corresponding to 1.75 V (b) For exact mass measurement Finnigan-MAT (Bremen, D) HSQ-30 (FAB, matrix, thioglycerol/dithiodiethanol) and 900ST (ESI).

0939–5075/99/0300–0175 $ 06.00

NMR: Bruker (Karlsruhe, D) 500 MHz instrument Advance DRX 500. Experimental details see Tables.

UV/Vis: Kontron (Zürich, CH) Uvicon 860. **4**: λ_{max} 286–287 (ε = 2.24x10^4 mol^{-1}cm^{-1}), 520–530 nm (H_2O/CH_3OH 4:1 v/v).

Chemicals

Dithiothreitol, iodoacetic and trifluoroacetic acid were obtained from Sigma (Deisenhofen, D), HPLC grade acetonitrile and H_2O from Burdick & Jackson (Muskegon MI, USA), the enzymes modified trypsin, Lys-C and Arg-C from Promega (Madison WI, USA), carboxypeptidase-B from Worthington (Lakewood NJ, USA).

Isolation, degradation and characterization of **4** ($[M + H]^+$ *m/z* 1065)

The purple compound **4** was obtained following an *E. coli* fermentation and subsequent purification of the heterologous recombinant peptide, hBNP. A description of hBNP can be found in the publications Seilhamer *et al.* (1989) and Seilhamer *et al.* (1997); hBNP is formed as a fusion protein. The unwanted leader protein is removed by acid cleavage. The purple compund **4** co-purifies with the 32-amino-acid hBNP and is separated from it by column chromatography as determined by monitoring the eluent at 286 nm.

Enzymatic degradation of **4** was performed by incubating 2 µg each with 0.1 µg of the respective enzyme in 100 mmol Tris (tris(hydroxymethyl)aminomethane) buffer at pH 8.0 for 19 hrs at ambient temperature. The reaction mixture was analyzed by LC/MS. Carboxypeptidase-B treatment was performed with 20 µg **4** and 10 µg enzyme, all other digest conditions were the same as noted above.

Reduction and S-carboxymethylation reaction of **4** was performed according to the method of Crestfield *et al.* (1963). 20 µg dried **4** were acetylated with acetic anhydride/methanol/pyridine (1.25:3.75:5) at ambient temperature for 6 min. The reaction product was dried on a speed vac. before MS analysis.

Protein sequencing was attempted on an Applied Biosystems 477A Protein Sequencer (Perkin Elmer, Applied Biosystems Division; Foster City CA, USA) using the Normal-1 reaction cycle.

Synthesis of **2**

To the Na^+ salt of Cys dissolved in water/methanol (1:1) 2,3-dibromo-1,4-naphthoquinone was added. The mixture was allowed to stand over night at room temperature. After acidification with acetic acid organic material was extracted with ethyl ether. Chromatography on silicagel plates yielded a series of colored fractions. The purple material was collected.

Results and Discussion

ESI-MS showed an $[M + H]^+$ at *m/z* 1065.5 for the purple contaminant. The $[M + H]^+$-ion yielded fragments at *m/z* 175 and 110, when it was subjected to CID. The first two fragments are consistant with a C-terminal His (^+H_2N=CHR-COOH and ^+H_2N=CHR, i.e. loss of COOH), while *m/z* 129 is chracteristic for an in-chain Lys (^+H_2N=CHR-CO). This suggests a peptidic structure. Acetylation gave a mono-acetyl derivative in agreement with a free amino group of Lys (α or ε). Attempted Edman degradation failed, hence the N-terminus is blocked. Amino acid analysis showed the presence of Arg, His, Leu, Lys and Val.

To investigate the nature of the purple compound, a set of digestions was performed with proteolytic enzymes that recognize basic amino acids in proteins. The purple contaminant was treated with the endoproteases trypsin, Lys-C, and Arg-C. Cleavage was observed in all cases. All *m/z*-values listed below are $[M + H]^+$-ions. LC/MS analysis of the trypsin digest gave cleavage products with the following *m/z*-values: 387, 404 and 773. The 287 nm UV absorbance tracks with those of *m/z* 404 and 773 indicating that they contain the chromophore. CID of the cleavage product with *m/z* 387 gave fragments indicative of the sequence Val-Leu-Arg. A CID-ion at *m/z* 175 ($[H_2N\text{-}CH(COOH)\text{-}CH_2\text{-}CH_2\text{-}(H_2N\text{-}C{=}NH) + H]^+$ places Arg at the C-terminus of the tripeptide. CID of the cleavage product with *m/z* 404 indicates the presence of a C-terminal Lys (for a detailed discussion see below). The minor cleavage product with *m/z* 773 appears to be comprised of the $[M + H]^+$-ion with *m/z* 404 plus Val-Leu-Arg. The mass difference between *m/z* 773 and 1065 allows for a second Arg and for His.

The Arg-C digest analyzed by LC/MS revealed cleavage products of *m/z* 773 (see above) and 924. The mass difference between these fragments adds another Arg to the sequence chromophore-Lys-Val-Leu-Arg. Hence this C-terminal amino acid has to be His (see above). This was confirmed by the Lys-C digest. In addition to a cleavage product *m/z* 404 (see above) one with *m/z* 680 was obtained, consistant with a pentapeptide Val-Leu-Arg-Arg-His. Its CID spectrum matches that of the same peptide obtained from a Lys-C digest of the recombinant 32-amino-acid protein, hBNP (Fig. 1).

For structure elucidation of the chromophore the cleavage product with *m/z* 404 (**1**) was investigated in detail by MS and NMR. The exact mass of **1** was determined by two MS techniques: FAB gave 404.126, ESI 404.125 for the $[M + H]^+$-ion. Elemental formulas were computed for a margin of + 0.005 u for both measured values allowing for the elements CHNOS. After excluding all elemental combinations with a smaller number of C and H than accounted for by NMR and those that contained an even number of N (because of the odd-numbered molecular mass – 403 -), there remained only one combination: $C_{19}H_{22}N_3O_5S$ (404.128). The presence of 1 S is in agreement with the increased intensity of *m/z* 406 (A + 2 ion) due to the 4.2% abundance of ^{34}S, in the isotope pattern (100:24:9:1.5, calculated 100:24:8:1.5) of the molecular ion region. CID of *m/z* 404 yielded the following fragment ions: *m/z* 387 (- NH_3), 386 (- H_2O), 341 (386 – COOH), 276 (- $H_2N(CH_2)_4CHCOOH$ + 2H), 275, 259, 258 (- Lys), 232 (276 – $CONH_2$), 230 (258 – CO), 198 (230 – S), 173, 147 (Lys + H), 129 (Lys – H_2O + H). This CID experiment confirmed the presence of Lys and further suggested the likelihood of a sulfur containing compound.

Cys is the amino acid that is N-terminal to the hBNP C-terminal sequence Lys-Val-Leu-Arg-Arg-

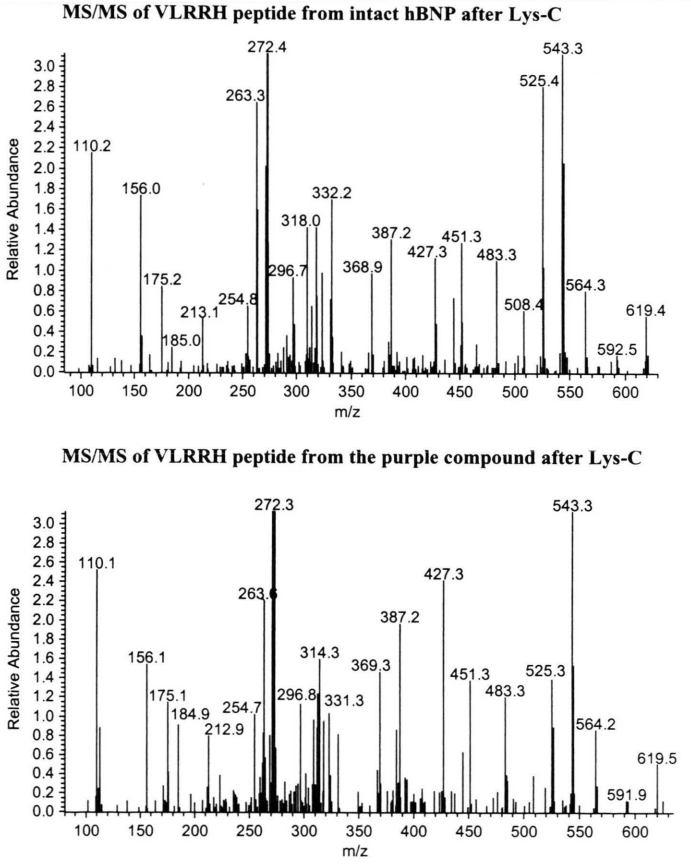
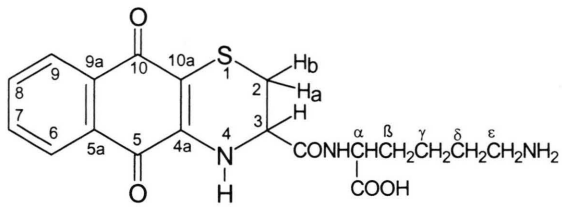


Fig. 1. Upper plot: the CID spectrum of Val-Leu-Arg-Arg-His from a partial degradation of the protein hBNP; lower plot: CID spectrum of Val-Leu-Arg-Arg-His from the Lys-C digest of **4**.

His. To see if a Cys was present, a reduction and S-carboxymethylation was performed on **4**, but no carboxymethylation product was obtained. This result indicates that if a Cys is present, the sulfur must be involved in a non-reducible covalent bond.



1

Signals for an *o*-disubstituted benzene ring, an ABX system ($CHCH_2$), and those for a Lys residue can be seen in the ^{1}H-NMR spectrum (solvent H_2O with suppression of the solvent signal by presaturation). The chemical shift values can be found in Table I. Those observed for α-CH (4.49 ppm) and ε−CH_2 (2.83 ppm) of Lys indicate an amidic connection via the α-NH_2 group (literature values ~4.4 and ~2.7 ppm as compared with ~4.0 and ~3.5 ppm for a connection via the ε-NH_2) (Budzikiewicz *et al.*, 1997; Budzikiewicz *et al.*, 1998). This is in agreement with the failure of an Edman degradation of the compound **4** that would have been expected to take place if a free α−NH_2 was present. By ^{1}H,^{13}C-HSQC (J^1 H,C-coupling) and by HMBC techniques (J^2 and higher coupling) ^{13}C-shift values (see Table I) and connectivities (see Table III) could be determined. The linkage of the α−NH_2 of Lys to the ring carboxyl group is confirmed by a cross signal of the Lys α-CH to the ring carboxyl CO group. The measurement in H_2O allowed to recognize also the NH protons (see Table I). Cross signals were observed between the Lys α-NH and the ring carboxyl CO, between the ring NH and the quaternary C-10a and the CO group at 179.8 ppm. As it cannot be differentiated between ^3J- and ^4J-coupling it is not possible to decide which of the two quinone carbonyls neighbors the NH and which one the S; also the differences in the chemical shifts do not allow a decision. The attribution of shift values for the two quinone carbonyls in Table I is therefore arbitrary; as a consequence, the benzene ring including the two carbonyl functions may be turned around with respect to the thiazoline ring. The shift values of the ring system were confirmed by comparison with those of synthetic **2** (Table II) measured in CD_3OD. Compound **2** has the structure 3-carboxy-3,4-dihydro-2*H*-naphtho[2,3-b][1,4]-thiazine-5,10-dione.

Treatment of **1** with carbopeptidase B yielded **2**. This further confirmed the presence of the C-terminal Lys in **1**. ESI MS analysis of **2** showed an $[M + H]^+$ ion at *m/z* 276. CID of *m/z* 276 gave

Table I. ^{1}H- and ^{13}C-chemical shifts of **1** in unbuffered H_2O with 10% D_2O, 278 K (δ ppm relative to [3,4,5-d_6]-2,2-dimethyl-2-silapentane-5-sulfonate – DSS – δ(^{1}H-CH_3) = 0.0 ppm, δ(^{13}C-CH_3) = −1.61 ppm).

Ring System Atom	^{1}H	^{13}C	Lys Atom	^{1}H	^{13}C
2	2.87 (a)	26.4	α-NH	8.27	
	3.47 (b)		α-CH	4.49	53.8
3	4.87	56,3	α-COOH		177.3
4	7.93		β-CH_2	1.69/1.92	31.2
4a		143.6	γ-CH_2	1.27	23.4
5		179.6*	δ-CH_2	1.50/1.57	27.3
5a		131.3	ε-CH_2	2.83	40.7
6	7.89	128.1	ε-NH_2	7.57	
7	7.65	135.0			
8	7.70	136.5			
9	7.80	127.1			
9a		133.2			
10		182.6			
10a		111.8			
3-CO		173.6			

* The shift assignments of the quinone system (positions 5 to 10) could be reversed (CO-5 182.8 ppm etc; see Text).

Table II. ^{1}H- and ^{13}C-chemical shifts of synthetic **2**.

Solvent Atom[3)]	CD_3OD[1)] ^{1}H	^{13}C	CD_3CN with 2 drops CF_3COOH[2)] ^{1}H	^{13}C
2	2.87 (a)	27.9	3.09	26.1
	3.31 (b)		3.20	
3	4.09	57.2	4.62	53.5
4			6.68	
4a				141.7
5				178.5
5a				131.4
6	8.00	127.3	8.00	126.9
7	7.63	133.8	7.67	133.8
8	7.69	135.7	7.72	135.4
9	7.95	133.8	7.95	126.4
9a				133.8
10				180.0
10a				111.9
3-CO				170.9

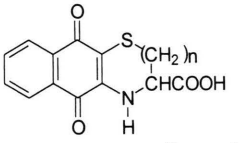
[1)] δ ppm relative to CD_3OD, δ(C^1HD_2) = 3.30 ppm, δ($^{13}CD_3$) = 49.0 ppm; [2)] δ ppm relative to CD_3CN, δ(C^1HD_2) = 1.93 ppm, δ(^{13}CN) = 118.2 ppm; [3)] The assignments were confirmed by ^{1}H, ^{1}H-COSY and by ^{1}H, ^{13}C-HMBC measurements.

Table III. ^{1}H, ^{1}H- and ^{1}H, ^{13}C-connectivities in the NMR-spectra of **1**.

^{1}H, ^{1}H-COSY
H-2a ↔ H-2b, H-2a ↔ H-3, H-2b ↔ H-3
H-6 ↔ H-7, H-7 ↔ H-8, H-8 ↔ H-9
H-α ↔ H-β, H-β ↔ H-γ, H-γ ↔ H-δ, H–δ ↔ H-ε
^{1}H, ^{13}C-HMBC
H-2a → 3-CO
H-2b → C-10a
H-3 → 3-CO, C-4a
H-4 → C-10a, → C-5 or C-10
H-6 → C-8, → C-9a, → C-5
H-7 → C-5a, → C-9a
H-8 → C-6, → C-9a
H-9 → C-10
α–NH → 3-CO
H-α → C-ß, → C-γ, → 3-CO, → α-COOH
H-β→ C-α, → C-γ, → C-δ, → α-COOH
H-γ → C-α, → C-ß, → C-δ, → C-ε
H-δ → C-ß, → C-γ, → C-ε
H-ε → C-ß, → C-γ

an ion *m/z* 230 ([M – COOH]$^+$ which in turn (MS^3) could be fragmented to *m/z* 202 (- CO), 197 (- SH), 186 and 105 ($C_6H_5CO^+$).

Compound **2** is mentioned in the literature (Eguchi *et al.*, 1977; Eguchi and Yokokawa., 1984) as a synthetic product obtained by reaction of 2,3-dibromo (or dichloro)-1,4-naphthoquinone with Cys, but no structural data are given. The ^{1}H- and ^{13}C-NMR shift values of synthetic **2** are assembled in Table II. The signals in the aliphatic portion are strongly solvent and pH dependant. Comparison of the data from *E. coli* derived and synthetic **2** in CD_3CN confirmed the identity of the two compounds.

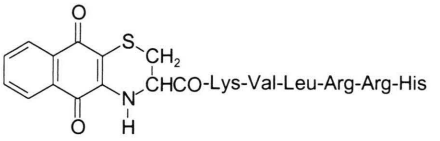


2: n = 1
3: n = 2

A higher homolog of **2** (**3**, derived from homocysteine) had been isolated from *Brevibacterium flavum* (*fluvum* in Euguchi and Yokokawa, 1984, is apparently a misprint), *B. lactofermentum*, *Corynebacterium acetoacidophilum* and *glutamicum* (Euguchi *et al.*, 1977). **2** is obviously an addition product of cysteine to naphthoquinone (both amines and thiols undergo 1,4-addition to the α,β-unsaturated carbonyl system (Finley, 1974). This explains why the reaction for a free SH-group (carboxymethylation) was negative. Kuhn and Hammer (1951) studied the reaction of 1,4-naphthoquinone with Cys ethyl ester. The ochre-colored main product was the result of an addition of the SH-group to one of the α,β-unsaturated carbonyl systems, re-oxidation of the resulting naphthoquinol to the quinone, and condensation of the amino group of Cys with the neighboring carbonyl group giving a 1,3-thiazine ring. As a side product black crystals were obtained, but not characterized. They could have been **2**-ethyl ester.

Conclusions

The structure **4** can be assigned to the purple compound with $[M + H]^+$ *m/z* 1065.



4

In this structure the carboxyl group of **2** forms a peptide bond with the α-amino terminus of Lys with the resulting loss of H_2O. Including Cys incorporated into **2**, it contains the seven C-terminal amino acids of the 32 amino acid hBNP (Seilhamer *et al.*, 1989; Seilhamer *et al.*, 1997). 1,4-Naphthoquinone is an intermediate in the bacterial biosynthesis of menaquinones (Weiss and Edwards, 1980; Nuhn, 1997) and therefore may be the direct precursor of **2**. Since menaquinones are typical secondary metabolites of bacteria it would be interesting to determine (1) whether the formation of derivatives of **2** is a general phenomenon when producing proteins with Cys at the N-terminus in some biogenetic intermediate and (2) what is the actual precursor of **2** and whether it is abundant in the inclusion bodies from which recombinant proteins are often recovered. Further research into the formation of this compound and possibly other similar compounds is planned.

Acknowledgement

The authors would like to thank Armin Ramel, and Peter Adriaenssens for fruitful discussions throughout the course of the investigation. We wish to specially acknowledge Steve Pollitt for offering the menaquinone hypothesis, and Olaf Kinzel for the synthsis of **2**.

Budzikiewicz H. (1997), Siderophores from fluorescent pseudomonads. Z. Naturforsch. **52c**, 713–720.

Budzikiewicz H., Kilz S., Taraz K. and Meyer J.-M. (1997), Identical pyoverdins from *Pseudomonas fluorescens* 9AW and from *Pseudomonas putida* 9BW. Z. Naturforsch. 1997, **52c**, 721–728.

Budzikiewicz H., Fuchs R., Taraz K., Marek-Kozaczuk M. and Skorupska, A. (1998), Dihydropyoverdin-7-sulfonic acids, unusual bacterial metabolites. Nat. Prod. Letters 1998, **12**, 125–130.

Crestfield A. M., Moore S. and Stein W. H. (1963), The preparation and enzymatic hydrolysis of reduced and S-carboxymethylated proteins. J. Biol. Chem. **238**, 622–627.

Eguchi Ch., Suzuki Y., Onuma T., Yokokawa Y., Hirata T. and Tosaka, O. (1977), Naphthoquinone amino acid derivatives. Japanese Patent 52–84 220.

Eguchi Ch. and Yokokawa, Y. (1984), New purple pigment produced by *Brevibacterium fluvum*. Heterocycl. **21**, 470.

Finley K. T. (1974), The addition and substitution chemistry of quinones. In: The Chemistry of Quinoid Compounds (S. Patai, ed.), chapter 17, Wiley, London.

Kuhn R. and Hammer I. (1951), Umsetzungsprodukte von Chinonen und anderen Carbonylverbindungen mit Cystein. Chem. Ber. **84**, 91–95.

Nuhn P. (1997), Naturstoffchemie, 3. Aufl. Hirzel, Stuttgart p. 390.

Ovchinnikov Yu. A., Abdulaev N. G., Feigina M. Yu., Kiselev A. V., Lobanov N. A. and Nazimov I. V. (1978), Amino acid sequence of bacteriorhodopsin, Bioorg. Khim. **4**, 1573–1574.

Seilhamer J. J., Arfsten A., Miller J. A., Lundquist P., Scarborough R. M., Lewicki J. A. and Porter J. G. (1989), Human and canine gene homologs of porcine brain natriuretic peptide. Biochem. Biophys. Res. Commun. **165**, 650–658.

Seilhamer J. J., Lewicki J., Scarborough, R. M. and Porter J. G. (1997), Recombinant techniques for production of human brain natriuretic peptide. US Patent 5,674,710.

Weiss U. and Edwards, J. M. (1980), The Biosynthesis of Aromatic Compounds, chapter 13, Wiley, New York.