THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 256, No. 9, Issue of May 10, pp. 4350-4356, 1981
Printed in U.S.A.

The Amino Acid Sequence of the D-Galactose-binding Protein from *Escherichia coli* B/r*

(Received for publication, December 18, 1980)

Walter C. Mahoney‡§, Robert W. Hogg¶‖, and Mark A. Hermodson‡§

From the ‡Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907 and the ¶Department of Microbiology, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106

The complete primary structure of the *Escherichia coli* B/r galactose-binding protein was determined by the automated sequencing of fragments produced by cleavage with cyanogen bromide, *o*-iodosobenzoic acid, limited trypsin digestion, mild acid hydrolysis, and *Staphylococcus aureus* strain V8 protease. The protein, which has 309 amino acids, is notable in the extent to which it differs from the L-arabinose-binding protein. Comparison of these two proteins indicates only about 18% homology despite the close structural resemblence of the molecules which they bind. The galactose-binding protein is the chemoreceptor initiating chemotaxis toward galactose, and it thus becomes the first protein component required for chemotaxis for which the primary structure is known.

The initial step leading to receptor-mediated high affinity solute transport and chemotaxis toward attractants in gram-negative bacteria is the recognition and binding of these solutes by specific proteins (see Refs. 1, 2, 3 for review). These proteins, called binding proteins or chemotaxis receptors (for those having that function), are water-soluble and can be released from the bacterium by osmotic shock or by treatment with lysozyme/EDTA. Thus these proteins appear to be dissolved in the periplasmic fluid or loosely associated with the membrane or peptidoglycan environment (1, 4). In the several systems studied to date, nutrient transport is accomplished by the interaction of the binding protein with nutrient bound with two membrane proteins in an energy-dependent process (4-6, 8). This has been demonstrated most clearly by Lo (5) by isolating the binding protein and the two membrane proteins involved in the transport of dicarboxylic acids in *Escherichia coli* K12 and reconstituting binding protein-dependent transport in mouse L-cells and rat myoblasts driven by a valinomycin-induced chemical gradient.

For those chemicals toward which the bacteria can chemotact, the interaction of the chemoreceptor-nutrient adduct with a different set of cytoplasmic membrane proteins ultimately leads to the directed movement of the organism. These latter membrane proteins, called methyl-accepting chemotaxis proteins, are composed of a single subunit of M_r = 56,000 to 65,000 and have been divided into three complementation groups (2). MCP-I[1] is encoded by the *tsr* gene, while MCP-II and MCP-III are the products of the *tar- and trg*-genes, respectively. While little is known about the flow of information from receptor to flagellum, including the number of steps involved, it has been inferred that the maltose-binding protein (in the presence of maltose) initiates the flow of signals through direct physical contact with MCP-II (9), and that the galactose-binding protein and ribose-binding protein which utilize MCP-III (3) employ similar mechanisms.

In order to characterize these protein-protein interactions at the molecular level, it is necessary to determine the chemical structures of the components. We present here the complete amino acid sequence of the *E. coli* B/r galactose-binding protein. Suitable crystals for high resolution x-ray crystallographic analysis of this protein are also available, and this analysis is proceeding in another laboratory (10).

METHODS AND RESULTS[2]

DISCUSSION

The primary structure of the *E. coli* B/r galactose-binding protein (Fig. 1) is the fifth complete binding protein sequence determined. The others are the *E. coli* B/r L-arabinose-binding protein (11), the *E. coli* K12 leucine, isoleucine, valine-binding protein (12), the *Salmonella typhimurium* LT2 sulfate-binding protein (13), and the *S. typhimurium* TA 1014 histidine-binding protein (14). X-ray crystallographic analyses are available for the arabinose-binding protein (15) and the galactose-binding protein (10) at a resolution of 2.4 Å and 4.1 Å, respectively. Preliminary crystallographic data are also available for the leucine, isoleucine, valine-binding protein (16), sulfate-binding protein and the maltose-binding protein (17). Of these proteins only the galactose-binding protein and the maltose-binding protein are chemotaxis receptors in addition to being membrane transport receptors. Thus, the galactose-binding protein is the first bacterial chemotaxis receptor for which the primary structure has been determined.

The sequence analysis of the galactose-binding protein was complicated due to the nonrandom distribution of methionine and tryptophan residues. Cleavage of the whole protein with cyanogen bromide produced seven peptides, two of which accounted for 77% of the total molecule. Cleavage at trypto-

* This is journal paper No. 8458 from the Purdue University Agricultural Experiment Station. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Recipient of United States Health Service Grant GM 24602.

‖ Recipient of United States Health Service Grant AM-13791.

[1] The abbreviations used are: MCP, methyl-accepting chemotaxis proteins; Pth, phenylthiohydantoin.

[2] Portions of this paper (including "Methods," "Results," Table II, Figs. 3 to 10, and additional references) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20014. Request Document No. 80M-2663, cite authors, and include a check for $6.40 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

phan with *o*-iodosobenzoic acid produced six peptides, three of which contained less than 25 residues. Five of the the six yielded information obtained from other fragments. Only one tryptophanyl peptide, W-3, obtained by a subdigest of the largest cyanogen bromide fragment (CB-3), was critical for the sequence determination.

The strategy used in the sequence analysis of the galactose-binding protein is outlined in Fig. 2. Heavy emphasis was placed on the isolation of peptides produced by subdigestion of the two large cyanogen bromide fragments. In many cases peptide purification was possible only by employing new high performance liquid chromatography methods for peptide isolation (18, 19). For example, the subdigestion of CB-3 at glutamyl residues with *S. aureus* strain V8 protease produced a single peptide peak on gel filtration chromatography which sequence analysis showed to be three peptides starting at residues Asn 39, Lys 79, and Ser 115. Reversed-phase high performance liquid chromatography cleanly separated the desired peptide beginning with Ser 115 from the other two which co-eluted (Fig. 8). Peptide E-3 was recovered in 90% yield. This technique yielded pure peptides quickly and reduced the total amount of protein required for the analysis, since low yield procedures, such as ion-exchange chromatography, were not necessary. The value of both 100 Å pore and 300 Å pore supports (19) for small and large peptides, respectively, was confirmed during the sequencing of the galactose-binding protein.

	5	10	15
1	Ala Asp Thr Arg Ile Gly Val Thr Ile Tyr Lys Tyr Asp Asp Asn		
16	Phe Met Ser Val Val Arg Lys Ala Ile Glu Gln Asp Ala Lys Ala		
31	Ala Pro Asp Val Gln Leu Leu Met Asn Asp Ser Gln Asn Asp Gln		
46	Ser Lys Gln Asn Asp Gln Ile Asp Val Leu Leu Ala Lys Gly Val		
61	Lys Ala Leu Ala Ile Asn Leu Val Asp Pro Ala Ala Ala Gly Thr		
76	Val Ile Glu Lys Ala Arg Gly Gln Asn Val Pro Val Val Phe Phe		
91	Asn Lys Glu Pro Ser Arg Lys Ala Leu Asp Ser Tyr Asp Lys Ala		
106	Tyr Tyr Val Gly Thr Asp Ser Lys Glu Ser Gly Ile Ile Gln Gly		
121	Asp Leu Ile Ala Lys His Trp Ala Ala Asn Gln Gly Trp Asp Leu		
136	Asn Lys Asp Gly Gln Ile Gln Phe Val Leu Leu Lys Gly Glu Pro		
151	Gly His Pro Asp Ala Lys Glu Arg Thr Thr Tyr Val Ile Lys Glu		
166	Leu Asn Asp Lys Gly Ile Lys Thr Glu Gln Leu Gln Leu Asp Thr		
181	Ala Met Trp Asp Thr Ala Gln Ala Lys Asp Lys Met Asp Ala Trp		
196	Leu Ser Gly Pro Asn Ala Asn Lys Ile Glu Val Val Ile Ala Asn		
211	Asn Asp Ala Met Ala Met Gly Ala Val Glu Ala Leu Lys Ala His		
226	Asn Lys Ser Ser Ile Pro Val Phe Gly Val Asp Ala Leu Pro Glu		
241	Ala Leu Ala Leu Val Lys Ser Gly Ala Leu Ala Gly Thr Val Leu		
256	Asn Asp Ala Asn Asn Gln Ala Lys Ala Thr Phe Asp Leu Ala Lys		
271	Asn Leu Ala Asp Gly Lys Gly Ala Ala Asp Gly Thr Asn Trp Lys		
286	Ile Asp Asn Lys Val Val Arg Val Pro Tyr Val Gly Val Asp Lys		
301	Asp Asn Leu Ala Glu Phe Ser Lys Lys		
		309	

FIG. 1. **The amino acid sequence of the *E. coli* B/r galactose-binding protein.**

The amino acid compositions of the galactose-binding protein and the various fragments used in the sequence analysis were determined (Table I). Reasonable agreement between the compositions and the sequence analysis was obtained for each.

Inspection of the amino acid sequence of the galactose-binding protein reveals no cysteine or cystine. The protein is rich in asparagine and aspartic acid residues, particularly in the COOH-terminal 100 residues of the molecule. Comparison of the amino acid sequence of the galactose-binding protein with the only other sugar-binding protein of known sequence, the arabinose-binding protein, reveals a striking lack of homology (approximately 18%). This is particularly noteworthy considering the structural similarity of the sugars galactose and arabinose. A detailed structural comparison of these binding proteins is presented in the accompanying paper.

The molecular details of how these binding proteins function remain to be solved, although it has been suggested that the event that triggers a binding protein to interact with the next protein(s) in the system is a conformational change brought about by the binding of the nutrient molecule (3). By

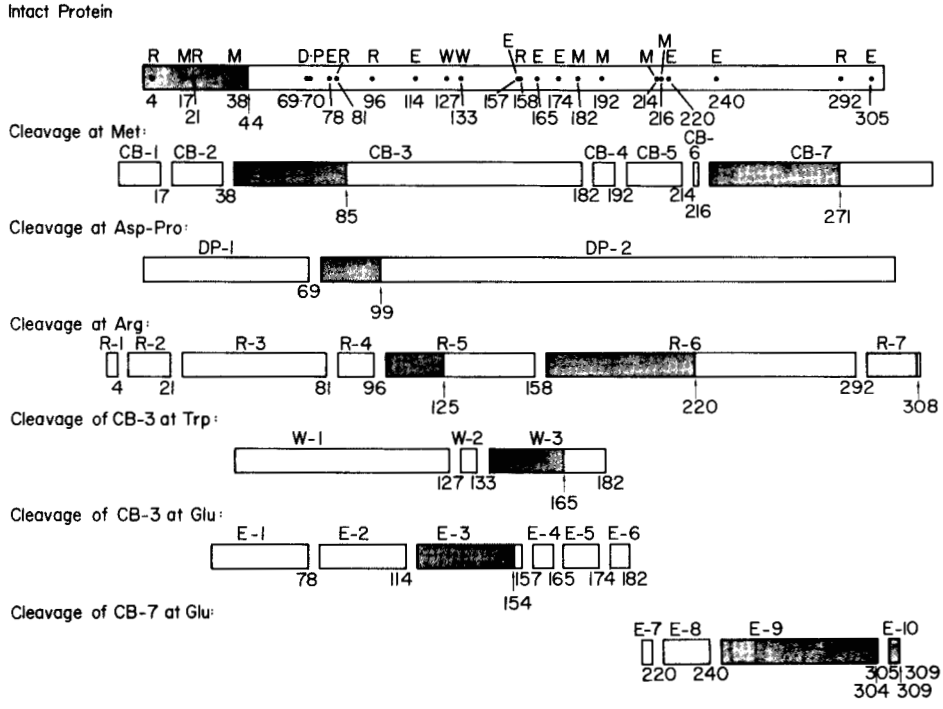

FIG. 2. **Summary of the sequencing strategy.** *Stippling* denotes those portions of the intact galactose-binding protein and selected fragments that were degraded. Standard one-letter abbreviations are used to designate residues of importance to the production of appropriate fragments.

TABLE I

Amino acid compositions of the galactose-binding protein and selected fragments[a]

Amino Acid	Residues	Whole Protein[b]	R-1	CB-1	R-2	R-3	E-1	CB-3	E-2	R-4	R-5	R-6	CB-4	CB-5	CB-7	T3	E-9	T4	DL-1	W-6	R-7
	yield(%)		1-4	1-17	5-21	22-81	39-78	39-182	79-114	82-96	97-158	159-292	183-192	193-214	217-309	228-246	241-305	247-263	268-305	285-309	293-309
			34%	90%	61%	48%	71%	95%	42%	31%	52%	40%	34%	72%	79%	12%	27%	18%	70%	85%	60%
Asp		52.6 (53)	1.0(1)	3.8(4)	3.1(3)	10.8(11)	8.8(9)	22.4(23)	5.0(5)	2.1(2)	9.0(9)	23.6(24)	2.0(2)	6.0(6)	16.1(16)	1.3(1)	14.0(14)	3.7(4)	8.6(9)	4.9(5)	2.8(3)
Thr		12[c] (12)	0.9(1)	1.9(2)	1.0(1)	1.0(1)	1.0(1)	5.5(6)	1.0(1)		0.9(1)	7.5(8)	1.3(1)		2.8(3)		2.8(3)	0.9(1)	0.5(1)		
Ser		12[c] (12)			0.8(1)	1.8(2)	1.9(2)	4.4(6)	2.6(3)	0.9(1)	3.0(3)	4.1(4)		1.1(1)	2.9(4)	2.0(2)	1.0(1)	0.9(1)	0.4	1.2(1)	0.9(1)
Glu		26.5 (27)		0.2		8.0(8)	5.1(5)	18.4(18)	3.0(3)	2.0(2)	7.3(7)	9.0(9)	1.3(1)	1.3(1)	5.7(4)	1.3(1)	2.0(2)	1.0(1)	1.6(1)	1.3(1)	1.1(1)
Pro		10.4 (10)				2.0(2)	1.2(1)	5.7(5)	2.1(2)	2.0(2)	2.5(2)	3.9(3)		1.0(1)	3.5(3)	1.8(2)	1.1(1)		1.4(1)	1.0(1)	0.8(1)
Gly		20.7 (21)		0.8(1)	0.9(1)	2.2(2)	2.2(2)	11.0(11)	2.0(2)	1.1(1)	6.7(7)	8.9(9)		1.1(1)	8.1(8)	1.4(1)	5.9(6)	1.9(2)	4.3(4)	0.9(1)	1.1(1)
Ala		42.8 (42)	0.9(1)	0.9(1)		10.8(11)	6.1(6)	14.9(14)	3.1(3)		5.5(6)	23.5(23)	2.1(2)	3.8(4)	17.3(16)	3.2(3)	11.6(12)	4.1(4)	5.4(5)	1.0(1)	1.1(1)
Val[d]		25.7[e] (26)		0.9(1)	2.5(3)	4.7(5)	3.8(4)	9.0(10)	3.7(4)	2.7(3)	1.9(2)	8.9(10)	0.4	1.8(2)	10.3(10)	3.0(3)	6.9(7)	1.0(1)	4.8(5)	4.6(5)	2.6(3)
Met		5.8 (6)		0.7[f] (1)	0.9(1)	0.5(1)		0.4[f] (1)				3.5(4)	1.0[f] (1)	0.8[f] (1)							
Ile[d]		16.3[e] (17)		2.0(2)	1.8(2)	3.6(4)	2.7(3)	8.2(9)			4.2(4)	5.0(6)		1.7(2)	2.4(2)	1.0(1)	1.0(1)		1.1(1)	1.2(1)	0.1
Leu		22.9 (23)				6.0(6)	3.9(4)	12.6(12)	1.0(1)		5.0(5)	11.5(12)		0.9(1)	10.7(9)	2.8(3)	6.9(7)	1.4(2)	3.2(3)	1.0(1)	1.0(1)
Tyr		7.2 (7)		1.8(2)	1.8(2)			3.6(4)	2.0(3)		2.6(3)	0.6(1)			0.6(1)		0.8(1)		0.6(1)	0.7(1)	0.5(1)
Phe		7.0 (7)		1.0(1)	1.0(1)			3.2(3)	2.0(2)	2.1(2)	1.6(1)	2.0(2)			3.2(3)	1.1(1)	1.0(1)			1.1(1)	1.0(1)
His		3.1 (3)						1.8(2)			1.4(2)	0.9(1)			0.9(1)						
Lys		32.5 (32)		1.1(1)	0.7(1)	5.9(6)	3.1(3)	14.3(15)	5.1(5)	1.1(1)	7.0(7)	14.0(14)	1.7(2)	1.0(1)	10.4(11)	1.0(1)	7.2(7)	0.9(1)	5.0(5)	4.9(5)	3.0(3)
Arg		6.0 (6)	0.9(1)	1.0(1)	1.0(1)	1.0(1)		2.8(3)	2.0(2)	1.0(1)	1.3(1)	0.8(1)			1.0(1)		1.0(1)		1.0(1)	0.9(1)	
Trp		4.5[g] (5)						ND[h] (2)			ND(2)	ND(3)	ND(1)	ND(1)	ND(1)		ND(1)		ND(1)		

[a] Values in parentheses indicate compositions as determined from the sequence.

[b] Averages of six separate 24-, 48-, 72-, and 96-h hydrolysates, except where otherwise noted.

[c] Determined by extrapolation to zero-time of hydrolysis.

[d] Low yields of valine and isoleucine attributable to the incomplete hydrolysis of the following bonds: Val_{19}-Val_{20}, Val_{76}-Ile_{77}, Val_{88}-Val_{89}, Ile_{117}-Ile_{118}, Val_{162}-Ile_{163}, Val_{206}-Val_{207}-Ile_{208}, and Val_{290}-Val_{291}.

[e] Averages of six 96-h hydrolysates.

[f] Detected as homoserine and homoserine lactone.

[g] Tryptophan was determined by base hydrolysis and was calculated as a ratio to histidine.

[h] ND, not determined.

radius of gyration measurements, an 18 Å movement of the two domains of the arabinose-binding protein has been found upon binding arabinose (20). Using fluorescent measurements (probably based upon an incorrect estimate of the number of tryptophan residues in the *S. typhimurium* galactose-binding protein), a movement of 30 Å upon binding galactose has been inferred (21). Clearly, more work needs to be done using direct methods (*e.g.* x-ray crystallography) before a clear understanding of these molecular events is gained.

REFERENCES

1 Adler, J. (1975) *Annu. Rev. Biochem.* **44,** 341–356
2. Springer, M. S., Goy, M. F., and Adler, J. (1979) *Nature* **280,** 279–284
3. Koshland, D. E., Jr. (1977) *Science* **196,** 1055–1063
4. Lo, T. C. Y. (1979) *Can. J. Biochem.* **57,** 289–301
5. Lo, T. C. Y. (1979) *J. Biol. Chem.* **254,** 591–594
6. Bavoil, P., Hofnung, M., and Nikaido, H. (1980) *J. Biol. Chem.* **255,** 8366–8369
7. Ames, G. F.-L., and Nikaido, K. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 5447–5451
8. Oxender, D. L., Anderson, J. J., Daniels, C. J., Landick, R., Gunsalus, R. P., Zurawski, G., Selker, E., and Yanofsky, C. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 1412–1416
9. Koiwai, O., and Hayashi, H. (1979) *J. Biochem. (Tokyo)* **86,** 27–34
10. Quiocho, F. A., and Pflugrath, J. W. (1980) *J. Biol. Chem.* **255,** 6559–6561
11. Hogg, R. W., and Hermodson, M. A. (1977) *J. Biol. Chem.* **252,** 5135–5141
12. Ovchinnikov, Yu. A., Aldanova, N. A., Grinkevich, V. A., Arzamazova, N. M., and Moroz, I. N. (1977) *FEBS Lett.* **78,** 313–316
13. Isihara, H., and Hogg, R. W. (1980) *J. Biol. Chem.* **255,** 4614–4618
14. Hogg, R. W. (1981) *J. Biol. Chem.* **256,** 1935–1939
15. Gilliland, G. L., and Quiocho, F. A. (1981) *J. Mol. Biol.*, in press
16. Meador, W. E., and Quiocho, F. A. (1978) *J. Mol. Biol.* **123,** 499–502
17. Quiocho, F. A., Meador, W. E., and Pflugrath, J. W. (1979) *J. Mol. Biol.* **133,** 181–184
18 Mahoney, W. C., and Hermodson, M. A. (1980) *J. Biol. Chem.* **255,** 11199–11203
19. Pearson, J. D., Mahoney, W. C., Hermodson, M. A., and Regnier, F. E. (1981) *J. Chromatogr* **207,** 325–332
20. Quiocho, F. A., Miller, D. M., Newcomer, M. E., Gilliland, G. L., and Pflugrath, J. W. (1980) *J. Supramol. Struct.* Suppl. 4, 76
21. Zukin, R. S., Hartig, P. R., and Koshland, D. E., Jr. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 1932–1936

Additional references are found on p. 4354.

Supplementary Material To
The Amino Acid Sequence of the D-Galactose-
Binding Protein from Escherichia coli B/r

Walter C. Mahoney, Robert W. Hogg, and Mark A. Hermodson

METHODS

Galactose Binding Protein Purification. The galactose-binding protein was purified from E. coli B/r (A-39) as described by Parsons and Hogg (22). Briefly, the E. coli cell paste (300-500 g) was resuspended in 0.1 M potassium phosphate, pH 7, (3 ml/gm cells) and homogenized in a Manton-Gaulin cell disrupter operating at 8000 psi. The mixture was treated for a short time with ribonuclease and deoxyribonuclease to reduce the viscosity and brought to 55% of saturation in ammonium sulfate. After stirring overnight the supernatant was isolated by centrifugation and dialyzed against several changes of 10 mM potassium phosphate (pH 7.8). It was then passed through a matrix of DEAE-cellulose (DE-52, Whatman) in a Buchler funnel equilibrated and eluted in the same buffer. The unretained material was dialyzed against several changes of 1 mM potassium phosphate (pH 7.8) and applied to a column (4.0 x 45-cm) of DEAE-cellulose (DE-53, Whatman) equilibrated with the same buffer. A linear 4 ℓ gradient from 0 to 30 mM KCl in that buffer yielded galactose-binding protein contaminated by only a slight amount of arabinose-binding protein and two other unidentified components. The galactose/glucose binding activity was collected, concentrated by vacuum dialysis, and applied to a column (2.5 x 90-cm) of Sephacryl S-200 developed with 50 mM imidazole, pH 6.9. This gave pure galactose binding protein as judged by disc gel electrophoresis (23), by SDS gel electrophoresis (24), and by sequence analysis. A typical yield of the galactose-binding protein from 500 g of E. coli wet cell paste was 95 mg.

Galactose Binding Protein Assay. Galactose-binding protein was assayed by equilibrium dialysis or the nitrocellulose filter assay (25), using D-[1-^{14}C] glucose and D-[1-^{14}C] galactose or D-[1-^{3}H] galactose (Amersham or New England Nuclear). The galactose-binding protein was also assayed by immunodiffusion (26).

Chemical and Enzymatic Cleavage Procedures. Galactose-binding protein was subjected to cyanogen bromide digestion (27), to acid hydrolysis at the aspartyl-prolyl bond (27,28), and to tryptic digestion at the arginyl residues of the citraconic anhydride-modified protein (29) as previously described. Digestion of peptides CB3 and CR7 at tryptophanyl residues was performed with o-iodosobenzoic acid (30) using the synthesis and p-cresol incubation methods of Mahoney, et al. (31). Peptides CB3 and CB7 were also digested with S. aureus V8 protease at pH 4.0 (32), and CB7 was subjected to total tryptic hydrolysis at pH 8.0 to generate peptides T3 and T4 (Table I) for compositional analysis.

Reversed-Phase High-Performance Liquid Chromatography. Many peptides were purified by reversed-phase high-performance liquid chromatography (18, 19). Lyophilized peptide mixtures were dissolved in 0.1% trifluoroacetic acid (1-2 mg/ml) and chromatographed on SynChropak RP-P columns (SynChrom, Inc., Linden, IN) or Brownlee LiChrosorb RP-8 columns in propanol or acetonitrile gradients. Both the water starting solvent and the organic limiting solvent contained 0.1% (v/v) trifluoroacetic acid (18).

Amino Acid Analysis. Whole protein and peptides were hydrolyzed in glass-distilled, 6 N HCl in evacuated sealed tubes at 110° C for 24 h (peptides) or for 24, 48, 72, and 96 h (whole protein). All analyses were performed on a Durrum D-500 amino acid analyzer according to the manufacture's instructions. Tryptophan content was determined from the Trp/His ratio following base hydrolysis in the presence of starch (33).

Sequence Analysis. The galactose-binding protein and selected fragments were degraded in a Beckman Model 890C sequencer according to Edman and Begg (34) as modified by Hermodson, et al. (35,36), and Nute, et al. (37). Pth-amino acids were identified and quantitated by high-performance liquid chromatography using a modification (38) of the method of Zimmerman, et al. (39). Pth-histidine and Pth-arginine were sometimes identified by spot tests (27). Each peptide yielded at least 70% of the expected amount of Pth-amino acid in the early cycles of the degradation, and only pure (>95%) peptides were degraded. When the ratio of peak to background fell below 3, no further identifications were made.

Designation of Peptides. Peptides (except those generated by subdigests) were numbered consecutively from those containing the amino-terminal sequence to those containing the carboxyl sequence of the galactose binding protein. Thus seven fragments, CB-1 (residues 1-17), CB-2 (18-38), CB-3 (39-182), CB-4 (183-192), CB-5 (193-214), CB-6 (215-216), and CB-7 (217-309), were produced by cleavage with cyanogen bromide. Cleavage at Asp-Pro produced fragments DP-1 (1-69) and DP-2 (70-309), while cleavage at Arg was expected to generate seven fragments, R-1 (1-4), R-2 (5-21), R-3 (22-81), R-4 (82-96), R-5 (97-158), R-6 (159-292), and R-7 (293-309). Designation of peptides generated by subdigest were numerically arranged in order of their position in the final sequence. For example, the expected fragments of CB3 digested at glutamyl residues were designated E-1 (39-78), E-2 (79-114), E-3 (115-157), E-4 (158-165), E-5 (166-174), and E-6 (175-182).

Identification of the Carboxyl-Terminal Residue. Carboxyl-terminal analyses were performed using carboxypeptidase B. CB7 (6 nmole) was dissolved in 2 ml of 5 mM NH_4HCO_3 (pH 8.0) containing 6 nmoles of norleucine. Digestion at 25° was initiated by the addition of 10 units of carboxypeptidase B. After 4 h the reaction was terminated by addition of 0.2 ml saturated trichloroacetic acid in water. The solution was clarified by centrifugation, and the supernatant was extracted three times with ether, evaporated, and subjected to amino acid analysis.

RESULTS

The complete amino acid sequence of the galactose-binding protein from E. coli B/r appears in Figure 1. Overlaps of partial sequences ranged from 2 to more than 10 residues and allowed the unambiguous placement of all fragments. In addition the close correspondence of amino acid compositions of the entire protein and of the fragments with compositions determined from the sequence indicates that this sequence is complete.

Isolation and Identification of Fragments. The elution profiles produced by gel filtration of CB, DP, and R fragments generated from the intact galactose-binding protein are presented in Figures 3, 4, and 5, respectively. Fragment R-7 was further purified by high-performance liquid chromatography (Figure 6). Fragment W-3, from a subdigest of CB-3 at tryptophan, was purified by high-performance liquid chromatography (Figure 7) as was E-3, produced by cleavage of CB-3 at Glu (Figure 8). Fragments E-9 and E-10, produced by cleavage of CB-7 at Glu, were isolated by gel filtration and high-performance liquid-chromatography, respectively (Figures 9 and 10). While these peptides provided the overlapping fragments required for the determination of the complete sequence of the galactose-binding protein, additional peptides were generated by the cleavage of CR 7 at tryptophanyl residues with o-iodosobenzoic acid and at lysyl and arginyl residues with trypsin in order to get accurate amino acid compositions through the CR7 sequence (Table I). Those peptides were purified by high performance liquid chromatography (not shown). Amino acid compositions of peptides overlapping the entire sequence of the galactose-binding protein are provided in Table 1.

Sequence Analysis. The strategy employed in analyzing the primary structure of the galactose binding protein is outlined in Figure 2. The residues occupying positions 1 through 44 were identified by the stepwise degradation of galactose-binding protein. Degradation of fragments CB-3 and CB-7 established the sequences of residues 39 through 84 and 217 through 271. The isolation and partial sequence of fragment DP-2 identified residues 70 through 98. Subsequent analyses of fragments R-5, R-6, and R-7 established the sequences of residues 97 through 124, 159 through 219, and 293 through 308. Sequence analysis of fragments W-3 and E-3, placed residues 134 through 162 and 115 through 155, respectively. Finally, the stepwise degradation of E-9 and E-10, which positioned residues 241 through 304 and residues 306 through 309, completed the primary structure of the galactose-binding protein. Carboxypeptidase B digest confirmed the presence of lysine as the penultimate and C-terminal residue by yielding 1.7 nmole lysine/nmole protein. Only traces of other amino acids could be discerned. A summary of the yields of the Pth-amino acids at each step of the degradations and the methods used in identifying the products generated by each cycle of the sequencer is presented in Table II.

REFERENCES

22. Parsons, R. G., and Hogg, R. W. (1974) J. Biol. Chem. 249: 3602-3607.
23. Davis, B. J. (1964) Ann. New York Acad. Sci. 121: 404-427.
24. Weber, K., and Osborn, M. (1969) J. Biol. Chem. 244: 4406-4412.
25. Lever, J. E. (1972) Anal. Biochem. 50: 73-83.
26. Mahoney, W. C. (1980) Biochem. Biophys. Res. Comm. 96: 1123-1127.
27. Mahoney, W. C., and Nute, P. E. (1980) Biochemistry 19: 4436-4442.
28. Jauregui-Adell, J. and Marti, J. (1975) Anal. Biochem. 69: 468-473.
29. Mahoney, W. C., and Nute, P. E. (1979) Arch. Biochem. Biophys. 196: 64-72.
30. Mahoney, W. C., and Hermodson, M. A. (1979) Biochemistry 18: 3810-3814.
31. Mahoney, W. C., Smith, P. K., and Hermodson, M. A. (1981) Biochemistry 20: (in press).
32. Houmard, J., and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. U.S.A. 69: 3506-3509.
33. Hugli, T. E., and Moore, S. (1972) J. Biol. Chem. 247: 2828-2834.
34. Edman, P., and Begg, G. (1967) Eur. J. Biochem. 1: 80-90.
35. Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972) Biochemistry 11: 4493-4502.
36. Hermodson, M. A., Schmer, G., and Kurachi, K. (1977) J. Biol. Chem. 252: 6276-6279.
37. Nute, P. E., Mahoney, W. C., and Boosman, A. (1981) Arch. Biochem. Biophys. (in press).
38. Nute, P. E., and Mahoney, W. C. (1980) Hemoglobin 4: 109-123.
39. Zimmerman, C. L., Appella, E., and Pisano, J. J. (1977) Anal. Biochem. 77: 569-573.

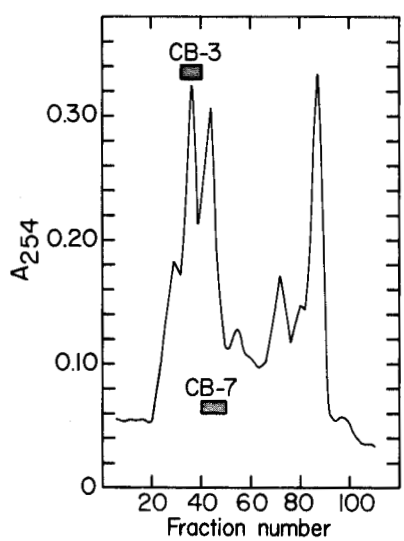

Figure 3: Gel filtration on a G-75 (fine) Sephadex column (2.5 x 90 cm) of fragments produced by cleavage of the galactose-binding protein at methionyl residues with cyanogen bromide. The column was equilibrated and run in 9% (v/v) formic acid, and 6 ml fractions were collected. Those fragments indicated by stippling were recovered, further purified by re-chromatography under the same conditions (not shown), and subjected to Edman-degradation and further subdigestion.

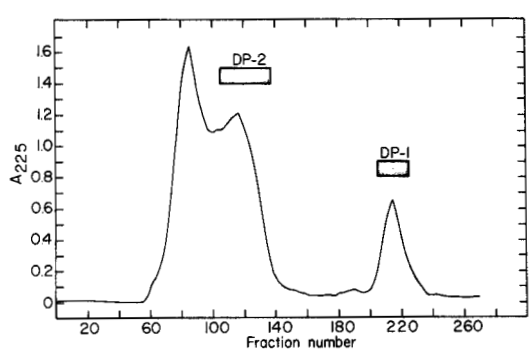

Figure 4: Isolation of fragment DP-2, generated by mild acid hydrolysis at the single aspartyl-prolyl linkage, by gel filtration on the G-75 column under identical conditions to Figure 3 except that 2.5 ml fractions were collected. Other products of the cleavage were not recovered.

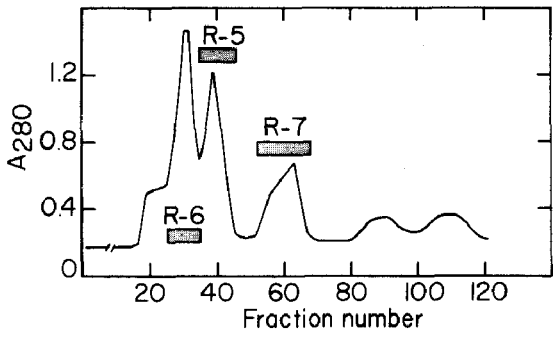

Figure 5: Elution profile produced by gel filtration on a G-50 (superfine) column (2.5 x 90 cm) of fragments generated by cleavage of intact galactose-binding protein at arginyl residues by limited trypsin digestion. The fractions were about 3.5 ml each with 130 ml of solvent (9% formic acid) collected before the fraction collector was started. Only the fragments indicated by stippled area were recovered and subjected to further analysis.

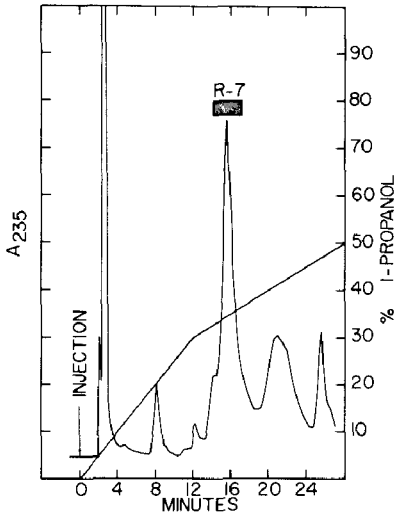

Figure 6: Fragment R-7 isolated by gel filtration (Figure 5) was further purified by reversed-phase high-performance liquid chromatography. The SynChropak RP-P column (0.46 x 25 cm) was eluted at a flow-rate of 0.7 ml/min with the indicated gradient of 1-propanol containing 0.1% (v/v) trifluoroacetic acid. The starting solvent was 0.1% (v/v) trifluoroacetic acid in water.

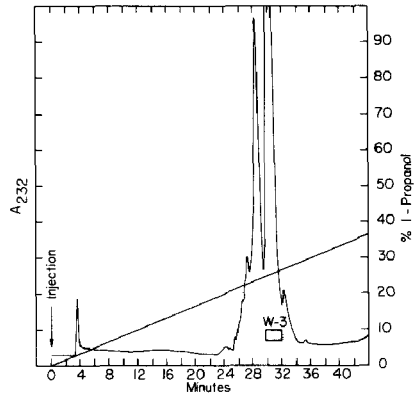

Figure 7: Purification of fragment W-3 from o-iodosobenzoic acid cleavage of CB-3. The column, flow rate, and solutions were the same as for Figure 6.

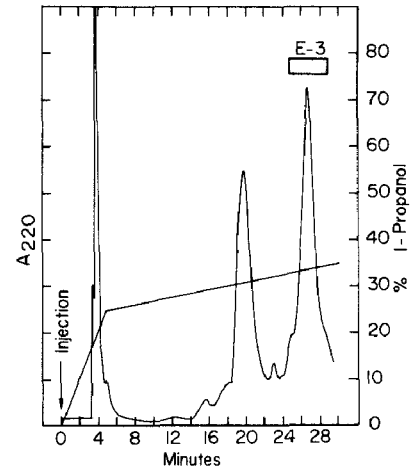

Figure 8: High-performance liquid chromatography of the products generated by the subdigestion of CB-3 at Glu with S. aureus strain V8 protease. The column (0.41 x 5-cm) was a Vydac (C-8) support, and the other conditions were as in Figure 6 except as indicated.

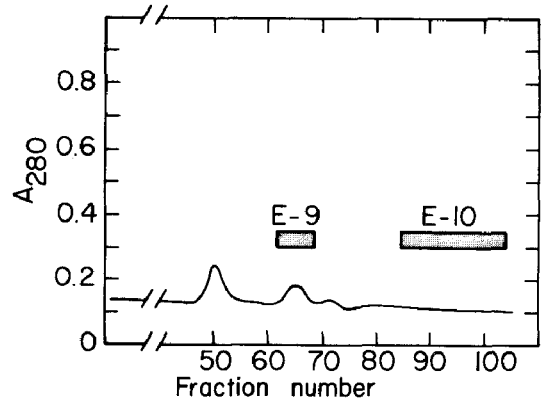

Figure 9: Gel filtration on the G-50 column (Fig. 5) of fragments produced by the subdigestion of CB-7 at Glu with S. aureus strain V8 protease. Peptides E-9 and E-10, indicated by the stippled areas, were recovered. The former was subjected to automatic degradation and the latter further purified by high-performance liquid-chromatography.

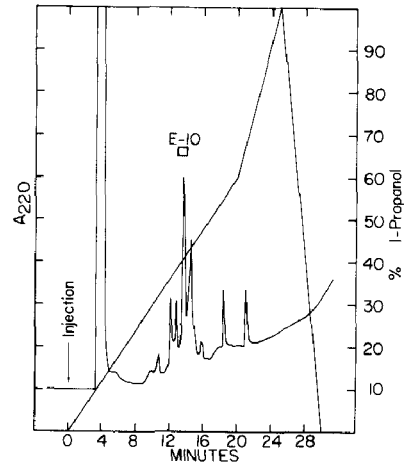

Figure 10: Purification of fragment E-10 (Figure 9) by high-performance liquid chromatography on a LiChrosorb (RP-8) reversed-phase column (0.46 x 25 cm) as described in the legend to Figure 6.

Table II: Yields and Methods of Identification of Products Generated by Automatic Degradation of Selected Galactose Binding Protein Fragments and Whole Protein

cycle	residue	yield (nmol)[a]	identified by[b]	cycle	residue	yield (nmol)[a]	identified by[b]
Whole Protein (375 nmol degraded)							
1	Ala	342	A	23	Ala	158	A
2	Asp	175	A	24	Ile	82	A
3	Thr	61	A	25	Glu	48	A
4	Arg	NDC	C	26	Gln	50	A
5	Ile	231	A	27	Asp	84	A
6	Gly	199	A	28	Ala	119	A
7	Val	285	A	29	Lys	78	A
8	Thr	45	A	30	Ala	116	A
9	Ile	217	A	31	Ala	123	A
10	Tyr	183	A	32	Pro	36	A
11	Lys	281	A	33	Asp	65	A
12	Tyr	165	A	34	Val	70	A
13	Asp	140	A	35	Gln	28	A
14	Asp	151	A	36	Leu	52	A
15	Asn	106	A	37	Leu	61	A
16	Phe	105	A	38	Met	41	A
17	Met	126	A	39	Asn	39	A
18	Ser	30	A	40	Asp	40	A
19	Val	159	A	41	Ser	ND	A
20	Val	162	A	42	Gln	19	A
21	Arg	ND	C	43	Asn	31	A
22	Lys	120	A	44	Asp	35	A
Fragment CB-3 (380 nmol degraded)							
1	Asn	240	A	24	Ala	166	A
2	Asp	290	A	25	Leu	89	A
3	Ser	81	A	26	Ala	168	A
4	Gln	190	A	22	Ile	110	A
5	Asn	200	A	28	Asn	77	A
6	Asp	215	A	29	Leu	90	A
7	Gln	185	A	30	Val	130	A
8	Ser	71	A	31	Asp	75	A
9	Lys	339	A	32	Pro	53	A
10	Gln	181	A	33	Ala	133	A
11	Asn	186	A	34	Ala	161	A
12	Asp	149	A	35	Ala	149	A
13	Gln	150	A	36	Gly	76	A
14	Ile	186	A	37	Thr	21	A
15	Asp	140	A	38	Val	89	A
16	Val	121	A	39	Ile	51	A
17	Leu	133	A	40	Glu	36	A
18	Leu	177	A	41	Lys	61	A
19	Ala	184	A	42	Ala	91	A
20	Lys	146	A	43	Arg	ND	C
21	Gly	139	A	44	Gly	53	A
22	Val	162	A	45	Gln	25	A
23	Lys	145	A	46	Asn	23	A
Fragment DP-2 (200 nmol degraded)							
1	Pro	44	A	16	Val	67	A
2	Ala	147	A	17	Pro	23	A
3	Ala	184	A	18	Val	67	A
4	Ala	156	A	19	Val	78	A
5	Gly	80	A	26	Phe	28	A
6	Thr	14	A	21	Phe	31	A
7	Val	94	A	22	Asn	20	A
8	Ile	75	A	23	Lys	46	A
9	Glu	72	A	24	Glu	40	A
10	Lys	79	A	25	Pro	15	A
11	Ala	97	A	26	Ser	ND	A
12	Arg	ND	C	27	Arg	ND	C
13	Gly	67	A	28	Lys	18	A
14	Gln	52	A	29	Ala	10	A
15	Asn	20	A				
Fragment R-5 (325 nmol degraded)							
1	Lys	168	A	15	Asp	66	A
2	Ala	291	A	16	Ser	12	A
3	Leu	139	A	17	Lys	50	A
4	Asp	156	A	18	Glu	20	A
5	Ser	37	A	19	Ser	8	A
6	Tyr	186	A	20	Gly	45	A
7	Asp	152	A	21	Ile	55	A
8	Lys	99	A	22	Ile	63	A
9	Ala	181	A	23	Gln	22	A
10	Tyr	174	A	24	Gly	40	A
11	Tyr	151	A	25	Asp	25	A
12	Val	110	A	26	Leu	40	A
13	Gly	113	A	27	Ile	34	A
14	Thr	27	A	28	Ala	36	A
Fragment E-3 (550 nmol degraded)							
1	Ser	61	A	21	Leu	164	A
2	Gly	300	A	22	Asn	88	A
3	Ile	516	A	23	Lys	180	A
4	Ile	491	A	24	Asp	82	A
5	Gln	411	A	25	Gly	139	A
6	Gly	316	A	26	Gln	66	A
7	Asp	158	A	27	Ile	109	A
8	Leu	320	A	28	Gln	47	A
9	Ile	395	A	29	Phe	71	A
10	Ala	282	A	30	Val	92	A
11	Lys	415	A	31	Leu	72	A
12	His	118	A/B	32	Leu	105	A
13	Trp	286	A	33	Lys	66	A
14	Ala	251	A	34	Gly	75	A
15	Ala	302	A	35	Glu	19	A
16	Asn	175	A	36	Pro	44	A
17	Gln	194	A	37	Gly	74	A
18	Gly	194	A	38	His	ND	B
19	Trp	139	A	39	Pro	40	A
20	Asp	104	A	40	Asp	21	A
				41	Ala	40	A
Fragment W-3 (550 nmol degraded)							
1	Asp	185	A	17	Pro	134	A
2	Leu	298	A	18	Gly	227	A
3	Asn	334	A	19	His	ND	B
4	Lys	486	A	20	Pro	81	A
5	Asp	152	A	25	Asp	75	A
6	Gly	340	A	22	Ala	122	A
7	Gln	150	A	23	Lys	75	A
8	Ile	323	A	24	Glu	87	A
9	Gln	152	A	25	Arg	ND	C
10	Phe	172	A	26	Thr	16	A
11	Val	220	A	27	Thr	24	A
12	Leu	167	A	28	Tyr	40	A
13	Leu	219	A	29	Val	63	A
14	Lys	268	A				
15	Gly	280	A				
16	Glu	142	A				

cycle	residue	yield (nmol)[a]	identified by[b]	cycle	residue	yield (nmol)[a]	identified by[b]
Fragment R-6 (850 nmol degraded)							
1	Thr	160	A	32	Asp	102	A
2	Thr	164	A	33	Lys	209	A
3	Tyr	391	A	34	Met	134	A
4	Val	535	A	35	Asp	97	A
5	Ile	513	A	36	Ala	185	A
6	Lys	>800	A	37	Trp	56	A
7	Glu	513	A	38	Leu	78	A
8	Leu	415	A	39	Ser	25	A
9	Asn	438	A	40	Gly	92	A
10	Asp	198	A	41	Pro	43	A
11	Lys	656	A	42	Asn	68	A
12	Gly	400	A	43	Ala	137	A
13	Ile	420	A	44	Asn	69	A
14	Lys	>600	A	45	Lys	94	A
15	Thr	70	A	46	Ile	48	A
16	Glu	321	A	47	Glu	70	A
17	Gln	300	A	48	Val	81	A
18	Leu	289	A	49	Val	93	A
19	Gln	270	A	50	Ile	43	A
20	Leu	231	A	51	Ala	105	A
21	Asp	148	A	52	Asn	80	A
22	Thr	110	A	53	Asn	94	A
23	Ala	292	A	54	Asp	49	A
24	Met	224	A	55	Ala	102	A
25	Trp	156	A	56	Met	30	A
26	Asp	106	A	57	Ala	101	A
27	Thr	50	A	58	Met	32	A
28	Ala	258	A	59	Gly	46	A
29	Gln	143	A	60	Ala	93	A
30	Ala	247	A	61	Val	51	A
31	Lys	214	A				
Fragment CR-7 (1100 nmol degraded)							
1	Gly	806	A	28	Leu	217	A
2	Ala	768	A	29	Val	285	A
3	Val	914	A	30	Lys	259	A
4	Glu	798	A	31	Ser	29	A
5	Ala	750	A	32	Gly	186	A
6	Leu	646	A	33	Ala	249	A
7	Lys	1070	A	34	Leu	139	A
8	Ala	718	A	35	Ala	239	A
9	His	ND	B	36	Gly	156	A
10	Asn	560	A	37	Thr	34	A
11	Lys	958	A	38	Val	178	A
12	Ser	88	A	39	Leu	114	A
13	Ser	84	A	40	Asn	120	A
14	Ile	424	A	41	Asp	81	A
15	Pro	294	A	42	Ala	163	A
16	Val	483	A	43	Asn	65	A
17	Phe	298	A	44	Asn	73	A
18	Gly	384	A	45	Gln	42	A
19	Val	466	A	46	Ala	133	A
20	Asp	216	A	47	Lys	74	A
21	Ala	400	A	48	Ala	126	A
22	Leu	273	A	49	Thr	11	A
23	Pro	193	A	50	Phe	27	A
24	Glu	198	A	51	Asp	51	A
25	Ala	363	A	52	Leu	48	A
26	Leu	234	A	53	Ala	104	A
27	Ala	345	A	54	Lys	61	A
Fragment E-9 (610 nmol degraded)							
1	Ala	341	A	33	Ala	214	A
2	Leu	570	A	34	Asp	97	A
3	Ala	342	A	35	Gly	192	A
4	Leu	559	A	36	Lys	167	A
5	Val	494	A	37	Gly	203	A
6	Lys	405	A	38	Ala	158	A
7	Ser	169	A	39	Ala	173	A
8	Gly	400	A	40	Asp	69	A
9	Ala	319	A	41	Gly	136	A
10	Leu	414	A	42	Thr	40	A
11	Ala	314	A	43	Asn	37	A
12	Gly	397	A	44	Trp	26	A
13	Thr	238	A	45	Lys	100	A
14	Val	405	A	46	Ile	48	A
15	Leu	398	A	47	Asp	51	A
16	Asn	357	A	48	Asn	31	A
17	Asp	199	A	49	Lys	68	A
18	Ala	263	A	50	Val	74	A
19	Asn	292	A	51	Val	98	A
20	Asn	346	A	52	Arg	ND	C
21	Gln	204	A	53	Val	74	A
22	Ala	241	A	54	Pro	19	A
23	Lys	388	A	55	Tyr	28	A
24	Ala	248	A	56	Val	67	A
25	Thr	126	A	57	Gly	83	A
26	Phe	244	A	58	Val	74	A
27	Asp	146	A	59	Asp	47	A
28	Leu	259	A	60	Lys	53	A
29	Ala	221	A	61	Asp	42	A
30	Lys	314	A	62	Asn	17	A
31	Asn	162	A	63	Leu	31	A
32	Leu	242	A	64	Ala	54	A
Fragment R-7 (275 nmol degraded)							
1	Val	233	A	9	Asp	73	A
2	Pro	86	A	10	Asn	53	A
3	Tyr	143	A	11	Leu	64	A
4	Val	181	A	12	Ala	103	A
5	Gly	151	A	13	Glu	70	A
6	Val	179	A	14	Phe	60	A
7	Asp	90	A	15	Ser	14	A
8	Lys	119	A	16	Lys	57	A
Fragment E-10 (550 nmol degraded)							
1	Phe	337	A	3	Lys	479	A
2	Ser	102	A	4	Lys	436	A

(a) Only 25% of each product generated by the sequence was analyzed by HPLC; yields listed above are normalized to 100% injection. (b) A, HPLC; B, spot test using Pauly reagent; C, spot test using phenanthrenequinone. (C) ND, not determined.