von Schütz, J. U., Zuclich, J., & Maki, A. H. (1974) *J. Am. Chem. Soc. 96,* 714.
Zuclich, J., Schweitzer, D., & Maki, A. H. (1972) *Biochem. Biophys. Res. Commun. 46,* 1764.
Zuclich, J., Schweitzer, D., & Maki, A. H. (1973) *Photochem. Photobiol. 18,* 161.
Zuclich, J., von Schütz, J. V., & Maki, A. H. (197) *J. Am. Chem. Soc. 96,* 710.

Primary Structure of the λ Repressor†

Robert T. Sauer* and Robert Anderegg

ABSTRACT: The complete covalent structure of the bacteriophage λ repressor has been determined by sequential Edman degradation, gas chromatographic–mass spectrometric peptide sequencing, and DNA sequencing of the repressor gene *c*I. The repressor is a single-chain, acidic protein containing 236 amino acids. The amino terminal 40 residues are highly polar and basic. Lysines and arginines in the sequence tend to be clustered.

There are many proteins which recognize specific nucleotide sequences within double stranded DNA (for review, see von Hippel & McGhee, 1972; Jovin, 1976). The λ repressor is a particularly well-studied example of this class of molecules.

λ repressor, the product of the *c*I gene of bacteriophage λ, binds strongly to multiple sites in the phage DNA. The repressor binding sites are organized into two independent operators, several thousand base pairs apart. Each operator contains three sequences which have been identified as repressor contact sites (Maniatis et al., 1975a; Humayun et al., 1977). A site consists of a 17 base pair sequence, with an axis of approximate twofold rotational symmetry through the ninth base. The binding sites are similar but not identical in nucleotide sequence and vary in their affinity for repressor (Flashman, 1976).

Repressor bound to certain of the sites sterically prevents RNA polymerase from binding to promoter sequences in the DNA (Steinberg & Ptashne, 1971; Meyer et al., 1975; Walz & Pirrotta, 1975). In this way, repressor can prevent transcription of two major phage operons, as well as transcription of its own gene *c*I. Under conditions in which only the strongest repressor binding sites are occupied, repressor stimulates *c*I transcription (Meyer et al., 1975; Ptashne et al., 1976; Walz et al., 1976). Thus repressor is both a positive and a negative gene regulator, and its function depends upon which operator sites or combination of sites are occupied. Site occupation, in turn, is a function of the concentration of repressor in the cell.

The repressor molecule is a single chain acidic protein with a monomer molecular weight of 27 000. Monomers of repressor are in equilibrium with dimers at typical intracellular concentrations (Pirrotta et al., 1970). The dimer is the species which binds strongly to the operators (Chadwick et al., 1970). Higher oligomeric forms of repressor have been observed at concentrations 100-fold greater than that in a single λ lysogen (Brack & Pirrotta, 1975). However, the relevance of these oligomers to repressor action is uncertain.

Repressor allows phage λ to reside inactively in the chromosome of its host bacterium by preventing synthesis of phage proteins necessary for the development and growth of the phage. Treatment of the bacterium with ultraviolet light, or a number of other agents (for review, see Witkin, 1976) results in inactivation of repressor and thus lytic growth of the phage. Roberts & Roberts (1975) have shown that during this "induction" process the λ repressor is cleaved, and presumably inactivated, by a protease. Many phages related to λ are also inducible. At present it is not known whether a single protease can inactivate the repressors of these phages, but it is possible that phage repressors share a common structural element which allows them to be induced via the same pathway.

Both the λ and P22 repressors can be inactivated in a noncovalent manner by the antirepressor protein of phage P22 (Susskind & Botstein, 1975). This recognition of λ repressor by a protein from a different but related phage argues that at least the lambda and P22 repressors must share some common structure.

Repressors from phages 434 and 21 have been isolated and partially characterized (Pirrotta & Ptashne, 1969; Ballivet et al., 1977). These molecules are similar to λ repressor in their monomer molecular weights, and both appear to oligomerize. Furthermore, limited homology between the λ and 434 *c*I genes has been shown by heteroduplex mapping (Westmoreland et al., 1969).

We would like to understand more fully how the phage repressors are related and how their structures are involved in the molecular interactions discussed. As a first step in this process, we have determined the amino acid sequence of the λ repressor. To do this we have used two independent methods of protein sequencing: Edman degradation and the newer method of peptide sequencing by a gas chromatographic–mass spectrometric technique. In addition, we have sequenced a portion of the DNA of the repressor *c*I gene in order to complete the protein sequence.

Materials and Methods

Repressor Purification. The repressor used for these studies was purified from the *E. coli* strain 294/pKB252 (Backman et al., 1976). Cells were suspended in buffer containing 100

† From The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138 (R.T.S.), and the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 (R.A.). *Received October 20, 1977.* This work was supported in part by Grant GM 05472 to Dr. Klaus Biemann, Department of Chemistry, Massachusetts Institute of Technology, and Grant GM 22526 to Dr. Mark Ptashne, The Biological Laboratories, Harvard University, from the National Institutes of Health, and by Grant BMS 74-09718 from the National Science Foundation. R.T.S. was a predoctoral fellow of the United States Public Health Service.

0006-2960/78/0417-1092$01.00/0

mM Tris-HCl (pH 8.0), 200 mM KCl, 0.1 mM dithiothreitol, 1 mM EDTA, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 5% glycerol, and 20 μg/mL phenylmethanesulfonyl fluoride. The cells were lysed by sonication, and debris was removed by centrifugation at 13 000*g*. The supernatant was diluted, made 0.6% in polyethylenimine, and the precipitate collected by centrifugation (Burgess & Jendrisak, 1975). Repressor was suspended from the pellet in SB buffer [10 mM Tris-HCl (pH 8.0), 2 mM $CaCl_2$, 0.1 mM EDTA, 0.1 mM dithiothreitol, 5% glycerol] plus 600 mM KCl. The resuspended material was precipitated with 80% saturated ammonium sulfate, resuspended in a small volume, and dialyzed into SB buffer plus 50 mM KCl. The dialyzed material was loaded onto a double-stranded DNA cellulose column (equilibrated in SB plus 50 mM KCl) and eluted with SB plus 200 mM KCl. The pooled activity was dialyzed against SB plus 50 mM KCl, loaded onto a CM-Sephadex column equilibrated in this buffer, and eluted with SB plus 200 mM KCl. At this point the repressor was greater than 95% pure as judged by $NaDodSO_4$[1] gel electrophoresis, and no contaminant comprised more than 1% of the total protein. Repressor purified in this manner was 30% to 50% active in binding operator DNA.

Chemical Modification. Repressor was reduced with dithiothreitol and alkylated with iodoacetic acid by the method of Konigsberg (1972). Following alkylation, the repressor sample was dialyzed against 200 mM NH_4HCO_3 and lyophilized.

Repressor was maleylated by treatment with a 100-fold molar excess of maleic anhydride. The protein was dissolved at 5 to 10 mg/mL in 100 mM $Na_2B_4O_7$ (pH 9.3). Maleic anhydride was dissolved at 150 mg/mL in 1,4-dioxane and added to the repressor in five aliquots. The pH of the reaction was adjusted to 8.5 with 1 N NaOH after each addition, when necessary. Maleylation was performed at room temperature (20–24 °C) for 10 to 15 min. Excess reagent was removed by dialysis against 200 mM NH_4HCO_3.

Mixtures of repressor peptides were maleylated using a fivefold molar excess of maleic anhydride. Other reaction conditions were the same as for native repressor. Following maleylation samples were loaded directly onto a Sephadex G-75 column for peptide separation. Peptides were demaleylated by treatment with 1 N formic acid, for 30 min, at 80 °C, under nitrogen.

Arginines were modified with 1,2-cyclohexanedione as described by Patthy & Smith (1975). Repressor was dissolved at 8 mg/mL in 200 mM $Na_2B_4O_7$ (pH 9.05). Cyclohexanedione was added to 20 mg/mL and the tube sealed under nitrogen. After incubation at 37 °C for 4 h, excess reagent was removed by dialysis against 100 mM $Na_2B_4O_7$ (pH 8.1).

Enzyme and Chemical Cleavage. Reduced and alkylated, maleylated repressor was digested with Tos-PheCH_2Cl-trypsin (E:S = 1:100) for 4 h, at 37 °C, in 200 mM NH_4HCO_3 (pH 8.2). Native repressor was digested for 18 h under the same conditions. Repressor with DHCH-blocked arginines was digested for 4 h in 100 mM $Na_2B_4O_7$ (pH 8.1).

Digestion of maleylated repressor or repressor peptides with staphylococcal protease was performed in 200 mM NH_4HCO_3 (pH 8.2) for 20 h, at 37 °C (E:S = 1:40).

Chymotryptic digests of repressor peptides were performed in 200 mM NH_4HCO_3 (pH 8.2), for 30 min, at 20 °C (E:S = 1:200).

Repressor was cleaved with cyanogen bromide in 70% formic acid, using a 300-fold molar excess of CNBr. The reaction was carried out in the dark, under nitrogen, for 20 h. Excess reagent was removed by lyophilization after a 20-fold dilution of the sample with distilled water.

Cleavage of repressor with hydroxylamine was performed by a slight modification of the method of Bornstein (1970). Native repressor was dissolved at 5 mg/mL in deionized 10 M urea. An equal volume of freshly prepared 4 M hydroxylamine in 200 mM Na_2CO_3 (pH 9.0) was added. The solution was mixed thoroughly and incubated at 37 °C for 3 h. The sample was then dialyzed against 200 mM NH_4HCO_3 to remove excess hydroxylamine and urea.

Peptide Purification. Most peptide mixtures were separated first by gel filtration on a 2.5 × 140 cm column of Sephadex G-75 (superfine). The column was equilibrated in 200 mM NH_4HCO_3 and run at a rate of 5 to 10 mL/h. Sample size was 0.5–3 μmol of cleaved repressor. Peptides greater than 80% to 90% pure by end-group analysis, following chromatography, were used directly for sequencing. Peptide mixtures generated by CNBr cleavage or tryptic cleavage of arginine blocked repressor were maleylated prior to chromatography to improve their solubility. This procedure has the additional advantage that all maleyl-peptides can be directly detected by absorbance at 225 nm. The α,ε-maleyl peptides were deblocked at acid pH prior to end-group analysis.

Following gel filtration, many peptides required further purification. In most cases, peptides from 10 to 30 amino acids were chromatographed on a 0.6 × 18 cm Dowex-1-X2 column at 37 °C. Peptides were adsorbed to the column at pH 11 and eluted with a four-step gradient, 100 mL each of 3% pyridine, 0.5 M pyridine-acetate (pH 6.0), 1.0 M pyridine-acetate (pH 6.0), and 2 M pyridine-acetate (pH 5.0). Maleylated peptides were usually demaleylated prior to ion-exchange chromatography. Fractions were monitored by the fluorescamine reaction (Udenfriend et al., 1972).

Cyanogen bromide peptides from 30 to 50 amino acids in length were purified by chromatography on CM-cellulose in 8 M urea, as described by Fowler (1975).

Amino Acid Analysis. Samples were hydrolyzed in evacuated sealed tubes containing 1 mL of 5.7 N HCl, 10 μL of phenol for 24 to 72 h, at 105–110 °C. A Beckman 121 M amino acid analyzer was used for analyses.

Edman Degradations. Manual Edman degradations were performed by a variation of the three-step Edman procedure as described (Sauer et al., 1974). A Beckman 890 B sequencer was used for automated Edman degradations. Native repressor and peptides larger than 35 amino acids were degraded using the 0.1 M Quadrol program described by Brauer et al. (1975). Phenylthiohydantoin-amino acid derivatives were identified by gas liquid chromatography (Pisano & Bronzert, 1969), thin-layer chromatography in the Edman H1 system (Edman, 1970), and two-dimensional thin-layer chromatography on polyamide (Kulbe, 1974; Summers et al., 1973). Cysteine was identified as Pth-*S*-carboxymethylcysteine. Pth-leucine and Pth-isoleucine were distinguished after treatment with *N,O*-bis(trimethylsilylacetamide) by gas liquid chromatography on a SP-400 column (Niall, 1973). Arginine and histidine residues were identified by amino acid analysis following back hydrolysis of the Pth derivatives.

In most Edman degradations, residue assignments could be made on the basis of a clear predominance by yield of a single Pth-amino acid. However, in the latter stages of some degradations several Pth-amino acids were observed in significant yield at each cycle. This resulted from the cumulative effect of nonspecific peptide bond hydrolysis during the cleavage

[1] Abbreviations used: Tos-PheCH_2Cl, L-1-tosylamido-2-phenylethyl chloromethyl ketone; DHCH-arginine, N^7,N^8-(1,2-dihydroxycyclohex-1,2-ylene)-L-arginine; GC-MS, gas chromatographic–mass spectrometric; MT, maleylated tryptic; $NaDodSO_4$, sodium dodecyl sulfate.

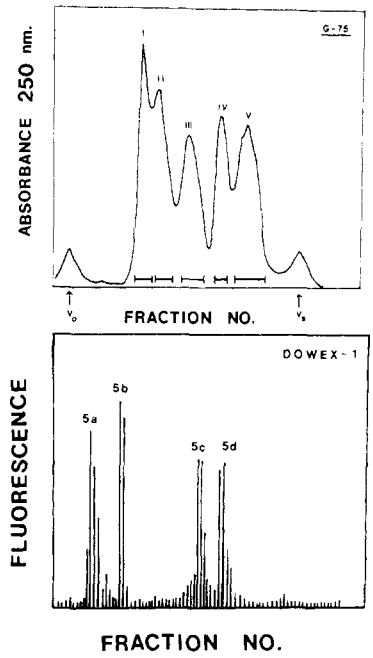

FIGURE 1: Purification of maleylated tryptic peptides. (Top) Chromatography of a maleylated tryptic digest of λ repressor on a 2.5 × 140 cm column of Sephadex G-75 (superfine). Horizontal bars indicate pooled fractions. (Bottom) Chromatography of demaleylated pool V (from top figure) on a 0.6 × 15 cm column of Dowex-1-X2. Peptides were located by fluorescence following reaction with fluorescamine.

reaction at each cycle of the degradation, and the lag or overlap of residues from one cycle to another caused by incomplete coupling or cleavage. Thus, in some cases, especially where a labile Pth derivative such as serine was involved, no Pth-amino acid could be unambiguously assigned on the basis of a predominant yield. At these cycles, assignments were based on at least a threefold rise of a single Pth-amino acid above background levels, with a subsequent fall to those levels in latter cycles. Additionally, the quantitative yields of chemically stable residues were expected to be consistent with the yield predicted from extrapolation of repetitive yield data, established during the early portion of the degradation.

GC–MS Sequencing. Selected peptides of repressor were hydrolyzed in 6 N HCl at 110 °C for 20 to 40 min to generate mixtures composed primarily of di-, tri-, and tetrapeptides. The oligopeptide mixtures were derivatized to poly(amino alcohol) related compounds and analyzed by the gas chromatographic–mass spectrometric method (GC–MS) (Nau & Biemann, 1976a–c; Anderegg, 1977).

DNA Sequencing. cI DNA was sequenced using the chemical method described by Maxam & Gilbert (1977). Restriction fragments were isolated from digests of pKB252 DNA by acrylamide gel electrophoresis as described (Maniatis et al., 1975b).

Reagents. Reagents were obtained from the following commercial sources: Tos-PheCH_2Cl-trypsin, α-chymotrypsin, and bacterial alkaline phosphatase (Worthington); staphylococcal protease and polyethylenimine (Miles); polynucleotide kinase (New England BioLabs); iodoacetic acid, 95% hydrazine, and hydroxylamine (Eastman); UltraPure urea (Schwarz/Mann); benzene, 1-chlorobutane, heptane, heptafluorobutyric acid, dimethylallylamine, trifluoroacetic acid, phenyl isothiocyanate, 1-propanol, 1,2-cyclohexanedione, cyanogen bromide, and maleic anhydride (Pierce); ethyl acetate and Quadrol buffer (Beckman Instruments); silica gel thin-layer plates (Analtech); dimethyl sulfate (Aldrich); polyamide thin-layer plates (Schleicher and Scheull); Dowex-1-X2 (Hamilton); γ-adenosine triphosphate (Amersham/Searle); [^{3}H]iodoacetate (New England Nuclear).

Results

Maleylated Tryptic Fragments. Cleavage of maleylated, reduced, and alkylated repressor with trypsin produces eight major peptides. The purification of these peptides by gel filtration on Sephadex G-75 and ion exchange on Dowex-1-X2 is shown in Figure 1. The results of Edman degradation on these peptides are presented in Table I. The yield in nanomoles of each Pth-amino acid derivative is subscripted below the residue. Superscripted numbers represent the residue number in the final sequence. These degradations determine the sequence of 124 amino acids of repressor.

Degradation of Native Repressor. Automated Edman degradation of repressor was carried out for 23 cycles. The results are presented in Table II. These results complete the sequence of MT-4 and establish the peptide order NH_2-MT4-Arg-MT1. There are eight arginines in repressor, and thus nine maleylated tryptic peptides are expected. The fact that only eight MT peptides were isolated is explained by the presence of the Arg^{16}-Arg^{17} sequence. The free arginine released by trypsin was not detected during the chromatographic procedures.

Cyanogen Bromide Peptides. Cleavage of repressor with cyanogen bromide results in a complex peptide mixture, due in part to incomplete cleavage at each methionine. However, several CNBr peptides were purified to homogeneity by a combination of gel filtration, and ion-exchange chromatography on carboxymethylcellulose in 8 M urea. Degradations on five of the CNBr peptides are depicted in Table II.

The sequence of CNBr-5a completes the sequence of MT-5d and established the order MT5d-MT5c. The sequence of CNBr-3c completes the sequence of MT-5c (see Table V for compositional data), includes all of MT-5b, and overlaps MT-2. The peptide order is thus MT5d-MT5c-MT5b-MT2.

The sequence of peptide CNBr-4 overlaps and completes the sequence of the MT-3 peptide. Following 23 cycles of Edman degradation, glycine was recovered as the free amino acid, establishing this residue as the 24th and carboxyl-terminal residue of CNBr-4. Sequence analysis of MT-3 by GC–MS techniques identified the tripeptide Thr^{234}-Phe^{235}-Gly^{236} but no peptides of sequence Phe-Gly-XXX, further confirming the assignment of glycine as the ultimate residue of both MT-3 and CNBr-4. Since MT-3 contains no arginine and CNBr-4 no homoserine or homoserine lactone, glycine is the likely candidate for the carboxyl terminus of the repressor molecule. Recently, Humayun (1977) has determined the sequence of a portion of the *cI* gene. His DNA sequence confirms our assignments from residues 19–24 of CNBr-4 and demonstrates the presence of a single termination codon TGA following the GGC codon for the terminal glycine. MT-3 is therefore the carboxyl terminal maleylated tryptic peptide.

Arginine-Blocked Tryptic Fragments. Following arginine blockade and tryptic digestion of repressor, three peptides were isolated by sizing on Sephadex G-75 in a form sufficiently homogeneous for sequencing. The amino terminal sequences of these fragments are shown in Table III.[2]

[2] The arginine-blocked tryptic peptides were purified using buffers containing relatively high concentrations of ammonium ion, and thus the DHCH-arginines may have been partially unblocked prior to sequencing (Patthy & Smith, 1975). For this reason we do not know whether the arginines identified by back hydrolysis during degradation of these peptides resulted from regeneration of the free amino acid from the phenylthiohydantoin derivative of arginine or from the DHCH adduct of arginine.

TABLE I: Maleylated Tryptic Peptides.

cycle number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
MT 1																							
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
	Leu-	Lys-	Ala-	Ile-	Tyr-	Glu-	Lys-	Lys-	Lys-	Asn-	Glu-	Leu-	Gly-	Leu-	Ser-	Gln-	Glu-	Ser-	Val-	Ala-	Asp-	Lys...	
	39	a	28	32	16	18	a	a	a	10	8	12	5	9	4	2	3	2	5	4	2	a	
MT 2																							
	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147				
	Trp-	Val-	Ser-	Thr-	Thr-	Lys-	Lys-	Ala-	Ser-	Asp-	Ser-	Ala-	Phe-	Trp-	Leu-	Glu-	Val-	Glu-	Gly...				
	14	43	28	17	21	a	a	25	13	18	9	17	14	2	13	10	10	10	4				
MT 3																							
	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219
	Asp-	Ser-	Gly-	Gln-	Val-	Phe-	Leu-	Gln-	Pro-	Leu-	Asn-	Pro-	Gln-	Tyr-	Pro-	Met-	Ile-	Pro-	Cys-	Asn-	Glu-	Ser-	Cys...
	50	42	37	25	55	39	45	19	12	35	30	7	13	8	5	16	18	5	6	8	10	3	2
MT 4																							
	1	2	3	4	5	6	7																
	Ser-	Thr-	Lys-	Lys-	Lys-	Pro-	Leu...																
	17	12	a	a	a	17	20																
MT 5a																							
	184	185	186	187	188	189	190	191	192	193	194	195	196										
	Leu-	Gly-	Gly-	Asp-	Glu-	Phe-	Thr-	Phe-	Lys-	Lys-	Leu-	Ile-	Arg-	COOH									
	25	18	20	17	8	7	6	6	a	a	2	2	b										
MT 5b																							
	120	121	122	123	124	125	126	127	128														
	Thr-	Phe-	Thr-	Lys-	Gly-	Asp-	Ala-	Glu-	Arg-	COOH													
	86	71	66	a	49	42	26	36	c														
MT 5c																							
	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115						
	Ser-	Glu-	Tyr-	Glu-	Tyr-	Pro-	Val-	Phe-	Ser-	His-	Val-	Gln-	Ala-	Gly-	Met-	Phe-	Ser...						
	70	90	55	83	62	41	70	64	31	49	47	27	47	11	32	20	7						
MT 5d																							
	83	84	85	86	87	88	89	90	91	92	93	94	95	96									
	Glu-	Ile-	Tyr-	Glu-	Met-	Tyr-	Glu-	Ala-	Val-	Ser-	Met-	Gln-	Pro-	Ser...									
	44	64	51	50	53	37	26	46	43	25	31	18	13	4									

[a] ε-Maleyl-α-Pth-lysine or ε-PTC-α-Pth-lysine identified by thin-layer chromatography. [b] 0.9 nM arginine identified by amino acid analysis after 12 cycles of degradation. [c] 15 nM arginine identified by amino acid analysis after 8 cycles of degradation. Numbers under residues are yields in nanomoles of the Pth-amino acid derivative.

TABLE II: Native Repressor and CNBr Peptides.

cycle number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
native																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
	Ser-	Thr-	Lys-	Lys-	Lys-	Pro-	Leu-	Thr-	Gln-	Glu-	Gln-	Leu-	Glu-	Asp-	Ala-	Arg-	Arg-	Leu-	Lys-	Ala-	Ile-	Tyr-	Glu...	
	9	17	a	a	a	12	14	12	5	10	9	10	10	5	8	3	5	6	a	6	5	2	3	
CNBr 3a																								
	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186		
	Leu-	Ile-	Leu-	Val-	Asp-	Pro-	Glu-	Gln-	Ala-	Val-	Glu-	Pro-	Gly-	Asp-	Phe-	Cys-	Ile-	Ala-	Arg-	Leu-	Gly-	Gly...		
	67	74	70	55	27	11	45	23	22	20	26	9	17	27	25	12	13	13	6	11	6	5		
CNBr 3c																								
	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130							
	Phe-	Ser-	Pro-	Glu-	Leu-	Arg-	Thr-	Phe-	Thr-	Lys-	Gly-	Asp-	Ala-	Glu-	Arg-	Trp-	Val...							
	40	19	16	31	33	16	18	25	11	a	11	17	21	9	6	1	11							
CNBr 4																								
	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236
	Ile-	Pro-	Cys-	Asn-	Glu-	Ser-	Cys-	Ser-	Val-	Val-	Gly-	Lys-	Val-	Ile-	Ala-	Ser-	Gln-	Trp-	Pro-	Glu-	Glu-	Thr-	Phe-	Gly-	COOH
	165	130	100	65	110	77	69	93	56	55	55	a	34	35	35	37	20	7	17	13	20	12	9	b	
CNBr 5a																								
	94	95	96	97	98	99	100	101																
	Gln-	Pro-	Ser-	Leu-	Arg-	Ser-	Glu-	Tyr...																
	19	12	13	28	14	7	9	3																
CNBr 5b																								
	151	152	153	154	155	156	157	158	159	160	161	162	163	164										
	Thr-	Ala-	Pro-	Thr-	Gly-	Ser-	Lys-	Pro-	Ser-	Phe-	Pro-	Asp	(Gly)(Met)	COOH										
	31	54	17	19	14	9	a	8	6	11	3	1	c	c										

[a] Identified as ε-PTC-α-Pth-lysine by thin-layer chromatography. [b] 3.5 nM glycine identified by amino acid analysis after 23 cycles of degradation. [c] Residues placed by difference with composition (see Table V). Numbers under residues are yields in nanomoles of the Pth-amino acid derivative.

TABLE III: DHCH-Tryptic and Mixture Degradations.

cycle number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
DHCH 1	136 Ala	137 Ser	138 Asp	139 Ser	140 Ala	141 Phe	142 Trp	143 Leu	144 Glu	145 Val	146 Glu	147 Gly	148 Asn	149 Ser	150 Met	151 Thr	152 Ala	153 Pro...						
	84	73	78	45	92	69	22	78	53	65	40	41	23	18	45	27	38	19						
DHCH 2	71 Val	72 Ser	73 Val	74 Glu	75 Glu	76 Phe	77 Ser	78 Pro	79 Ser	80 Ile	81 Ala	82 Arg	83 Glu	84 Ile	85 Tyr	86 Glu	87 Met	88 Tyr	89 Glu	90 Ala	91 Val	92 Ser	93 Met	94 Gln...	
	207	120	197	85	156	171	132	22	131	139	97	46	54	103	81	62	76	71	30	19	24	12	9	a
DHCH 3	194 Leu	195 Ile	196 Arg	197 Asp	198 Ser	199 Gly...																		
	113	158	37	130	36	42																		
SP 2e 1	218 Ser	219 Cys	220 Ser	221 Val	222 Val	223 Gly	224 Lys	225 Val	226 Ile	227 Ala	228 Ser	229 Gln	230 Trp	231 Pro	232 Glu	233 Glu...								
	71	62	58	55	53	75	b	38	43	38	29	19	8	c	17	25								
SP 2e 2	163 Gly	164 Met	165 Leu	166 Ile	167 Leu	168 Val	169 Asp	170 Pro	171 Glu	172 Gln	173 Ala	174 Val	175 Glu	176 Pro	177 Gly	178 Asp...								
	81	90	64	48	53	43	32	27	47	29	31	22	13	c	22	19								

degraded as mixture

[a] Glutamine identified by thin-layer chromatography only. [b] Identified as ε-PTC-α-Pth-lysine by thin-layer chromatography. [c] 16 nM Pth-proline found at cycle 14 of SP 2e mixture degradation. Numbers under residues are yields in nanomoles of the Pth-amino acid derivative.

TABLE IV: DNA Sequence and Partial Peptide Sequence.[c]

SP ↓ CT ↓

34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57

glu ser val ala asp lys met gly met gly gln ser gly val gly ala leu phe asn gly ile asn ala leu

AA TCT GTC GCA GAC AAG ATG GGG ATG GGG CAG TCA GGC GTT GGT GCT TTA TTT AAT GGC ATC AAT GCA TTA

SP-1 Ser-Val-Ala-Asp-Lys-Met-Gly-Met-Gly-Gln-Ser-Gly-Val-Gly-Ala-Leu-Phe

19 61 55 37 a 46 24 41 28 17 13 9 26 16 18 12 3

hydroxylamine cleavage site

CT ↓ CT ↓ CT ↓ SP ↓

58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

asn ala tyr asn ala ala leu leu ala lys ile leu lys val ser val glu glu phe ser pro ser ile

AAT GCT TAT AAC GCC GCA TTG CTT GCA AAA ATT CTC AAA GTT AGC GTT GAA GAA TTT AGC CCT TCA ATC

Asn(Ala,Tyr) 6 Asn-Ala-Ala(Leu,Leu) 12 9 4 Ala-Lys(Ile,Leu) 15 a Lys-Val-Ser-Val-Glu-Glu-COOH a 22 8 12 16 b

[a] Identified by TLC. [b] 12 nM glutamic acid identified on AAA after 5 cycles of Edman degradation. [c] SP, sites of cleavage with staphylococcal protease. CT, sites of cleavage with chymotrypsin. Numbers under residues are yields in nanomoles of the Pth-amino acid derivative.

The sequence of peptide DHCH-3 provides a direct overlap between the MT-5a and MT-3 sequences. DHCH-1 bridges the MT-2 sequence and the CNBr-5b sequence, extending the MT-2 sequence through 36 amino acids. The remainder of the MT-2 sequence, and its overlap with MT-5a, are established by abutting CNBr-5b and CNBr-3a. The mixture degradation of two staphylococcal protease peptides presented in Table III provides evidence for the joining of CNBr-5b and CNBr-3a. Even though two Pth-amino acids were identified at each cycle of the degradation, one of the sequences (SP-2e-1) was already known from sequence analysis of MT-3 and CNBr-4, and thus the other sequence could be generated by difference. The SP-2e-2 sequence provides a two residue overlap between CNBr-5b and CNBr-3a. This proposed arrangement of peptides is further supported by the amino acid composition (Table V) of the large tryptic peptide which encompasses residues 136–183 in the final repressor sequence (Figure 3). Together, these degradations establish the peptide order MT2-MT5a-MT3-COOH.

Completed MT-1 Sequence. Digestion of MT-1 with staphylococcal protease generates a 41 amino acid peptide which is completely separated from other digestion products by filtration on Sephadex G-50. The sequence of this peptide (Table IV) extended the MT-1 sequence through Phe[51].

TABLE V: Selected Peptide Compositions.[a]

	MT-5a	MT-5b	MT-5c	MT-5d	T-1	SP-1	CNBr-5b
SCMC	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.75 (1)	0.00 (0)	0.00 (0)
Asx	1.06 (1)	0.98 (1)	0.22 (0)	0.20 (0)	4.96 (5)	5.07 (5)	1.14 (1)
Thr	0.94 (1)	2.13 (2)	0.12 (0)	0.13 (0)	1.96 (2)	0.07 (0)	1.74 (2)
Ser	0.00 (0)	0.03 (0)	2.92 (3)	2.12 (2)	4.86 (5)	2.52 (3)	1.65 (2)
Glx	1.22 (1)	1.04 (1)	4.04 (4)	4.00 (4)	5.18 (5)	3.22 (3)	0.15 (0)
Pro	0.06 (0)	0.00 (0)	2.09 (2)	0.99 (1)	5.07 (5)	0.11 (0)	2.71 (3)
Gly	2.16 (2)	1.00 (1)	1.10 (1)	0.19 (0)	4.08 (4)	5.22 (5)	2.04 (2)
Ala	0.22 (0)	0.98 (1)	1.06 (1)	1.10 (1)	4.77 (5)	6.96 (7)	1.10 (1)
Val	0.09 (0)	0.11 (0)	1.94 (2)	1.07 (1)	3.22 (3)	3.81 (4)	0.09 (0)
Met	0.00 (0)	0.00 (0)	0.97 (1)	1.65 (2)	2.02 (2)	1.89 (2)	*b* (1)
Ile	0.89 (1)	0.02 (0)	0.15 (0)	0.91 (1)	2.07 (2)	1.93 (2)	0.00 (0)
Leu	1.90 (2)	0.07 (0)	1.10 (1)	1.05 (1)	3.07 (3)	4.91 (5)	0.13 (0)
Tyr	0.09 (0)	0.05 (0)	1.84 (2)	1.77 (2)	0.08 (0)	1.12 (1)	0.05 (0)
Phe	1.70 (2)	0.94 (1)	1.77 (2)	0.14 (0)	2.80 (3)	1.06 (1)	0.95 (1)
His	0.00 (0)	0.00 (0)	0.89 (1)	0.05 (0)	0.02 (0)	0.00 (0)	0.04 (0)
Lys	1.92 (2)	1.01 (1)	0.12 (0)	0.11 (0)	1.18 (1)	3.00 (3)	1.20 (1)
Arg	0.92 (1)	0.97 (1)	1.01 (1)	0.97 (1)	0.95 (1)	0.05 (0)	0.00 (0)
Trp	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.23 (1)	0.00 (0)	0.00 (0)
Sequence position	184–196	120–128	99–119	83–98	136–183	35–75	151–164

[a] Analyses based on 24-h hydrolysis. The data have not been corrected for Ser, Thr, Val, Ile, or Trp. [b] One residue of methionine assumed on the basis of homoserine recovery. Numbers in parentheses are residues found by sequence analysis.

TABLE VI: Hydroxylamine Mixture Degradation.

cycle number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	Ser-	Thr-	Lys-	Lys-	Lys-	Pro-	Leu-	Thr-	Gln-	Glu-	Gln-	Leu-	Gln-	Asp-	Ala-	Arg-	Arg-	Leu-	Lys...
	34	60	a	a	a	50	63	55	36	48	35	b	21	32	38	12	15	26	a
	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
	Gly-	Ile-	Asn-	Ala-	Leu-	Asn-	Ala-	Tyr-	Asn-	Ala-	Ala-	Leu-	Leu-	Ala-	Lys-	Ile-	Leu-	Lys-	Val...
	41	58	27	51	46	31	39	22	17	32	36	b	37	24	a	18	23	a	16

[a] Identified as ε-PTC-α-Pth-lysine by thin-layer chromatography. [b] 77 nM of Pth-leucine identified at step 12. Numbers under residues are yields in nanomoles of the Pth-amino acid derivative.

However, in several degradations no residues could be identified following Phe[51], even though the degradations had progressed well up to this point. Analysis of a partial acid hydrolysate of MT-1 by GC–MS identified several peptides (Figure 3) including Ala[49]-Leu[50]-Phe[51]-Asx[52] which indicated that residue 52 was either an asparagine or an aspartic acid residue. This raised the strong possibility of an α–β shift of the peptide bond carboxy to the Asx residue, thus blocking progression of the Edman degradation.

The amino acid sequence directly following Phe[51] was first deduced from DNA sequencing of the *Hinf*/Alu* 254 restriction fragment from *c*I (Sauer, in preparation). This sequence is shown in Table IV. Some of the protein sequence predicted from the DNA sequence was corroborated by the compositions and partial sequences of the chymotryptic peptides isolated from SP-1 (Table IV). Subsequently, we succeeded in cleaving the Asn[52]–Gly[53] peptide bond with hydroxylamine. The results of a mixture degradation on repressor cleaved with hydroxylamine confirm the proposed sequence from residues 53–71 (Table VI).

The extended MT-1 sequence overlaps the DHCH-2 sequence beginning at residue 71. This overlap is confirmed by DNA sequence data (Table IV) and by the composition of SP-1 (Table V). Since the DHCH-2 sequence provides an overlap with MT-5d, the maleylated tryptic fragments are completely ordered as follows: NH_2-MT4-Arg-MT1-MT5d-MT5c-MT5b-MT2-MT5a-MT3-COOH. Figure 2 schematically illustrates the positions of peptides used to establish the sequence. The complete amino acid sequence is shown in Figure 3.

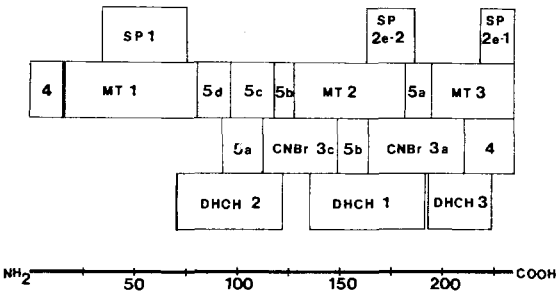

FIGURE 2: Schematic representation of the positions of peptides used for sequence analysis.

The structure proposed for repressor contains three free SH groups and no disulfides. During the course of sequence analysis, no evidence was obtained for peptides crosslinked by a disulfide bond. When denatured repressor was alkylated with [^{3}H]iodoacetate in the absence of reducing agent, radioactivity was incorporated equally into all three cysteine positions. The incorporation, however, was not stoichiometric (data not shown). We therefore cannot rigorously exclude the possibility

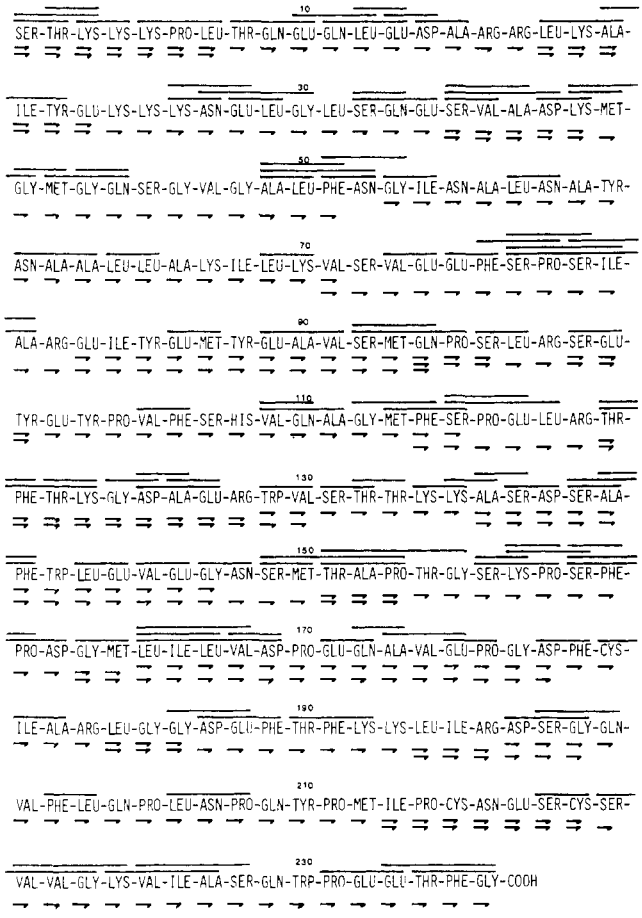

FIGURE 3: Repressor sequence. Peptides located by GC–MS sequencing are shown as bars above the amino acids. Leucine/isoleucine, aspartic acid/asparagine, and glutamic acid/glutamine were not differentiated by this method. The arrows under the residues represent Pth-amino acid derivatives identified in the Edman degradations depicted in Tables I–IV and VI.

of a heterogeneous population of free and partially oxidized SH groups.

Confirming Data. The sequence proposed here agrees well with the calculated amino acid composition (Table VII), and the calculated molecular weight of 26 228 is in good agreement with the molecular weight of ~27 000 determined by Na-DodSO$_4$ gel electrophoresis.

Sequence analysis by the GC–MS technique was carried out on the maleylated tryptic peptides and on SP-1. The peptides identified by this method are shown schematically in Figure 3. In all, 179 of the 236 amino acids of repressor were located as di-, tri-, tetra-, and pentapeptides by the GC–MS method. Because of the confirming role of this method, further analysis was deemed unnecessary.

Discussion

We have determined the complete amino acid sequence of the phage λ repressor, using classical methods of protein and peptide sequencing, GC–MS sequencing, and DNA sequencing. Although the final sequence can be proved by composition and Edman data alone, the GC–MS and DNA sequence data support and increase our confidence in the proposed structure. During the course of sequence determination, data from the GC–MS or DNA methods often preceded the Edman data and allowed us to revise our strategy accordingly.

For example, the failure of several degradations to proceed beyond Phe^{51} suggested either cyclization of a glutamine residue, or an α–β transpeptidation involving either an aspartate or asparagine residue. Identification of the peptide Ala^{49}-Leu^{50}-Phe^{51}-Asx^{52} by GC–MS indicated the latter possibility. DNA sequencing revealed the Asn^{52}–Gly^{53} sequence, and we were subsequently able to concentrate our efforts on cleaving this bond with hydroxylamine (Bornstein, 1970) in order to complete the amino acid sequence.

TABLE VII: Amino Acid Composition.

	Calculated[a]	Found by sequence
SCMC	2.66	3
Asn	16.89	8
Asp		9
Thr	9.97[b]	10
Ser	20.85[b]	22
Gln	33.49	11
Glu		22
Pro	14.94	15
Gly	17.03	17
Ala	19.11	19
Val	14.98[c]	15
Met	7.77	8
Ile	9.66[c]	10
Leu	19.35	19
Tyr	7.23	7
Phe	12.10	12
His	1.18	1
Lys	16.70	17
Arg	8.05	8
Trp	2.70[d]	3

[a] Average of duplicate hydrolyses at 24, 48, 72, and 96 h. [b] Values extrapolated to zero time. [c] 96-h value used. [d] Spectrophotometric determination.

The recent development of rapid methods for DNA sequencing (Sanger & Coulson, 1975; Maxam & Gilbert, 1977) provides an extremely powerful tool for the sequence analysis of proteins where sufficient quantities of structural DNA are available. By itself, DNA sequencing cannot unambiguously determine the amino acid sequence of a protein, since enzymic processing, disulfide formation, and covalent modifications can occur following synthesis of the protein. More seriously, recent studies have suggested that some messenger RNA molecules are synthesized from noncontiguous DNA sequences (Berget et al., 1977). Nevertheless, the combination of DNA sequencing with limited protein sequencing, either by the GC–MS technique or by Edman degradation, may provide the fastest route to a complete protein sequence.

The DNA sequences at both termini of the *c*I gene are now known (Ptashne et al., 1976; Humayun, 1977). Comparison with the protein sequence presented here reveals that a single formyl methionine residue is enzymically removed from the amino terminus following repressor synthesis, while no processing occurs at the carboxyl terminus.

Beyreuther & Gronenborn (1976) have reported the sequence of the first 50 amino acids of the λ repressor. The sequence reported here differs from theirs at three positions. They report Asn^{14}, Ser^{40}, and Thr^{42}, whereas we have found Asp^{14}, Met^{40}, and Met^{42}. We have identified the methionines at 40 and 42 in multiple Edman degradations, by GC–MS sequencing (Figure 3), and by DNA sequencing (Table IV). We have also identified aspartate rather than asparagine at residue 14 in four separate degradations. We cannot, however, rule out the possibility that our repressor was preferentially deamidated at this position. Beyreuther (personal communication) has reinvestigated the sequence positions in question, and his findings are now in agreement with our own.

There are several striking features of the λ repressor se-

quence. The amino terminal 26 residues contain nine basic amino acids which occur in clusters: Lys^{3}-Lys^{4}-Lys^{5}, Arg^{16}-Arg^{17}-Leu^{18}-Lys^{19}, and Lys^{24}-Lys^{25}-Lys^{26}. Two additional dilysine sequences occur later in the sequence at residues 134–135 and 192–193. Thus, although the repressor is an acidic molecule, the amino terminus is very basic, and the basic amino acids throughout the sequence tend to be clustered. The repressor also contains regions of sequence with a high negative charge density. For example, the sequence from residues 74–102 contains seven glutamic acids. It is possible that these regions, rich in charged amino acids, play some role in either the orientation or binding of repressor to DNA.

There are several regions of repressor with unusual residue distributions. For example, the sequence Asn^{52}-X-X-Asn^{55}-X-X-Asn^{58}-X-X-Asn^{61} occurs in the hydrophobic region of repressor, and the sequence Pro^{205}-X-X-Pro^{208}-X-X-Pro^{211}-X-X-Pro^{214} is found near the carboxyl terminus. Repressor contains seven tyrosine residues, and five of these are found in conjunction with glutamic acids including the Glu^{100}-Tyr^{101}-Glu^{102}-Tyr^{103} and Tyr^{85}-Glu^{86}-Met^{87}-Tyr^{88}-Glu^{89}. Several other examples of this type can be adduced, but further structural characterization will be necessary before the functional significance of such sequences can be assessed.

The clustering of basic amino acids in the amino terminus of λ repressor invites speculation that these residues might interact with the negatively charged phosphate backbone of DNA. The charge distribution of residues in repressor superficially resembles that in some of the histones, and the amino termini of both histones and the *lac* repressor have been implicated in the binding of these molecules to DNA (Weintraub & Van Lente, 1974; Adler et al., 1972). However, direct inspection reveals little if any sequence homology between these proteins and the λ repressor.

Several lines of biochemical and genetic evidence indicate that the amino terminal 60 amino acids of the *lac* repressor are involved in specific binding to operator DNA (for review, see Meuller-Hill, 1975; Weber et al., 1975). Mutants of *lac* repressor which do not bind operator DNA exist, and many mutations responsible for this phenotype have been located, both genetically and by protein sequencing, in the amino terminus. These mutants are transdominant and presumably inactivate wild-type repressor by forming mixed tetramers (Meuller-Hill et al., 1968).

Oppenheim & Salomon (1970) have described a similar class of transdominant mutations in the *c*I gene of λ repressor. These λ *CP* mutations map in the amino terminal third of the *c*I gene (Oppenheim & Salomon, 1972; Oppenheim & Noff, 1975). If the interpretation of these mutations as DNA binding mutations is correct, then the amino terminal 80 or so amino acids of repressor are implicated in operator binding. Beyreuther & Gronenborn (1976) have pointed out that many of the amino-terminal residues of λ repressor could in principle form hydrogen bonds with either the nucleotide bases or the DNA backbone. In fact, 50% of the first 28 residues of repressor are charged amino acids, and 75% are polar amino acids. Unfortunately, little is known about the kind of forces which stabilize DNA–protein complexes, and local frequencies of either hydrophilic or hydrophobic amino acids may be misleading. The amino acid sequence in conjunction with the genetic analysis provides intriguing clues, but the question of how λ repressor recognizes and specifically binds to its operators remains unsolved.

Acknowledgments

We thank Mrs. Loretta Yeung and Mr. Mehmet Pamuku for excellent technical assistance. This work was performed in the laboratory of Dr. Mark Ptashne (The Biological Laboratories, Harvard University) and Dr. Klaus Biemann (Department of Chemistry, Massachusetts Institute of Technology) to whom we are grateful for financial and intellectual support.

References

Adler, K., Beyreuther, K., Fanning, E., Geisler, E., Gronenborn, B., Klemm, A., Mueller-Hill, B., Pfahl, M., & Schmitz, A. (1972) *Nature (London) 237,* 322.

Anderegg, R. (1977) Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass.

Backman, K., Ptashne, M., & Gilbert, W. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73,* 4174.

Ballivet, M., Reichardt, L. F., & Eisen, H. (1977) *Eur. J. Biochem. 73,* 601.

Berget, S. M., Moore, C., & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 3171.

Beyreuther, K., & Gronenborn, B. (1976) *Mol. Gen. Genet. 147,* 115.

Bornstein, P. (1970) *Biochemistry 9,* 2408.

Brack, C., & Pirrotta, V. (1975) *J. Mol. Biol. 96,* 139.

Brauer, A. W., Margolies, M. N., & Haber, E. (1975) *Biochemistry 14,* 3029.

Burgess, R. R., & Jendrisak, J. J. (1975) *Biochemistry 14,* 4634.

Chadwick, P., Pirrotta, V., Steinberg, R., Hopkins, N., & Ptashne, M. (1970) *Cold Spring Harbor Symp. Quant. Biol. 35,* 283.

Edman, P. (1970) in *Protein Sequence Determination* (Needleman, S., Ed.) p 211, Springer-Verlag, New York, N.Y.

Flashman, S. (1976) Ph.D. Thesis, Harvard University, Cambridge, Mass.

Fowler, A. V. (1975) in *Solid Phase Methods in Protein Sequence Analysis* (Laursen, R., Ed.) p 169, Pierce Chemical Co., Rockford, Ill.

Humayun, Z. (1977) *Nucleic Acids Res. 4,* 2137.

Humayun, Z., Kleid, D., & Ptashne, M. (1977) *Nucleic Acids Res. 4,* 1595.

Jovin, T. M. (1976) *Annu. Rev. Biochem. 45,* 889.

Konigsberg, W. (1972) *Methods Enzymol. 25,* 185.

Kulbe, K. D. (1974) *Anal. Biochem. 59,* 564.

Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A., & Maurer, R. (1975a) *Cell 5,* 109.

Maniatis, T., Jeffrey, A., & van de Sande, H. (1975b) *Biochemistry 14,* 3787.

Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74,* 560.

Meyer, B. J., Kleid, D. G., & Ptashne, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72,* 4785.

Mueller-Hill, B. (1975) *Prog. Biophys. Mol. Biol. 30,* 227.

Mueller-Hill, B., Crapo, L., & Gilbert, W. (1968) *Proc. Natl. Acad. Sci. U.S.A. 59,* 1259.

Nau, H., & Biemann, K. (1976a) *Anal. Biochem. 73,* 139.

Nau, H., & Biemann, K. (1976b) *Anal. Biochem. 73,* 154.

Nau, H., & Biemann, K. (1976c) *Anal. Biochem. 73,* 175.

Niall, H. D. (1973) *Methods Enzymol. 27,* 942.

Oppenheim, A. B., & Noff, D. (1975) *Virology 64,* 553.

Oppenheim, A. B., & Salomon, D. (1970) *Virology 41,* 151.

Oppenheim, A. B., & Salomon, D. (1972) *Mol. Gen. Genet. 115,* 101.

Patthy, L., & Smith, E. L. (1975) *J. Biol. Chem. 250,* 557.

Pirrotta, V., & Ptashne, M. (1969) *Nature (London) 222,* 541.

Pirrotta, V., Chadwick, P., & Ptashne, M. (1970) *Nature (London) 227,* 41.
Pisano, J. J., & Bronzert, T. J. (1969) *J. Biol. Chem. 244,* 5597.
Ptashne, M., Backman, K., Humayun, M. Z., Jeffrey, A., Maurer, R., Meyer, B., & Sauer, R. T. (1976) *Science 194,* 156.
Roberts, J. W., & Roberts, C. W. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72,* 147.
Sanger, F., & Coulson, A. R. (1975) *J. Mol. Biol. 94,* 441.
Sauer, R. T., Niall, H. D., Hogan, M. L., Keutmann, H. T., O'Riordan, J. L. H., & Potts, J. T., Jr. (1974) *Biochemistry 13,* 1994.
Steinberg, R. A., & Ptashne, M. (1971) *Nature (London) 230,* 76.
Summers, M. R., Smythers, G. W., & Oroszlan, S. (1973) *Anal. Biochem. 53,* 624.
Susskind, M. M., & Botstein, D. (1975) *J. Mol. Biol. 98,* 413.
Udenfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W., & Weigele, M. (1972) *Science 178,* 871.
von Hippel, P. H., & McGhee, J. D. (1972) *Annu. Rev. Biochem. 41,* 231.
Walz, A., & Pirrotta, V. (1975) *Nature (London) 254,* 118.
Walz, A., Pirrotta, V., & Ineichen, K. (1976) *Nature (London) 262,* 665.
Weber, K., Files, J. G., Platt, T., Ganem, D., & Miller, J. H. (1975) in *Protein Ligand Interactions* (Sund, H., Blauer, G., Eds.) p 228, Walter de Gruyter, Berlin.
Weintraub, H., & Van Lente, F. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71,* 4249.
Westmoreland, B. C., Szybalski, W., & Ris, H. (1969) *Science 163,* 1343.
Witkin, E. W. (1976) *Bacteriol. Rev. 40,* 869.

Variation in the Glycosylation Pattern of Bovine κ-Casein with Micelle Size and Its Relationship to a Micelle Model†

Charles W. Slattery

ABSTRACT: Bovine casein micelles were separated into seven size classes and the supernatant by differential centrifugation. The κ-casein was isolated from each class, and $s^0_{20,w}$, the weight-average molecular weight (M_w), the hexose content, and the proportion glycosylated were determined for each. Glycosylation appears to follow complete micelle formation so that surface or near-surface κ-casein is modified while that incorporated into the interior is unchanged. The largest micelles seem to have little interior κ-casein but instead act as a coat-core system with κ-casein all on the surface. The single class of largest micelles, fraction 1, contained κ-casein with the highest M_w, the largest amount of associated hexose and a proportion glycosylated approaching that of supernatant κ-casein. The proportion of glycosylated κ-casein in the remaining micelle size classes appeared to be inversely related to micelle size with the exception of fraction 2 which probably contained some fraction 1-type micelles. The hexose to glycosylated protein ratio generally followed the M_w and $s^0_{20,w}$ values. Further fractionation of the κ-caseins according to polymer size by gel filtration revealed that, except for the κ-casein from the largest micelles, hexose content and the hexose to glycosylated protein ratio were greatest for the larger sized polymers and decreased slightly as polymer size decreased. A higher hexose content for κ-casein polymers of high molecular weight than for those of lower molecular weight would favor a previously proposed model [Slattery, C. W., & Evard, R. (1973) *Biochim. Biophys. Acta 317,* 529].

The casein micelle system of bovine milk depends for its stability on the presence of κ-casein (Waugh & Von Hippel, 1956), but the stabilization mechanism has not yet been completely determined. Like the other caseins, κ-casein is a phosphoprotein, but dephosphorylation has little, if any, effect on its stabilizing ability (Pepper & Thompson, 1963). It is the only major casein that contains sulfhydryl groups and purified preparations contain disulfide-linked polymers ranging in size from trimers (Swaisgood & Brunner, 1963; Swaisgood et al., 1964) to decamers or larger (Talbot & Waugh, 1970). However, the extent of its association into covalent polymers has very little effect on the stabilizing capacity of the κ-casein (Talbot & Waugh, 1970). A constituent of κ-casein which does seem to have an effect on its stabilizing ability is the carbohydrate which may be associated with it. When κ-casein is glycosylated, the terminal sugar residue is very often sialic acid (Fiat et al., 1972). The sialic acid is, of itself, not necessary for micelle stability (Gibbons & Cheeseman, 1962) but desialylyzed κ-casein has less ability to stabilize α_s-casein against precipitation by Ca^{2+} ion than does intact κ-casein (Thompson & Pepper, 1962).

The relationship of carbohydrate to micelle structure was also explored by Creamer et al. (1973) who separated milk micelles into "large" and "small" by sedimentation and measured the extent of glycosylation of the κ-casein for these two size classes. They found a greater proportion of the κ-casein to be glycosylated in the large micelles than in the small and suggested therefore that the glyco-κ-casein probably was mainly in the micelle interior while the carbohydrate-free κ-casein was mainly on the micelle surface. This concept appears to be at variance with what is usually seen as a mechanism for stabilization by carbohydrate and calls for further investigation. This paper reports the results of more extensive

† From the Department of Biochemistry, Loma Linda University School of Medicine, Loma Linda, California 92354. *Received July 15, 1977.*

0006-2960/78/0417-1100$01.00/0