Amino Acid Sequence of Rabbit Skeletal Muscle Myosin Light Chain Kinase†

Koji Takio,*,‡ Donald K. Blumenthal,§ Kenneth A. Walsh,‖ Koiti Titani,‖ and Edwin G. Krebs‡
Howard Hughes Medical Institute and Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received June 25, 1986; Revised Manuscript Received August 29, 1986

ABSTRACT: The amino acid sequence of the amino-terminal, 235-residue segment of rabbit skeletal muscle myosin light chain kinase has been determined. Together with the carboxyl-terminal segment previously described [Takio, K., Blumenthal, D. K., Edelman, A. M., Walsh, K. A., Krebs, E. G., & Titani, K. (1985) *Biochemistry 24*, 6028], the present work completes the 603-residue sequence of this protein. The amino-terminal segment that has been analyzed herein corresponds to a domain reported to be of highly asymmetrical shape and as yet unknown function. Secondary structure calculations failed to provide any evidence of α-helix or β-structures, but polyproline II like helical structure is possible. Sequence analysis indicates the presence of approximately equal quantities of two isoforms differing in a single amino acid replacement. Unexpected difficulties were encountered in the present sequence analysis due to the presence of acid-labile Asp–Pro bonds and to five separable variants of a blocked 21-residue amino-terminal peptide, arising from rearrangement at an Asn–Gly bond.

Protein kinases are known to perform widespread regulatory roles in response to both intracellular and extracellular signals [cf. reviews by Krebs & Beavo (1979), Flockhart & Corbin (1982), Klee & Vanaman (1982), Nishizuka (1984), and Stull et al. (1985, 1986)]. The earliest studies of the protein kinases revealed that their phosphorylation of serine or threonine residues in target proteins controlled enzyme activities involved in glycogen metabolism. More recently, several proteins encoded by viral oncogenes and several cellular receptors for hormones or mitogens have been shown to possess protein kinase activity directed toward tyrosine residues.

Sequence analysis of several members of this diverse group of enzymes has indicated that they possess homologous catalytic domains fused to segments that may be responsible for mediating the regulatory responses of the kinases or participating in the specificity of their interactions. The sequence of the catalytic subunit of cAMP-dependent protein kinase was the first described (Shoji et al., 1981), and it was found to be homologous to the Rous sarcoma tyrosine kinase, oncogene product $pp60^{src}$ (Barker & Dayhoff, 1982). Sequence analysis of the cGMP-dependent protein kinase from bovine lung (Takio et al., 1984a) and of the calcium-dependent catalytic subunit of phosphorylase *b* kinase from rabbit skeletal muscle (Reimann et al., 1984) extended the homologous relationship among the protein kinases and provided the chemical basis for the hypothesis that all protein kinases may have evolved from a common ancestral prototype. Recent reports of receptor sequences (Ullrich et al., 1984, 1985) and of a variety of oncogene products (Sefton & Hunter, 1984; Pike & Krebs, 1986) provide a wealth of evidence in support of the concept that protein kinases comprise a family of evolutionarily related enzymes.

Myosin light chain kinase (MLCK)[1] catalyzes the phosphorylation of a specific light chain of myosin in skeletal muscle (Frearson & Perry, 1975). Other enzymes in smooth muscle and nonmuscle tissue share with the skeletal muscle enzyme a high degree of substrate specificity and a complete dependence of activity on calcium/calmodulin [reviewed in Stull et al. (1986)]. Phosphorylation of the light chain in skeletal muscle is involved in modulating the tension produced during contraction (Manning & Stull, 1982), whereas in smooth muscle it appears to be required for initiation of contraction (Kamm & Stull, 1985). The role of myosin phosphorylation in nonmuscle function is not well understood.

We have undertaken sequence analysis of rabbit skeletal muscle MLCK in order to understand better the molecular basis of its regulation and action and to compare it with the growing family of related protein kinases. We reported previously the amino acid sequence of the 368-residue carboxyl-terminal segment of MLCK (Takio et al., 1985). That segment included a catalytically active, calcium/calmodulin-regulated fragment produced by limited proteolysis of the protein (Edleman et al., 1985) and a short carboxyl-terminal segment that bound calmodulin in the presence of calcium (Blumenthal et al., 1985). Together these two domains account for both the catalytic activity of MLCK and its responsiveness to calcium/calmodulin. The present study is a continuation of that work and provides the remaining 235-residue amino acid sequence at the amino terminus of the protein.

Materials and Methods

Procedures for preparation of reduced, carboxymethylated (*S*-CM) MLCK, the citraconylated protein, the cyanogen bromide cleavage product M1, and the large product of limited tryptic hydrolysis (Tα) were described by Takio et al. (1985). Proteolytic enzymes were obtained from the following sources: *Staphylococcal aureus* V8 protease from Miles, TPCK–trypsin from Worthington, and thermolysin from Calbiochem-Behring. *Achromobacter* protease I was a gift of Dr. T. Masaki (Ibaraki University, Japan). Bacterial collagenase, a product of Advance Biofactures Corp., was a gift from Dr. Paul Bornstein, University of Washington.

The columns used during reversed-phase HPLC were from Beckman Instruments (Altex Ultrapore RPSC, with C3

† This work was supported in part by a grant from the National Institutes of Health (GM-15731). D.K.B. was supported as a postdoctoral fellow of the Muscular Dystrophy Association.
* Address correspondence to this author.
‡ Howard Hughes Medical Institute.
§ Present address: Department of Biochemistry, University of Texas Health Center at Tyler, Tyler, TX 75710.
‖ Department of Biochemistry.

[1] Abbreviations: MLCK, myosin light chain kinase; *S*-CM, *S*-carboxymethyl; HPLC, high-performance liquid chromatography; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; Tris·HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/86/0425-8049$01.50/0

groups, and Ultrasphere, with C8 groups), Varian (Micropak M CH5n, an end-capped C18 column), and SynChrom (SynChropak RP-P, with C18 groups). With each column, gradients of acetonitrile were generated in dilute aqueous trifluoroacetic acid (Mahoney & Hermodson, 1980). LKB was the source of TSK G3000SW and G2000SW columns used for molecular exclusion chromatography in 6 M guanidine hydrochloride/10 mM phosphate, pH 6.0.

Enzymatic digestions were performed at 37 °C with substrate/enzyme weight ratios of 20–100 for 2–6 h in 0.1 M NH_4HCO_3 (pH 8.0) unless otherwise specified. Digestion with *S. aureus* V8 protease used 50 mM potassium phosphate buffer, pH 7.8. Digestion with *Achromobacter* protease at lysyl residues utilized 50 mM Tris·HCl (pH 9.0) with a substrate/enzyme weight ratio of 300 (Masaki et al., 1981). Collagenase digestion took place in 10 mM calcium acetate, 10 mM *N*-ethylmaleimide, and 50 mM Tris·HCl at pH 7.4 for 8 h.

Subdigests of fragment Rc with trypsin and of fragment M1 with either thermolysin or *Achromobacter* protease were separated on Synchropak RP-P columns. A thermolytic digest of Rc was fractionated on an Ultrapore RPSC column. The tryptic digest of Ra and the collagenase digest of M1 were first separated by size on TSK columns; pooled fractions were then purified on Ultrapore RPSC columns.

Amino acid analyses, sequenator analyses with an Applied Biosystems sequencer, and identifications of phenylthiohydantoins followed previously published procedures (Takio et al., 1983, 1984b; Glajch et al., 1985). A Beckman Model 990B synthesizer and standard procedures (Glass, 1983) were used to prepare an N^{α}-acetylated peptide corresponding to residues 1–18, but with Asn-4 replaced by Asp.

RESULTS

Our prior study of MLCK (Takio et al., 1985) established the amino acid sequence of the 368-residue, carboxyl-terminal segment, including a 359-residue proteolytic fragment Tβ (Figure 1) that retained calmodulin-dependent kinase activity. At that time, a large amino-terminal segment remained unsolved, but as illustrated in Figure 1, several large peptides generated during analysis of the carboxyl-terminal segment provided appropriate source material for completion of the sequence determination. One of these, denoted M1, was the largest product of cyanogen bromide cleavage. It was derived from the amino-terminal half of the protein, and it overlapped Tβ by 61 residues. Another fragment, Tα, the largest product of limited tryptic proteolysis, also included a portion of the unsolved segment. Finally, three products of arginyl cleavage, Ra, Rb, and Rc, had been assigned to the amino-terminal segment. The present study completes the determination of the amino acid sequence of fragment M1 and establishes the general relation of these five fragments to each other and to the whole MLCK molecule.

The specific strategy for the determination of the amino acid sequence is summarized in Figure 2. A major portion of that sequence was analyzed in a set of peptides resulting from cleavage at glutamyl residues and overlapped by a set cleaved at lysyl residues. As will be described, major difficulties were encountered in the amino-terminal 40-residue segment due to a blocked amino terminus and to an unusual combination of chemically labile dipeptide sequences.

Glutamyl Cleavage Products. Peptides generated by cleavage of fragment M1 (30 nmol) at Glu residues are designated by the prefix M1-E in Figure 2. The purification protocol employed is illustrated in Figure 3, and the compositions of the peptides are listed in Table I. Some peptides

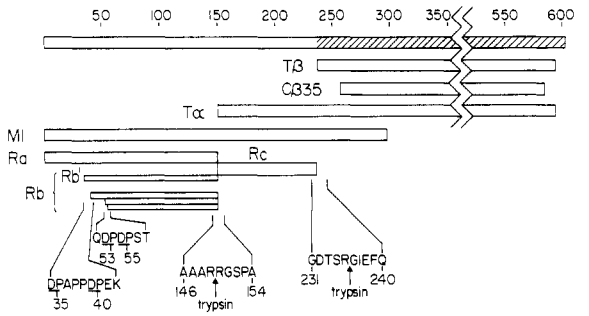

FIGURE 1: Relative alignment of major fragments isolated from MLCK. The amino acid sequence of residues 236–603 (hashed segment of upper bar) was reported by Takio et al. (1985). That segment included the calmodulin-dependent active fragment Tβ (residues 236–594) and the calmodulin-independent active fragment Cβ35 (residues 256–584; Edelman et al., 1985). The present report describes the proof of sequence of M1, the large amino-terminal fragment generated by cleavage at Met-296 (Takio et al., 1985). Cleavage of citraconylated *S*-CM-MLCK with trypsin at Arg-149 and Arg-235 excised peptide Rc and generated the five overlapping fragments labeled Ra and Rb. The various forms of Rb result from incomplete random cleavage of acid-labile Asp–Pro bonds to yield artifactual fragments with amino-terminal prolyl residues 35, 40, 53, or 55. Fragment Tα is a large active fragment, derived by limited tryptic hydrolysis (Takio et al., 1985) and found to have the same amino-terminal sequence as Rc.

in this set were also isolated from an analogous digest of the mixture of related arginyl fragments designated Rb in Figure 1; these bear Rb-E prefixes in Figure 2. Edman degradation of 11 of these peptides placed 179 residues in nonoverlapping sequences (Figure 2). One of them, peptide M1-E28 (residues 219–238), appeared to overlap the amino terminus of segment Tβ (Gly-Ile-Glu). Others (not shown) also corresponded to sequences within the M1 portion of Tβ (i.e., residues 239–296). A pair of peptides designated M1-E10g and Rb-E10v had identical sequences except that residue 92 was glycine in M1-E10g and valine in Rb-E10v. These two peptides were recovered in low but similar yields (Table I), suggesting that they were derived from equally distributed isoforms of the enzyme.

Arginyl Cleavage Products. Cleavage of citraconylated *S*-CM-MLCK at arginyl residues (Takio et al., 1985) had generated a virtually complete set of arginyl peptides from fragment Tβ (residues 236–594) and in addition three large arginyl peptides denoted Ra, Rb, and Rc. In the present study, a slightly modified isolation procedure was used (Figure 4). Amino-terminal analysis of each fragment revealed that Ra was largely blocked, although the phenylthiohydantoin of proline was observed in about 5% yield in the first cycle of Edman degradation. The amino terminus of fragment Rc (residues 150–236) corresponded to that of segment Tα (Figure 1). Edman degradation of Rb indicated amino-terminal proline, but in the second cycle, Ala, Glu, and Asp were all observed. Subsequent cycles suggested a minimum of four sequences. Attempts to resolve this mixture (Figure 4B) led to the isolation of the component denoted Rb′. When the complete sequence was established, it became clear that the four components of Rb (of which only Rb′ was isolated) corresponded to overlapping artifactual products of random cleavage of four Asp–Pro bonds (Figure 1) during the acidic decitraconylation process.

To clarify the relationship of Rb to Ra, and to provide a set of small peptides overlapping the glutamyl peptides, 1 nmol each of Ra and Rb was digested with *Achromobacter* protease, and their peptide profiles were compared (Figure 5). Many peptides were found to be common to both Ra and Rb, but some (e.g., Kα and K2) were observed only in Ra. Others,

Table I

Amino Acid Compositions of Peptides Resulting from Cleavage of Fragment M1 with S. Aureus Protease[a]

Peptide	E1	E2[b]	E3	E4	E8	E9[c]	E10v[c]	E10g	E15	E17[c]	E19	E23	E25	E28
Residues	1-3	4-8	9-18	19-27	42-68	69-88	89-114	89-114	118-121	122-143	144-169	187-203	204-211	219-238
Asx (D/N)		0.7 (1)	0.8 (1)		1.9 (5)	1.3 (1)				0.7 (1)	0.6 (0)		0.7 (1)	0.7 (2)
Glx (E/Q)	1.0 (1)	1.0 (1)	1.9 (2)	1.9 (2)	2.2 (2)	2.8 (3)	4.9 (6)	5.4 (6)	0.7 (1)	2.3 (4)	2.0 (2)	1.8 (2)	0.8 (1)	4.0 (4)
Cys (C)											0.3 (1)			
Ser (S)			1.6 (2)	1.3 (2)	2.1 (2)	1.7 (2)	1.2 (1)	1.3 (1)		0.9 (1)	3.0 (4)			1.3 (1)
Gly (G)		1.4 (1)	1.5 (1)	1.1 (1)	3.2 (2)	4.5 (5)	5.1 (4)	5.3 (5)		3.1 (4)	2.6 (2)	3.0 (3)	2.0 (2)	4.0 (3)
His (H)										0.5 (1)	0.5 (1)			
Arg (R)											1.7 (2)			0.6 (1)
Thr (T)	0.7 (1)		0.8 (1)		1.4 (3)		2.0 (2)	1.9 (2)		0.9 (1)	0.9 (1)	1.7 (2)		0.8 (1)
Ala (A)	1.0 (1)	1.0 (1)		3.0 (3)	2.0 (2)	3.0 (3)	7.0 (7)	7.0 (7)		2.0 (2)	6.0 (6)	3.0 (3)	1.1 (1)	3.0 (3)
Pro (P)					8.7 (7)	3.7 (3)	4.6 (3)	4.3 (3)	1.3 (1)	3.2 (3)	4.2 (3)	6.5 (5)		2.5 (1)
Val (V)		1.0 (1)				1.5 (2)	1.5 (1)	0.5 (0)		1.4 (2)		1.1 (1)	0.9 (1)	1.1 (1)
Ile (I)			1.0 (1)								2.6 (2)			0.8 (1)
Leu (L)			2.0 (2)				1.5 (1)	1.3 (1)			1.5 (1)		2.0 (2)	
Phe (F)											1.4 (1)			
Lys (K)				0.9 (1)	1.8 (4)	0.7 (1)	1.3 (1)	1.4 (1)	2.0 (2)	2.4 (3)		0.7 (1)		1.6 (2)
Total	(3)	(5)	(10)	(9)	(27)	(20)	(26)	(26)	(4)	(22)	(26)	(17)	(8)	(20)
Yield (%)	25	N.D.	27	40	14	40	5	8	33	30	80	53	45	60

[a]Residues/molecule by Picotag amino acid analysis or (in parentheses) from the sequence. These peptides have the prefix M1- in Figure 2.

[b]From a digest of Kα (corresponds to peptide b in Figure 7).

[c]The corresponding peptide, isolated from a subdigest of fragment Rb, was identified by the prefix Rb- and sequenced (Figure 2).

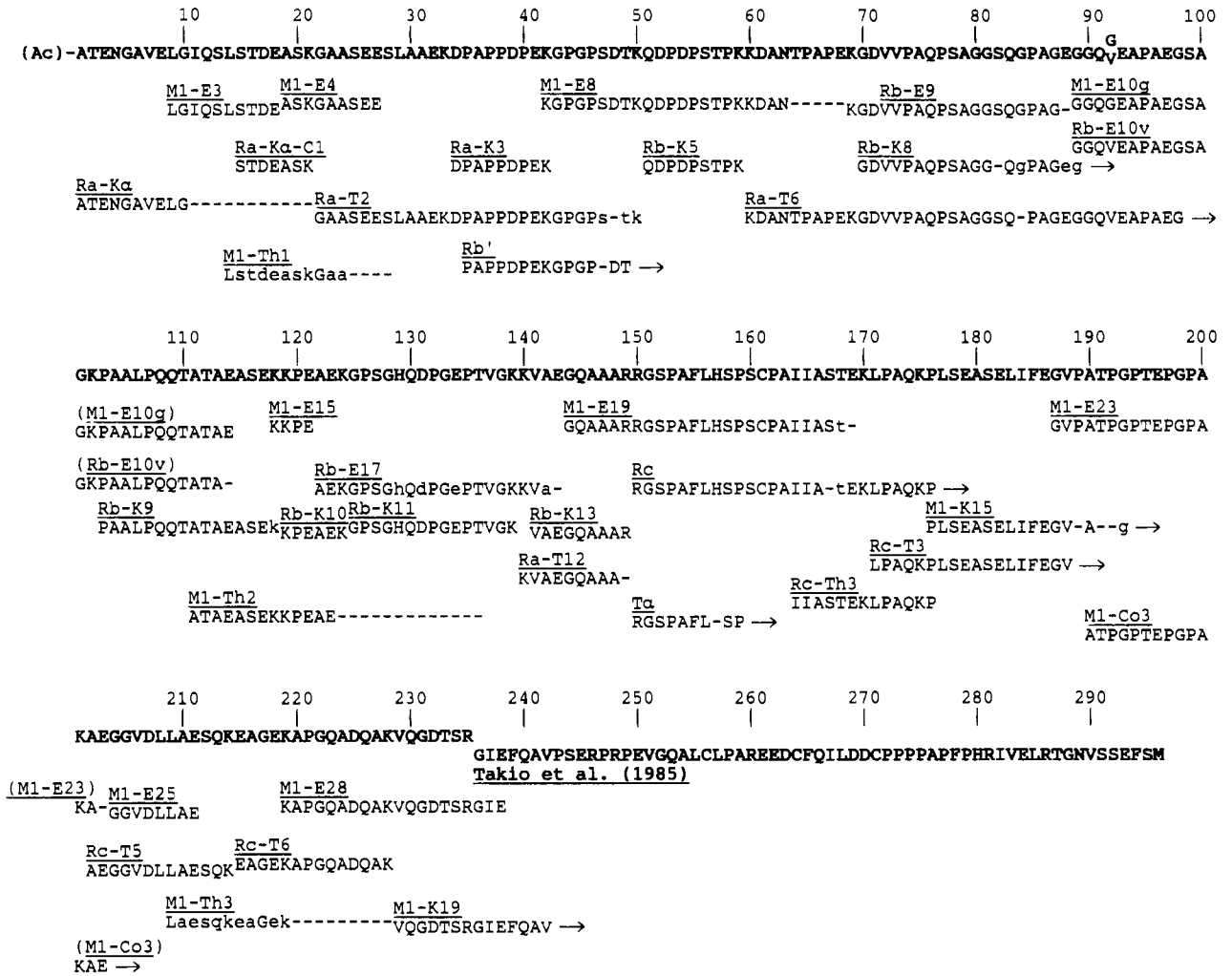

FIGURE 2: Detailed summary of the proof of sequence of MLCK. Sequences derived by Edman degradation of specific peptides (names are underlined) are given below the summary sequence (bold type). Prefixes R and M refer to peptides derived by cleavage at arginyl and methionyl residues. Subpeptides are identified by hyphenated suffixes where E, T, K, Th, C, and Co denote cleavage by *S. aureus* V8 protease, trypsin, *Achromobacter* protease I, thermolysin, chymotrypsin, or collagenase, respectively. The large fragments M1, Ra, Rb, Rb′, Rc, and Tα are identified in Figure 1. Within each specified sequence, lower case letters denote tentative identification of phenylthiohydantoins. Those not identified are indicated by dashes or by an arrow for a long unidentified sequence. The sequence of M1-Th3 was deduced from the degradation of its mixture with M1-Th1. The sequence of residues 236–296 was reported previously (Takio et al., 1985). At residue 92, both Val and Gly were observed (in about equal quantities). The amino-terminal alanyl residue is N-acetylated. Proof of sequence of Ra-Kα (residues 1–21) is summarized in Figure 7.

unique to Rb (marked with asterisks in Figure 5), corresponded to shorter segments of peptides missing from the Ra profile. An explanation of these patterns was suggested by the results of Edman degradation of Ra-K3 and Rb-K5, where a total of four Asp-Pro sequences were found between residues 34 and 55 (Figure 2). These dipeptide sequences are known to be acid labile (Landon, 1977), thus accounting for the artifactual generation of fragments with an amino-terminal proline such as fragment Rb′ that corresponds to the product of Asp-Pro cleavage in Ra-K3 (residues 34–35).

Other Overlapping Peptides. Sequence analyses of the various lysyl cleavage products of Ra and Rb, their counterparts from tryptic cleavage of Ra and Rc, and lysyl cleavage products of M1 provided the various peptides identified with K or T suffixes in Figure 2 and, in turn, overlaps of the glutamyl peptides that aligned residues 19–118, 119–202, and 215–238. It remained only to determine the sequence from the amino terminus to residue 19, to overlap residues 118–119, and to provide a bridge from residue 202 to residue 215.

A thermolytic digest of 3 nmol of M1 provided a peptide (M1-Th2, residues 111–136) that overlapped Rb-K9 and Rb-K10, completing the sequence of residues 19–202. Another peptide from the same digest (M1-Th3, residues 209–228) was found as an equimolar mixture with M1-Th1 (residues 14–28). Edman degradation of the mixture identified 11 residues of the sequence of M1-Th3 after subtracting the sequence (described below) of residues 14–24. The deduced sequence overlapped Rc-T5 and Rc-T6, extending the peptide chain to Glu-238. This sequence, and one derived from M1-K19, provided unambiguous overlaps to the previously determined structure of fragment Tβ (residues 236–603).

Amino-Terminal Region. The amino terminus of MLCK was blocked as were fragments Ra, E1, and Kα. As a consequence, the amino-terminal sequence could not be determined by Edman degradation of the whole protein. As with Ra, Edman degradation of the large cyanogen bromide fragment M1 released only the phenylthiohydantoin of proline in about 20% yield in the first cycle, and subsequent cycles indicated a mixture of four sequences in even lower yields. Although preparations of M1 thus appeared to comprise a blocked chain contaminated with four minor chains, each with amino-terminal proline, M1 was treated as if it were a single fragment in the subdigestions already discussed. All of the resulting data fit a single sequence, with "ragged ends" (Figure

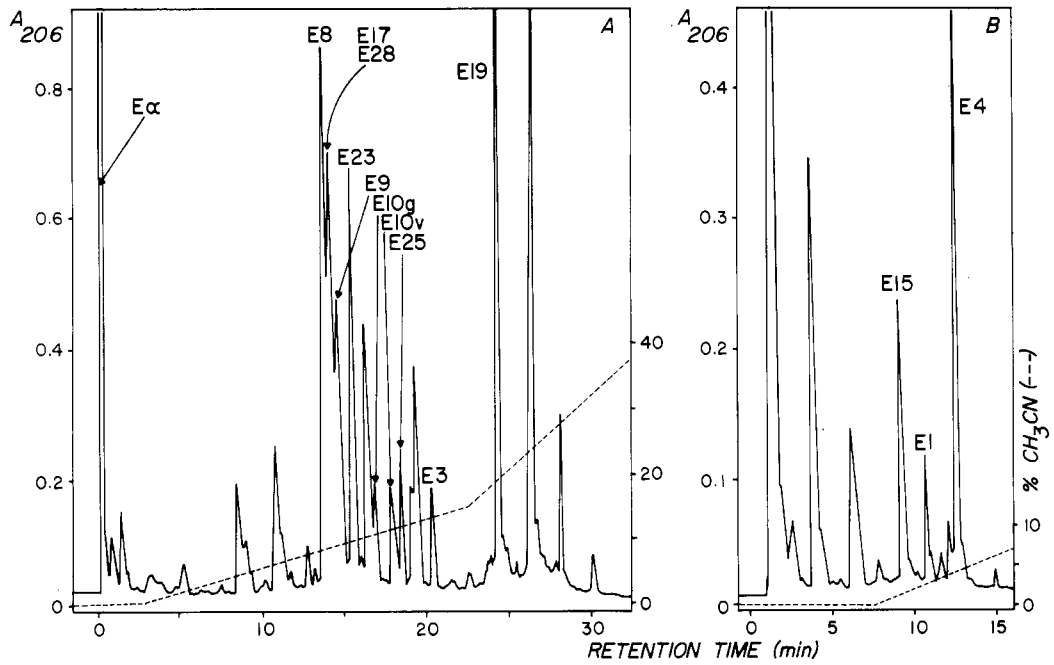

FIGURE 3: Separation of peptides generated by cleavage of fragment M1 (30 nmol) with the *S. aureus* V8 protease. (A) Reversed-phase HPLC at 2 mL/min used an acetonitrile gradient on an Ultrapore RPSC column. (B) Pooled breakthrough fraction Eα was resolved on a Varian MCH5n column at 1 mL/min. The only peptides identified are those used in the strategy illustrated in Figure 2 (and in Table I) with the appropriate prefix M1 or Rb. Unnamed fractions contain peptides redundant with segments of sequence more simply derived in other digests. Pooled fractions containing E8, E10v, E17, and E28 were further fractionated by HPLC.

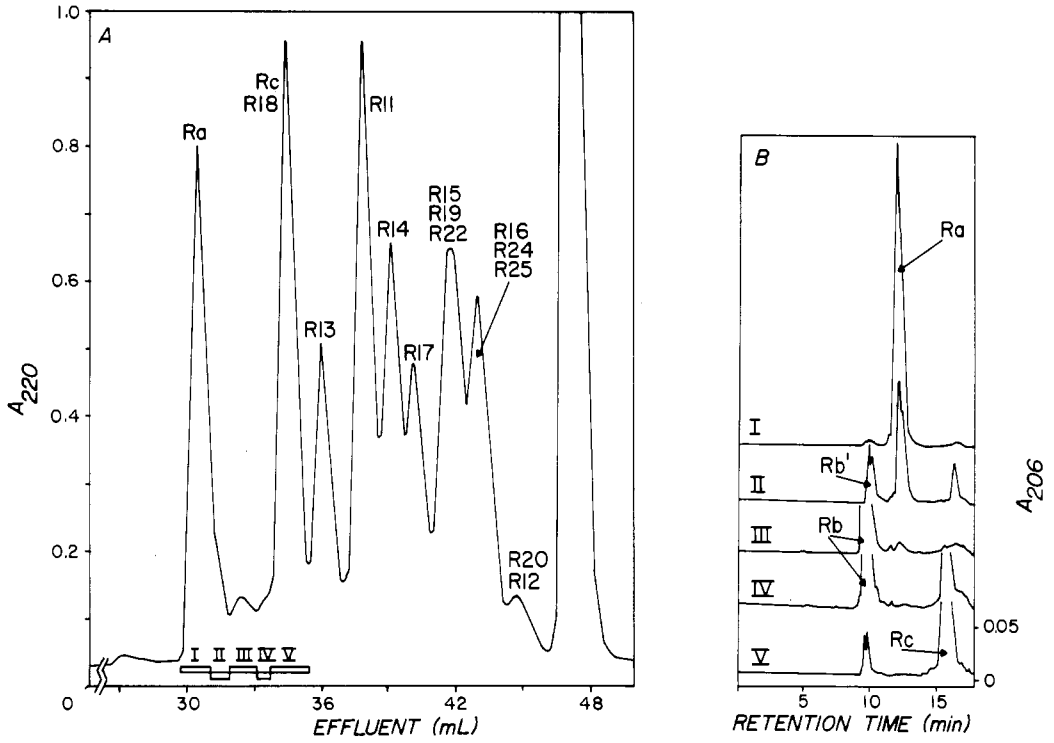

FIGURE 4: (A) Products of arginyl cleavage of 20 nmol of citraconylated *S*-CM-MLCK. The digest was decitraconylated in 9% formic acid for 3 h (37 °C) and then separated on two TSK G3000SW columns in series using 10 mM phosphate (pH 6) containing 6 M guanidine hydrochloride at a flow rate of 0.4 mL/min. Peptides R11 through R25 were analyzed previously in the carboxyl-terminal segment, residues 236–603 (Takio et al., 1985). Pooled fractions I through V were each separated further on an Ultrapore RPSC column (B) (2 mL/min, gradient 2%/min) to generate fragments Ra, Rb′, Rb, and Rc. Fractions denoted Rb contained some Rb′. Peptide R18 (not shown) eluted after Rc. Only 20% and 25% of pooled fractions I and V, respectively, were applied to the Ultrapore column.

1) artifactually generated by Asp–Pro cleavage during the various acidic steps in the preparative procedures.

Attempts to analyze the sequence of residues 1–19 encountered the enigma of five separable variants of peptide Kα (Figure 5), all with identical amino acid compositions (cf. Kα4 in Table II). Moreover, all five peptides were blocked, preventing direct comparison of their sequences. Treatment of Kα5 with acid (12 N HCl, 15 h, 20 °C) promoted an N to O acyl shift that revealed the sequence TEDGAVELG, tentatively overlapping residues 9 and 10 in M1-E3. Control experiments with Asn and Gln indicated that >60% of these amino acids remained as amides during this treatment. Each of the five Kα variants (3 nmol of each) was then subdigested at glutamyl residues with the *S. aureus* V8 protease to generate

Table II: Amino Acid Composition of MLCK and of Its Key Fragments[a]

residues	MLCK[b] 1–603	Ml[c] 1–296	Ra[d] 1–149	Rc[d] 150–235	Kα4[e] 1–21
Asx (D/N)	53.8 (53)	17.7 (18)	9.9 (11)	3.2 (3)	1.7 (2)
Glx (E/Q)	85.3 (82)	48.5 (50)	26.1 (26)	13.8 (14)	4.0 (4)
Cys (C)	10.1 (10)	3.8 (4)		0.6 (1)	
Ser (S)	39.7 (40)	20.1 (24)	10.2 (12)	6.4 (8)	2.6 (3)
Gly (G)	48.9 (48)	30.2 (31)	18.9 (19)	8.6 (9)	2.6 (2)
His (H)	10.5 (10)	2.5 (3)	0.9 (1)	0.6 (1)	
Arg (R)	24.5 (21)	8.5 (8)	1.0 (1)	2.0 (2)	
Thr (T)	27.1 (28)	11.9 (13)	6.9 (8)	3.5 (4)	1.6 (2)
Ala (A)	60.0 (62)	40.7 (43)	25.7 (26)	11.8 (13)	3.0 (3)
Pro (P)	47.0 (50)	37.9 (42)	20.7 (21)	9.7 (11)	
Tyr (Y)	8.5 (8)				
Val (V)	31.2 (34)	11.1 (13)	6.0 (6)	2.9 (3)	0.9 (1)
Met (M)	11.6 (13)	0.6 (1)			
Ile (I)	23.3 (25)	5.0 (7)	1.0 (1)	2.3 (3)	1.0 (1)
Leu (L)	49.0 (49)	14.0 (14)	4.0 (4)	6.0 (6)	2.0 (2)
Phe (F)	18.4 (20)	6.5 (6)		2.0 (2)	
Lys (K)	44.2 (46)	18.7 (19)	11.7 (13)	5.6 (6)	0.9 (1)
Trp (W)	ND (4)				
total	(603)	(296)	(149)	(86)	(21)

[a] Residues per molecule by ion-exchange amino acid analysis, unless otherwise stated, or (in parentheses) from the sequence. [b] From analysis of a time course of hydrolysates of *S*-CM-MLCK. [c] Isolated by Takio et al. (1985); recalculated for 296 residues. [d] From Figure 4. [e] One of five peptides of similar composition denoted Kα in Figure 5. Analyzed by Picotag amino acid analysis.

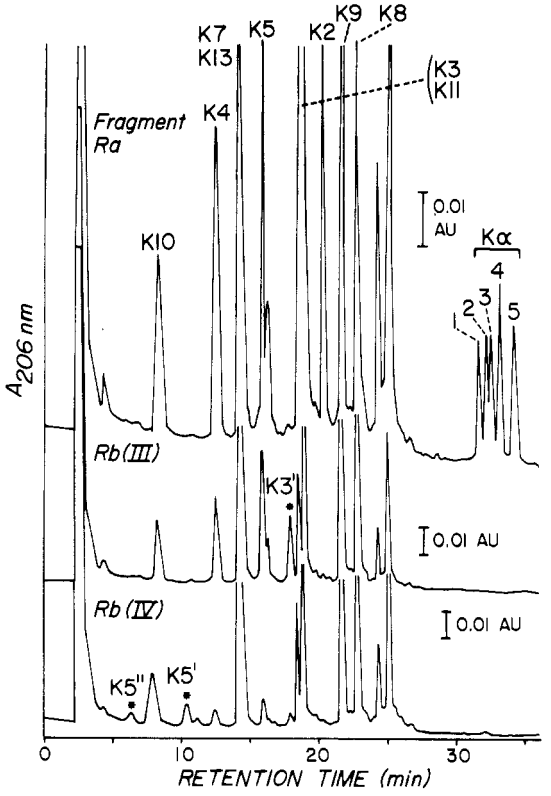

FIGURE 5: Products of lysyl cleavage of fragments Ra (from Figure 4B, chromatogram I) and Rb (from Figure 4B, chromatograms III or IV) as resolved on a SynChropak RP-P column with an acetonitrile gradient (1% increase/min) flowing at 1 mL/min. The upper chromatogram displays the peptides K2 through K13 comprising residues 22–149. Many of these peptides are placed in Figure 2 with appropriate prefixes (Ra or Rb). Whereas several peptides (e.g., K8 and K9) are found in digests of both Ra and Rb, others (e.g., K2, K4, and K5) are absent from Rb, where new peptides (indicated by asterisks) represent truncated versions. The five fractions designated Kα are found only in the digest of Ra. Each Kα variant has a composition corresponding to residues 1–21 (Table II). The differences between these variants of Kα are accounted for in Figure 7.

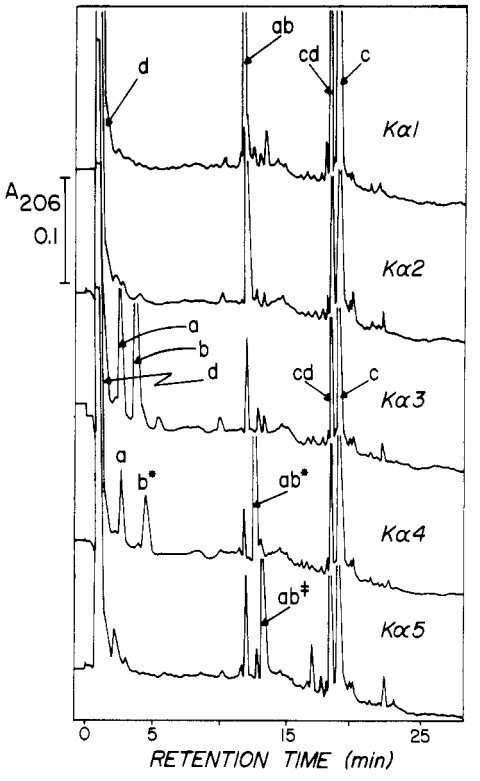

FIGURE 6: HPLC separation (on an Ultrasphere C8 column with a 40% gradient in 30 min at 1.0 mL/min) of the products of cleavage at glutamyl residues of Kα fractions 1–5 (Figure 5). The middle chromatogram illustrates the mobility of four primary products from Kα3, labeled a through d, that correspond to the four segments illustrated in Figure 7. Peptide c, from each of the five forms of Kα, corresponds to M1-E3 (residues 9–18) in Figure 2. Peptide d (residues 19–21) and an overlapping peptide cd (residues 9–21) also occur throughout. The differences between Kα fractions 1–5 are related to differences in peptide b (residues 4–8), which by composition is recognized as b* from Kα4, and in the overlapping peptides ab (from Kα1 and Kα2), ab* (Kα4), and ab‡ (Kα5). Peptide a is found both free (from Kα4) and in the various forms of peptide ab.

five sets of peptides (Figure 6). Peptides c and d and the overlapping peptide cd were found in such digests of each of the Kα variants and correspond to residues 9–21 (Figure 7). Peptide a was recovered from peptides Kα3 and Kα4 but found as an overlap (ab) with peptide b from the three other forms of Kα. Peptides a and ab were placed at the amino termini by virtue of their blocked ends. As with Kα5, Edman degradation of peptides a and ab required acid pretreatment to

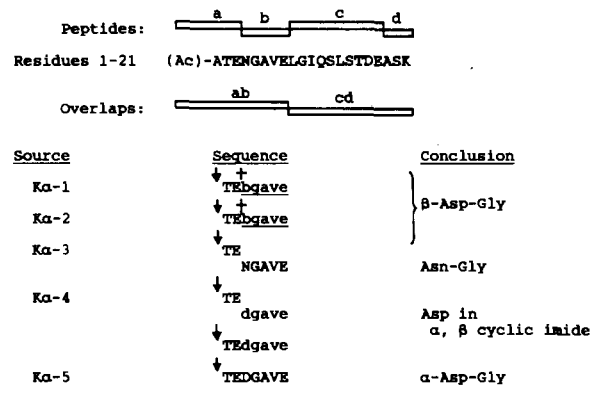

FIGURE 7: Analysis of the amino-terminal region of MLCK. The five variants of the amino-terminal peptide Kα (Figure 5) were each digested at glutamyl residues (Figure 6) to form four primary peptides (as defined by their compositions) indicated by bars a through d at the top of the diagram (or their overlaps ab and cd). The vertical arrow indicates that peptides a (and ab) were deblocked by treating with acid to promote an N to O acyl migration (releasing acetyl-Ala) prior to Edman degradation. Results of the Edman degradation are indicated with the same code used in Figure 2. The dagger symbol indicates that Edman degradations of peptides Kα1-ab and Kα2-ab became blocked after two cycles; the underlined portions of these peptides are defined by composition only. Degradation of Kα4-b was characterized by a low yield in the first cycle and an unusually large overlap in subsequent cycles. Similarly, the third cycle of Kα4-ab suffered a marked drop in yield, and there were large subsequent overlaps.

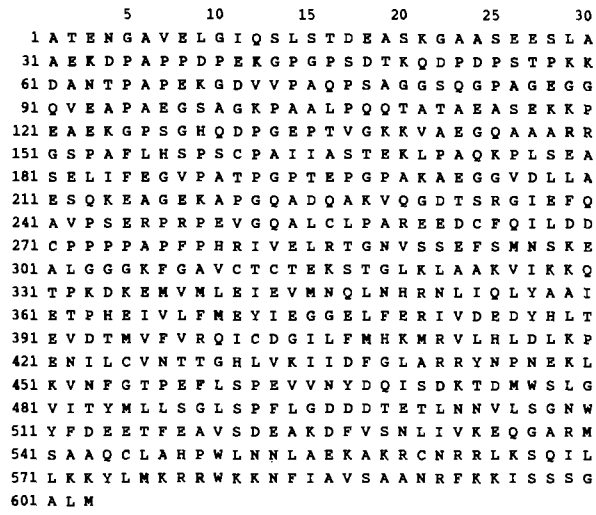

FIGURE 8: Amino acid sequence of MLCK. Residues 1–235 were placed in the present study; the sequence of residues 236–603 was reported previously (Takio et al., 1985). There are two isoforms of the enzyme, one (illustrated) with valine at residue 92 and the other with glycine at that position. The amino terminus is acetylated. From considerations of homology, the catalytic domain appears to comprise residues 302–508 (Takio et al., 1985) and the calmodulin-interacting segment residues 577–595 (Blumenthal et al., 1985).

expose amino-terminal Thr-2 by the N to O acyl shift.

The differences between the five forms of Kα appear to be explained by differences in peptide b, which was determined to have the sequence NGAVE in Kα3 but DGAVE within ab from Kα5 (where the Glu–Asp bond is not cleaved by the V8 protease). The sequence of residues 2–8 in native MLCK must be TENGAVE (as in Kα3), and rearrangement at the Asn–Gly bond gives rise to β-aspartyl-Gly in Kα1 and Kα2, to α-aspartyl-Gly in Kα5, and to the cyclic imide intermediate in Kα4 [cf. Bornstein & Balian (1977)]. These conclusions are in accord with the blockage of the Edman degradation at the third cycle of Kα1-ab and Kα2-ab and with the anomalous degradations of Kα4-b and Kα4-ab (Figure 7), where partial opening of the cyclic imide would be expected to generate some α-Asp-Gly in each cycle. The difference between Kα1 and Kα2 is not apparent. Perhaps the α-carboxyl group in one has formed an ester bond with the side chain of Thr-2. Jörnvall (1974) described the formation of analogous diketopiperazine structures from peptides containing amino-terminal X-Asn-Gly– structures.

The amino acid compositions of peptides M1-E1 (Table I), Kα4 (Table II), Kα3-a, and Kα4-a indicate that the amino-terminal residue in each case must be alanine. The following evidence indicated that the α-amino group was acetylated, as in many other proteins. The alanine-containing fragment released by the N to O shift coeluted with acetyl-Ala on a Varian MicroPak MCH-5n reversed-phase column. In addition, peptide M1-E1 coeluted with synthetic acetyl-Ala-Thr-Glu on both Ultrasphere C8 and MicroPak MCH-5n columns.

The four small peptides from the Kα variants are aligned as indicated in Figure 7 because the overlaps ab and cd allow for no other arrangement. The cd overlap was confirmed by the sequence of a chymotryptic fragment of Kα (denoted Ra-Kα-C1), and the overlap of Kα to Ra-T2 is provided by M1-E4 (Figure 2).

Search for Homology. It was shown previously (Takio et al., 1985) that a portion of the carboxyl-terminal segment of the protein [residues 302–508, using the numbering system in this report (Figure 8)] was homologous with the extended family of protein kinases exhibiting catalytic activity toward serine, threonine, or tyrosine residues, whatever their mode of regulation. A search for other homologous domains within the amino-terminal segment of MLCK was carried out by using similar procedures. Two proteins of possible interest were identified by the search, a proline-rich phosphoprotein from human saliva (Wong & Bennick, 1980) and the family of collagens. However, detailed statistical analysis of the significance of their relationship to the amino-terminal segment of MLCK indicated that there was no homology. Apparently, the Gly-Pro-rich collagen chains and the Pro-rich parotid protein were selected from the data bank on the basis of their unusual composition rather than their similar sequences.

Discussion

Analysis of the amino acid sequence of the amino-terminal 235-residue segment of rabbit skeletal muscle MLCK proved to be a more difficult problem than would be expected from consideration of size alone. The difficulties related in part to a blocked amino terminus and to a nearby Asn-Gly structure that rearranged and interfered with conventional isolation and degradation procedures. In addition, five acid-labile Asp–Pro bonds added an element of instability to the segment, and the lack of methionyl residues precluded the usual cleavage technique of choice. As a result, it was necessary to employ an unusually wide range of less satisfactory fragmentation and subdigestion protocols.

Nonetheless, virtually every residue in the sequence was identified during Edman degradation of more than one peptide (Figure 2); each phenylthiohydantoin was examined in two complementary HPLC systems except those from residues 114–118, 165–175, and 203–215, which were identified only in the system of Glajch et al. (1985). The weakest portions of the proof are found in the marginal overlaps at residues 186–188, 201–204, and 214–215. In addition, the amino-terminal acetyl group is identified only indirectly by the behavior of small peptides on HPLC. The final result is a sequence which, together with the carboxyl-terminal 368-residue

segment, is in accord with the amino acid composition of the 603-residue *S*-CM protein (Table II, Figure 8).

Two different amino acids, valine and glycine, were found at residue 92 in approximately equimolar proportions. Other than the rearrangements and lability identified with the Asn-Gly and Asp-Pro sequences, there were no other indications of microheterogeneity or polymorphism in the preparation.

The molecular weight of the whole protein is calculated to be 65 040 (for the valine isoform), and this can be compared with reported values ranging from 77 000 to 103 000 [cf. review by Stull et al. (1985)]. The consistently higher, but quite variable, prior estimates were based on gel electrophoresis or sedimentation data. One of the more recent estimates came from a detailed study of the shape and substructural characteristics of the molecule using both hydrodynamic and electrophoretic techniques (Mayr & Heilmeyer, 1983). The structure was reported to be highly asymmetric with an apparent molecular weight of 70 300. The protein could be fragmented into a globular, catalytically active "head" (M_r ~36 000) and a highly asymmetric, enzymatically inactive "tail" (M_r ~31 700) with an axial ratio >10. The tail was reported to be anomalously rich in Glx, Pro, Gly, and Ala. We have pointed out previously that the Mayr/Heilmeyer head fragment must correspond closely to our carboxyl-terminal fragment Tβ (Takio et al., 1985) and that their tail fragment represents the remaining amino-terminal segment of the protein. The precise carboxyl-terminus of their tail fragment was not determined. However, considering the patterns of limited proteolytic digestion obtained with trypsin and chymotrypsin (Edelman et al., 1985; Takio et al., 1985), it is likely that the tail segment of Mayr and Heilmeyer is not longer than 235–255 residues.

One might anticipate that the asymmetric tail segment of MLCK would have extensive secondary structure or repeating segments typical of fibrous proteins. Mayr and Heilmeyer (1983) estimated its dimensions to be 270 Å by 18 Å. However, analysis using the RELATE program of Dayhoff et al. (1983) failed to find evidence of internal homology. The hydropathic profile (Kyte & Doolittle, 1982) indicated a relatively even distribution of hydrophilic residues. Circular dichroism measurements of the tail segment by Mayr and Heilmeyer (1983) indicate only 10% α-helix and even less β-structure. Segments of secondary structure were also sought by using the predictive methods of Chou and Fasman (1978). The predominance of prolyl and glycyl residues in the composition argues against both α-helical and β-structures, as pointed out previously by Mayr and Heilmeyer (1983). However, residues 23–33 and 206–216 lack these helix breakers, and those sequences are consistent with predictions of α-helix in each case.

The segment between these putative helices is 172 residues long with the following unusual composition of amino acids (in decreasing order of frequency): Pro_{31}, Ala_{27}, Gly_{22}, Glu_{17}, Lys_{14}, Ser_{13}, Thr_9, Gln_9, Asp_7, Val_7, Leu_5, Ile_3, Phe_2, Arg_2, His_2, Asn_1, Cys_1. Just three amino acids, proline, alanine, and glycine, make up 46% of this segment, and hydrophilic residues provide another 43%. However, their distribution in the sequence fits no discernible repetitive pattern. The high proline content is reminiscent of collagen, but the distribution of glycine is incompatible with its unique triple-helical structure. Assuming that there are intrachain interactions, one is led to consider a hairpin structure with ~86 residues per limb in an antiparallel arrangement as in the polyproline II conformation suggested by Mayr and Heilmeyer (1983). In this structure, two paired antiparallel 86-residue segments would span approximately 275 Å (Segal & Traub, 1969), in approximate agreement with the dimensions of the tail fragment estimated by Mayr and Heilmeyer (1983).

Whatever the three-dimensional arrangement that accounts for the asymmetry of the tail segment, its internal sequence ARRGSP (residues 148–153) appears to be accessible to trypsin in the native enzyme. Cleavage of the Arg–Arg bond gives rise to the active fragment Tα (Takio et al., 1985). Moreover, the neighboring residue Ser_{152} is a likely site for phosphorylation by cAMP-dependent protein kinase (Edelman & Krebs, 1982). The sequence corresponds to the substrate specificity of that kinase, and the phosphoryl group is known to remain with fragment Tα (residues 150–603, the product of limited tryptic proteolysis) but not with the smaller fragment Tβ (Figure 1, residues 236–595). Further work is needed to test this assertion.

The role of the tail fragment in the function of MLCK is not clear. Its removal is known to leave behind an enzyme with catalytic activity toward myosin light chains and its characteristic response to calcium/calmodulin. The fibrous nature of the tail fragment may well contribute to the intracellular localization of MLCK in the muscle cell. Detailed knowledge of its three-dimensional structure could follow from the present solution of its primary structure and provide more definitive clues to its cellular role.

ACKNOWLEDGMENTS

We thank Maria Harrylock for able technical assistance, Roger D. Wade for the amino acid analyses, Barbara Flug and Floyd Kennedy for their help with the preparation of the protein, and Mary K. Woods for help with preparation of the manuscript.

Registry No. MLCK, 51845-53-5; MLCK (rabbit skeletal muscle isoenzyme 1 reduced), 104910-55-6; MLCK (rabbit skeletal muscle isoenzyme 2 reduced), 104910-54-5.

REFERENCES

Barker, W. C., & Dayhoff, M. O. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2836.

Blumenthal, D. K., Takio, K., Edelman, A. M., Charbonneau, H., Titani, K., Walsh, K. A., & Krebs, E. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3187.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45.

Dayhoff, M. O., Barker, W. C., & Hunt, L. C. (1983) *Methods Enzymol. 91*, 524.

Edelman, A. M., & Krebs, E. G. (1982) *FEBS Lett. 138*, 293.

Edelman, A. M., Takio, K., Blumenthal, D. K., Hansen, R. S., Walsh, K. A., Titani, K., & Krebs, E. G. (1985) *J. Biol. Chem. 260*, 11275.

Flockhart, D. A., & Corbin, J. D. (1982) *CRC Crit. Rev. Biochem. 12*, 133.

Frearson, N., & Perry, S. V. (1975) *Biochem. J. 151*, 99.

Glajch, J. L., Gluckman, J. C., Charikofsky, J. G., Minor, J. M., & Kirkland, J. J. (1985) *J. Chromatogr. 318*, 23.

Glass, D. B. (1983) *Methods Enzymol. 99*, 119.

Jörnvall, H. (1974) *FEBS Lett. 38*, 329.

Kamm, K. E., & Stull, J. T. (1985) *Annu. Rev. Pharmacol. Toxicol. 25*, 593.

Klee, C. B., & Vanaman, T. C. (1982) *Adv. Protein Chem. 35*, 213.

Krebs, E. G., & Beavo, J. A. (1979) *Annu. Rev. Biochem. 48*, 923.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105.

Landon, M. (1977) *Methods Enzymol. 47*, 145.
Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199.
Manning, D. M., & Stull, J. T. (1982) *Am. J. Physiol. 242*, C234.
Masaki, T., Tanabe, M., Nakamura, K., & Soejima, M. (1981) *Biochim. Biophys. Acta 660*, 44.
Mayr, G. W., & Heilmeyer, L. M. G., Jr. (1983) *Biochemistry 22*, 4316.
Nishizuka, Y. (1984) *Nature (London) 308*, 693.
Pike, L. J., & Krebs, E. G. (1986) in *The Receptors* (Conn, P. M., Ed.) Vol. 4, Academic Press, New York (in press).
Reimann, E. M., Titani, K., Ericsson, L. H., Wade, R. D., Fischer, E. H., & Walsh, K. A. (1984) *Biochemistry 23*, 4185.
Sefton, B. M., & Hunter, T. (1984) *Adv. Cyclic Nucleotide Protein Phosphorylation Res. 18*, 195.
Segal, D. M., & Traub, W. (1969) *J. Mol. Biol. 43*, 487.
Shoji, S., Parmelee, D. C., Wade, R. D., Kumar, S., Ericsson, L. H., Walsh, K. A., Neurath, H., Long, G. L., Demaille, J. G., Fischer, E. H., & Titani, K. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 848.
Stull, J. T., Nunnally, M. H., Moore, R. L., & Blumenthal, D. K. (1985) *Adv. Enzyme Regul. 23*, 123.
Stull, J. T., Nunnally, M. H., & Michnoff, C. H. (1986) *Enzymes (3rd Ed.) 17*, 114.
Takio, K., Smith, S. B., Walsh, K. A., Krebs, E. G., & Titani, K. (1983) *J. Biol. Chem. 258*, 5531.
Takio, K., Wade, R. D., Smith, S. B., Krebs, E. G., Walsh, K. A., & Titani, K. (1984a) *Biochemistry 23*, 4207.
Takio, K., Kominami, E., Bando, Y., Katunuma, N., & Titani, K. (1984b) *Biochem. Biophys. Res. Commun. 121*, 149.
Takio, K., Blumenthal, D. K., Edelman, A. M., Walsh, K. A., Krebs, E. G., & Titani, K. (1985) *Biochemistry 24*, 6028.
Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Liberman, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D., & Seeburg, P. H. (1984) *Nature (London) 309*, 418.
Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petruzzelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y.-C., Tsubokawa, M., Mason, A., Seeburg, P. H., Grunfeld, C., Rosen, O. M., & Ramachandran, J. (1985) *Nature (London) 313*, 756.
Wong, R. S. C., & Bennick, A. (1980) *J. Biol. Chem. 255*, 5943.

Amino Acid Specific ADP-Ribosylation: Specific NAD:Arginine Mono-ADP-Ribosyltransferases Associated with Turkey Erythrocyte Nuclei and Plasma Membranes

Robert E. West, Jr.,* and Joel Moss

Laboratory of Cellular Metabolism, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892

Received March 4, 1986; Revised Manuscript Received July 8, 1986

ABSTRACT: Turkey erythrocytes contain NAD:arginine mono-ADP-ribosyltransferases which, like cholera toxin and *Escherichia coli* heat-labile enterotoxin, catalyze the transfer of ADP-ribose from NAD to proteins, to arginine and other low molecular weight guanidino compounds, and to water. Two such ADP-ribosyltransferases, A and B, have been purified from turkey erythrocyte cytosol. To characterize further the class of NAD:arginine ADP-ribosyltransferases, the particulate fraction was examined; 40% of erythrocyte transferase activity was localized to the nucleus and cell membrane. Transferase activity in a salt extract of a thoroughly washed particulate preparation was purified 36 000-fold by sequential chromatography on phenyl-Sepharose, (carboxymethyl)cellulose, concanavalin A–Sepharose, and NAD–agarose. Subsequent DNA–agarose chromatography separated two activities, termed transferases C and A′, which were localized to the membrane and nucleus, respectively. Transferase C, the membrane-associated enzyme, was distinguished from the cytosolic enzymes by a relative insensitivity to salt and histone; transferase C was stimulated 2-fold by 300 mM NaCl in contrast to a 20-fold stimulation of transferase A and a 50% inhibition of transferase B. Similarly, histones, which stimulate transferase A 20-fold, enhanced transferase C activity only 2-fold. Transferase A′, the nuclear enzyme, was retained on DNA–agarose. It was similar to transferase A in salt and histone sensitivity. Gel permeation chromatography showed slight molecular mass differences among the group of enzymes: A, 24 300 daltons (Da); B, 32 700 Da; C, 26 000 Da; and A′, 25 500 Da. The affinities of transferase C for NAD and agmatine were similar to those of the cytosolic transferases A and B. Thus, mono-ADP-ribosyltransferases of the turkey erythrocyte occur as a group of similar enzymes whose multiple forms can be distinguished by their specific intracellular localization and their regulatory properties.

Mono-ADP-ribosylation, the transfer of the ADP-ribose moiety from NAD to specific protein substrates, is a mechanism of action common to several bacterial toxins and bacteriophage enzymes, some of which effect profound changes in cellular metabolism (Honjo & Hayaishi, 1973; Iglewski et al., 1977, 1978; Moss & Vaughan, 1977; Gill & Meren, 1978; Moss et al., 1979; Katada & Ui, 1982; Goff, 1974; Pesce et al., 1976). Among these are diphtheria toxin and *Pseudo-*

* Address correspondence to this author.