Protease digestion studies of an equilibrium intermediate in the unfolding of creatine kinase

Timothy WEBB*, Philip. J. JACKSON† and Glenn E. MORRIS*‡

*MRIC Biotechnology Group, N.E. Wales Institute, Plas Coch, Wrexham LL11 2AW, U.K., and †Applied Biosystems, PE Applied Biosystems, Kelvin Close, Birchwood Science Park North, Warrington WA3 7PB, U.K.

Protease digestion experiments have been used to characterize the structure of an equilibrium intermediate in the unfolding of creatine kinase (CK) by low concentrations (0.625 M) of guanidine hydrochloride (GdnHCl). Eighteen of the major products of digestion by trypsin, chymotrypsin and endoproteinase Glu-C have been identified by microsequencing after separation by SDS/PAGE and electroblotting on poly(vinylidene difluoride) membranes. The C-terminal portion (Gly^{215} to Lys^{380}) was much more resistant to digestion than the N-terminal portion (Pro^1 to Gly^{133}), although the area most sensitive to proteolysis was in the middle of the CK sequence (Arg^{134} to Arg^{214}). These experiments are consistent with the two-domain model for the CK monomer. The structure of the intermediate is proposed to consist of a folded C-terminal domain and a partly folded N-terminal domain separated by an unfolded central linker. Protease susceptibility is clustered within two N-terminal regions and one central region. These regions are evidently exposed as a result of the partial unfolding and/or separation of the N-terminal domain. Further evidence for the structure of this intermediate comes from gel filtration studies. Treatment of CK with 0.625 M GdnHCl resulted in slow aggregation at 37 °C, but not at 12 °C, a phenomenon previously reported for phosphoglycerate kinase. The aggregation did not occur at higher GdnHCl concentrations and was unaffected by a reducing agent. It is proposed that aggregation is a consequence of non-specific interactions between hydrophobic regions, possibly domain/domain interfaces, which become exposed in the intermediate.

INTRODUCTION

The sites at which proteins are digested by proteases are determined both by the amino acid specificity of the protease (target residues) and by the structure of the target protein. For proteolysis to occur at a target residue, the surrounding sequence must be both accessible to the protease and flexible enough to adopt the conformation required by the active site of the protease [1–4]. Susceptibility to digestion is therefore predominantly associated with sites within unstructured, flexible regions of proteins, for example, the hinge regions of immunoglobulin [5] or the mobile surface loops of thermolysin [1]. Domain linkers in multidomain proteins are particularly susceptible to digestion because they are often less stable than the adjoining domains [2]. Furthermore they often contain α-helical segments that when denatured can adopt the conformation necessary for proteolysis more readily than other more extended secondary structures [3]. Digestion of native proteins with proteases is therefore a useful tool for the identification and isolation of intact functional domains [6].

The usefulness of proteolysis might not, however, be confined to the study of native proteins. Digestion could also be used for the identification of locally unfolded regions of partly denatured proteins and to provide information on the pathway of unfolding within specific regions, thus complementing physical techniques, such as intrinsic fluorescence and CD, that reflect on the global state of the protein molecule. Products of digestion can be identified at their N-terminus by automated protein microsequencing.

Creatine kinase (CK; ATP:creatine N-phosphotransferase, EC 2.7.3.2) is a member of the guanidino-phosphotransferase family, the primary structures of which are highly conserved

[7,8], although they show no sequence identity with other proteins [7]. However, differences are evident in the tertiary and quaternary structures of the individual members [9]. So although the related enzyme lobster arginine kinase is monomeric, and evidence exists that the CK monomer is enzymically active [10,11], the cytosolic isoenzymes function as dimers and the mitochondrial isoenzymes function both as dimers and octamers. The native state of CK is resistant to digestion by proteolytic enzymes [12], except for a region near Ala^{323} that is thought to be a surface loop [13]. Digestion of CK has been used to characterize an inactive, low-temperature intermediate in the refolding process after denaturation with 8 M urea [14], but these studies were performed under dynamic conditions in which the degree of refolding was difficult to control. It has been shown that an inactive equilibrium intermediate of CK is formed at low concentrations of guanidine hydrochloride (GdnHCl) [9,15–17]. The transition from native to intermediate is highly co-operative and, for muscle CK (MM-CK), has a midpoint at 0.6 M GdnHCl [9]: above this concentration there is a progressive loss of tertiary and then secondary structure [17]. The midpoint of the transition from the native to the intermediate state differs between members of the guanidino-phosphotransferase family [9]. The stabilization of the protomer in the quaternary structure of the two CKs might account for the increased stabilization in comparison with arginine kinase. The transition from the intermediate to the unfolded state is similar in the guanidino-phosphotransferases and it is proposed that the intermediate state possesses a similar core structure. The very fast changes observed on refolding [11] have been interpreted in light of the present evidence in terms of the hydrophobic collapse of the protein to the intermediate state [9]. This is the situation thought to occur when a refolding intermediate was trapped at 0 °C for earlier digestion studies [14].

Here we describe proteolytic digestion studies of this equilibrium intermediate, in which we have attempted to identify all the major digestion products by protein microsequencing. Gel filtration studies of the aggregation behaviour of the intermediate have also been performed. A brief account of some of these results was presented at the 654th meeting of the Biochemical Society at the University of Leicester in April 1995 [18].

MATERIALS AND METHODS

Materials

Chicken MM-CK was purified from chick muscle by the method of Eppenberger et al. [19]. Rabbit MM-CK, endoproteinase Glu-C from *Staphyloccus aureus* strain V8, 1-chloro-4-phenyl-3-L-toluene-*p*-sulphonamidobutan-2-one-treated trypsin (bovine pancreas), chymotrypsin (bovine pancreas), (3-cyclohexylamino)-propane-1-sulphonic acid (CAPS), PMSF, dithioerythritol (DTE) and GdnHCl (molecular biology grade) were purchased from the Sigma Chemical Company (Poole, Dorset, U.K.). The poly(vinylidene difluoride) (PVDF) membranes used for protein transfer (ProBlott) were from PE Applied Biosystems (Warrington, Cheshire, U.K.). Superdex 75, jacketed column, injector, UV-M monitor and chart recorder were all purchased from Pharmacia (St. Albans, U.K.).

Proteolytic digestion

CK (2.5 mg/ml) was incubated in a 0.625 M GdnHCl solution (50 mM Tris/HCl, pH 7.6) for 1 h at 37 °C. The partly denatured protein was then digested with protease at a substrate-to-protease mass ratio of 100:1. Digestion by trypsin and chymotrypsin at 37 °C was terminated at different times by the addition of PMSF (100 mM stock solution in ethanol) to give a final concentration of 2.5 mM. Endoproteinase Glu-C digestion was stopped by freezing with liquid nitrogen. A zero-time control was used to check the efficiency of protease inactivation. Residual GdnHCl was removed by dialysis against 50 mM Tris/HCl, pH 6.8, on Millipore hydrophobic VS filters (0.025 μm) at 0 °C for 20 min.

Electrophoresis, blotting and microsequencing

Samples were prepared for electrophoresis by adding 5 × SDS loading buffer [1 × loading buffer: 67.5 mM Tris/HCl, pH 6.8, containing 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) 2-mercaptoethanol] and boiling for 2 min. Electrophoresis was performed in a 20% (w/v) polyacrylamide separating gel [20] until the marker dye had reached the bottom. Samples were transferred to a solid support membrane (PVDF) for the identification of individual bands by microsequencing based on the Edman degradation. Methanol was used to wet the membrane before it was soaked in transfer buffer [10 mM CAPS/NaOH, pH 11.0, with 10% (v/v) methanol, analytical grade]. Gels were removed, soaked briefly in distilled water and then soaked in transfer buffer for 30 min. The transfer apparatus was assembled with the gel and membrane sandwiched between two sheets of presoaked Whatman 3MM paper. Transfer was performed for 1 h at 200 mA with the Transblot apparatus (Bio-Rad). After transfer, the membrane was rinsed with distilled water and stained for 5 min with Amido Black [filtered solution of 0.1% (v/v) Amido Black/1% (v/v) acetic acid/40% (v/v) methanol]. Distilled water was used to destain the membrane before it was dried between two strips of Whatman 3MM paper. Bands were excised from the membrane and washed with 20% (v/v) ethanol before being sequenced on PE Applied Biosystems 476A or Procise 494 automated microsequencers.

Gel filtration

A jacketed column of Superdex 75 (diameter 9 mm, bed height 460 mm) and water bath were used, experiments being performed at either 37 or 12 °C. Samples were incubated with denaturant under varying conditions, for 30 min to 18 h. Samples (50 μl; 0.94 mg/ml) were introduced through an injector.

RESULTS

Figure 1(a) shows the time course of digestion of 0.625 M GdnHCl-treated chick MM-CK with trypsin. After electrophoretic separation the protein bands were transferred to PVDF membranes. The major digestion products, which were subjected to N-terminal microsequencing and identified by comparison with the known chick M-CK amino acid sequence, are shown in Figure 1(b). The apparent M_r of each fragment after SDS/PAGE was determined relative to M_r markers and then compared with values calculated for the same fragment assuming an intact C-terminus (Table 1). All C-terminal M-CK fragments identified in Table 1 migrated faster than predicted on SDS/PAGE, as previously observed for full-length M-CK [21,22]. Figure 1(a) shows that the digestion of GdnHCl-treated CK results in the accumulation of discrete metastable fragments, all of which seem to be C-terminal fragments of various sizes; no stable N-terminal fragments were identified (Figure 1b). Protein microsequencing sometimes revealed multiple digestion sites that were not evident from SDS/PAGE alone. Four cycles of automated Edman degradation of an apparently homogeneous electrotransferred CK fragment (T3 in Figure 1) indicated the presence of two major components (RAVE and AVEK) and one minor com-
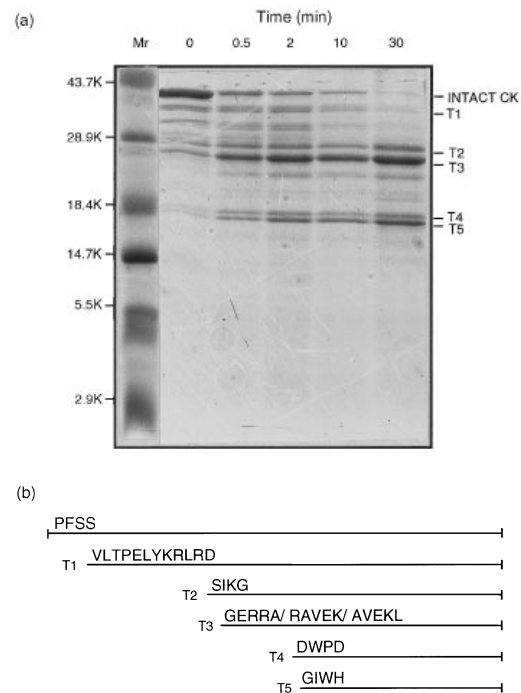
Figure 1   Time course of digestion of a chick MM-CK intermediate with trypsin (a) and diagram identifying the resultant fragments (b)

CK (2.5 mg/ml) treated with 0.625 M GdnHCl was digested with trypsin (1:100, w/w) for the times shown and the reaction stopped by the addition of 2.5 mM PMSF. After SDS/PAGE and electroblotting, the fragments were identified by N-terminal protein microsequencing. The M_r values (K, thousands) of the standard protein markers are shown at the left.

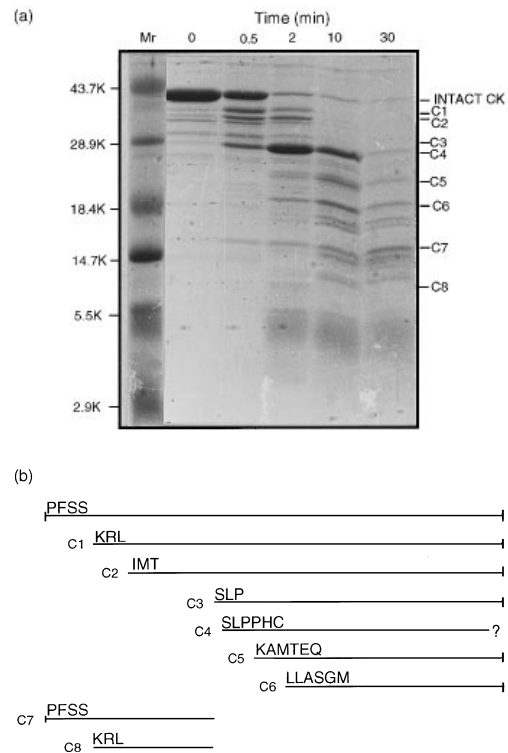**Table 1   Digestion fragments identified by protein microsequencing**

The observed M_r values are given only for those fragments in which M_r was observed to be linearly related to migration. The calculated M_r values were calculated on the assumption that each fragment contained the intact C-terminus, except for those marked with an asterisk, for which M_r was calculated from a knowledge of the probable C-terminal digestion site suggested from Figures 1, 2 and 3.

	N-terminal sequence	Observed M_r	Calculated M_r
Chick MM-CK			
Trypsin			
Intact CK	PFSS		43 202
T1	VLTPELYKRLRD		39 582
T2	SIKG	25 200	27 968
T3	GERRA	23 900	26 541
T3	RAVEK	23 900	26 198
T3	AVEKL	23 900	26 042
T4	DWPD	17 700	19 606
T5	GIWH	16 900	18 865
Chymotrypsin			
Intact CK	PFSS		43 202
C1	KRL		38 766
C2	IMT		35 543
C3	SLP		27 419
C4	SLPPHC	25 200	–
C5	KAMTEQ	21 100	23 347
C6	LLASGM	18 900	20 406
C7	PFSS		15 637*
C8	KRL		11 201*
Endoproteinase Glu-C			
Intact CK	PFSS		43 202
V1	LYKRLRDK		39 042
V2	SYEVFK		34 338
V3	GEFKGRY	22 200	24 558
V4	PFSS	20 100	18 531*
V5	GEFK	18 500	–
V6	EQEQ/QQQL	16 900	–
V7	LYKR	14 300	14 372*
V8	IFKKAG	13 300	13 249
Rabbit MM-CK			
Endoproteinase Glu-C			
1	PFGNTH		20 542*
2	SYTVFK		20 758*
3	QQQLID		22 332
4	SYTVFKDLFDPII		11 622*
5	IFKKAGHPPFMWT		13 066

Figure 2   Time course of digestion of a chick MM-CK intermediate with chymotrypsin (a) and diagram identifying the resultant fragments (b)

CK (2.5 mg/ml) treated with 0.625 M GdnHCl was digested with chymotrypsin (1:100, w/w) for the times shown and the reaction stopped by the addition of PMSF (2.5 mM). The M_r values (K, thousands) of the standard protein markers are shown at the left.

ponent (GERR) arising from digestion at three arginine residues close together in the sequence (RGERRAVEK).

The patterns of digestion obtained after treatment with chymotrypsin (Figure 2 and Table 1) and endoproteinase Glu-C (Figure 3 and Table 1) are also consistent with digestion as a discrete process resulting in the accumulation of large C-terminal portions of the CK molecule. Unlike trypsin, however, chymotrypsin and endoproteinase Glu-C produced some digestion close to the C-terminus, because fragments with the same N-terminus but a lower M_r have been identified (fragments C3 and C4 in Figure 2, and fragments V3 and V5 in Figure 3; see also Table 1). Furthermore N-terminal fragments have been identified after digestion with chymotrypsin and endoproteinase Glu-C (fragments C7 and C8 in Figure 2, and fragments V4 and V7 in Figure 3) although they were not seen with trypsin.

Figure 4 shows the locations in the chick M-CK sequence of all the sites of proteolysis identified by protein microsequencing. All the peptide bonds cut by trypsin followed Arg or Lys residues, whereas chymotrypsin was found to cut after Tyr, Leu and Phe residues, and endoproteinase Glu-C only after Glu

residues. Digestion was observed at only a small proportion of the target residues (Figure 4). Table 2 shows that the greatest sensitivity to proteolysis is in the middle of the CK sequence (Arg^{134} to Arg^{214}), where 34 % of all target residues are cleaved, compared with 10 % in the N-terminal portion and only 2 % in the C-terminal portion. The results indicate that treatment of CK with 0.625 M GdnHCl exposes only a limited number of target residues to attack by proteases (13 %). The exposed sites most sensitive to proteolysis are all in the N-terminal portion of the molecule up to Arg^{214} and seem to form three regions (Figure 4). Region I, Lys^{31} to Tyr^{38}, shows susceptibility to all three proteases. Region II, Phe^{67} to Glu^{79}, shows susceptibility to both chymotrypsin and endoproteinase Glu-C. Finally, Region III, the central tract of 81 residues from Arg^{134} to Arg^{214}, is particularly sensitive to digestion by all three proteases. If we separate the digestion sites into 'major' sites that are evident early in digestion and give prominent protein fragments (large, bold arrows in Figure 4) and 'minor' sites that appear later in digestion or give only minor protein bands (smaller arrows in Figure 4), then the central protease-sensitive region seems much smaller (32 residues; Arg^{134} to Glu^{165}). The results are consistent with the initial production of a large C-terminal fragment that can then be attacked, albeit more slowly, at its new N-terminal end. Clearly it is the initial digestion sites that best reflect the structure of the intact intermediate because later digestion might be occurring within protein fragments that may be less well structured. The remaining C-terminal portion of CK seems to be quite resistant to proteolysis except for a minor endoproteinase Glu-C-susceptible site at Glu^{261}.
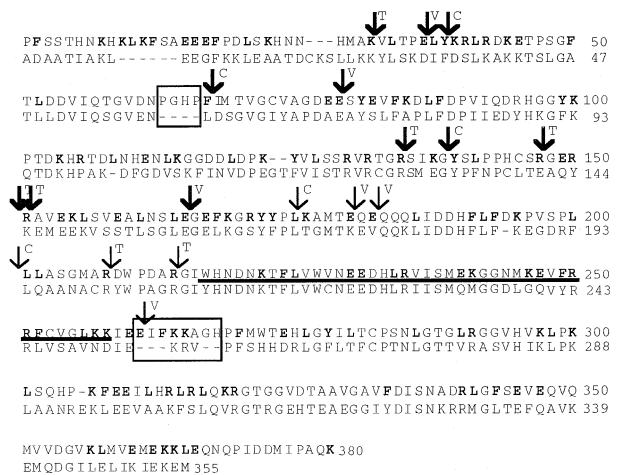
(a)
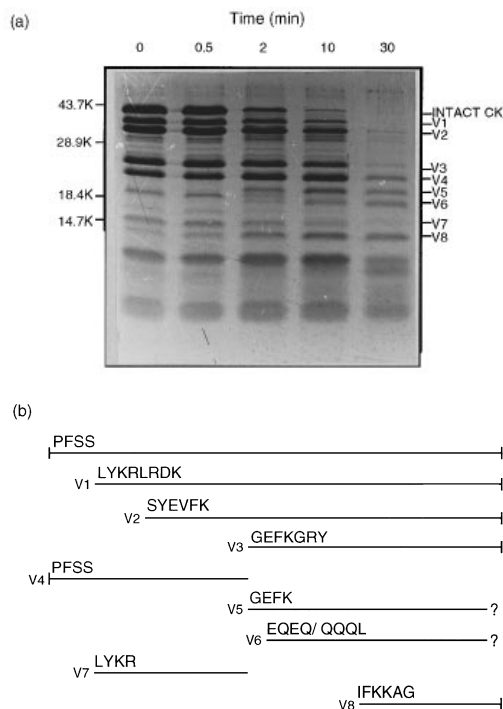
Figure 3   Time course of digestion of a chick MM-CK intermediate with endoproteinase Glu-C (a) and diagram identifying the resultant fragments (b)

CK (2.5 mg/ml) treated with 0.625 M GdnHCl was digested with endoproteinase Glu-C (1:100, w/w) for the times shown and the reaction was stopped by freezing with liquid nitrogen (PMSF does not inhibit this enzyme). Inhibition by this method was not complete because some digestion was evident even without incubation at 37 °C. The M_r values (K, thousands) of the standard protein markers are shown at the left.

Rabbit MM-CK lacks two of the three major susceptible Glu residues identified in chick MM-CK (Glu[36] and Glu[165]). The digestion of GdnHCl-treated rabbit MM-CK yielded a pattern consistent with the Glu residues still present in the rabbit M-CK sequence, the most susceptible site in the centre of the protein being Glu[182], which is only a minor digestion site in chick MM-CK (Figure 4 and Table 1). Digestion of rabbit MM-CK was slower than that of chick MM-CK (results not shown), presumably because the two major digestion sites (Glu[36] and Glu[165]) are absent from the rabbit M-CK sequence.

Figure 5(a) shows the aggregation behaviour of CK when incubated in the denaturing buffer used in our digestion studies [0.625 M GdnHCl/0.1 M Tris/HCl (pH 8); 37 °C]. Some aggregation was observed after 1 h and the aggregate predominated after 16 h. The temperature dependence of aggregation is illustrated in Figure 5(b). No aggregation was observed when CK was incubated at 12 °C rather than 37 °C for 18 h in 0.625 M GdnHCl buffer. When incubated at 37 °C but at higher concentrations of GdnHCl (1.5 or 4 M), no aggregation of CK was observed (Figure 5c). These samples were analysed by gel filtration in 0.3 M GdnHCl buffer to render the unfolded protein globular [9]; CK incubated with 0.625 M GdnHCl buffer still showed aggregation when analysed under these gel-filtration conditions (Figure 5d). The presence of 10 mM DTE had no affect on the aggregation of CK at 37 °C in 0.625 M GdnHCl buffer (Figure 5d). This suggests that aggregation at 37 °C is not a consequence of disulphide bond formation.

```
PFSSTHNKHKLKFSAEEEFPDLSKHNN---HMAKVLTPELYKRLRDKETPSGF  50
ADAATIAKL------EEGFKKLEAATDCKSLLKKYLSKDIFDSLKAKKTSLGA  47

TLDDVIQTGVDNPGHPFIMTVGCVAGDEESYEVFKDLFDPVIQDRHGGYK  100
TLLDVIQSGVEN----LDSGVGIYAPDAEAYSLFAPLPDPIIEDYHKGFK   93

PTDKHRTDLNHENLKGGDDLDPK--YVLSSRVRTGRSIKGYSLPPHCSRGER  150
QTDKHPAK-DFGDVSKFINVDPEGTFVISTRVRCGRSMEGYPFNPCLTEAQY  144

RAVEKLSVEALNSLEGEFKGRYYPLKAMTEQEQQQLIDDHFLFDKPVSPL  200
KEMEEKVSSTLSGLEGELKGSYFPLTGMTKEVQQKLIDDHFLF-KEGDRF  193

LLASGMARDW PDARGIWHNDNKTFLVWVNEEDHLRVISMEKGGNMKEVFR  250
LQAANACRYW PAGRGIYHNDNKTFLVWCNEEDHLRIISMQMGGDLGQVYR  243

RFCVGLKKIEEIFKKAGHPFMWTEHLGYILTCPSNLGTGLRGGVHVKLPK  300
RLVSAVNDIE---KRV--PFSHHDRLGFLTFCPTNLGTTVRASVHIKLPK  288

LSQHP-KFEEILHRLRLQKRGTGGVDTAAVGAVFDISNADRLGFSEVEQVQ  350
LAANREKLEEVAAKFSLQVRGTRGEHTEAEGGIYDISNKRRMGLTEFQAVK  339

MVVDGVKLMVEMEKKLEQNQPIDDMIPAQK 380
EMQDGILELIKIEKEM 355
```

**Figure 4   Summary of digestion sites and other results suggesting possible surface exposure of residues**

Sequence of chick MM-CK (top sequence) showing the clustering of the major digestion sites (bold arrows) in three regions of sequence (Lys[31] to Tyr[38]; Phe[67] to Glu[79]; Arg[134] to Glu[165]). Minor digestion sites (arrows) occur, although these (except Glu[261]) can also be grouped into an extended central region (Arg[134] to Arg[214]). Potential as well as actual target residues are indicated by bold type. The exon conserved in all isoenzymes of CK is underlined. Residues encoded by this exon are known to contribute to nucleotide binding; the sequence is highly conserved throughout the guanidinokinase family. Arg[214] at the N-terminal end of this sequence was digested by trypsin and Glu[261] at the C-terminal end was digested by endoproteinase Glu-C, although these are only minor digestion sites. Comparison with the sequence of arginine kinase (bottom sequence) shows the location of insertions (boxed) into the guanidino-phosphotransferase family consensus sequence, suggesting a possible surface location for these residues.

Table 2   Distribution of actual and potential digestion sites within the chick M-CK sequence

Target residues	Subregions			Total
	Pro[001]–Gly[133]	Arg[134]–Arg[214]	Gly[215]–Lys[380]	Pro[001]–Lys[380]
Glu	2/9 (22%)	3/7 (43%)	1/14 (7%)	6/30 (20%)
Arg	0/5 (0%)	6/7 (86%)	0/8 (0%)	6/20 (30%)
Lys	1/12 (8%)	0/5 (0%)	0/15 (0%)	1/32 (3%)
Leu	0/11 (0%)	2/10 (20%)	0/15 (0%)	2/36 (6%)
Tyr	1/4 (25%)	1/3 (33%)	0/1 (0%)	2/8 (25%)
Phe	1/7 (14%)	0/3 (0%)	0/8 (0%)	1/18 (6%)
All targets	5/48 (10%)	12/35 (34%)	1/61 (2%)	18/144 (13%)

DISCUSSION

The present study represents a considerable advance on earlier proteolytic studies of CK in that (1) digestions were performed on an equilibrium intermediate in the unfolding process, overcoming the problem of native resistance to digestion and avoiding the problem of refolding during digestion that occurred in earlier studies; (2) the wider range of proteases used in this study decreased the influence of the distribution of protease target residues in the sequence on the final picture of protease sensitivity; (3) the sequencing of all major fragments gives a more complete picture; and (4) time courses of digestion allow a distinction to be made between initial sites of digestion and secondary sites that are attacked either more slowly or only after some initial digestion has occurred elsewhere. The combined results with all three proteases show that initial digestion occurs in three regions, one
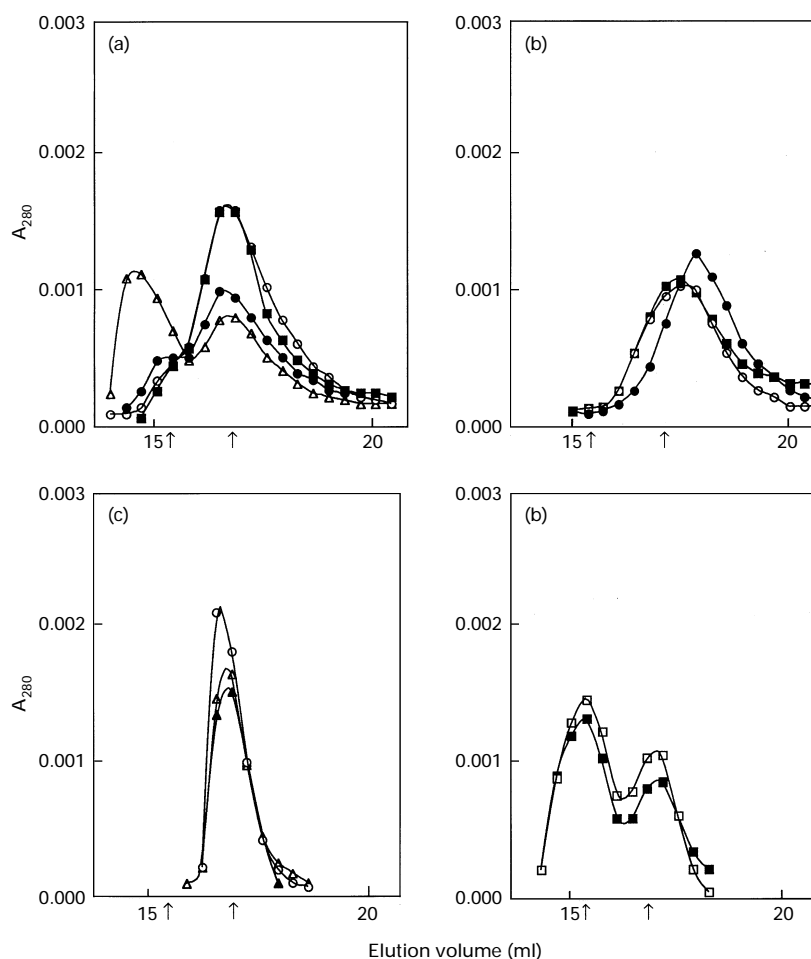
Figure 5   Aggregation of the CK intermediate studied by gel filtration

CK was incubated under various conditions (see below) before gel filtration on Superdex 75. The gel-filtration buffer was 0.1 M Tris/HCl (pH 8)/1 mM DTE containing either 0.625 M GdnHCl (**a**) or 0.3 M GdnHCl (**b–d**). The temperature of the jacketed column was maintained at either 37 °C (**a**) or 12 °C (**b–d**). Arrows indicate the positions of the aggregated and non-aggregated forms of the CK intermediate respectively. (**a**) Samples were injected directly (○), after 1 h of treatment in buffer at 37 °C (■), after 1 h of treatment in 0.625 M GdnHCl buffer at 37 °C (●) or after 16 h of treatment in 0.625 M GdnHCl buffer at 37 °C (△). (**b**) Samples were injected directly (○), after 18 h of treatment in 0.625 M GdnHCl buffer at 12 °C (■) or after 18 h of treatment in 0.625 M GdnHCl buffer at 12 °C (●). (**c**) Samples were injected directly (○), after 1 h of treatment in 1.5 M GdnHCl buffer at 37 °C (▲) or after 30 min of treatment in 4 M GdnHCl buffer at 37 °C (△). (**d**) Samples were injected after 10 h of treatment in 0.625 M GdnHCl buffer at 37 °C (■) or after 10 h of treatment in 0.625 M GdnHCl/10 mM DTE buffer at 37 °C (□).

containing a putative domain linker and two regions in the N-terminal domain. They support a model for CK in which a stable C-terminal domain is separated from a structurally less stable N-terminal domain by a protease-sensitive linker region.

Other kinases, such as phosphoglycerate kinase, have their binding sites for the two substrates on separate domains [23], the C-terminal domain possessing a conserved nucleotide-binding fold [24]. Although the sequence of CK does not display the classic ATP-binding motifs, the residues Trp[210], Trp[227], Cys[282] and Arg[291] have been shown to lie close to the ATP substrate, supporting the idea that CK possesses a C-terminal nucleotide-binding domain [25–27].

The digestion of the partly unfolded intermediate of CK with trypsin, chymotrypsin and endoproteinase Glu-C (Figure 4) shows that a C-terminal portion is more resistant (Table 2) and hence by implication more stable than the rest of the molecule. That large regions of sequence are initially protected from digestion, allowing the accumulation of the corresponding fragments, suggests that regions within this partly denatured CK intermediate retain a compact structure. The sites that show susceptibility to more than one protease (Region I, Lys[31] to

Tyr[38]; Region II, Phe[67] to Glu[79]; the shorter Region III, Arg[134] to Glu[165]) are likely to be unconstrained, that is either already locally denatured or relatively unstable compared with other regions of structure. The importance of using several proteases is illustrated by the slower digestion by endoproteinase Glu-C of rabbit MM-CK, which lacks two of the three main target residues for this protease in chick MM-CK (results not shown). It is possible that the stretch of sequence that shows particular sensitivity to digestion (Arg[134] to Glu[165]; Table 2) consists in part of a domain-linker.

Several of the sites of digestion in Figure 4 can be correlated with other results which corroborate the localized loss of structure. (1) Antibody binding sites are often associated with exposed flexible segments and so it is significant that four monoclonal antibodies that bind only to denatured forms of CK (including the refolding intermediate) have been mapped to segments that coincide with all three regions of proteolysis in the intermediate [14,28]. These regions are special because the only other monoclonal antibodies raised against 'native' CK that recognize denatured forms map to the extreme C-terminus [29] and the extreme N-terminus [30] of the protein. The only antibody

obtained against the C-terminal domain recognizes both the intermediate and the native protein [21]. (2) The regions around the endoproteinase Glu-C digestion site at Glu^{261} and around the chymotrypsin-sensitive site at Phe^{67} (Region II) comprise large insertions into the guanidino-phosphotransferase consensus sequence. The significance of an insertion within a consensus sequence is that they are difficult to accommodate within elements of secondary structure other than loops. In addition, proteolysis occurs at both boundaries of the CK region encoded by the only exon that is conserved throughout all avian and mammalian CK genes [31] and that contains residues important for binding the nucleotide substrate.

Further evidence for the exposure of new surface regions comes from our gel-filtration study of the aggregation behaviour of CK (Figure 5). These results show the temperature-dependence of aggregation of the intermediate, suggesting the predominance of hydrophobic interactions in this process [32], and confirming the existence of nascent hydrophobic patches. Gross et al. [9] report that aggregation of rabbit MM-CK is a result of the formation of disulphide bonds. Aggregation of chick MM-CK is not a consequence of the formation of intermolecular disulphide bridges because the presence of high concentrations of DTE has no affect on aggregate formation (Figure 5).

The enhancement of 8-anilino-1-naphthalene-sulphonic acid (ANS) binding in the intermediate shows no protein concentration dependence [17] and this suggests that although hydrophobic clusters are revealed they are not exposed at the monomer/monomer interface. From the results presented here we suggest that hydrophobic regions are revealed as a result of the unfolding/separation of the N-terminal domain. These results lend further support to our earlier proposals that CK consists of two domains (Pro^1 to Glu^{165} and Gly^{166} to Lys^{380}), the N-terminus of which is less stable than the C-terminus [14,22]. Expression of two similar fragments (1–167 and 168–380) of mitochondrial CK in *Escherichia coli* by Gross et al. [33] has confirmed our digestion results because the N-terminal fragment folds only in the presence of the autonomously folding C-terminal fragment. Grossman and Mixon [34] have also presented evidence suggesting that the nucleotide-binding region folds before the rest of CK. It is possible that some of the sites of proteolysis, antibody binding and chemical modification in the intermediate of CK are a result of the exposure of loops at the interface after separation/unfolding of the N-terminal domain.

Our present studies show that proteolysis and microsequencing might be a useful general approach to obtaining preliminary structural information on proteins that have not yet been analysed by X-ray crystallography, even if the proteins are largely resistant to proteolysis in their native conformations, and can also provide information on dynamic changes in proteins during, for example, folding and unfolding.

After this paper was submitted, the three-dimensional structure of chick mitochondrial CK was published [35]. Because the muscle isoform of CK is likely to have a similar general structure, this enables us to evaluate the blind predictions made in this study from proteolysis data. Most of the proteolytic sites identified lie in linkers between elements of secondary structure, confirming the prediction that most of the native secondary structure is retained in the partly unfolded intermediate. As predicted, the Phe^{67} and Glu^{261} digestion sites do form parts of surface loops, and the N-terminal and C-terminal halves of the CK molecule do indeed seem to come together to form the catalytic site, although the interaction is more complex than was expected in the original two-domain model [14]. Thus the C-terminal half is an anti-parallel β-sheet with its end strand ($\beta2$) starting, as predicted, near Glu^{165}, but strand $\beta1$ (Val^{125} to Arg^{134}) integrates into the centre of the β-sheet to form the nucleotide-binding site. The region that is highly sensitive to protease (Arg^{134} to Glu^{165}) links $\beta1$ with $\beta2$ and would be exposed if $\beta1$ became separated from the β-sheet after treatment with low concentrations of GdnHCl. This would be consistent with both our data and the fragment refolding data of Gross et al. [33]. However, separation and partial unfolding of a smaller N-terminal domain (Pro^1 to Gly^{117}) might also be sufficient to expose the major proteolytic sites in the intermediate. In either case, hydrophobic sequences (such as Phe^{67} to Gly^{76} at the domain/domain interface) would be exposed, which might be responsible for the observed aggregation of the intermediate.

REFERENCES

1   Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M. and Zambonin, M. (1986) Biochemistry **25**, 1847–1851
2   Wilson, J. E. (1991) Methods Biochem. Anal. **35**, 207–250
3   Hubbard, S. J., Eisenmenger, F. and Thornton, J. M. (1994) Protein Sci. **3**, 757–768
4   Hubbard, S. J. and Thornton, J. M. (1993) Faraday Disc. **93**, 13–25
5   Porter, R. R. (1959) Biochem. J. **73**, 119–126
6   White, T. K. and Wilson, J. E. (1989) Arch. Biochem. Biophys. **274**, 375–393
7   Babbitt, P. C., Kenyon, G. L., Kuntz, I. D., Cohen, F. E., Baxter, J. D., Benfield, P. A., Buskin, J. D., Gilbert, W. A., Hauschka, S. D., Hossle, J. P. et al. (1986) J. Prot. Struct. **5**, 1–14
8   Muhlebach, S. M., Gross, M., Wirz, T., Walliman, T., Perriard, J. C. and Wyss, M. (1994) Mol. Cell. Biochem. **133/134**, 245–262
9   Gross, M., Lustig, A., Wallimann, T. and Furter, R. (1995) Biochemistry **34**, 10350–10357
10 Bickerstaff, G. F. and Price, N. C. (1976) FEBS Lett. **64**, 319–322
11 Grossman, S. H., Pyle, J. and Steiner, R. J. (1981) Biochemistry **20**, 6122–6128
12 Williamson, J., Greene, J., Cherif, S. and Milner-White, E. J. (1977) Biochem. J. **167**, 731–737
13 Wyss, M., James, P., Schlegel, J. and Wallimann, T. (1993) Biochemistry **32**, 10727–10735
14 Morris, G. E. and Cartwright, A. J. (1990) Biochim. Biophys. Acta **1039**, 318–322
15 Zhou, H.-M, Zhang, X.-H, Yin, Y. and Tsou, C.-L. (1993) Biochem. J. **291**, 103–107
16 Grossman, S. H. (1994) Biochim. Biophys. Acta **1209**, 19–23
17 Couthon, F., Clottes, E., Ebel, C. and Vial, C. (1995) Eur. J. Biochem. **234**, 160–170
18 Webb, T., Jackson, P. J. and Morris, G. E. (1995) Biochem. Soc. Trans. **23**, 477S
19 Eppenberger, H. M., Dawson, D. M. and Kaplan, N. O. (1967) J. Biol. Chem. **242**, 204–209
20 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
21 Morris, G. E. (1989) Biochem. J. **257**, 461–469
22 Morris, G. E. and Jackson, P. J. (1991) Biochem. J. **280**, 809–811
23 Banks, R. D., Blake, C. C. F., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. W., Merrett, M. and Phillips, A. W. (1979) Nature (London) **279**, 773–777
24 Thornton, J. M. and Taylor, W. R. (1989) in Protein Sequencing: A Practical Approach (Findlay, J. B. C. and Geisgow, M. J., eds.), pp. 147–190, IRL Press/Oxford University Press, Oxford
25 Furter, R., Furter-Graves, E. M. and Wallimann, T. (1993) Biochemistry **32**, 7022–7029
26 Lin, L. J., Perryman, M. B., Friedman, D., Roberts, R. and Ma, T. S. (1994) Biochim. Biophys. Acta **1206**, 97–104
27 Gross, M., Furter-Graves, E. M., Wallimann, T., Eppenberger, H. M. and Furter, R. (1994) Protein Sci. **3**, 1058–1068
28 Morris, G. E., Frost, L. C., Newport, P. A. and Hudson, N. (1987) Biochem. J. **248**, 53–59
29 Nguyen Thi Man, Cartwright, A. J., Osborne, M. and Morris, G. E. (1991) Biochim. Biophys. Acta **1076**, 245–251
30 Morris, G. E. and Nguyen Thi Man. (1992) Biochim. Biophys. Acta **1120**, 233–238
31 Haas, R. C., Korenfeld, C., Zhang, Z., Perryman, B., Roman, D. and Strauss, A. W. (1989) J. Biol. Chem. **264**, 2890–2897
32 Mitraki, A. and King, J. (1989) Bio/Technology **7**, 690–696
33 Gross, M., Wyss, M., Furter-Graves, E. M., Wallimann, T. and Furter, R. (1996) Protein Sci. **5**, 320–330
34 Grossman, S. H. and Mixon, D. (1985) Arch. Biochem. Biophys. **236**, 797–806
35 Fritz-Wolf, K., Schnyder, T., Wallimann, T. and Kabsch, W. (1996) Nature (London) **381**, 341–345