The Journal of Experimental Biology 206, 1545-1556

doi:10.1242/jeb.00299

Functional consequences of a gene duplication and fusion event in an arginine kinase

Deanne M. Compaan* and W. Ross Ellington†

Department of Biological Science and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-4370, USA

*Present address: Genentech, 1 DNA Way, South San Francisco, CA 94080-4990, USA

†Author for correspondence (e-mail: elling@bio.fsu.edu)

Accepted 19 February 2003

Summary

Arginine kinase (AK) from the foot of the razor clam *Ensis directus* consists of two full-length AK domains, denoted D1 and D2, fused in a single polypeptide chain. The full-length cDNA for *Ensis* AK was obtained and its deduced amino acid sequence was analyzed in the context of the X-ray crystal structure of a typical, monomeric AK. Both domains of *Ensis* AK contain most of the residues currently thought to be critical in catalysis, suggesting that both AK domains are catalytically competent. The full-length *Ensis* AK, a D2–NusA–His-tag fusion protein and a D2-truncated AK (enterokinase cleavage product of the fusion protein) were expressed in *Escherichia coli* and purified. All recombinant AK constructs displayed high enzyme activity. Attempts at expressing active D1 alone, D2 alone or a D1–NusA–His-tag fusion protein were unsuccessful. The catalytic properties of the active proteins were compared with the corresponding properties of recombinant AK from the horseshoe crab *Limulus polyphemus*, which is a typical monomeric AK. In contrast to expectations, the kinetic results strongly suggest that *Ensis* AK has only one active domain, namely D2. The K_{cat} values for all *Ensis* constructs were roughly twice that of typical AKs, indicating higher overall catalytic throughput at the competent active site. Furthermore, both the full-length and truncated D2 *Ensis* AKs showed no synergism of substrate binding unlike typical AKs. The D2–NusA–His-tag fusion construct actually displayed negative synergism of substrate binding, which means that, in effect, the first substrate bound acts as a competitive inhibitor of the second. The conservation of the structure of the apparently inactive D1 may be related to constraints imposed by structural changes that could potentially impact substrate binding in D2 and/or possibly influence the proper folding of the enzyme during synthesis. Overall, the results from the present study indicate that the AK contiguous dimer from *Ensis directus* functions with activity in only the second domain. Although lacking activity in D1, D2 appears to compensate by having a higher intrinsic catalytic throughput than typical 40-kDa monomeric AKs.

Key words: arginine kinase, gene duplication, *Ensis directus*, D1, D2, gene fusion, domain.

Introduction

Phosphagen kinases catalyze the reversible transfer of a high-energy phosphoryl group from phosphorylated guanidino storage compounds known as phosphagens to ADP in the following general reaction: phosphagen + MgADP + $H^+ \leftrightarrow$ guanidino acceptor + MgATP. These reactions are typically found in cells that display high and variable rates of energy turnover, such as muscle fibers, neurons, spermatozoa and transport epithelia (Ellington, 2001). Phosphagen kinases comprise a discrete enzyme family consisting of creatine kinase (CK), arginine kinase (AK), glycocyamine kinase (GK), taurocyamine kinase (TK), hypotaurocyamine kinase (HTK), lombricine kinase (LK), opheline kinase (OK) and thalassamine kinase (ThK). CK and AK have been the most extensively studied. These reactions play a critical role in maintaining high values of the effective free energy change of ATP hydrolysis (ΔG_{ATP}) during burst contractile activity in muscle (Newsholme et al., 1978; Meyer et al., 1984; Grieshaber et al., 1994; Ellington, 2001). To mediate this role, CK and AK activities in burst muscles are kept at very high levels so as to maintain the reaction near equilibrium with its substrates over a broad range of rates of ATP turnover (Meyer et al., 1984; Ellington, 2001).

Sequence analyses of members of the phosphagen kinase family reveal a high degree of homology (at least 30–40% amino acid identity; Babbitt et al., 1986; Dumas and Camonis, 1993; Mühlebach et al., 1994; Suzuki and Furukohri, 1994), indicating a shared evolution from the same ancestral gene (Watts, 1971). AK has traditionally been regarded as being most closely related to the ancestral phosphagen kinase (Watts, 1971, 1975). Current evidence suggests that these enzymes

evolved by gene duplication events presumably from a monomeric ancestor. Duplication of the ancestral phosphagen kinase gene provided the enzymatic redundancy for modification of the specificity of the guanidine binding site.

Duplication of phosphagen kinase genes has led to more than just the acquisition of different guanidine specificities; it has also resulted in the formation of multiple enzyme isoforms and oligomerization of the simple monomer. CK in vertebrates, for example, exists as four different isoforms: two cytoplasmic isoforms, brain (B) and muscle (M), which form homo- (BB, MM) and hetero- (MB) dimers, and two mitochondrial isoforms, ubiquitous and sarcomeric, which form primarily homo-octamers. These isoforms appear to be the result of two distinct gene duplication events (Qin et al., 1998). GK, LK, TK and HTK are all dimer-like cytoplasmic CKs.

In addition to forming enzymes with new specificities and/or quaternary structures, duplication events in phosphagen kinase genes have also produced rather unusual multiple-domain enzymes. Sea urchin sperm flagellar CK consists of three fused CK domains (Wothe et al., 1990). In effect, this CK is a contiguous trimer that was probably the result of a gene duplication/fusion event and subsequent unequal crossing-over (Wothe et al., 1990). A similar three-domain CK appears to be present in the spermatozoa of the polychaete *Chaetopterus variopedatus* (Ellington et al., 1998).

Nearly all characterized protostome AKs exist as monomers of approximately 40 kDa, with the exception of what appear to be unique 80-kDa AKs found in certain molluscs. These latter molluscan AKs were observed to be roughly twice the size typical of most AKs and were thought to be dimers consisting of two 40-kDa subunits (Moreland and Watts, 1967). However, large AKs from the marine bivalve mollusks *Pseudocardium sachalinensis* (Suzuki et al., 1998), *Solen strictus* and *Corbicula japonica* (Suzuki et al., 2002) and the sea anemone *Anthopleura japonicus* (Suzuki et al., 1997) were recently shown to be monomers consisting of two fused domains, each with a purported active site. The prevalence of these contiguous-domain phosphagen kinases in a variety of different species may be indicative of the proclivity of these genes for duplication and/or possibly a greater tolerance in the overall design of the proteins to accommodate multiple catalytic domains.

The presence of two or more catalytic domains on a single polypeptide may impose important functional restrictions, especially in phosphagen kinases. Spectroscopic studies (Reed and Cohn, 1972) and X-ray solution scattering results (Forstner et al., 1998, 1999) have demonstrated that AK and CK undergo large conformational changes that are induced upon substrate binding. In the case of the horseshoe crab (*Limulus polyphemus*) AK, upon binding of the substrates a 100-residue amino-terminal region undergoes a large rotation to clamp down upon a larger 257-residue carboxy-terminal region (Zhou et al., 2000; Yousef et al., 2003). This conformational change aligns the two substrates for catalysis, properly configures the active site and protects the reactants from contact with the solvent.

Furthermore, both AK and CK display a rapid equilibrium, random-order reaction mechanism (Cleland, 1967; Blethen 1971, 1972). They typically display synergism of substrate binding, which means that the binary association constant is higher than the ternary association constant for these bi-substrate reactions. In other words, binding of one co-substrate enhances the binding of the second substrate, presumably through conformational movements adjusting the configuration of the catalytic pocket. Substrate binding synergy is most likely critical from a physiological perspective as it tunes the catalytic properties of the AK or CK to prevailing cellular concentrations of reactants.

The presence of multiple domains in AK and CK could clearly impact the large conformational movements described above, as the multiple domains are not independent but are present on the same polypeptide chain. In the case of the contiguous dimeric AKs from bivalve mollusks, the gene duplication event occurred early in the divergence of bivalves, between 250 and 400 million years ago (Suzuki et al., 2002). Thus, there has been considerable time for sequence divergence. Since each polypeptide contains two distinct catalytic domains, presumably there would be latitude for random mutation, and potential loss of activity, in one of the domains. However, the deduced amino acid sequences for all the molluskan contiguous dimers appear to display the full suite of catalytically important residues identified thus far, implying activity in both domains (Suzuki et al., 2002).

In the present report, we describe studies of the contiguous dimeric AK from the razor clam *Ensis directus*. We have cloned and sequenced the cDNA for this AK, expressed recombinant protein and created truncated recombinant constructs to probe the impact of domain fusion on activity and kinetic properties. Our results show that the duplication and fusion events have produced a protein that has altered functional properties but evidently still appears to be able to fulfill its role in temporal ATP buffering.

Materials and methods

Animals and chemicals

Specimens of *Ensis directus* Conrad were purchased from the Marine Biological Laboratory, Woods Hole, MA, USA. The foot muscle of each animal was removed and either used immediately or stored frozen at –70°C. Unless otherwise stated, chemicals and reagents were purchased from Roche Molecular Biochemicals (Indianapolis, IN, USA), Sigma Chemical Co. (St Louis, MO, USA) or Fisher Scientific (Pittsburgh, PA, USA).

Properties of AK from Ensis directus

Fresh or previously frozen tissue was homogenized in 19 volumes of extraction buffer [50 mmol l^{-1} imidazole-HCl (pH 7), 1 mmol l^{-1} EDTA, 14 mmol l^{-1} 2-mercaptoethanol] using a Brinkmann polytron (Brinkmann Insruments, Westbury, NY, USA). The homogenate was then clarified by centrifugation (12 000 ***g***, 20 min) and the resulting supernatant used

immediately for activity assays or further purified for additional analysis. AK activity was determined in the reverse direction by the protocol of Strong and Ellington (1993). AK was initially purified using 50–85% saturated ammonium sulfate fractionation. The resulting pellets were exhaustively dialyzed against diethylaminoethyl (DEAE) running buffer [10 mmol l^{-1} Tris-HCl (pH 8.1), 0.5 mmol l^{-1} EDTA, 1 mmol l^{-1} dithiothreitol (DTT)] and then applied to a DEAE Sephacel column (Amersham Biotech, Piscataway, NJ, USA) and eluted with a salt gradient of 0–400 mmol l^{-1} KCl. Peak DEAE fractions were used for native and subunit relative molecular mass (M_r) determinations by fast-preparative liquid chromatography (FPLC) and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), respectively, as described by Ellington et al. (1998).

Total RNA isolation and reverse transcription

Total RNA was isolated from fresh muscle using the Trizol reagent (Invitrogen, Carlsbad, CA, USA) extraction protocol. Approximately 2 g of *Ensis* foot muscle was homogenized in 27.5 ml Trizol reagent using a Brinkmann polytron. Cellular debris was pelleted and the supernatant containing the RNA was subject to chloroform extraction and further isopropanol purification. Purified RNA was pelleted and stored in 75% ethanol [in diethylpyrocarbonate (DEPC)-treated water] at –70°C. Reverse transcription was performed on total RNA templates using Ready-To-Go You-Prime First-Strand Beads (Amersham Biotech) and a lock-docking oligo-dT primer [5′-GGCCACGCGTCGACTAGTAC$(T)_{17}$(A,C,G)(A,C,G)-3′; Borson et al., 1992].

3′ RACE

RACE (rapid amplification of cDNA ends) protocols were patterned after those of Frohman et al. (1988) using kits from Invitrogen. PCR amplifications were conducted using a Hybaid (Middlesex, UK) Omn-E thermocycler. The 3′ end of the cDNA was amplified using Taq DNA polymerase (Invitrogen), the lock-docking oligo-dT primer and a ‘universal’ phosphagen kinase primer that corresponds to a highly conserved sequence (W[V/I]NEEDH) in phosphagen kinases (Suzuki and Furukohri, 1994). PCR amplification was performed using a touchdown PCR protocol (Graber and Ellington, 2001). The resulting product was gel-purified using the Concert Rapid Gel Extraction System (Invitrogen) and subcloned into the puC19 TA cloning vector (Invitrogen). Blue/white screening was used to identify recombinant clones. Plasmids were isolated (Wizard Mini Prep; Promega, Madison, WI, USA) and digested with *Bam*HI and *Eco*RI (Invitrogen) to determine insert size. Plasmids from three independent clones were sequenced on an automated PE-ABI model 373A DNA sequencer (Applied Biosystems, Foster City, CA, USA).

5′ RACE

For amplifying the 5′ ends of the cDNA, two gene-specific primers were synthesized based on the 3′ cDNA sequence. The 5′ RACE was carried out using the 5′ RACE system (Invitrogen) and the resulting products were reamplified by PCR and gel-purified. The purified product was then cloned using the TOPO TA cloning kit (Invitrogen). Three independent clones were identified and sequenced.

Creation of a full-length Ensis *AK cDNA expression construct*

A full-length cDNA was produced by PCR using two gene-specific primers. The product was then incubated with 0.5 µl Taq DNA polymerase (Invitrogen) for 20 min at 72°C to ensure an A overhang. The product was purified using the Promega Wizard PCR Preps DNA purification system, ligated into the pETBlue-1 AccepTor vector (Novagen, Madison, WI, USA) and transformed into NovaBlue Cells (Novagen). Recombinant colonies were identified by blue/white screening. Plasmids from these clones were purified, and insert orientation was established by *Eco*RI and *Bgl*II digestion. Plasmids containing the correct insert were then transformed into Tuner (DE3) pLacI-competent cells (Novagen). 20% glycerol stocks of recombinant clones were made and stored at –70°C.

Expression and purification of the full-length recombinant Ensis *AK*

Expression was carried out using Tuner (DE3) pLacI-transformed cells grown in LB media containing 1% glucose, 50 µg ml^{-1} carbenicillin and 34 µg ml^{-1} chloramphenicol (CalBiochem, La Jolla, CA, USA). Because full-length *Ensis* AK was expressed as an insoluble protein at 37°C, expressions were carried out at 15°C for 24 h to produce soluble, active protein. Cells were grown at 37°C to an OD_{600} of 0.4 then shifted to 15°C and induced at an OD_{600} of 0.7. Cells were centrifuged and pellets stored at –70°C. Frozen pellets were thawed and cells were lysed using bpER bacterial protein extraction reagent (Pierce, Rockford, IL, USA) containing 14 mmol l^{-1} 2-mercaptoethanol. AK was then purified by anion exchange chromatography on a DEAE Sephacel column and by size exclusion chromatography using a Sephacryl S-200 column. Purity of the AK was established by SDS-PAGE.

Truncated Ensis *D1 and D2 cloning, expression and refolding*

Truncated mutants of D1 and D2 were first generated in the same manner as the full-length construct and ligated into the pETBlue-1 vector. However, neither truncated construct produced soluble protein when expressed at either 37°C for 4 h or at 15°C for 24 h. To increase soluble expression of these mutants, they were instead designed to be inserted into the pET43.1a vector (Novagen) in order to produce a fusion protein with a NusA-tag (NusA protein) to facilitate protein folding, a His-tag for ease of isolation and an enterokinase cleavage site for removal of the NusA–His-tag polypeptide. D1 and D2 NusA–His-tag constructs were engineered using two primers each designed from the full-length cDNA sequence using PFU DNA polymerase (Stratagene, La Jolla, CA, USA) to ensure blunt ends. The resulting D1 and D2 products were digested with *Bam*HI to create a 3′ sticky end and were then gel purified.

The inserts were then directionally ligated into the pET43.1a vector, previously linearized by digestion with *Psh*A1 and *Bam*HI to create blunt and sticky ends. The digested vector was gel purified and transfected into NovaBlue (Novagen) hosts, which were used for initial blue/white screening of recombinant hosts. Plasmids were purified from positive clones and insert size was determined by colony PCR with two vector-specific primers. Three plasmids for each construct with the correct insert were transformed into the BL21(DE3) expression host (Novagen).

The transformed BL21(DE3) expression hosts were grown in 1 liter of LB media containing 1% glucose and 50 μg ml^{-1} carbenicillin. Cells were grown at 37°C to an OD_{600} of 0.6, induced with 1 mmol l^{-1} isopropyl-β-D-galactopyranoside (IPTG), incubated for 4.5 h, harvested by centrifugation and frozen at –70°C. As with the original truncated constructs, minimal soluble protein was produced for both the D1 and D2 NusA–His-tag constructs. Thus, inclusion bodies were isolated with BugBuster protein extraction reagent (Novagen) according to the protocol and either used fresh or stored at –70°C.

Inclusion bodies from 250 ml of culture were unfolded in 7 ml unfolding buffer, consisting of 100 mmol l^{-1} NaH_2PO_4, 10 mmol l^{-1} Tris-HCl (pH 8.0), 8 mol l^{-1} urea and 14 mmol l^{-1} 2-mercaptoethanol, and stirred slowly at room temperature for 1 h. The sample was briefly centrifuged to remove any particulates and then added to 10 ml of a 50% Ni–NTA agarose (Qiagen, Valencia, CA, USA) slurry pre-equilibrated in unfolding buffer. The mixture was gently mixed at room temperature for 1–1.5 h then loaded into a column and washed with four column volumes of unfolding buffer. Protein was refolded on the column by washing with four column volumes of solubilization buffer, consisting of 50 mmol l^{-1} NaH_2PO_4, 300 mmol l^{-1} NaCl, 10 mmol l^{-1} Tris-HCl (pH 8.0) and 14 mmol l^{-1} 2-mercaptoethanol. Protein was then eluted from the column with 50 mmol l^{-1} imidazole elution buffer [50 mmol l^{-1} imidazole (pH 7.0), 300 mmol l^{-1} NaCl, 50 mmol l^{-1} NaH_2PO_4, 14 mmol l^{-1} 2-mercaptoethanol] followed by 100 mmol l^{-1} imidazole elution buffer [100 mmol l^{-1} imidazole (pH 7.0), 300 mmol l^{-1} NaCl, 50 mmol l^{-1} NaH_2PO_4, 14 mmol l^{-1} 2-mercaptoethanol]. Large amounts of soluble protein were produced for both constructs. Interestingly, substantial AK enzymatic activity was measured for D2 fusion protein while no activity was detected for D1 fusion protein.

Purification of fusion protein and enterokinase cleavage and purification of product

Soluble fusion protein was obtained from inclusion bodies as described above. Fractions eluted from the Ni–NTA column containing the lowest amount of contaminating protein (established by SDS-PAGE) were pooled and concentrated by pressure ultrafiltration. A small amount of glycerol was added, and the sample was then loaded onto a Sephacryl S-300 column; fractions were analyzed by SDS-PAGE and those with the highest activity and minimal contamination were pooled and concentrated. The D2–NusA–His-tag fusion protein retained high activity and was used for subsequent kinetic assays.

The NusA-tag was removed from the fusion protein by digestion with recombinant enterokinase (Novagen). Pooled fractions from the S-300 column were dialyzed *versus* 800 ml of cleavage buffer, containing 20 mmol l^{-1} Tris-HCl (pH 7.4), 50 mmol l^{-1} NaCl, 2 mmol l^{-1} $CaCl_2$, 0.4 mmol l^{-1} DTT and 5% glycerol, for 2 h with one buffer change. Approximately 30 units of enterokinase were added per mg target protein. The cleavage reaction proceeded at room temperature for 8.5 h. After incubation, the sample was centrifuged (12 000 ***g***, 20 min) to pellet precipitated protein. No activity in the D1 fusion protein was observed after cleavage, so this product was not treated any further. Enterokinase was removed from the D2 fusion protein reaction with an enterokinase cleavage capture kit (Novagen). Pefabloc (Roche Molecular Biochemicals) was added to a final concentration of 0.5 mg ml^{-1} to quench any residual enterokinase activity. The sample was dialyzed *versus* 700 ml Ni–NTA solubilization buffer [50 mmol l^{-1} NaH_2PO_4, 300 mmol l^{-1} NaCl, 10 mmol l^{-1} Tris-HCl (pH 8.0), 14 mmol l^{-1} 2-mercaptoethanol] for 2 h with one buffer change. Dialyzed sample was added to 4 ml of 50% Ni–NTA agarose (Qiagen) and rotated for 1.5 h at 4°C to remove NusA-tag and any uncleaved full-length AK. Flow-through was collected, and remaining cleaved D2 was eluted with a solubilization buffer wash. Fractions from the Ni–NTA flow-through were analyzed by SDS-PAGE, and fractions containing cleaved D2 were pooled and further purified by FPLC on a Superdex 200HR column. Superdex fractions of over 99% pure D2 were pooled, concentrated and used for kinetic assays.

*Expression and purification of horseshoe crab (*Limulus*) AK*

The cDNA for *L. polyphemus* AKWTrev was based on the original construct of Strong and Ellington (1996) and had previously been cloned into *E. coli*. Cells transformed with AKWTrev were kindly provided by the lab of M. Chapman (Florida State University, Institute for Molecular Biophysics). Cells were grown in 500 ml of LB medium with 50 μg ml^{-1} carbenicillin (CalBiochem) and induced at an OD_{595} of 0.6 with 1 mmol l^{-1} IPTG (Sigma). After 5 h of expression, cells were harvested by centrifugation (5000 ***g***, 15 min) and stored at –70°C. Pellets were thawed and homogenized (with a Brinkman polytron) in 50 ml of buffer containing 50 mmol l^{-1} imidazole (pH 7), 1 mmol l^{-1} EDTA and 14 mmol l^{-1} 2-mecaptoethanol. The homogenate was then sonicated with a Microson ultrasonic cell disrupter (Heat Systems Ultrasonics, Farmingdale, NY, USA) to ensure cell lysis. The cell lysate was clarified by high-speed centrifugation (12 000 ***g***, 20 min). The resulting supernatant was dialyzed *versus* 1 liter DEAE running buffer [10 mmol l^{-1} Tris-HCl (pH 8.1), 0.5 mmol l^{-1} EDTA and 7 mmol l^{-1} 2-mercaptoethanol] for 4 h with one buffer change. AK was then separated by ion exchange on a DEAE Sephacel column run with a 0–200 mmol l^{-1} KCl gradient. Purity of the AK was determined by SDS-PAGE.

Enzyme kinetics

Kinetic assays were run on a UV–visible spectrophotometer (3E; Varian Cary, Walnut Creek, CA, USA) using manufacturer's software. Initial velocity values were determined in the reverse reaction by varying the concentration of one substrate *versus* six fixed concentrations of the second substrate and *vice versa*, resulting in a 6×6 matrix. Reaction mixtures contained 4 mmol l^{-1} D-glucose, 1 mmol l^{-1} NADP, 7.5 EU ml^{-1} hexokinase and 2 μg ml^{-1} glucose-6-phosphate dehydrogenase. Six concentrations of ADP (0.03, 0.06, 0.12, 0.24, 0.48 and 0.96 mmol l^{-1}) were varied *versus* each of six concentrations of arginine phosphate (0.2, 0.3, 0.5, 0.8, 1.2 and 1.8 mmol l^{-1}). Actual concentrations of both substrates were empirically determined by enzymatic standardization (for arginine phosphate) and spectrophotometric standardization (for ADP). Magnesium acetate was added to a concentration of 1 mmol l^{-1} above the concentration of ADP to ensure full saturation of ADP with Mg^{2+}. Assay buffer (100 mmol l^{-1} imidazole-HCl, pH 7) was added to each 3 ml cuvette to bring the total reaction volume to 2.5 ml. All assays were run at 25°C. Inverse values of initial reaction velocities (determined between 1.5 min and 2.5 min reaction time) were plotted *versus* the inverse substrate concentrations to create Lineweaver–Burke primary plots. Secondary replots were then created by plotting the slopes and *y*-intercepts of the primary plots *versus* inverse substrate concentrations. K_a, K_{ia} and V_{max} were determined by extrapolating from the secondary plots (kinetic parameters defined in Results). Data analyses were conducted using SigmaPlot (SPSS, Chicago, IL, USA).

Results

AK from Ensis *foot muscle*

Activity assays using crude tissue homogenates indicated that the foot muscle of *Ensis directus* contained a high amount of enzyme activity, averaging 1190.89 μmoles min^{-1} g^{-1} tissue. This level is equivalent to AK activities observed in the foot muscle of the congenor *Ensis ensis* (Zammit and Newsholme, 1976). AK from tissue homogenates was partially purified by ammonium sulfate fractionation and DEAE Sephacel chromatography. Native M_r, determined by the AK elution profile from FPLC against standards, was calculated to be 82.11 kDa. This value is virtually identical to the subunit M_r of 82.33 kDa determined by reducing SDS-PAGE, indicating that this AK is a monomer of roughly twice the size of typical monomeric AKs.

Sequence analyses

RT-PCR (reverse-transcription PCR) of *Ensis* total RNA was carried out with a universal phosphagen kinase primer and an oligo-dT adapter. Despite the presence of two potential NEEDH binding sites for the universal phosphagen kinase primer, only the second site was utilized, yielding an intense 3′ product of approximately 900 bases. Two new oligonucleotide primers were designed and used to amplify the 5′ end of the cDNA, yielding a 1.8-kb product. The 3′ and 5′ sequences were assembled using the GCG (Genetics Computer Group, Madison, WI, USA) software package. The 2715-bp cDNA has an open reading frame of 2171 nucleotides flanked by a 5′ untranslated region consisting of 58 nucleotides and a 3′ untranslated region consisting of 486 nucleotides. The open reading frame codes for a 723 amino acid protein with an estimated isoelectric point (pI) of 6.66 and a calculated M_r of 81.3 kDa. This M_r is consistent with the empirically determined values of 82.11 kDa and 82.33 kDa. The cDNA sequence is not reported here but has been deposited in and is available through GenBank (accession number AF404508).

The deduced amino acid sequence of *Ensis* AK clearly indicates a two-domain structure. The division between domain 1 and domain 2 is included in Fig. 1. Domain 1 (D1) consists of 363 amino acids with an estimated pI of 7.18 and a calculated M_r of 40.49 kDa. Domain 2 (D2) consists of 360 amino acids with an estimated pI of 6.31 and a calculated M_r of 40.78 kDa. Percent identities and percent similarities of *Ensis* AK D1 and D2 and other selected AKs are shown in Table 1. The two domains of *Ensis* AK share 59.94% amino acid identity with one another. Each domain, however, shares a much higher degree of identity (over 80%) with its corresponding domain of the contiguous dimeric AK from the bivalve *Pseudocardium sachalinensis*. *Ensis* AK also shows a high percent identity (up to 57%) to 40-kDa monomeric AKs. The lowest percent identity (44% and 45%) occurred between the two domains of *E. directus* and the two domains of the contiguous dimer from the sea anemone *Anthopleura japonicus*, respectively.

In all of the sequence comparisons, neither domain in the *Ensis* contiguous dimer displays a consistently higher percent

Table 1. *Percent amino acid sequence identity of D1 and D2 from* Ensis directus *arginine kinase (AK) to each other and to other selected AK sequences*

	Ensis D1	*Ensis* D2
Ensis D1	–	59.94 (67.61)
Ensis D2	59.94 (67.61)	–
Pseudocardium D1	84.57 (87.60)	61.65 (68.18)
Pseudocardium D2	59.94 (67.04)	81.67 (85.28)
Nautilus	57.27 (65.11)	57.31 (64.76)
Nordotis	55.33 (62.54)	59.49 (65.16)
Liolophura	55.52 (61.63)	58.05 (65.52)
Limulus	53.28 (59.54)	51.29 (61.32)
Homarus	49.71 (58.00)	49.29 (60.40)
Anthopleura D1	46.76 (54.37)	50.71 (58.97)
Anthopleura D2	44.02 (54.52)	45.53 (55.62)

Percent similarities of sequences are shown in parentheses. Values were determined using the GCG software package and the following sequences (GenBank accession numbers are in parentheses): *Pseudocardium sachalinensis* – cockle (AB017255); *Nautilus pompilius* – cephalopod (AB042332), *Nordotis madka* – gastropod mollusc (D26104), *Liolophura japonica* – chiton (O15990); *Limulus phoyphemus* – horseshoe crab (S52098); *Homarus gammarus* – lobster (P14208); *Anthopleura japonicus* – sea anemone (O015992).

identity to the other selected AK sequences. To determine conservation of key residues in the sequence of *Ensis* AK, sequence comparisons were made between *Ensis* AK D1 and D2, monomeric AK from the horseshoe crab *Limulus polyphemus* and an AK consensus sequence. These alignments are shown in Fig. 1. A 1.8-Å crystal structure of *Limulus* AK in the transition state analog complex (TSAC) was recently reported (Zhou et al., 1998). This structure revealed a variety of residues interacting with substrates as well as those stabilizing the catalytic pocket. The catalytic mechanism and fine details of the interaction of enzyme/substrates for AK, and in fact CK and other phosphagen kinases, remain to be fully elucidated (Zhou et al., 1998, 2000). However, the *Limulus* AK structure provides strong insight into residues critical for AK function, although other residues are clearly involved.

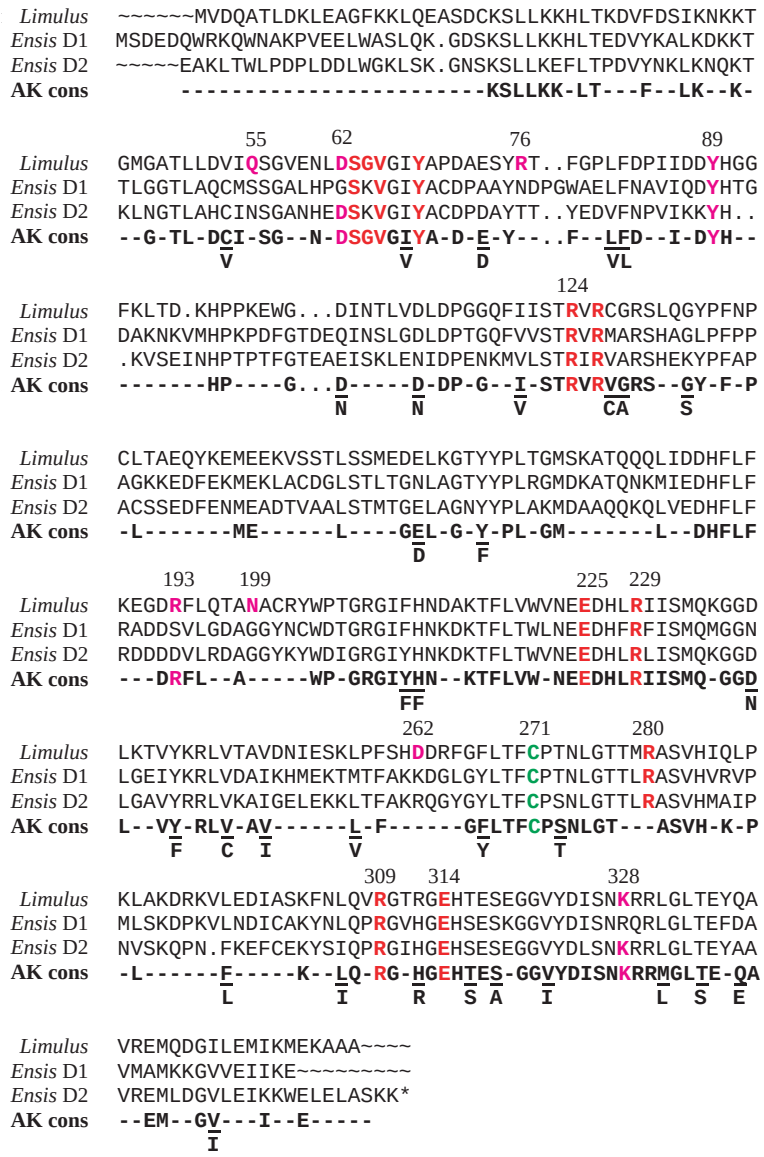

Fig.·1. Multiple sequence alignment of *Ensis directus* arginine kinase (AK) domains 1 and 2 (D1 and D2, respectively), *Limulus polyphemus* AK and an AK consensus sequence (Ellington and Bush, 2002). Numbering is according to that of *L. polyphemus* AK. The 18 sequences used for the AK consensus are sea anemone *Anthopleura* domains 1 and 2 (O015992), crab *Eriocheir* (AAF43438), crab *Carcinus* (AAD48470), lobster *Homarus* (P14208), shrimp *Penaeus* (P51545), locust *Schistocerca* (P91798), bee *Apis* (PC6506), horseshoe crab *Limulus* (S52098), fruit fly *Drosophila* (P48610), trypanosome *Trypanosoma* (AAF23164), snail *Turbo* (O015989), abalone *Suluculus* (S46407), snail *Cellana* (BAB41096), sea hare *Aplysia* (BAB41095), squid *Seipiateuthis* (BAA95610), octopus *Octopus* (BAA95609) and chiton *Lilophura* (O15990). Catalytically important residues are shown in red. These residues contact and stabilize the reactants during catalysis. With the exception of Gly64, all the catalytically important residues are conserved in both domains of the *Ensis* contiguous dimer, suggesting two active domains. Residues shown in pink are those that interact with one another by forming salt bridges when the enzyme is in the closed state. The 'reactive' cysteine characteristic of all phosphagen kinases is shown in green.

Comparison of the *Ensis* contiguous dimer with the *Limulus* AK and AK consensus sequences shows that residues important to functional activity and conserved throughout the AK family are also conserved in the two domains of the *Ensis* contiguous dimer. Residue positions that follow correspond to those of *Limulus* AK. With the exception of Gly64 (which is absent in both D1 and D2), all residues that are known to contact the substrates are present in both domains. Other residues conserved among AKs, such as Asp62 and Arg193, are those suggested to form salt bridges stabilizing the closed state of the enzyme. These residues are not present in *Ensis* AK except for the Asp62 equivalent in D2. This is consistent with the hypothesis that contiguous dimeric AKs have a unique method of maintaining favorable topology of the amino-terminal and carboxy-terminal regions during catalysis (Suzuki et al., 2002). On balance, it appears that both *Ensis* D1 and D2 have the requisite residues to be catalytically competent.

Recombinant full-length Ensis *AK*

Soluble, full-length protein was expressed at 15°C for 24 h in *Escherichia coli*. Cells were lysed chemically and the 82-kDa protein was then purified *via* size exclusion and ion exchange chromatography to greater than 99% homogeneity (established by SDS-PAGE) for kinetic studies. Fig. 2 gives a pictorial summary of the nature of the recombinant proteins produced in this study and whether solubility and AK activity were retained.

Analyses of truncated AK

Early attempts to express truncated D1 and D2 were unsuccessful – all recombinant protein formed inclusion bodies when expressed at either 37°C or 15°C (Fig. 2). To facilitate proper folding, truncated forms of *Ensis* AK were expressed as fusion proteins with a NusA-tag. The 100-kDa fusion proteins were highly expressed in *E. coli* but, again, formed inclusion bodies at 37°C. When expressed at 15°C for 24 h,

some soluble protein could be detected, and activity was evident for the D2 fusion protein although not for the D1 fusion construct. Attempts to purify the small amount of protein expressed in soluble form at 15°C were unsuccessful as the folded protein did not bind well to the nickel column.

Truncated fusion constructs were instead purified from inclusion bodies harvested after 4 h of expression at 37°C. Unfolded protein bound with high affinity to the nickel column, and refolding on the column resulted in approximately 75% refolded protein that was easily separated from improperly folded protein by size exclusion chromatography. Refolded D2 fusion protein displayed significant activity, while no activity could be detected from refolded D1 fusion protein nor when the product was digested with enterokinase (Fig. 2). Active D2 fusion protein was purified to homogeneity by size exclusion chromatography and used for kinetic studies. Cleavage of D2 from the NusA-tag was achieved by digestion with enterokinase. The D2 was then purified by Ni–NTA chromatography and FPLC and used for kinetic studies (Fig. 2).

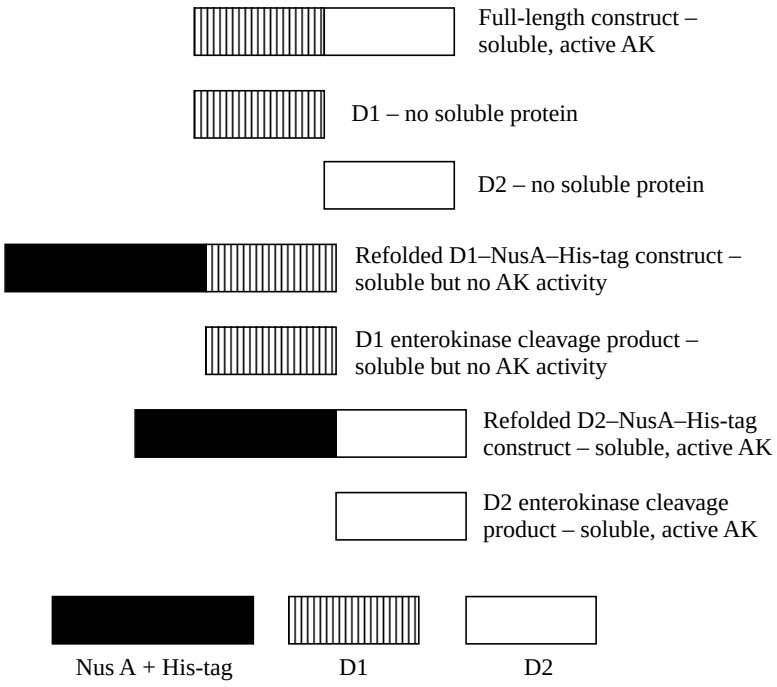

Fig. 2. Summary of the nature and outcome of the expression of *Ensis* full-length arginine kinase (AK), truncated D1 and D2 and the various AK NusA–His-tag fusion constructs and enterokinase cleavage products.

Catalytic properties of recombinant proteins

Kinetic parameters were determined for four AK constructs: (1) full-length recombinant *Ensis* AK, (2)

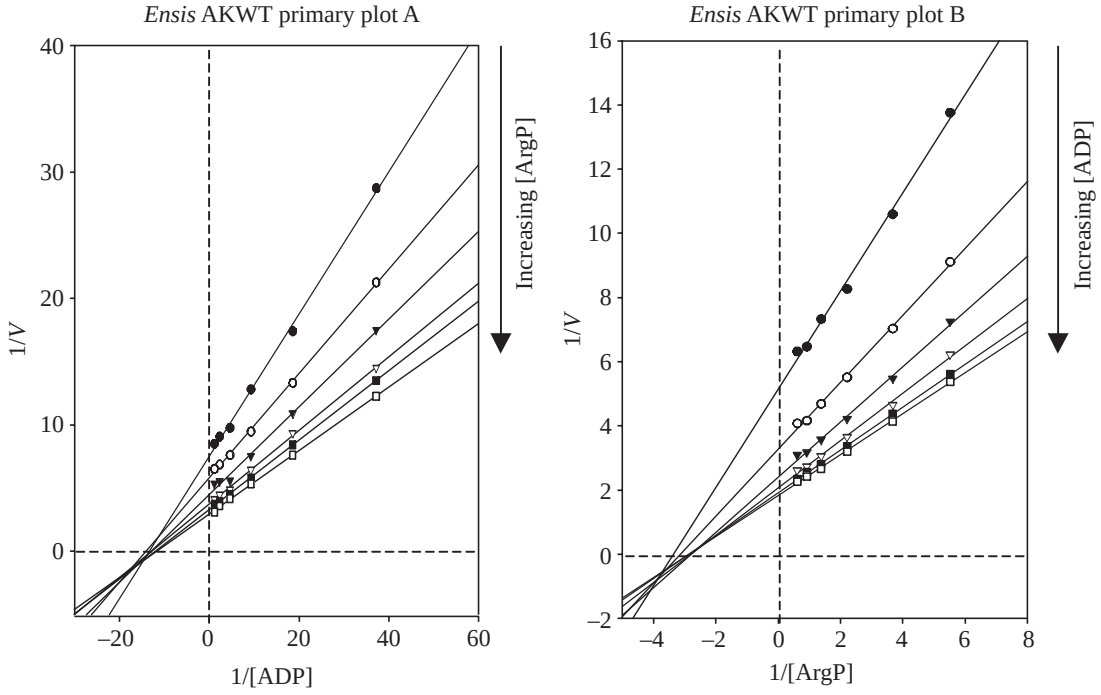

Fig. 3. Primary kinetic plots from *Ensis directus* full-length arginine kinase (AK). At each of six different fixed concentrations of substrate A, substrate B is varied through six concentrations. The inverse of reaction velocities (V) from this 6×6 matrix are plotted *versus* the inverse substrate concentration (ADP or arginine phosphate). Slopes and y-intercepts from each of the six lines in the primary plot are replotted as secondary plots to yield the kinetic constants (K_a and K_{ia}) for each substrate as well as the V_{max} of the enzyme. Note that the six lines in the primary plots intersect at or below the origin indicating no, or even negative, synergy of substrate binding.

truncated *Ensis* AK D2–NusA–His-tag fusion protein, (3) truncated *Ensis* AK D2 and (4) *Limulus* AK for comparison of typical monomeric AK kinetics. Inverse reaction velocities were plotted *versus* the inverse substrate concentration, creating the primary plots. An example of primary plots for full-length *Ensis* AK is shown in Fig. 3. Interestingly, the six lines in both plots intersect at or below the *x*-axis, indicating that this enzyme does not display the synergy of substrate binding that is characteristic of both AKs and CKs (Cleland, 1967; Maggio, 1977; Blethen 1971, 1972). In this context, synergy of substrate binding occurs when the binding of one substrate facilitates the binding of the co-substrate (see discussion below on binding constants). For comparison, kinetic assays were also completed on the typical ~40-kDa monomeric AK from *Limulus*. Primary plots from *Limulus* AK show that the six lines intersect well above the *x*-axis, which is indicative of significant substrate binding synergy (Fig. 4).

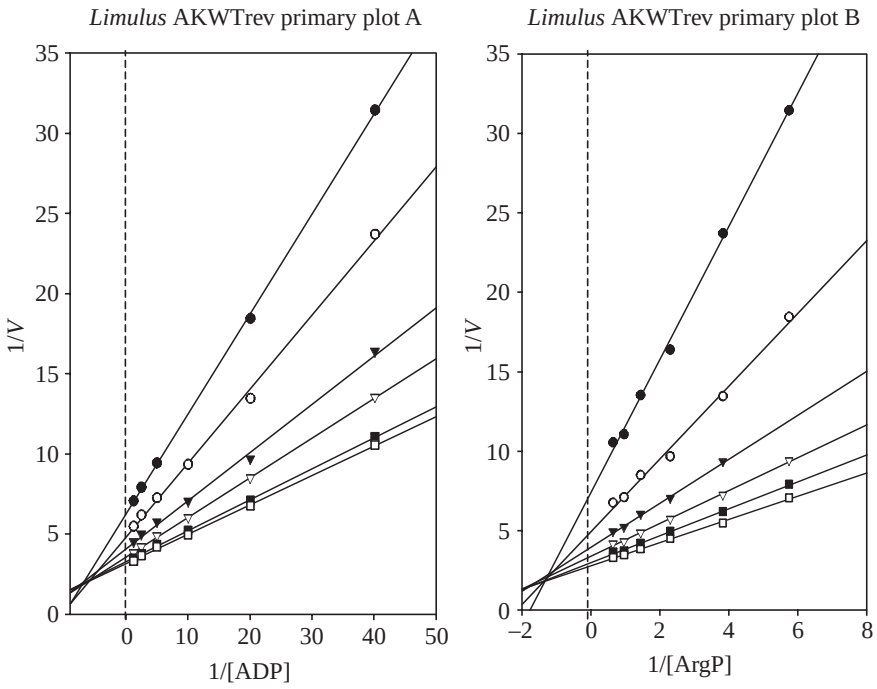

Fig. 4. Primary kinetic plots from *Limulus polyphemus* arginine kinase (AK) WTrev. This AK is a typical 40-kDa monomer. Its kinetic parameters were determined to serve as a source of comparison for *Ensis* AK. *Limulus* AK displays significant synergy of substrate binding, which is characteristic of both arginine kinase and creatine kinase. The intersection of the six lines in each plot above the *x*-axis is indicative of this substrate binding synergy. To determine kinetic constants, the slopes and *y*-intercepts from these primary plots were replotted in secondary plots.

Because the AK reaction is bi-reactant, kinetic constants could not be obtained directly from primary plots; instead, secondary plots were created from the slopes and *y*-intercepts of the primary plots. The kinetic binding constants (K_{ia} and K_a) as well as the true V_{max} of each AK construct were extrapolated from the secondary plots (Table 2). Results from the kinetic analyses indicate considerable differences in the substrate-binding constants of *Ensis* AK *versus Limulus* AK. For a random bi–bi substrate binding mechanism, the K_{ia} can be defined as the binding constant of the enzyme to the initial substrate (binary binding constant; in some literature referred to as the K_s), whereas the K_a value is the binding constant of

Table 2. *Kinetic parameters determined for four arginine kinase (AK) constructs:* Ensis *full-length AK, D2–NusA–His-tag fusion protein, D2 alone and* Limulus *AKWTrev*

Parameter	*Ensis* full-length	*Ensis* D2–NusA–His-tag	*Ensis* D2	*Limulus* WT
Arginine phosphate				
K_{ia} (mmol l^{-1})	0.333±0.046	0.618±0.083	0.592 (0.578, 0.606)	0.804±0.038
K_a (mmol l^{-1})	0.313±0.059	1.222±0.166	0.645 (0.625, 0.665)	0.226±0.004
K_{ia}/K_a	1.06	0.51	0.92	3.56
ADP				
K_{ia} (mmol l^{-1})	0.050±0.008	0.087±0.012	0.076 (0.078, 0.074)	0.156±0.010
K_a (mmol l^{-1})	0.053±0.003	0.160±0.009	0.086 (0.084, 0.087)	0.043±0.002
K_{ia}/K_a	0.94	0.54	0.88	3.62
V_{max} (μmol min^{-1} mg^{-1})	196.65±17.41 **(D2 only: 391.79)**	175.57±7.45 **(D2 only: 434.50)**	446.07 (460.48, 431.65)	223.32±8.09
K_{cat} (s^{-1})	266.31	295.33	303.21	156.32

With the exception of *Ensis* D2, all kinetic parameters are given as a mean value of at least three kinetic runs ± S.D. *Ensis* D2 values are given as a mean of two kinetic runs, with values from each run given in parentheses. Values shown in bold are V_{max} values for *Ensis* full-length AK and D2–NusA–His-tag fusion protein if only the mass of D2 is taken into consideration.

the binary enzyme–substrate complex to the second substrate (ternary binding constant; in some literature referred to as the true K_m).

For *Ensis* full-length AK, as well as *Ensis* D2, the K_{ia} and K_a values were similar, indicating that this enzyme displays the same binding capacity for the substrate regardless of whether it binds first or second. By contrast, for *Limulus* AK, the binding capacity of the enzyme for the second substrate is nearly four times higher than that for the first substrate. This increased binding capacity for the second substrate is quantified by the K_{ia}/K_a ratios, as shown in Table 2. A K_{ia}/K_a value greater than one indicates positive synergy of substrate binding where the binding of the first substrate facilitates the binding of the second substrate (Segel, 1993). For the majority of CKs and AKs, K_{ia}/K_a is greater than one, as in the case of *Limulus* AK, where K_{ia}/K_a is approximately 3.6 for both substrates. For *Ensis* full-length AK, however, K_{ia}/K_a is approximately equal to 1, indicating that the binding of one substrate has no effect on the binding of the other, and therefore there exists no synergy of substrate binding. Interestingly, the D2 fusion construct clearly displays a paradoxical negative binding synergy, with K_{ia}/K_a values approximating 0.5 for both substrates. In effect, the first substrate functions as a competitive inhibitor for the second substrate!

The V_{max} of a typical 40-kDa AK monomer is illustrated by *Limulus* AK at 223.32 μmoles min^{-1} mg^{-1} protein. While the V_{max} of the full-length *Ensis* AK is in the same range as the V_{max} of the *Limulus* AK, the V_{max} of *Ensis* D2 (446.07 μmol min^{-1} mg^{-1} protein) is nearly twice the value of *Limulus* AK. This high V_{max} of D2 accounts for all of the activity of the full-length enzyme (196.65 μmol min^{-1} mg^{-1} protein). If the V_{max} of the full-length contiguous dimer is determined by taking into account the mass of D2 only, the V_{max} is 391.79 μmol min^{-1} mg^{-1} protein, a value similar to that of D2. In other words, if D2 in the full-length contiguous dimer retains a similar level of activity as it displays as a monomer, then the activity of D2 is high enough to account for the activity of the contiguous dimer as a whole.

The activity of the four AK constructs can also be quantified by the K_{cat} values or turnover numbers. K_{cat} is a measure of an enzyme's maximal catalytic activity, or the number of substrate molecules converted to product per enzyme molecule per unit time. For the full-length *Ensis* AK and D2 and D2–NusA–His-tag fusion constructs, the K_{cat} values are nearly twice that of the monomeric *Limulus* AK. In other words, the *Ensis* AK and D2 constructs are able to produce twice as much product as the same number of *Limulus* AK enzyme molecules. The catalytic throughput is considerably higher (Table 2).

Discussion

In this study, the cDNA and deduced amino acid sequences of the unusual 80-kDa AK from the marine bivalve *Ensis directus* have been determined. The 723 amino acid sequence clearly indicates the presence of two distinct domains, each composed of approximately 360 amino acids. The kinetic constants for the recombinant full-length AK as well as two truncated constructs were determined and compared with the kinetic constants for the typical 40-kDa monomeric AK from *Limulus polyphemus*. The present results suggest that, despite high sequence conservation in both domains of the protein, the *Ensis* contiguous dimeric AK appears to contain only one active domain, namely D2.

Typically, when gene duplication events occur there is an opportunity for divergence and, potentially, the adoption of some novel function by one of the copies. However, divergence of this sort has not taken place in the fused, contiguous phosphagen kinases, suggesting that some evolutionary constraints exist in retaining these multiple-domain enzymes. A comparison of the sequence data of *Ensis* AK with that of *Limulus* AK – the monomeric AK for which the TSAC crystal structure has been solved – revealed that both domains contain the amino acid residues thought to be critical for catalytic activity. With one exception, all residues conserved in AKs and known to contact the substrates during catalysis are present in the two domains.

Other residues purported to be important in stabilizing the closed state of AK are not, however, conserved in the contiguous dimer. For example, Asp62 and Arg193 in *Limulus* AK are two residues suggested to form a hydrogen bond stabilizing the substrate-bound AK (Zhou et al., 1998). These two residues are highly conserved in typical monomeric AKs but are often found divergent in contiguous dimers (Suzuki et al., 1998, 2002), including *Ensis* AK. Site-directed mutagenesis of these two residues in monomeric AK results in a considerable loss of activity (Suzuki et al., 2000a,b); therefore, the lack of either of these two residues in the contiguous dimer was assumed to create an inactive domain (Suzuki et al., 1998).

In the *Ensis* contiguous dimer, however, the Asp62 equivalent is conserved only in the second domain and the Arg193 equivalent is not conserved in either domain. That neither domain of *Ensis* AK can form this hydrogen bond, yet the protein still retains active, supports the proposition that the closed state in contiguous dimeric AKs is stabilized in a different manner to in typical monomeric AKs. Recently, Suzuki et al. (2002) suggested that the hydrogen bond formed between Asp62 and Arg193 in 40-kDa monomeric AKs occurs between His60 and Asp197 in contiguous dimers. Both of these residues are conserved in the two domains of *Ensis* AK. Since contiguous dimeric AKs probably form different bonds to stabilize the substrate-bound enzyme form than do monomeric AKs, lack of conservation of salt-bridge-forming residues in the *Ensis* AK sequence does not imply inactivity in a contiguous dimer.

Retention of catalytic activity would certainly serve as a strong driving force for conservation of amino acid sequence. However, even though sequence data seem to provide clear evidence for the likelihood of two active domains in the *Ensis* contiguous dimer, kinetic results strongly contradict this point of view, suggesting, instead, that the contiguous dimer

functions with only an active second domain. Despite its sequence conservation and the retention of key catalytic residues, D1 appears to have no activity. Initial indications suggesting that the first domain was inactive came from the unsuccessful attempts to express active D1 independently or as a fusion protein. Low temperature expression of D1–NusA–His-tag fusion protein did result in some level of soluble protein, but no activity could be measured. Likewise, purification of the D1 fusion construct from inclusion bodies and refolding on the nickel column produced considerable amounts of soluble, refolded protein, but this protein too was inactive. The lack of measurable activity in the D1 constructs provides strong evidence of an inactive domain. However, we cannot exclude the possibility that the lack of activity could be a result of the inability of D1 to properly fold into an active conformational state in the absence of D2. Recently, Suzuki et al. (2003) have reported that they were also unable to express D1 of the contiguous dimeric AK from the clam *Corbicula japonica* but obtained significant soluble activity with a D2 construct.

The kinetic analyses in this study provide additional evidence supporting an inactive D1. Kinetic parameters were determined for three *Ensis* AK constructs: full-length native AK, D2 alone and a D2 fusion protein construct. In addition, kinetic values were also measured for the typical 40-kDa AK from *Limulus polyphemus*. As seen in Table 2, the V_{max} values for D2 indicate that this 40-kDa truncated protein has a V_{max} that is double that of typical monomeric AK. If only the mass of D2 is taken into consideration, both the D2 fusion protein and the native, full-length AK display similarly high V_{max} values. In other words, the V_{max} of D2 is high enough to account for all activity measured in the full-length enzyme. With its remarkably high specific activity, D2 seems to compensate, so to speak, for the apparent lack of activity in D1 of the contiguous dimer.

Again, the high specific activity observed in D2 can be seen in the K_{cat} values measured for each of the four AK constructs. The high K_{cat} of D2 implies an inactive first domain. The turnover number of the full-length *Ensis* AK, the D2–NusA–His-tag fusion protein and the D2 construct are nearly twice that of a typical 40-kDa monomer. This means that native enzyme and the engineered constructs have a much higher catalytic throughput than do *Limulus* AKs. K_{cat} values for particular enzymes are plastic in an adaptive sense, as has been exquisitely shown for orthologs of lactate dehydrogenase from fish adapted to different thermal environments (Fields and Somero, 1998). Enhancements in K_{cat} are most likely related to adaptive increases in conformational flexibility (Fields and Somero, 1998; Zavodsky et al., 1998). The mechanistic basis for the apparent, elevated K_{cat} of the *Ensis* AK D2 may be due to subtle changes in primary structure, which serve to increase such conformational flexibility.

Although *Ensis* AK has a higher turnover number than monomeric AK from *Limulus*, substrate binding synergism appears to be lacking in the contiguous dimer. The synergistic substrate binding displayed by *Limulus* AK is characteristic of the phosphagen kinase family of enzymes. K_{ia}/K_a values for full-length *Ensis* AK are less than or equal to one, indicating no, or even slightly negative, substrate binding synergy. Initially, this was thought to be a result of movement constraints placed on the enzyme by a second domain. This is not the case, however, as the single domain D2 also displays no substrate binding synergy. The lack of synergistic substrate binding may not, however, be detrimental to the contiguous dimeric AK. While the secondary (K_{ia}) and ternary (K_a) binding constants for full-length *Ensis* do not markedly differ from one another, they also do not differ greatly from the true K_a values measured for *Limulus* AK. In other words, despite the lack of synergistic substrate binding, the *Ensis* contiguous dimer displays nearly the same binding capacity for both substrates as *Limulus* AK displays for the second substrate. Thus, this enzyme is probably not functionally compromised in spite of what appears to be an inactive AK domain.

While the available evidence presented in the present work supports the contention that D1 is inactive, clearly an unsettling paradox remains: namely, the high degree of sequence conservation of the D1 portion of the *Ensis* AK compared with all other AKs. Recall that we failed to express D2 alone; expression was only possible when D2 was expressed as the contiguous dimer or the D2–NusA–His-tag fusion protein. The former yielded considerable soluble activity while expression of the D2–NusA–His-tag fusion protein produced some soluble activity, although a much higher yield was obtained by unfolding and refolding protein from inclusion bodies. It is possible that D2 can only fold properly with an attached amino-terminal domain (either D1 or an amino-terminal fusion construct) and, therefore, the first domain of the protein cannot be truncated or lost. Furthermore, folding of D2 into the active conformation seems to be more efficient when it is attached to D1 as opposed to a heterologous construct. The most recent models for protein folding revolve around a simple ‘nucleation–condensation’ mechanism in which some secondary and tertiary structural elements are formed early to reach a transition state followed by a rapid formation of the native structure (Daggett and Ferscht, 2003). Thus, a properly configured *Ensis* D1 may be critical in this nucleation process.

We are arguing here that D1 has been retained to facilitate proper folding of D2. However, this suggestion does not explain the extreme conservation of the sequence of the first domain. If D1 is inactive, as the kinetic data suggest, then maintenance of catalytic activity is not a constraint for retention of sequence identity and, presumably, the domain should be free to diverge. The kinetic properties of the D2 fusion protein may serve to explain this conservation of D1 sequence. Although the D2–NusA–His-tag fusion protein was able to fold properly and displayed the same V_{max} and K_{cat} values as the truncated D2, the K_a values for the fusion protein were nearly double those of D2. In other words, the presence of an amino-terminal NusA-tag caused a large reduction in the binding capacity of the enzyme for the second substrate. The impact of the NusA-tag domain on the fusion protein is an

indication that improper structure of the amino-terminal domain negatively affects the ability of D2 to bind substrates. This may explain why the first domain of the contiguous dimer has not diverged over time, since small changes in sequence have the potential to considerably alter a protein's structure.

Site-directed mutagenesis studies of true dimeric CKs illustrate the effect of a single amino acid substitution in one subunit on the second subunit. For instance, Hornemann et al. (2000) demonstrated that mutation of the reactive cysteine in one B subunit (denoted B*) of an MB*-CK heterodimer and a BB*-CK homodimer not only resulted in inactivation of the B* subunit but also resulted in a decreased binding capacity for the first substrate in the unaltered domain (evidenced by increased K_{ia} values) as well as a reduction of the V_{max} to one-third that of the unmodified dimer. Similar results for V_{max} were observed by Lin et al. (1994) on a BB* construct.

The above observations clearly show that the structure and dynamics of one subunit impact the functional properties of the second subunit in a dimeric enzyme. While the two subunits of dimeric CKs are not fused, as is the case with *Ensis* contiguous dimeric AK, it seems possible that interactions of domains in the same polypeptide chain might be even more profound. Large scale mutations of key residues in D1 of *Ensis* AK would probably have a detrimental effect on the kinetics and activity of D2. If mutations in the sequence of the first domain of the contiguous dimer result in structural changes negatively impacting the substrate binding properties of the protein as a whole, then sequence divergence would be constrained, possibly resulting in the high percent identity seen in D1.

Gene duplications provide opportunities for adaptive evolution of enzymatic systems (Hughes, 2002). In the case of the AK from *Ensis* and certain other bivalve molluscs, the duplication event was followed by fusion of the genes, resulting in unusual two-domain proteins, which has greatly constrained the degrees of freedom for evolutionary divergence. Overall, the results of the present study indicate that the AK contiguous dimer from *Ensis* functions with activity in only the second domain. Although lacking activity in D1, D2 appears to compensate by having a higher intrinsic catalytic throughput than typical 40-kDa monomeric AKs. The retention of the inactive first domain may limit the potential for deleterious negative synergism of substrate binding as seen in the heterologous construct and may also serve to facilitate proper folding of the protein during synthesis. The net effect is to produce an AK with high catalytic activity in spite of what might appear to be a rather cumbersome evolutionary accident in terms of the contiguous domain structure.

This research effort was supported by grants from the National Science Foundation (IBN-9631907 and IBN-0130024) to W.R.E.

References

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, H. P., Ordhal, C. P., Pearson, M. L., Perriard, J. C., Pickering, L. A., West, B. L. and Zivin, R. A. (1986). Comparisons of creatine kinase primary structures. *J. Protein Chem.* **5**, 1-14.

Blethen, S. L. (1971). Are the arginine kinase isoenzymes of *Limulus polyphemus* the product of two separate genes? *Biochem. Genet.* **5**, 275-286.

Blethen, S. L. (1972). Kinetic properties of the arginine kinase isoenzymes of *Limulus polyphemus*. *Arch. Biochem. Biophys.* **149**, 244-251.

Borson, N. D., Salo, W. L. and Drewes, L. R. (1992). A lock-docking oligo(dT) primer for 5′ and 3′ RACE PCR. *PCR Methods Appl.* **2**, 144-148.

Cleland, W. W. (1967). The statistical analysis of enzyme kinetic data. *Adv. Enzymol.* **29**, 1-32.

Daggett, V. and Ferscht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18-25.

Dumas, C. and Camonis, J. (1993). Cloning and sequence analysis of the cDNA for arginine kinase of lobster muscle. *J. Biol. Chem.* **268**, 21599-21605.

Ellington, W. R., Roux, K. and Pineda, A. O., Jr (1998). Origin of octameric creatine kinases. *FEBS Lett.* **425**, 75-78.

Ellington, W. R. (2001). Evolution and physiological roles of phosphagen systems. *Annu. Rev. Physiol.* **63**, 289-325.

Ellington, W. R. and Bush, J. (2002). Cloning and expression of a lombricine kinase from an echuiroid worm: insights into structural correlates of substrate specificity. *Biochem. Biophys. Res. Commun.* **291**, 939-944.

Fields, P. A. and Somero, G. N. (1998). Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A_4 orthologs of Antarctic notothenid fishes. *Proc. Natl. Acad. Sci. USA* **95**, 11476-11481.

Forstner, M., Kriechbaum, M., Laggner, P. and Wallimann, T. (1998). Structural changes of creatine kinase upon substrate binding. *Biophys. J.* **75**, 1016-1023.

Forstner, M., Berger, C. and Wallimann, T. (1999). Nucleotide binding to creatine kinase: an isothermal titration microcalorimetry study. *FEBS Lett.* **461**, 111-114.

Frohman, M. A., Dush, M. K. and Martin G. R. (1988). Rapid production of full length cDNAs from rare transcripts: amplification using single gene-specific oligonucleotide primers. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.

Graber, N. A. and Ellington, W. R. (2001). Gene duplication events producing muscle (M) and brain (B) isoforms of cytoplasmic creatine kinase: cDNA and deduced amino acid sequences from two lower chordates. *Mol. Biol. Evol.* **18**, 1305-1314.

Grieshaber, M. K., Hardewig, I., Kruetzer, U. and Portner, H. O. (1994). Physiological and metabolic responses to hypoxia in invertebrates. *Rev. Physiol. Biochem. Pharmacol.* **125**, 43-147.

Hornemann, T., Rutishauser, D. and Wallimann, T. (2000). Why is creatine kinase a dimer? Evidence for cooperativity between the two subunits. *Biochim. Biophys. Acta* **1480**, 365-373.

Hughes, A. L. (2002). Adaptive evolution after gene duplication. *Trends Genet.* **18**, 433-434.

Lin, L., Perryman, M. B., Friedman, D., Roberts, R. and Ma, T. S. (1994). Determination of the catalytic site of creatine kinase by site-directed mutagenesis. *Biochim. Biophys. Acta* **1206**, 97-104.

Maggio, E. T. and Kenyon, G. L. (1977). Properties of a CH3-blocked creatine kinase with altered catalytic activity. Kinetic consequences of the presence of the blocking group. *J. Biol. Chem.* **252**, 1202-1207.

Meyer, R. A., Sweeny, H. L. and Kushmerick, M. J. (1984). A simple analysis of the phosphocreatine shuttle. *Am. J. Physiol.* **246**, C365-C377.

Moreland, B. and Watts, D. C. (1967). Molecular weight isoenzymes of arginine kinase in the Mollusca and their association with muscle function. *Nature* **215**, 1092-1094.

Mühlebach, S. M., Gross, M., Wirz, T., Wallimann,T., Perriard, J. C. and Wyss, M. (1994). Sequence homology and structure predictions of the creatine kinase isoenzymes. *Mol. Cell. Biochem.* **133/134**, 245-262.

Newsholme, E. A., Beis, I., Leech, A. R. and Zammit, V. A. (1978). The role of creatine and arginine kinase in muscle. *Biochem. J.* **172**, 533-537.

Qin, W., Khuchua, Z., Cheng, J., Boero, J., Payne, R. M. and Strauss, A. W. (1998). Molecular characterization of the creatine kinases and some historical perspectives. *Mol. Cell. Biochem.* **184**, 153-167.

Reed, G. H. and Cohn, M. (1972). Structural changes induced by substrates and anions at the active site of creatine kinase. Electron paramagnetic resonance and nuclear magnetic relaxation rate studies of the manganous complexes. *J. Biol. Chem.* **247**, 3073-3081.

Segel, I. (1993). *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems.* New York: John Wiley & Sons.

Strong, S. J. and Ellington, W. R. (1993). Horseshoe crab sperm contains a unique isoform of arginine kinase that is present in the midpiece and flagellum. *J. Exp. Zool.* **267**, 563-571.

Strong, S. J. and Ellington, W. R. (1996). Expression of horseshoe crab arginine kinase in *Escherichia coli* and site-directed mutations of the reactive cysteine peptide. *Comp. Biochem. Physiol. B* **113**, 809-816.

Suzuki, T. and Furukohri, T. (1994). Evolution of phosphagen kinase. Primary structure of glycocyamine kinase and arginine kinase from invertebrates. *J. Mol. Biol.* **237**, 353-357.

Suzuki, T., Kawasaki, Y. and Furukohri, T. (1997). Evolution of phosphagen kinase. Isolation, characterization and cDNA-derived amino acid sequence of two-domain arginine kinase from the sea anemone *Anthopleura japonicus*. *Biochem. J.* **328**, 301-306.

Suzuki, T., Kawasaki, Y., Unemi, Y., Nishimura, Y., Soga, T., Kamidochi, M., Yazawa, Y. and Furukohri, T. (1998). Gene duplication and fusion have occurred frequently in the evolution of phosphagen kinases – a two-domain arginine kinase from the clam *Pseudocardium sachalinensis*. *Biochim. Biophys. Acta.* **1388**, 253-259.

Suzuki, T., Fukuta, H., Nagato, H. and Umekawa, M. (2000a). Arginine kinase from *Nautilus pompili*us, a living fossil. Site-directed mutagenesis studies on the role of amino acid residues in the guanidino specificity region. *J. Biol. Chem.* **275**, 23884-23890.

Suzuki, T., Yamamoto, Y. and Umekawa, M. (2000b). *Stichopus japonicus* arginine kinase: gene structure and unique substrate recognition system. *Biochem. J.* **351**, 579-585.

Suzuki, T., Sugimura, N., Taniguchi, T., Unemi, Y., Murata, T., Hayashida, M., Yokouchi, K. and Furukohri, T. (2002). Two-domain arginine kinases from the clams *Solen strictus* and *Corbicula japonica*. Exceptional amino acid replacement of the functionally important D62 by G. *Int. J. Biochem. Cell Biol.* **34**, 1221-1229.

Suzuki, T., Tomoyuki, T. and Uda, K. (2003). Kinetic properties and structural charcteristics of an unusual two-domain arginine kinase from the clam *Corbicula jaonica*. *FEBS Lett.* **533**, 95-98.

Watts, D. C. (1971). Evolution of phosphagen kinases. In *Biochemical Evolution and the Origin of Life* (ed. E. Schoffeniels), pp. 150-173. Amsterdam: North-Holland Publishing.

Watts, D. C. (1975). Evolution of phosphagens along the chordate line. *Symp. Zool. Soc. Lond.* **36**, 105-127.

Wothe, D. D., Charbonneau, H. and Shapiro, B. M. (1990). The phosphocreatine shuttle of sea urchin sperm: flagellar creatine kinase resulted from a gene triplication. *Proc. Natl. Acad. Sci. USA* **87**, 5203-5207.

Yousef, M. S., Clark, S., Fabiola, F., Ellington, W. R. and Chapman, M. S. (2003). Induced fit in guanidino kinases – comparison of substrate-free and transition state analog structures of arginine kinase. *Protein Sci.* **12**, 103-111.

Zammit, V. A. and Newsholme, E. A. (1976). The maximum activities of hexokinase, phosphorylase, phosphofructokinase, glycerolphosphate dehydrogenases, lactate dehydrogenase, octopine dehydrogenase, phosphoenolpyruvate carboxykinase, nucleoside diphosphatekinase, glutamte-oxaloacetate transaminase and arginine kinase in relation to carbohydrate utilization in muscles from marine invertebrates. *Biochem. J.* **160**, 447-462.

Zavodsky, P., Kardos, J., Svingor, A. and Petsko, G. A. (1998). Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. USA* **95**, 7406-7411.

Zhou, G., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W. R. and Chapman, M. S. (1998). Transition state structure of arginine kinase: implications for catalysis of bimolecular reactions. *Proc. Natl. Acad. Sci. USA* **95**, 8449-8454.

Zhou, G., Ellington, W. R. and Chapman, M. S. (2000). Induced fit in arginine kinase. *Biophys. J.* **78**, 1541-1550.