The EMBO Journal Vol.1 No.8 pp.945-951, 1982

Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold

John E.Walker*, Matti Saraste, Michael J.Runswick, and Nicholas J.Gay

Laboratory of Molecular Biology, The MRC Centre, Hills Road, Cambridge CB2 2QH, UK

Communicated by R.Henderson
Received on 5 July 1982

The α- and β-subunits of membrane-bound ATP synthase complex bind ATP and ADP: β contributes to catalytic sites, and α may be involved in regulation of ATP synthase activity. The sequences of β-subunits are highly conserved in *Escherichia coli* and bovine mitochondria. Also α and β are weakly homologous to each other throughout most of their amino acid sequences, suggesting that they have common functions in catalysis. Related sequences in both α and β and in other enzymes that bind ATP or ADP in catalysis, notably myosin, phosphofructokinase, and adenylate kinase, help to identify regions contributing to an adenine nucleotide binding fold in both ATP synthase subunits.

Key words: ATP synthase/myosin/kinases/adenine nucleotides

Introduction

The ATP synthase complex plays a central role in energy transduction in living cells. According to Mitchell's theory of chemiosmosis it uses the proton gradient of bacterial, mitochondrial, and chloroplast membranes to promote the final step in oxidative phosphorylation (or photophosphorylation) of ADP (Mitchell, 1981). According to Kayalar *et al.* (1977) catalysis may occur by an energy-dependent binding change mechanism. In a current proposal (consistent with 3α- and 3β-subunits) this involves three interconvertible and co-operative catalytic sites (Grubmeyer and Penefsky, 1981; Cross and Nalin, 1982). Each may assume three states: 'open', 'loose', or 'tight'. An 'open' site has very low affinity for ligands (ADP and inorganic phosphate) and a 'loose' site has a greater affinity than an 'open' site. Both 'open' and 'loose' sites are catalytically inactive. It is envisaged that binding of substrates ADP and Pi to an 'open' site converts it to a 'loose' site. The proton-motive force generated by oxidation causes substrates to bind tightly and subsequently ATP, formed in the previous cycle, to be released from a 'tight' site. Finally ATP is formed from ADP and Pi in the 'tight' site. The cycle is then repeated (for a review, see Cross, 1981). This model for ATP synthesis has features in common with the mechanism of ATP hydrolysis by the muscle actomyosin complex in that large energy changes are needed for the release (in synthesis) or binding (in hydrolysis) of ATP rather than for formation (or scission) of the P-O bond (see below) (Chappell, 1977; Matsuoka *et al.*, 1981).

The catalytic sites are to be found in an assembly F_1 which is bound to a structure, F_o, buried in the lipid bilayer. In *Escherichia coli* F_o is made up from three proteins *a, b,* and *c* (see Downie *et al.*, 1979; Fillingame, 1981). It contains a trans-membrane proton channel linking the proton-motive force of the membrane to the catalytic sites in F_1. In all species examined F_1 contains two major subunits, α and β, which together constitute ~80% by weight of F_1 and three minor ones, γ, δ, and ε (see Futai and Kanazawa, 1980). These proteins are probably assembled with a stoichiometry of $3\alpha:3\beta:1\gamma:1\delta:1\epsilon$ (Catterall and Pedersen, 1971; Bragg and Hou, 1975; Fillingame, 1981). An adenine nucleotide binding site appears to be present in each of the 3α- and 3β-subunits, making six in all (Slater *et al.*, 1979; Wagenvoord *et al.*, 1980). Covalent modification experiments indicate that β contributes to catalysis (Ferguson *et al.*, 1975; Esch and Allison, 1978; Drutsa *et al.*, 1979; Yoshida *et al.*, 1981a, 1981b); also it is able to bind both ADP and ATP as required for catalysis (Harris, 1978; Wagenvoord *et al.*, 1980). Also, ATP binds to isolated α-subunits with high affinity, thereby causing a large conformational change in the protein (Dunn, 1980).

Recently we established the sequence of the α and β proteins by nucleotide sequence analysis of the *unc* operon which codes for all eight subunits of the *E. coli* complex (Gay and Walker, 1981a, 1981b; Saraste *et al.*, 1981; Walker *et al.*, 1982a). In addition, we determined the protein sequence of the bovine β protein. As described below, α and β are homologous to each other throughout most of their sequences. In two particular regions they are also homologous to other ATP-binding proteins. This implies the presence of a common adenine nucleotide binding fold in all of these proteins.

Results

Conservation of α- and β-subunits

The sequence of the bovine β-subunits was established *de novo* by sequence analysis of peptides isolated from cyanogen bromide and tryptic arginine digests of the protein. All overlaps were experimentally determined by protein sequence analysis. The *E. coli* sequences were determined in the course of DNA sequencing of the *unc* operon (Gay and Walker, 1981a, 1981b; Saraste *et al.*, 1981); the genes for the α- and β-subunits were identified by N-terminal protein sequence analysis of isolated subunits (Dunn, 1980; Walker *et al.*, 1982a).

The sequences of the *E. coli* and bovine mitochondrial β-subunits aligned in Figure 1B are identical in 70% of their amino acids. In maize chloroplasts this protein is also highly conserved (Krebbers *et al.*, 1982). Protein sequence analysis of the bovine α-subunits shows that it also is very similar to the *E. coli* protein.

The α- and β-subunits are homologous

The protein sequences of the *E. coli* α- and β-subunits were compared as described in Materials and methods. The comparison matrix in Figure 2 shows a persistent homology throughout the common sequences of the two proteins. The homology is strongest in the N-terminal regions, particularly around residues 60-130 of the α protein. When aligned as suggested by Figure 2, the α-subunit extends beyond the β-subunit at both ends of the protein chains and a deletion of 15

*To whom reprint requests should be sent.

 0261-4189/82/0108-0945$2.00/0.

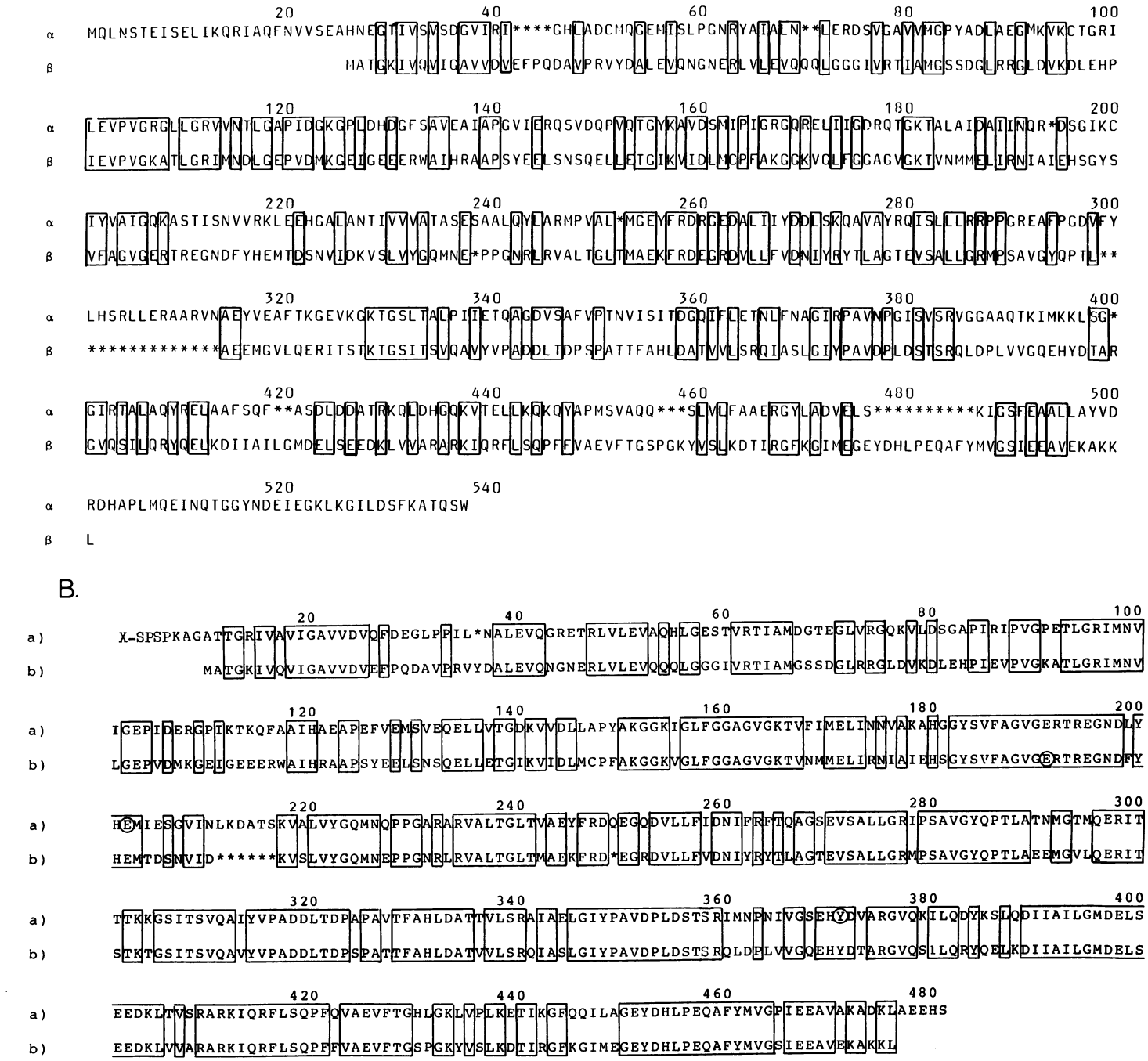

Fig. 1. Amino acid sequences of α- and β-subunits of *E. coli* ATP synthase and of the β-subunit of bovine mitochondrial enzyme. The *E. coli* sequences are derived from DNA sequences (Gay and Walker, 1981; Saraste *et al.*, 1981); the bovine sequence was established by protein sequence analysis as described in Materials and methods. In **A**, *E. coli*, α- and β-subunits are aligned according to the comparison matrix (Figure 2). Identities and conservative substitutions are boxed; * are deletions that correspond to shifts from one diagonal to another in Figure 2. In **B**, comparison of the sequences of the β-subunits of (**a**) *E. coli* and (**b**) bovine mitochondrial ATP synthase. Identities are boxed. In the bovine sequence glutamic acid-202 (circled) is specifically covalently labelled when dicyclohexylcarbodiimide is reacted with F_1 (Yoshida *et al.*, 1981a). In the *E. coli* protein a different but adjacent residue, glutamic acid-191 (circled), reacts (Yoshida *et al.*, 1981b). Tyrosine-371 (circled) in the bovine protein was covalently labelled with *p*-fluorosulphonyl-[^{14}C]benzoyl-5′-adenosine (Esch and Allison, 1978).

amino acids is found around residue 300 of the β chain (see Figure 1A). Approximately 22% of amino acid residues along the common sequence are identical and a further 15% conservatively substituted. Sequence homology between bovine α and β had been suspected earlier on the basis of immunological evidence (N.J.Gay, unpublished data) and from the similarity of their amino acid compositions (Harris, 1981). Taken with the high conservation of heterologous α- and β-subunits this shows that the gene duplication that gave rise to α and β occurred before the emergence of mitochondria, i.e., $>1.2 \times 10^9$ years ago (Cloud *et al.*, 1969).

Homology between α and β and other enzymes that use ATP

To try and delineate sequences in α and β that might be in-

volved in binding adenine nucleotides, we used DIAGON to compare their sequences with a number of enzymes that employ ATP in catalysis. These include notably two enzymes of known three-dimensional structure, phosphofructokinase (Evans and Hudson, 1979; Evans *et al.*, 1981) and adenylate kinase (Pai *et al.*, 1977) in which nucleotide binding sites have been identified. In a third case, myosin, chemical studies (see below) have localised nucleotide binding sites to a specific region of the protein.

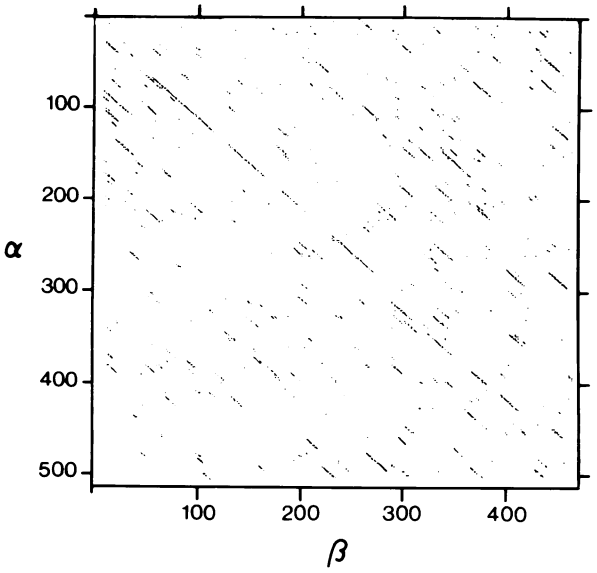

Fig. 2. Comparison matrix of the sequences of the α-subunit (ordinate) and the β-subunit (abcissa) of *E. coli* ATP synthase (see Figure 1). Spans of 19 were used in the calculation. Points on the figure represent homology falling into the best 1% at each point (midpoint of span).

First we searched for the most highly conserved regions in these proteins. Figure 3a, c, e, and g, for example, shows the regions of greatest homology between adenylate kinase and α, β, myosin, and phosphofructokinase, respectively. The sequences labelled A are homologous to each other as are the sequences labelled B (see Table I). These aligned sequences in adenylate kinase, myosin, and β were then used to search for similar sequences in other proteins listed in Table I. In this way two sets of related sequences, A and B in Table I, were detected in some of these proteins but not in others. The sequences A have a number of conserved features notably the sequence G-X-X-X-X-G-K(T)X-X-X-X-X-X-I/V additionally preceded (except in myosin) by a basic amino acid (encircled). The sequences B contain the common sequence R/K-X-X-X-G-X-X-X-L-hydrophobic-hydrophobic-hydrophobic-hydrophobic followed by an aspartic acid residue (encircled). The homologies in myosin, phosphofructokinase, and adenylate kinase are of particular interest as discussed below. Not shown in Table IA is a region of phosphofructokinase homologous to adenylate kinase, but more distant from the other proteins tabulated. Additionally, sequence homologies were not found in proteins thought to show structural homology with adenylate kinase in the region of the nucleotide binding sites (e.g., alcohol dehydrogenase).

Adenylate kinase

Adenylate kinase catalyses phosphorylation of AMP by ATP. The AMP site is very specific for adenine nucleotides, the adenine moiety being deeply buried in the protein. In contrast, the ATP site can bind GTP and is less specific (Secrist *et*

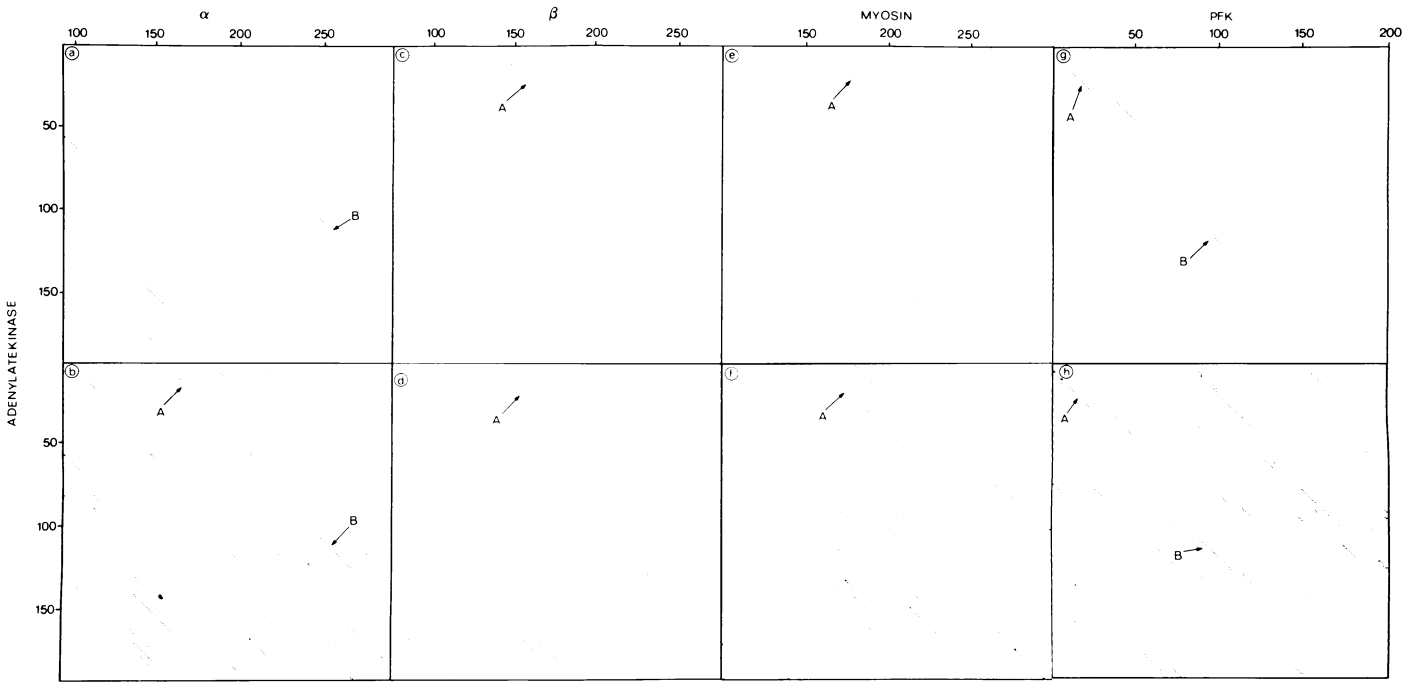

Fig. 3. Comparison of the sequence of adenylate kinase with those of the α- (panels **a** and **b**) and β- (**c** and **d**) subunits of ATP synthase and that of myosin (**e** and **f**), and phosphofructokinase (**g** and **h**). The residues used for the comparisons are numbered on the axes. Comparisons were made with a span of 25. **a**, **c**, **e**, and **g** have a score of 275 with a double matching probability of 6×10^{-4}, **b**, **d**, **f**, and **h** score 265, double matching probability 6×10^{-3}. Thus, in the four upper panels **a**, **c**, **e**, and **g** the double matching probability is one order of magnitude lower than in the lower panels. This means that the most significant homologies are revealed in the upper panels and in the lower panels additional statistically less significant homologies are detected. In the upper panels the points shown are the best 0.1% and in the lower panel the best 0.5% of matches. A and B refer to two segments in Table II which in β are homologous to a continuous myosin sequence designated (a), (b), and (c) (see Table II).

Table I. Alignment of homologous sequences in α- and β-subunits of ATP synthase and other adenine nucleotide binding proteins.

	Protein	Residues	Sequences	References
			1 10 20 30	
A	Bovine ATPase β	144–173	L A P Y A K G G (K) I G L F – [G] G A G V [G K T] V F I M E L I N	a
	E. coli ATPase β	137–165	M C P F A K G G (K) V G L F – [G] G A G [V] G K T V N M M E L I R	b
	E. coli ATPase α	156–184	M I P I G R G Q (R) E L I I – [G] D R G [T G K T] A L A I D A I I	c
	Adenylate kinase	1–30	M E E K L K K S (K) I I F V V [G] G P G S G K [G] T Q C E K I V Q	d
	*Rec*A-protein	52–84	G A G G L P M G (R) I V E I Y [G] P E S [S] G K T T L T L Q V I A	e
	Myosin, nematode	156–184	M L Q D H E N Q S M L I T – [G] E S G A [G K T] E N T K K V I C	f
	Myosin, rabbit	165–193	M L T D R E N Q S I L I T – [G] E S G A [G K T] V N T K R V I Q	g
			1 10 20 30	
B.	Bovine ATPase β	241–270	V A E Y F [R] D Q E [G] Q D V [L L F I] (D) N I F R F T Q A G S E V	a
	E. coli ATPase β	227–255	M A E K F [R] D – E [G] R D V [L L F V] (D) N I Y R Y T L A G T E V	b
	E. coli ATPase α	265–293	M G E Y F [R] D – R [G] E D A [L I I Y] (D) D L S K Q A V A Y R Q I	c
	ATP/ADP translocase	275–297	S N V L – [R] G M G [G] A F V [L V L Y] (D) E I K K F V	h
	Adenylate kinase	102–130	G E E F E [R] K – I [G] Q P T [L L L Y] V (D) A G P E T M T K R L L	f
	Phosphofructokinase	85–113	G I E Q L [K] K – H [G] I Q G [L V V I] G G (D) G S Y Q G A K K L T	i

In A boxed sequences are identical; in B the boxes contain either conserved basic or conserved hydrophobic residues. The significance of encircled residues is discussed in the text. References: a, this work; b, Saraste *et al.* (1981); c, Gay and Walker (1981); d, Pai *et al.* (1977); e, Sancar *et al.* (1980); f, J.Karn, personal communication; h, Aquila *et al.* (1982); i, Kolb *et al.* (1980).

al., 1972; Price *et al.*, 1973; Slotin and Hampton, 1975). The two sequences in adenylate kinase listed in Table I both make important contributions to the AMP binding site (Pai *et al.*, 1977). Sequence A (residues 16–23) forms a flexible loop structure with the sequence Gly-X-Gly-X-Gly. A similar sequence is to be found in β-ATP synthase. In adenylate kinase the conformation of the loop that may interact with AMP changes dramatically when AMP is bound with the phosphate of AMP. Thus it is noteworthy that a lysine following the loop may interact with the α-phosphate in adenylate kinase (Pai *et al.*, 1977) and is conserved throughout the examples in Table IA. Sequence B (residues 110–120 of adenylate kinase) forms a hydrophobic β-sheet structure at the back of the AMP pocket. Close to it as the other part of this structure is a β-sheet including the hydrophobic sequence Ile-Ile-Phe-Val-Val in sequence A (Table I). Also, it contains, at position 119, an aspartic acid residue that could bind magnesium in a way similar to that proposed in phosphofructokinase (see below).

Phosphofructokinase

The region of homology (residues 82–110) contains a β-sheet (residues 97–102, Figure 5) which is part of the substrate binding pocket. It makes important contacts with an α-helix (residues 105–114) which provides direct contacts with the substrate, Mg-ATP. Interestingly, the conserved sequence (see Table IB) includes an aspartic acid residue (Asp-103 in phosphofructokinase) which is close to, and may be important for, binding magnesium. This aspartic acid residue, encircled in Table IB, is found in a similar position in all cases listed. It is interesting to note that the effector molecule, ATP, is bound into the regulatory site of phosphofructokinase predominantly by its phosphate groups (Evans and Hudson, 1979). This site was not picked up in our survey.

Myosin

The sequence detected in myosin listed in Table IA is located in a region of the nematode sequence (J. Karn, personal communication) corresponding to the 25-kilodalton (kd) fragment generated by limited tryptic proteolysis of rabbit heavy meromyosin (Bálint *et al.*, 1975). This fragment is derived from the S_1 head portion of myosin (Bálint *et al.*, 1975) containing the site of ATP hydrolysis. Several other pieces of evidence point to part of the 25-kd fragment being involved in nucleotide binding; it is labelled by photoaffinity labelling with ATP analogues (Szilagyi *et al.*, 1979; R.G.Yount, personal communication) and limited tryptic hydrolysis of the hinge region between the 25-kd fragment and the adjacent 50-kd fragment is influenced by Mg-ATP (Muhlrad and Hozumi, 1982).

The suggested analogy between the enzymic mechanism of ATP synthase and actomyosin (Chappell, 1977) adds further interest to the comparisons between myosin and α and β. Comparison of myosin and α (Figure 4) reveals sequence A and an additional adjacent sequence containing a cysteine residue (see Table II). The β/myosin comparison shows both sequences A and B and an additional homology adjacent to A in the myosin sequence and containing part of the region A of β. This suggests that the structural elements represented by A and B in α or β are in the order BA in myosin; in addition A in myosin is followed by a third element related to A (see Table II).

ADP/ATP translocase

This protein is embedded in the inner mitochondrial membrane. It forms an essential element in ATP synthesis by removing ATP from the mitochondrial matrix and by replenishing the matrix with ADP from the cytoplasm. In contrast with all other examples in Table I its action does not involve scission of phosphate-phosphate bonds. A weak homology (type B, Table I) is found near its C terminus suggesting that this part of the sequence is involved forming a nucleotide binding pocket. Of greater significance, however, is the presence of a strong 3-fold repeat in the sequence described elsewhere. One repeat includes residues 275–297, Table IB. The two other homologous regions were not detected in the survey described here and may form transmembrane α-helices (Saraste and Walker, 1982).

Discussion

Significance of homologies

In making these comparisons in the first instance we have

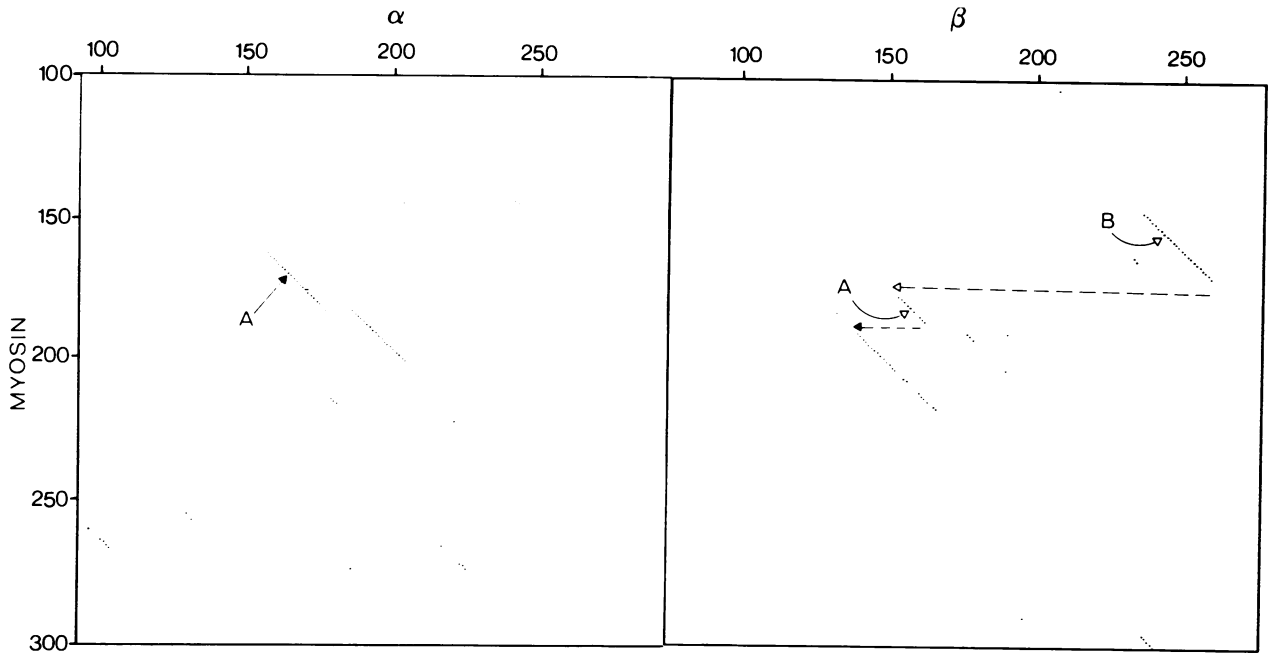

Fig. 4. Comparison of segments of myosin with α- and β-subunits of ATP synthase. Span 25; score 272; double matching probability 1.7×10^{-3}. B and A refer to two segments in Table II which in β are homologous to a continuous myosin sequence designated (a), (b), and (c) (see Table II).

Table II. Part of the sequence of nematode (J.Karn, unpublished results) and rabbit (M. Elzinga, unpublished results) myosin aligned with segments of the sequences of the β-subunits of *E. coli* and bovine mitochondrial ATP synthase.

	(a)	(b)	(c)
Bovine β	A E Y F R D Q E G Q D V L L F I * D N I F R F T Q A G S E V * S A L L	G G A G V G K T V F I	L A P Y A K G G K I G L F G G A G V G
E. coli β	A E K F R D * E G R D V L L F V * D N I Y R Y T L A G T E V * S A L L	G G A G V G K T V N M	M C P F A K G G K V G L F G G A G V G
Myosin, nematode (139–206)	M G K R K * T E M P P H L F A V S D E A Y R N M L Q D H E N Q S M L I T	G E S G A G K T E N T K	K V I C Y F A A V G A S Q Q E G G A E V D
Myosin, rabbit (143–204)	R G K K R * Q E A P P H I F S I S D N A Y Q F M L T D R E N Q S I L I T	G E S G A G K T V N T K	R V I Q Y F A T I A I/V T G D K
	B	A	

A and B correspond to Table IA and IB. (a) Bovine β residues 243–275, *E. coli* β residues 228–259; (b) bovine β residues 159–169, *E. coli* β residues 151–160; and (c) β residues 146–164, *E. coli* β residues 137–155. Conservative substitutions are boxed.

sought out the most similar sequences in these proteins based on the assumption that conservative constraints will be strongest at the substrate (i.e., nucleotide) binding sites. Employing a similar tactic Jörnvall *et al.* (1981) recently demonstrated a similarity between pre-albumin (which binds thyroxine) and several pre-hormones. They found a particular segment of pre-albumin that always had the highest relative similarity to the pre-hormones.

Doolittle (1981) has pointed out that any alignment of two proteins will show some accidental homology making difficult the interpretation of the significance of weakly homologous sequences. However, alignment of several proteins with each other avoids this problem, because it is much less likely that several proteins will have the same conserved sequence. This is one reason for thinking that the sequences in Table I are significantly homologous. Perhaps a stronger reason is that independent information implicates the sequences we have found as being in regions which effect nucleotide binding in myosin and the phosphofructo- and adenylate-kinases.

Secondary structure of adenine nucleotide fold

Comparison of the secondary structures of the regions covered by sequences A and B in adenylate kinase and sequence B in phosphofructokinase with the predicted secondary structures of the corresponding regions in the α- and β-subunits is illustrated in Figure 5. The structures predicted for α and β are not entirely concordant although given the sequence homology between α and β it is most likely that they will have closely related structures as has been found in other families of proteins related by sequence homologies (serine proteases, globins, cytochromes). It tends to support the view, however, that the ATP synthase subunits will have a fold to bind adenine nucleotides related to those in the kinases. The three-dimensional structures of the two segments A and B in adenylate kinase are homologous to the structures in the region of nucleotide binding sites in a number of enzymes (Rossman *et al.*, 1975). Adenine nucleotide binding folds have recently been identified in tubulin (Krauhs *et al.*, 1981) and methionyl-tRNA synthetase (Risler *et al.*, 1981) and related to the Rossman fold for mono- and dinucleotides. Although these and other proteins were included in our survey, convincing sequence homologies with the proteins in Table I were not detected. The fold in α and β described here and the examples shown in Table I extend the range of distribution of the fold to include ATPases (myosin, *rec*A, ATP synthase) and the ATP-ADP translocase which is embedded in the inner mitochondrial membrane, possibly in close

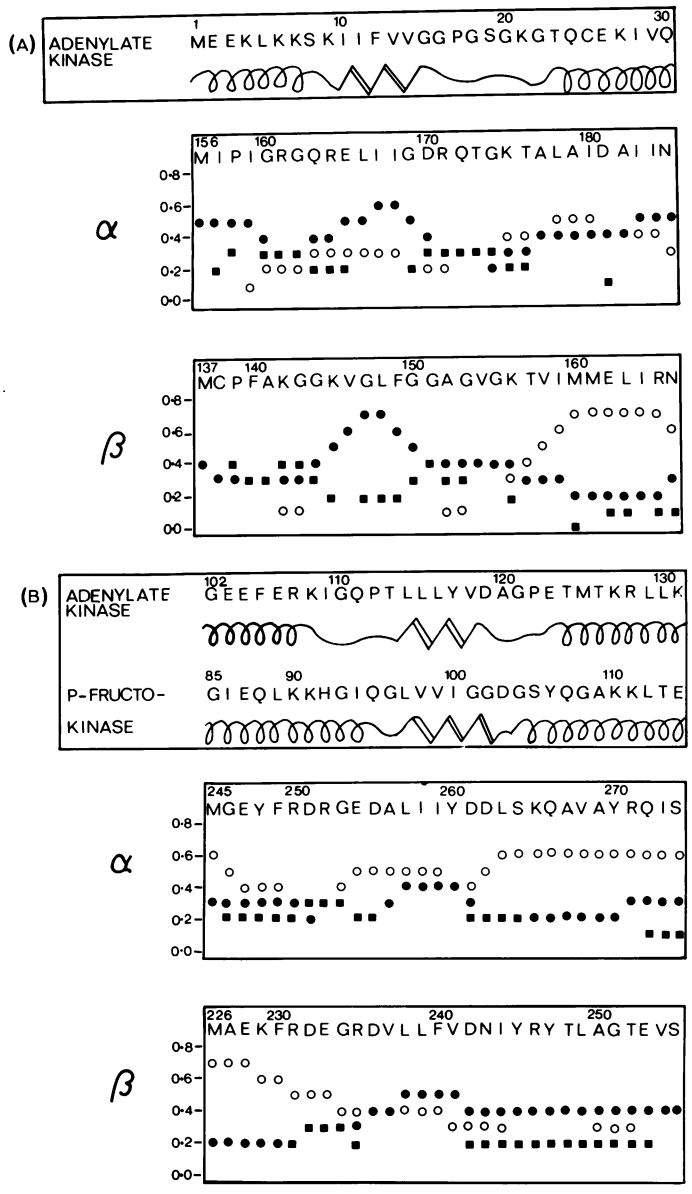

Fig. 5. Predicted secondary structures in regions proposed to be contributing to the nucleotide binding fold in *E. coli* ATP synthase α- and β-subunits. They are compared with secondary structures of homologous sequences in adenylate kinase and phosphofructokinase in regions shown in Table I. The secondary structures in α- and β- were predicted by McLachlan's procedure (McLachlan, 1977). The vertical axis is a probability scale. ●, β-sheet, ○, α-helix; ■, β-bend. In the secondary structures in adenylate kinase and phosphofructokinase, α-helices are denoted by ⌇⌇, β-sheet by ⩙ and β-bends by ~. (A) and (B) correspond to the sequences in Table IA and IB, respectively.

association with the ATP synthase complex (M.Saraste *et al.*, unpublished data).

Implications for ATP synthase

Chemical labelling studies have shown that the α- and β-subunits of ATP synthase do not have identical functions in the enzyme complex. For instance, the β-subunit contains specific amino acids which have been implicated in catalysis for which function counterparts are not found in the α-subunit. Thus, Yoshida *et al.* (1981b) showed that a particular glutamic acid residue in the β-subunit of a thermophile F_1 reacts with dicyclohexylcarbodiimide, thereby inactivating the enzyme. The inhibition is stimulated by ADP and reduced by Mg^{2+} and so it is suggested that this amino acid participates in Mg^{2+} binding. A different, but adjacent, glutamic acid reacts in the bovine mitochondrial protein (Yoshida *et al.*, 1981b). Counterparts of both residues are to be found in the *E. coli* β-protein as indicated in Figure 1B. A second example concerns β-Tyr-371 in Figure 1B identified by Esch and Allison (1978) as the tyrosyl residue modified by the adenine nucleotide analogue, *p*-fluorosulphonylbenzoyl-5'-adenosine which irreversibly inactivates F_1. This tyrosine is also conserved in the *E. coli* β-protein. However, again the α-subunit does not have equivalent counterparts. Nonetheless, the observed persistent homology between α and β probably is a reflection of common functions including their ability to bind adenine nucleotides. The sequences proposed to form part of the nucleotide binding sites in α and β that we describe are insufficient to explain the extensive overall homology between these two proteins. Other parts of the sequence must also contribute to the nucleotide binding fold. For example, photoaffinity labelling experiments with 8'-azido-ATP have shown that lysines 303 and 304 in β (Figure 1B) react and presumably are in the ATP pocket (M.Hollemans and J.E.Walker, unpublished data). Also it seems likely that other common constraints must operate. One such constraint might be related to the associations between α and β and other proteins in the enzyme complex, particularly γ, δ, and ε and protein *b*, the membrane-bound constituent of F_0 which protrudes from the membrane, forming important helix-helix interactions with the F_1 assembly (Gay and Walker, 1981b; Walker *et al.*, 1982c).

Materials and methods

Proteins and peptides

Beef heart mitochondrial F_1-ATPase was prepared by chloroform extraction of submitochondrial particles (Beechey *et al.*, 1975) followed by gel filtration through Ultrogel AcA34 (Saraste *et al.*, 1982). The β-subunit was purified from urea-solubilised bovine F_1 by gel filtration on Sepharose 6B to remove the γ-, δ-, and ε-subunits and ion-exchange chromatography of the fraction containing the α and β proteins on DEAE-cellulose in 8 M urea, 10 mM Tris HCl (pH 7.2), 1 mM β-mercaptoethanol, and 0.1% phenylmethyl sulphonyl fluoride with a NaCl gradient of 0–0.15 M. The α-subunit was not retained by the column under these conditions whilst the β-subunit eluted at 0.06–0.12 M NaCl (see Knowles and Penefsky, 1972).

Tryptic arginine peptides were isolated from a tryptic digest of succinyl-β-protein (250 nmol) by fractionation first by gel filtration on Sephadex G-75 (superfine) in 0.5% ammonium bicarbonate. Pooled fractions were subjected to h.p.l.c. on a radially compressed column (Waters C_{18} Radialpak) in 10 mM sodium acetate pH 4.5 with linear acetonitrile gradients up to 60% or 100% using an Altex chromatography system. The pumping rate was 2 ml/min.

Peptide sequence analysis

Automated analysis of peptides derived from the β-subunit was performed in either a modified Beckman 890B Sequencer or in a solid phase microsequencer (Walker *et al.*, 1982a, 1982b). Phenylthiohydantoins were identified by reverse phase chromatography on Zorbax-ODS (Brock and Walker, 1980). Manual sequence analysis of peptides was performed by a scaled down version of the method described by Chang *et al.* (1978). With this procedure 2–10 nmol of peptide could be sequenced up to the twentieth residue.

Amino acid analyses of acid hydrolysates were obtained with the aid of a Durrum D-500 analyser.

Computer analyses

Secondary structures of the *E. coli* α- and β-subunits of ATP synthase were predicted from their primary sequences using a program described by McLachlan (1977). This method applies the laws of statistical inference to amino acid frequency data from proteins of known structure. It calculates the probability of α-helix, β-sheet, β-bends, and irregular structure along the length of a given sequence.

Amino acid sequences of proteins were compared with an interactive graphics program DIAGON (Staden, 1982) using the proportional matching option (McLachlan, 1971). This involves calculating a score at each position in

the comparison matrix by summing points found when looking forwards and backwards along a diagonal line of given length. This length is called the span. The algorithm uses a score matrix for every possible pair of amino acid substitutions. This matrix, MDM78, was calculated from accepted point mutations of 71 families of related proteins and found to be the most powerful score matrix for the detection of distant relationships (Dayhoff, 1978). High scores correspond to good matches, low scores to poor (insignificant) matches.

Statistical significance of matches is assessed by the double matching probability (see legends to Figures 3 and 4). This is the probability of finding a particular score, given two infinitely long sequences of the same amino acid compositions as the sequences being compared, with the same span length and score matrix (McLachlan, 1971; Staden, 1982). A double matching probability of 6×10^{-4}, for example (see Figure 3 legend), means that in a random sequence of the same amino acid composition about one sequence in 1700 should reach this scoring level.

Acknowledgements

We thank L.Bogorad, M.Elzinga, and J.Karn for access to unpublished sequences, R.G.Yount for unpublished information about myosin, and A.D.McLachlan for help with secondary structure predictions. M.S. is supported by a long-term EMBO fellowship and N.J.G. by an MRC studentship.

References

Aquila,H., Misra,D., Eulitz,M., and Klingenberg,M. (1982) *Hoppe-Seyler's Z. Physiol. Chem.*, **363**, 345-349.

Bálint,M., Sréter,F.A., Wolf,I., Nagy,B., and Gergely,J. (1975) *J. Biol. Chem.*, **250**, 6168-6177.

Beechey,R.B., Hubbard,S.A., Linnett,P.E., Mitchell,A.D., and Munn,E.A. (1975) *Biochem. J.*, **148**, 533-537.

Bragg,P.D., and Hou,C. (1975) *Arch. Biochem. Biophys.*, **167**, 311-321.

Brock,C.J., and Walker,J.E. (1980) *Biochemistry (Wash.)*, **19**, 2873-2882.

Catterall,W.A., and Pedersen,P.L. (1971) *J. Biol. Chem.*, **246**, 4987-4994.

Chang,J.H., Brauer,D., and Wittmann-Liebold,B. (1978) *FEBS Lett.*, **93**, 205-214.

Chappell,J.B. (1977) in *ATP, Carolina Biology Reader 50,* published by Carolina Biological Supply Co., Burlington, NC.

Cloud,P.E., Licari,G.R., Wright,L.A., and Troxel,B.W. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 623-630.

Cross,R.L. (1981) *Annu. Rev. Biochem.*, **50**, 681-714.

Cross,R.L., and Nalin,C.M. (1982) *J. Biol. Chem.*, **257**, 2874-2881.

Dayhoff,M.O. (1978) in *Atlas of Protein Sequence and Structure,* Vol. **5**, Suppl. 3, published by National Biomedical Research Foundation, Washington, DC, pp. 356-358.

Doolittle,R.F. (1981) *Science (Wash.)*, **214**, 149-158.

Downie,J.A., Gibson,F., and Cox,G.B. (1979) *Annu. Rev. Biochem.*, **48**, 103-131.

Drutsa,V.L., Kozlov,I.A., Milgrom,Y.M., Shabarova,Z.A., and Sokolova, N.I. (1979) *Biochem. J.*, **182**, 617-619.

Dunn,S.A. (1980) *J. Biol. Chem.*, **255**, 11857-11860.

Esch,F.S., and Allison,W.S. (1978) *J. Biol. Chem.*, **253**, 6100-6106.

Evans,P.R., and Hudson,P.J. (1979) *Nature,* **279**, 500-504.

Evans,P.R., Farrants,G.W., and Hudson,P.J. (1981) *Phil. Trans. R. Soc. Lond. B,* **293**, 53-62.

Ferguson,S.J., Lloyd,W.J., Lyons,M.H., and Radda,G.K. (1975) *Eur. J. Biochem.*, **54**, 117-126.

Fillingame,R.H. (1981) *Curr. Top. Bioenerg.*, **11**, 35-106.

Futai,M., and Kanazawa,H. (1980) *Curr. Top. Bioenerg.*, **10**, 181-215.

Gay,N.J., and Walker,J.E. (1981a) *Nucleic Acids Res.*, **9**, 2187-2194.

Gay,N.J., and Walker,J.E. (1981b) *Nucleic Acids Res.*, **9**, 3919-3926.

Grubmeyer,C., and Penefsky,H.S. (1981) *J. Biol. Chem.*, **256**, 3728-3734.

Harris,D.A. (1978) *Biochim. Biophys. Acta,* **463**, 245-273.

Harris,D.A. (1981) *BioSystems,* **14**, 113-121.

Jörnvall,H., Carlström,A., Pettersson,T., Jacobsson,B., Persson,M., and Mutt,V. (1981) *Nature,* **291**, 261-263.

Kayalar,C., Rosing,J., and Boyer,P.D. (1977) *J. Biol. Chem.*, **252**, 2486-2491.

Knowles,A.F., and Penefsky,H.S. (1972) *J. Biol. Chem.*, **247**, 6624-6630.

Kolb,E., Hudson,P.J., and Harris,J.I. (1980) *Eur. J. Biochem.*, **108**, 587-597.

Krauhs,E., Little,M., Kempf,T., Hofer-Warbinek,R., Ade,W., and Postingl, H. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 4156-4160.

Krebbers,E.T., Larrinna,I.M., McIntosh,L., and Bogorad,L. (1982) *Nucleic Acids Res.*, **10**, 4985-5002.

McLachlan,A.D. (1971) *J. Mol. Biol.*, **61**, 409-424.

McLachlan,A.D. (1977) *Int. J. Quantum Chem.*, **12**, Suppl. 1, 371-385.

Matsuoka,I., Watanabe,T., and Tonomura,Y. (1981) *J. Biochem.*, **90**, 967-989.

Mitchell,P.D. (1981) in Lee,C.P., Schatz,G., and Dallner,G. (eds.), *Mitochondria and Microsomes,* Addison-Wesley, Reading, MA, pp. 427-457.

Muhlrad,A., and Hozumi,T. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 958-962.

Pai,E.G., Sachsenheimer,W., Schirmer,R.H., and Schultz,G.E. (1977) *J. Mol. Biol.*, **114**, 37-45.

Price,N.C., Reed,G.H., and Cohn,M. (1973) *Biochemistry (Wash.)*, **12**, 3322-3327.

Risler,J.L., Zelwer,C., and Brunie,S. (1981) *Nature,* **292**, 384-386.

Rossmann,M.G., Liljas,A., Branden,C.I., and Babaszak,L.J. (1975) in Boyer,P.D. (ed.), *The Proteins,* Vol. **11a**, Academic Press, NY/London, pp. 62-103.

Sancar,A., Stachelek,C., Konigsberg,W., and Rupp,W.D. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 2611-2615.

Saraste,M., and Walker,J.E. (1982) *FEBS Lett.*, 250-254.

J.E. (1982) *Proceedings 2nd European Bioenergetics Symposium,* 629-630.

Saraste,M., and Walker,J.E. (1982) *FEBS Lett.*, in press.

Saraste,M., Eberle,A., Gay,N.J., Runswick,M.J., and Walker,J.E. (1981) *Nucleic Acids Res.*, **9**, 5287-5296.

Secrist,J.A., Barrio,J.R., Leonard,N.J., and Weber,G. (1972) *Biochemistry (Wash.)*, **11**, 3499-3506.

Slater,E.C., Kemp,A., van der Kraan,I., Muller,J.L.M., Roveri,O.A., Verschoor,G.J., Wagenvoord,R.J., and Wielders,J.P.M. (1979) *FEBS Lett.*, **103**, 7-11.

Slotin,L.A., and Hampton,A. (1975) *Biochemistry (Wash.)*, **14**, 5438-5446.

Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951-2961.

Szilagyi,L., Bálint,M., Sreter,F.A., and Gergely,J. (1979) *Biochem. Biophys. Res. Commun.*, **87**, 936-945.

Wagenvoord,R.J., Kemp,A., and Slater,E.C. (1980) *Biochim. Biophys. Acta,* **593**, 204-211.

Walker,J.E., Auffret,A.D., Carne,A., Gurnett,A., Hanisch,P., Hill,D., and Saraste,M. (1982a) *Eur. J. Biochem.*, **123**, 253-260.

Walker,J.E., Eberle,A., Gay,N.J., Hanisch,P., Saraste,M., and Runswick, M.J. (1982b) in Elzinga,M. (ed.), *Methods in Peptide and Protein Sequence Analysis,* Humana, Clifton, NJ, pp. 337-354.

Walker,J.E., Saraste,M., and Gay,N.J. (1982c) *Nature,* 867-869.

Yoshida,M., Poser,J.W., Allison,W.S., and Esch,F.S. (1981a) *J. Biol. Chem.*, **256**, 148-153.

Yoshida,M., Allison,W.J., and Esch,F.S. (1981b) *Fed. Proc.*, **40**, 1734.