Hypothesis

A common topology of proteins catalyzing ATP-triggered reactions

Masasuke Yoshida*, Toyoki Amano

Research Laboratory of Resources Utilization, R-1, Tokyo Institute of Technology, Nagatsuta 4259, Yokohama 227, Japan

Received 11 November 1994

Abstract A protein fold, six parallel β strands surrounding the central α helix, is likely to be a common structure in protein families known to have a typical set of nucleotide binding consensus sequence motifs A and B and to catalyze ATP-triggered reactions. According to this ATP-triggered protein fold, the conserved Glu (or Asp), which acts as a general base to activate a water molecule for an in-line attack of the γ-phosphate, is at the exit of the second β strand. The fifth β strand may be involved in propagation of conformational change triggered by ATP hydrolysis.

Key words: ATP binding domain; ABC transporter; MDR folding; SecA folding; Lon protease; Proteasome

1. Introduction

Since Walker and his colleagues first reported two consensus sequence motifs A (GXXXXGKT/S; X can be varied) and B (ZZZZD; Z is a hydrophobic residue) associated with nucleotide binding [1], a large number of proteins have been shown to have these motifs. In all cases, motif A resides on the amino-terminal side relative to motif B and the distance in the primary structure between two motifs is 50–130 residues. During the study of essential residues of the F_1-ATPase β subunit, which has a typical set of motifs A and B, we found that a conserved Glu located at 26 residues from the Lys in motif A is essential for ATPase activity; the elimination of the carboxyl group from this Glu by specific chemical modification with dicyclohexylcarbodiimide or by mutational analysis always resulted in almost complete loss of ATPase activity [2,3]. The modified or mutated enzymes retained the ability to bind substrate but were unable to catalyze even a single catalytic turnover. On the other hand, based on crystallography of the *Escherichia coli* recA protein which also has a typical set of motifs A and B, Story and Steitz proposed that Glu^{96} of the recA protein is in a position to serve as a general base to activate a water molecule for an in-line attack of the γ-phosphate during ATP hydrolysis [4]. This Glu^{96} is located at 24 residues from the Lys in motif A. Although there is no overall amino acid sequence similarity between the F_1-ATPase β subunit and the recA protein, it appeared more than coincidental that the catalytic Glu's are located at almost the same position in both proteins and a similar topology of their ATP-binding domains was indicated [5]. Indeed, the recently unraveled crystal structure of mitochondrial F_1-ATPase (MF_1) [6] clearly showed that the folding topology of the ATP binding domain of MF_1-β subunit is almost the same as that of the recA protein, and that the essential Glu (Glu^{188}) of MF_1-β subunit above described is really located at the right position to act as a general base. Here, we propose that proteins containing a typical set of Walker's motifs A and B have the conserved Glu (or Asp) between the two motifs and that their topologies of the ATP binding domains are probably all common.

2. Conserved 'catalytic carboxylate' in various protein families

When the amino acid sequences of the region covering motifs A and B are aligned for several protein families which have a set of typical motifs A and B (Fig. 1), it becomes clear that occurrence of a Glu or Asp (Fig. 1, catalytic carboxylate, *E*) at 24 ± 2 residues from the Lys of motif A (Fig. 1, motif A, *K*) is a common feature (Fig. 1, distance of *E-K*). These Glu's (or Asp's) are well conserved in each of the protein families. Fig. 1 includes a variety of proteins which are supposed to catalyze ATP-triggered reactions and are expected to undergo conformational change upon hydrolysis (or binding) of ATP. Therefore, it is highly likely that the Glu's (or Asp's) indicated in Fig. 1 play the same functional role, activation of a water molecule to attack γ-phosphate of ATP in each of proteins as is proposed for the recA protein and F_1-ATPase β subunit. Interestingly, F_1-ATPase α subunit, the overall amino acid sequence of which is similar to the β subunit, has Gln instead of Glu at this position. This is quite reasonable since it has been well established that the ATP-binding site of the α subunit does not participate in ATP hydrolysis [7]. Similarly, TBP1, in which Glu is replaced by Gln, can probably bind ATP but hydrolyze it only very slowly, if at all. Examination of the primary and tertiary structure of nucleotide binding proteins which do not have a typical set of motifs A and B, such as *ras* p21 [8], EF-Tu [9], transducin-α [10], $G_{i\alpha 1}$ [11], adenylate kinases [12,13], Hsc70 ATP binding domain [14], actin [15], and myosin [16] revealed that such a catalytic Glu (or Asp) is not present at the corresponding position. Although Glu^{203} of transducin-α is assigned as a catalytic residue, its position in the primary structure is different from the Glu (or Asp) described here and the architecture of its ATP binding domain [10] is not similar to those of the recA protein and MF_1-β subunit.

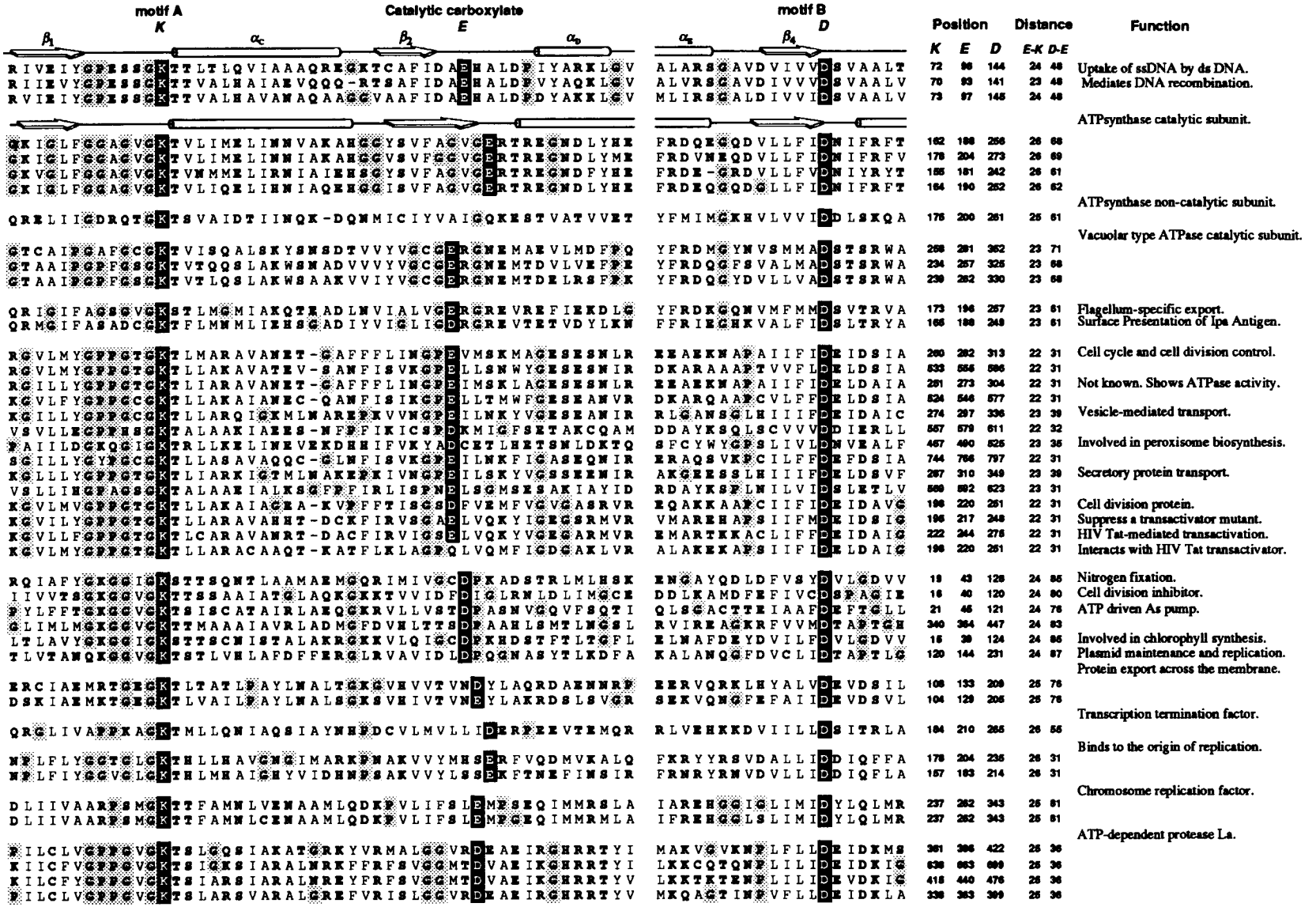
3. Putative 'catalytic carboxylate' in ABC transporters

In members of the ABC (ATP-binding cassette) transporter family [17] which have a typical set of motifs A and B, the

*Corresponding author. Fax: (81) (45) 924-5277.

Abbreviations: MF_1, F_1-ATPase from a bovine heart mitochondria.

SSDI 0014-5793(94)01438-8


motif A
K
Catalytic carboxylate
E
motif B
D
β1
αc
β2
αD
αE
β4
Position
Distance
K E D E-K D-E
Function
RecA
E.coli 72 96 144 24 48 Uptake of ssDNA by ds DNA.
B.subtilis 70 93 141 23 48 Mediates DNA recombination.
M. tuberculosis 73 97 145 24 48
H+-ATPase
F1-β
ATPsynthase catalytic subunit.
bovine 162 186 256 26 68
spinach 176 204 273 26 69
E.coli 185 181 242 26 61
Bacillus PS3 164 190 252 26 62
F1-α
ATPsynthase non-catalytic subunit.
Bacillus PS3 176 200 261 25 61
V1 69k subunit
Vacuolar type ATPase catalytic subunit.
Carrot 258 281 352 23 71
T.thermophilus 234 257 325 23 68
S.acidocaldarius 239 262 330 23 68
FliI family
Fli I 173 196 267 23 61 Flagellum-specific export.
Spa47 166 188 249 23 61 Surface Presentation of Ipa Antigen.
Putative ATPase
CDC48 (N) 260 282 313 22 31 Cell cycle and cell division control.
CDC48 (C) 533 556 586 22 31
p97 (N) 281 273 304 22 31 Not known. Shows ATPase activity.
p97 (C) 524 546 577 22 31
NSF (N) 274 297 336 23 39 Vesicle-mediated transport.
NSF (C) 557 579 611 22 32
Pas1 (N) 467 490 525 23 35 Involved in peroxisome biosynthesis.
Pas1 (C) 744 766 797 22 31
Sec18 (N) 287 310 349 23 39 Secretory protein transport.
Sec18 (C) 599 592 623 23 31
FtsH 196 220 251 22 31 Cell division protein.
SUG1 196 217 248 22 31 Suppress a transactivator mutant.
MSS1 222 244 276 22 31 HIV Tat-mediated transactivation.
TBP1 196 220 251 22 31 Interacts with HIV Tat transactivator.
NifH family
NifH 19 43 128 24 85 Nitrogen fixation.
MinD 16 40 120 24 80 Cell division inhibitor.
ArsA (N) 21 45 121 24 76 ATP driven As pump.
ArsA (C) 340 364 447 24 83
FrxC 16 39 124 24 85 Involved in chlorophyll synthesis.
IncC 120 144 231 24 87 Plasmid maintenance and replication.
SecA protein
Protein export across the membrane.
E.coli 108 133 209 25 76
P.luteril 104 129 205 25 76
Rho Protein
Transcription termination factor.
E.coli 184 210 265 26 55
DnaA protein
Binds to the origin of replication.
E.coli 178 204 236 26 31
B.subtilis 157 183 214 26 31
DnaB protein
Chromosome replication factor.
E.coli 237 262 343 25 81
S.typhimurium 237 262 343 25 81
Lon Protease
ATP-dependent protease La.
E.coli 361 386 422 25 36
S.cerevisiae 638 663 699 25 36
H.sapiens 415 440 476 25 36
B.brevis 338 363 399 25 36

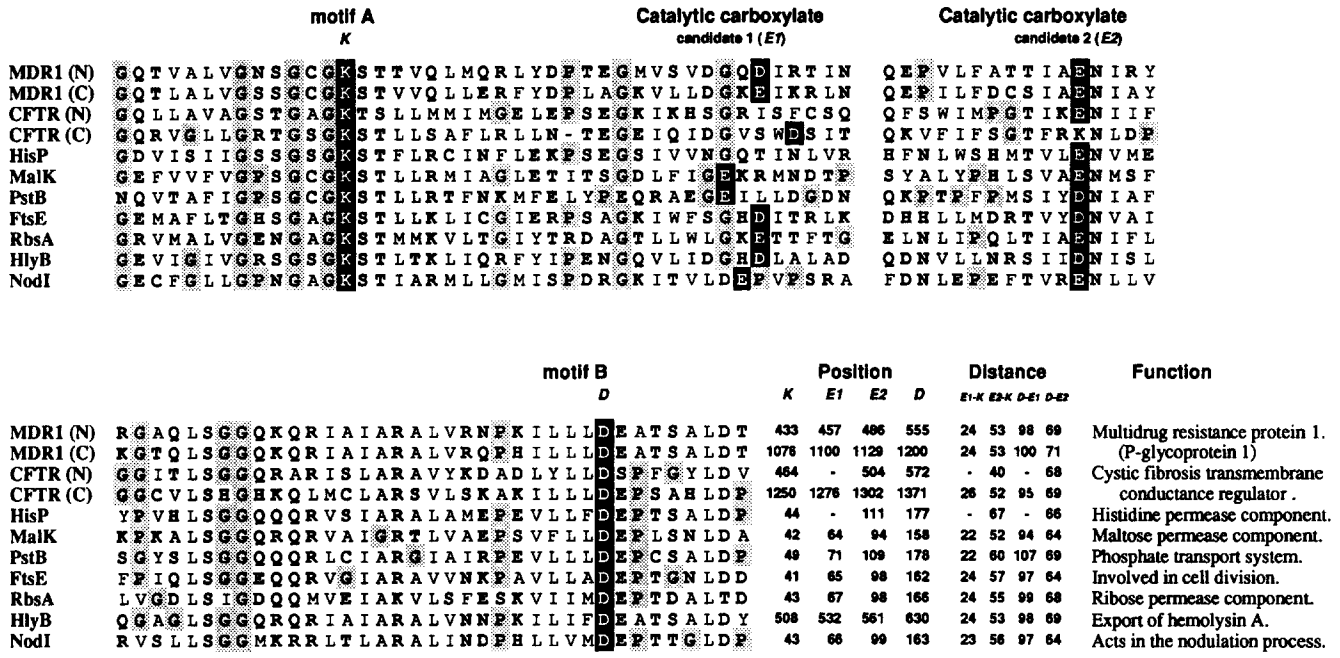


Fig. 2. Comparison of amino acid sequences of proteins of the ABC transporter family. In addition to the first candidate for catalytic carboxylate (*E1*), the region containing another candidate of catalytic carboxylate (*E2*) is also shown. In CFTR(N) and HisP sequences, candidate Glu (or Asp) is not found in the first candidate region. CFTR(C) has a replaced residue (Lys) at the second candidate position.

distance between motif A and B in the primary structure is 133–166 residues, much longer than other proteins listed in Fig. 1. Although we can find Glu (or Asp) in a region 24 ± 2 (Fig. 2, catalytic carboxylate candidate 1, *E1*) from the Lys of motif A in most cases, higher conservation was observed for a Glu (or Asp) in a region 59 ± 8 residues from the Lys (Fig. 2, catalytic carboxylate candidate 2, *E2*). There is a possibility that, in the case of the ABC transporter family, some additional structure comprising 32–41 residues is inserted in the loop region between the α_C helix and the β_2 strand (see next paragraph). The question of which one of the two conserved Glu's (or Asp's) acts as an essential residue will be easily answered by mutagenesis studies.

4. A common topology of ATP-triggered proteins

Conservation of three functional elements (motifs A, B, and a conserved Glu or Asp) among various proteins which catalyze ATP-triggered reactions might be an indication for the presence of the common structural motif in their ATP-binding domains. The following structure of the ATP-binding domain of the recA protein, which is made up of six repetitive β–α structures, might be a prototype fold*.

$$-\beta_1\alpha_C\beta_2\alpha_D\beta_3\alpha_E\beta_4\alpha_F\beta_5\alpha_G\beta_6-$$

The folding topology of the ATP binding domain of the MF_1-β subunit is the same as that of the recA protein except that there are two α helices between the β_4 and β_5 strands [6]. The essential feature of the structure is that the six parallel β strands and the α_C helix form the ATP binding domain (Fig. 3A). This folding motif is likely to be common to proteins listed in Figs. 1 and 2. Although there is no obvious overall sequence similarity among the protein families, distribution of amino acid residues in the sequences shown in Figs. 1 and 2 is consistent with the assumption of this protein fold. The first thirteen amino acid sequences of the proteins listed in Fig. 1 are consistent with a β strand–loop structure; several hydrophobic residues are followed by a cluster of Pro and Gly (shaded letters).

*Numbering of the α helices or β strands is corresponding to *Escherichia coli* recA protein.

←

Fig. 1. Comparison of amino acid sequences* of ATP binding domain in proteins which catalyze ATP-triggered reactions containing a set of typical nucleotide-binding motifs A and B. The sequence covering motif A and putative catalytic Glu (or Asp) and the sequence around motif B are shown. Lys in motif A, catalytic Glu (or Asp), and Asp in motif B are shown as *K*, *E*, and *D* (letters in black boxes), respectively. Pro and Gly, residues which tend to interrupt secondary structure, are shown by shaded letters. Hydrophilic residues are indicated by bold letters. Others are hydrophobic residues. In the case when a protein has duplicate sets of motifs A and B, a set at the amino-terminal side and at the carboxyl-terminal side are defined as (N) and (C), respectively. Secondary structures of the recA protein and MF_1-β subunits determined from X-ray crystallography are shown (column, α helix; arrow, β strand). Rho protein and Lon protease have another nearby conserved Glu which can also be candidates of catalytic carboxylate. The description of function for each protein is not complete; for example, SUG1, MSS1, and TBP1 probably play a regulatory role in the 26 S proteasome complex. *References from which sequence data were obtained are not cited here to avoid too many references.

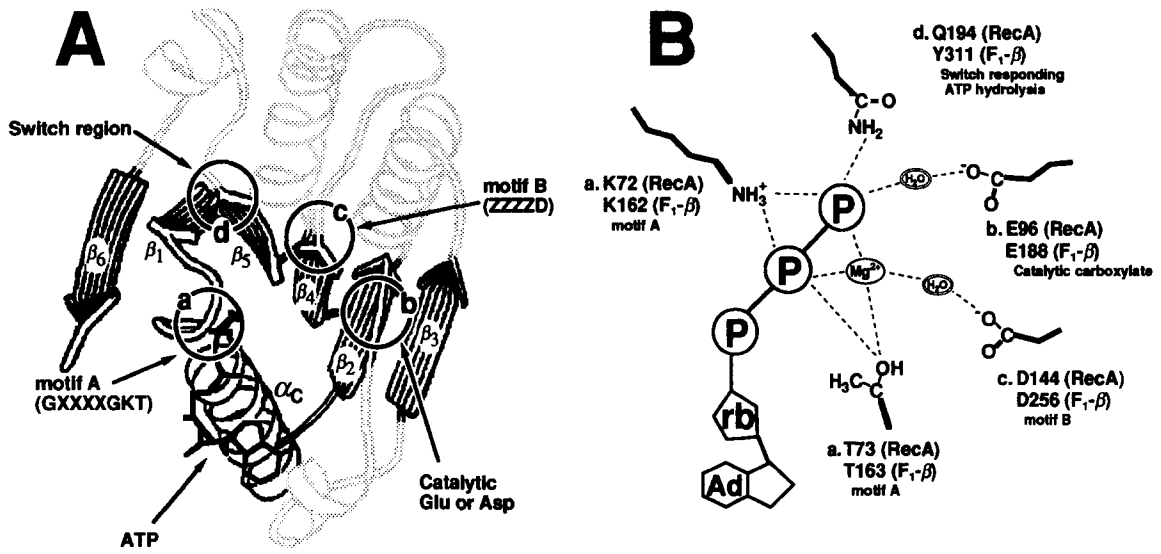


Fig. 3. (A) A common topology of the ATP binding domain of proteins which catalyze ATP-triggered reactions. Circle **a**, **b**, **c**, and **d** indicate the positions of motif A, catalytic carboxylate, motif B, and a 'switch' region, respectively. (B) Schematic illustration of residues interacting with a bound ATP molecule in this fold. Numbering of the residues in the recA protein and F_1-β is according to *E. coli* recA protein and bovine mitochondrial F_1-ATPase β subunit, respectively.

The α_C helix starts from the Thr in motif A and shows amphiphilic characteristics in recA protein and MF_1-β subunit [4,6]. It might also be the case in most proteins listed; hydrophilic residues (bold letters) appear periodically, 2nd, 5th (or 6th), and 9th (or 10th) positions counting from the Lys of motif A. As are observed for the recA protein and MF_1-β subunit, the hydrophilic side of the helix is facing to the outside. There is no Pro in putative α_C helix region except secA protein**. After the α_C helix, a loop region, which tends to contain Gly and Pro, and then the relatively hydrophobic β_2 strand follow. The catalytic carboxylate, Glu (or Asp), is located immediately after the β_2 strand [4,6].

Except for the region from the β_1 strand to the β_2 strand, the length (or even number) of α helices and loops connecting each β strand might be varied from one protein family to another, thus accounting for the variable distances in the primary structures between conserved Glu (or Asp) and motif B (Fig. 1, distance of *D–E*). The motif B sequence in the recA protein and MF_1-β subunit belongs to the β_4 strand which is located in the center of the ATP binding domain. The side chains of the residues in the β_4 strand are all hydrophobic and half of them are facing the hydrophobic side of the α_C helix. As demonstrated by high resolution X-ray crystallography for the Asp^{57} in the non-typical motif B in *ras* P21 protein [8], the Asp in motif B (Fig. 1, motif B, *D*) is likely to be involved in binding of active-site Mg^{2+} through a water molecule and to also be hydrogen-bonded to the Thr of motif A (Fig. 3B).

5. 'Switch region'

In the recA protein structure, Gln^{194} is located at the carboxyl end of the β_5 strand and is proposed to play a crucial role in transmission of conformational changes; Gln^{194}, when interacting with the γ-phosphate of ATP, stabilizes the crystallographically disordered loop which connects β_5 strand and α_G helix, and hydrolysis of ATP destroys this interaction resulting in a change in conformation of the loop [4]. Also in the structure of MF_1, conformation of the carboxyl end region of the β_5 strand in each of three copies of β subunits in a MF_1 molecule is substantially different from each other depending on the presence of bound nucleotide and on the interaction with the γ subunit [6]. Modification of Tyr^{311} in the β_5 strand by 7-chloro-4-nitrobenzofrazan in a single copy of β subunit in a MF_1 molecule is sufficient to inactivate catalytic turnover, probably by interfering with the inter-subunit conformational interaction [18,19]. Thus, it is reasonable to propose that the region at the carboxyl end of the β_5 strand of this fold is a switch region involved in propagating conformational change triggered by ATP hydrolysis. Successive conformational change occurs to the loop connecting the β_5 strand and the α_G helix which are interacting with another molecule, subunit, or functional domain of the protein. The β_6 strand provides the anchor for the α_G helix and two flanking loops, and stabilizes the β_1 strand.

6. Discussion

The occurrence of a similar folding topology among proteins with non-homologous amino acid sequences has recently been recognized for hexokinase [20], Hsc70 ATP binding domain [14], and actin [15]. Here, we point out that the proteins, the activities of which are coupled to or triggered by ATP hydrolysis, may have a common topology of the ATP-binding domain. The arrangement of catalytically important residues in the right positions in their tertiary structures may have been a consequence of convergent evolution. These residues include Lys and Thr in motif A, the catalytic Glu (or Asp), Asp in motif B, and the residue responsible for conformational transmission

**SecA protein is different from others in that it has a Pro residue in the middle of putative α_C helix and, therefore, the α_C helix of secA protein is likely to be interrupted and bent at this point.

(Fig. 3B). The validity of this postulation can be easily tested. For example, the mutagenesis of the putative catalytic Glu (or Asp) in these proteins would be easy to test.

References

[1] Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) EMBO J. 1, 945–951.
[2] Amano, T., Tozawa, K., Yoshida, M. and Murakami, H. (1994) FEBS Lett. 348, 93–98.
[3] Ohtsubo, M., Yoshida, M., Ohta, S., Kagawa, Y., Yohda, M. and Date, T. (1987) Biochem. Biophys. Res. Commun. 146, 705–710.
[4] Story, R.M. and Steiz, T.A. (1992) Nature 355, 374–376.
[5] Amano, T., Yoshida, M., Matsuo, Y. and Nishikawa, K. (1994) FEBS Lett. 351, 1–5.
[6] Abrahams, J.P., Leslie, A.G.W., Lutter, R. and Walker, J.E. (1994) Nature 370, 621–628.
[7] Boyer, P.D. (1993) Biochim. Biophys. Acta 1140, 215–250.
[8] Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W. and Wittinghofer, A. (1990) EMBO J. 9, 2351–2359.
[9] Berchtold, H., Reshetnikova, L., Reiser, C.O., Schirmer, N.K., Sprinzl, M. and Hilgenfield, R. (1993) Nature 365, 126–132.
[10] Noel, J.P., Hamm, H.E. and Sigler, P.B. (1993) Nature 366, 654–663.
[11] Coleman, D.E., Berghuis, A.M., Lee, E., Linder, M.E., Gilman A.G. and Sprang S.R., (1994) Science 265, 1405–1412.
[12] Fry, D.C., Kuby, S.A. and Mildvan, A.S. (1985) Biochemistry 24, 4680–4694.
[13] Mueller, C.W. and Schulz, G.E. (1992) J. Mol. Biol. 224, 159–177.
[14] Flaherty, K.M., DeLuca-Flaherty, C. and McKay, D.B. (1990) Nature 346, 623–628.
[15] Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F. and Holmes, K.C. (1990) Nature 347, 37–44.
[16] Rayment, I., Rypniewski, W.R., Schmidt-Bäse, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G. and Holden, H.M. (1993) Science 261, 50–58.
[17] Hyde, S.C. et al. (1990) Nature 364, 362–365.
[18] Andrews, W.W., Hill, F.C. and Allison, W.S. (1984) J. Biol. Chem. 259, 8219–8225.
[19] Ferguson, S.J., Lloyd, W.J. and Radda, G.K. (1975) Eur. J. Biochem. 54, 127–133.
[20] Anderson, C.M., Stenkamp, R.E. and Steitz, T.A. (1978) J. Mol. Biol. 123, 15–33.