An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins

(structural comparison/property pattern/remote homology)

PEER BORK, CHRIS SANDER, AND ALFONSO VALENCIA

European Molecular Biology Laboratory, D-6900 Heidelberg, Federal Republic of Germany

ABSTRACT   The functionally diverse actin, hexokinase, and hsp70 protein families have in common an ATPase domain of known three-dimensional structure. Optimal superposition of the three structures and alignment of many sequences in each of the three families has revealed a set of common conserved residues, distributed in five sequence motifs, which are involved in ATP binding and in a putative interdomain hinge. From the multiple sequence alignment in these motifs a pattern of amino acid properties required at each position is defined. The discriminatory power of the pattern is in part due to the use of several known three-dimensional structures and many sequences and in part to the "property" method of generalizing from observed amino acid frequencies to amino acid fitness at each sequence position. A sequence data base search with the pattern significantly matches sugar kinases, such as fuco-, glucono-, xylulo-, ribulo-, and glycerokinase, as well as the prokaryotic cell cycle proteins MreB, FtsA, and StbA. These are predicted to have subdomains with the same tertiary structure as the ATPase subdomains Ia and IIa of hexokinase, actin, and Hsc70, a very similar ATP binding pocket, and the capacity for interdomain hinge motion accompanying functional state changes. A common evolutionary origin for all of the proteins in this class is proposed.

In spite of their different biological functions, actin, Hsc70, and hexokinase contain similar three-dimensional structures (Fig. 1) (1–4). No overall sequence similarity between these three protein families can be detected with standard pairwise sequence alignment algorithms, so the structural similarity came as a surprise (1, 3, 4). All three bind and hydrolyze ATP. The ATPase activity of actin is involved in the control of polymerization (5); that of Hsc70 (a member of the hsp70 family of heat shock proteins) is involved in a variety of "chaperonin" functions, such as keeping protein chains translocation competent, preventing aggregation, or aiding in their refolding from aggregated states (6); and that of hexokinase is involved in phosphorylation of glucose at the entry to the glycolytic pathway (7).

The common structural feature is that of two domains of similar fold on either side of a large cleft with an ATP binding site at the bottom of the cleft. Each of these (I and II in Fig. 1) is composed of two subdomains. Subdomains Ia and IIa have the same basic fold (Fig. 1, lower left and right, shaded ribbons): a central β-sheet surrounded by helices, with identical topology of loop connections, probably a result of gene duplication (1, 3, 4). The phosphate tail of ATP is bound by residues on two β-hairpins, one from each of the subdomains Ia and IIa, and on other nearby segments (1–3) (not yet proven for hexokinase). This ATP binding motif is distinctly different from the single phosphate binding loop in, e.g.,
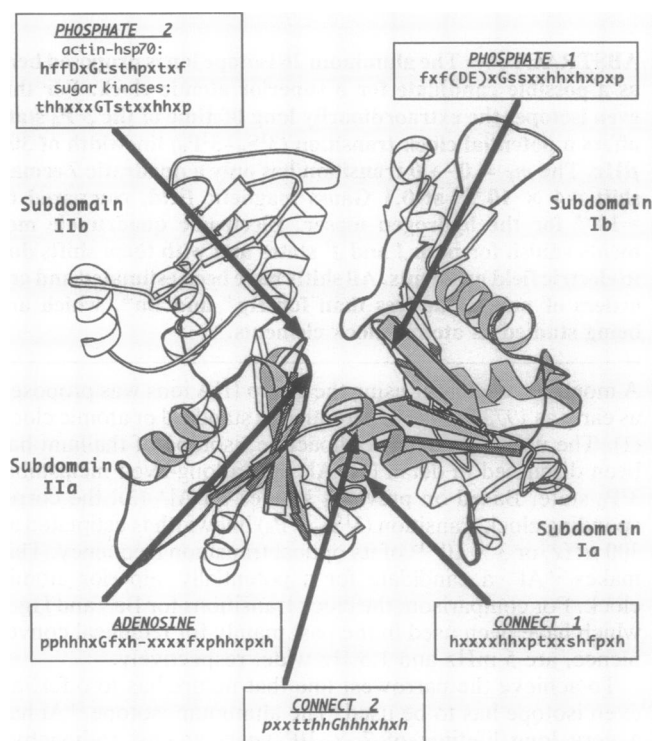
FIG. 1.   Diagramatic representation of actin. The five parts of the sequence pattern characteristic for the actin/hexokinase/hsc70 ATPase domain are mapped onto the three-dimensional structure (solid arrows). The sequence motif in each of the five regions is given here in terms of the following amino acid groups: h, purely hydrophobic (VLIFWY); f, partly hydrophobic (VLIFWYMCGATKHR); t, tiny (GSAT); s, small (GSATNDVCP); p, tiny plus polar (GSATNDQEKHR); x, any amino acid. Note that the groups overlap and that their names are merely mnemonic. The five-motif pattern used for the search cannot be fully represented in this simple form. The motifs are all located in the conserved substructure (dark shading) common to the three known crystal structures. We denote the subdomains by Ia, Ib, IIa, and IIb, as in the Hsc70 crystal structure (4).

adenylate kinase, recA protein, elongation factor Tu, or ras oncogene protein p21 (8).

These three structures are not only similar in three-dimensional fold but probably also in some aspects of mechanism: their ATPase active sites are lined with identical or similar residues and the overall domain structure is suggestive of the capacity for interdomain motion, perhaps directly coupled to the ATPase activity. A rotation of ≈30° of one of the domains is required to superimpose optimally the nucleotide-bound [actin or Hsc70 (1, 3)] onto the nucleotide-free [hexokinase (2)] structure, suggesting that each of the structures can be in an open and closed state. For hexokinase, there is direct crystallographic evidence for a large conformational change when a nucleotide binds (9, 10). Thus, the structural similarity of the three proteins is correlated with

partial similarity of function. A common evolutionary origin of this ATPase fold is therefore likely.

Here, we exploit the similarities of the three remotely, but clearly, related three-dimensional structures and the information contained in the multiple sequence alignment within each of the three families and define a common sequence pattern. The pattern appears to be characteristic of the tertiary structure and ATPase properties of this class of domains and can be used to scan data bases for structural related proteins.

MATERIALS AND METHODS

Alignment of Three-Dimensional Structures. The common structural core of the three proteins is based on optimal superposition of the actin crystal structure onto that of hsc70 and of a model of hexokinase. The superposition is done one subdomain at a time, using an algorithm that defines pairs of residues equivalent in three dimensions (11). The common structural core is defined here as those residue pairs that fall within a 2.5-Å cutoff on C_α–C_α distances between equivalent residues. For actin and Hsc70, the superposition is compatible with that of Flaherty *et al.* (4). The structural core—i.e., the equivalenced residues—covers as much as 45% (90 of 199) of the residues in subdomains Ia and IIa.

The hexokinase model was derived from the coordinate data set 2YHX (Brookhaven Protein Data Bank) after alignment of the isoform b sequence (Swiss-Prot identifier HXKB_YEAST), using an automatic side-chain building procedure (12).

Combining Multiple Sequence Alignment and Structure Alignment. To obtain the overall alignment of the sequences in all three families, we make use of the three-dimensional alignment of actin/Hsc70/hexokinase in the structural core. The many additional sequences of unknown three-dimensional structure are aligned separately in each family by a multiple sequence alignment procedure (13). Finally, the multiple sequence alignments for the three families are brought into register by the three-way structural alignment. Without the knowledge of equivalent residue positions in the three-dimensional structure, an overall alignment of the three families would have been exceedingly difficult, if not impossible.

Definition of Conserved Regions. The multiple sequence alignment reveals regions with conserved residues; several of these are near the ATP binding site, indicating a functional role. The boundaries of the conserved regions were delineated by trial and error, optimizing discriminatory power. Each of the five conserved regions chosen was ≈20 residues long and each was used to define a motif. Together, the five motifs define the overall 93-residue search pattern, covering 47% of all residues in subdomains Ia and IIa.

Pattern Definition. The problem of pattern definition is one of extrapolating from the known frequencies of residue types at each position, to the spectrum of structurally permissible residue types. This process of generalization is a key feature of our pattern approach and goes beyond the direct use of residue frequencies in standard profile search methods. The pattern derivation was carried out by the approach of Bork and Grunwald (14). The method is based on the definition of an 11 by 20 table of amino acid properties, indicating which of the 20 standard amino acids has which of 11 amino acid properties (hydrophobic, positive, negative, polar, charged, small, tiny, aliphatic, aromatic, proline, glycine).

The property pattern is defined for the five sequence motifs or "boxes" (Figs. 1 and 2). Each of the five motifs (PHOSPHATE1, CONNECT1, PHOSPHATE2, ADENOSINE, and CONNECT2) is derived from the multiple sequence alignment of the known members of the actin/hexokinase/hsp70 class as follows (14). (*i*) The multiple alignment of all available sequences is translated to a set of amino acid properties observed at each position. (*ii*) The set of properties allowed at each position is translated to a mismatch penalty for each potential amino acid at that position. (*iii*) A variable gap penalty is assigned to each position, depending on whether a gap was observed in any one protein. The full property pattern can be reduced to a simplified pattern shown in Fig. 1, in terms of the binary language of allowed/disallowed amino acids at specific positions. This simpler pattern can be used with the pattern search option in standard sequence analysis packages but it is less discriminating (data not shown) compared to the more sophisticated property pattern used here.

The mismatch score for a pattern is the sum over the mismatch scores for each of the positions in the constituent motifs. The selectivity of the procedure can be estimated by
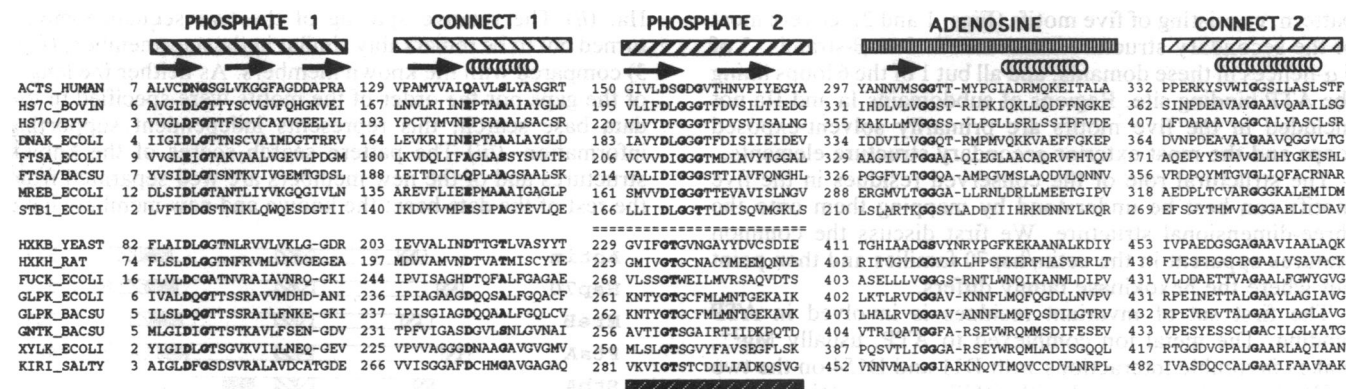
		PHOSPHATE 1	CONNECT 1		PHOSPHATE 2	ADENOSINE	CONNECT 2
ACTS_HUMAN	7	ALVCD**N**G**S**GLVKGGFAGDDAPRA	129 VPAMYVAIQAVL**S**LYASGRT	150	GIVLD**S**G**D**GVTHNVPIYEGYA	297 YANNVMSG**G**TT-MYPGIADRMQKEITALA	332 PPERKYSVWIGGSILASLSTF
HS7C_BOVIN	6	AVGID**L**G**T**TYSCVGVFQHGKVEI	167 LNVLRIIN**E**PTAAAIAYGLD	195	VLIFD**L**G**G**TFDVSILTIEDG	331 IHDLVLVG**G**ST-RIPKIQKLLQDFFNGKE	362 SINPDEAVAYGAAVQAAILSG
HS70/BYV	3	VVGLD**F**G**T**TFSCVCAYVGEELYL	193 YPCVYMVN**E**PSAAALSACSR	220	VLVYDFG**G**TFDVSVISALNG	355 KAKLLMVG**G**SS-YLPGLLSRLSSIPFVDE	407 LFDARAAVAGGCALYASCLSR
DNAK_ECOLI	4	IIGID**L**G**T**TNSCVAIMDGTTRRV	164 LEVKRIIN**E**PTAAALAYGLD	183	IAVYD**L**G**G**GTFDISIIEIDEV	334 IDDVILVG**G**QT-RMPMVQKKVAEFFGKEP	364 DVNPDEAVAIGAAVQGGVLTG
FTSA_ECOLI	9	VVGL**S**I**G**TAKVAALVGEVLPDGM	180 LKVDQLIFAGLA**S**SYSVLTE	206	VCVVDIG**G**GTMDIAVYTGGAL	329 AAGIVLTG**G**AA-QIEGLAACAQRVFHTQV	371 AQEPYYSTAVG**L**IHYGKESHL
FTSA/BACSU	7	YVSID**L**G**T**SNTKVIVGEMTGDSL	188 IEITDICLQPLAAGSAALSK	214	VALIDIG**G**STTIAVFQVNGHL	326 PGGFVLTG**G**QA-AMPGVMSLAQDVLQNNV	356 VRDPQYMTGVG**L**IQFACRNAR
MREB_ECOLI	12	DLSID**L**G**T**ANTLIYVKGQGIVLN	135 AREVFLIE**E**PMAAAIGAGLP	161	SMVVDIG**G**GTTEVAVISLNGV	288 ERGMVLTG**G**GA-LLRNLDRLLMEETGIPV	318 AEDPLTCVARGGGKALEMIDM
STB1_ECOLI	2	LVFIDD**G**STNIKLQWQESDGTII	140 IKDVKVMP**E**SIPAGYEVLQE	166	LLIIDLG**G**TTLDISQVMGKLS	210 LSLARTKG**G**SYLADDIIIHRKDNNYLKQR	269 EFSGYTHMVIGGGAELICDAV
		====================			====================		
HXKB_YEAST	82	FLAIDL**G**GTNLRVVLVKLG-GDR	203 IEVVALIND**T**TGTLVASYYT	229	GVIF**G**TGVNGAYYDVCSDIE	411 TGHIAADG**S**VYNRYPGFKEKAANALKDIY	453 IVPAEDGSGAGAAVIAALAQK
HXKH_RAT	74	FLSLDL**G**GTNFRVMLVKVGEGEA	197 MDVVAMVND**T**VATMISCYYE	223	GMIV**G**TGCNACYMEEMQNVE	403 RITVGVDG**G**VYKLHPSFKERFHASVRRLT	438 FIESEEGSGRGAALVSAVACK
FUCK_ECOLI	16	ILVLDC**G**ATNVRAIAVNRQ-GKI	244 IPVISAGHDTQFALFGAGAE	268	VLSSG**T**WEILMVRSAQVDTS	403 ASELLLVG**G**GS-RNTLWNQIKANMLDIPV	431 VLDDAETTVAGAALFGWYGVG
GLPK_ECOLI	6	IVALD**Q**GTTSSRAVVMDHD-ANI	236 IPIAGAAGDQQSALFGQACF	261	KNTYG**T**GCFMLMNTGEKAIK	402 LKTLRVDG**G**AV-KNNFLMQFQGDLLNVPV	431 RPEINETTALGAAYLAGIAVG
GLPK_BACSU	5	ILSLD**Q**GTTSSRAILFNKE-GKI	237 IPISGIAGDQQAALFGQLCV	262	KNTYG**T**GCFMLMNTGEKAVK	403 LHALRVDG**G**AV-KNNFLMQFQSDILGTRV	432 RPEVREVTALGAAYLAGLAVG
GNTK_BACSU	5	MLGIDI**G**TTSTKAVLFSEN-GDV	231 TPFVIGASDGVL**S**NLGVNAI	256	AVTIG**T**SGAIRTIIDKPQTD	404 VTRIQAT**G**GFA-RSEVWRQMMSDIFESEV	433 VPESYESSCLGACILGLYATG
XYLK_ECOLI	2	YIGIDL**G**TSGVKVILLNEQ-GEV	225 VPVVAGGGDNAA**G**AVGVGMV	250	MLSLG**T**SGVYFAVSEGFLSK	387 PQSVTLIGG**G**A-RSEYWRQMLADISGQQL	417 RTGGDVGPALGAARLAQIAAN
KIRI_SALTY	3	AIGLD**F**G**S**DSVRALAVDCATGDE	266 VVISGGAFDCHM**G**AVGAGAQ	281	VKVIG**T**STCDILIADKQSVG	452 VNNVMALG**G**IARKNQVIMQVCCDVLNRPL	482 IVASDQCCALGAAIFAAVAAK

FIG. 2. Multiple alignment of representative sequences for each motif. Residue numbers and secondary structure of actin (1) are given, where → is β-strand and ∞ is α-helix. Boldface indicates positions that correspond to the most conserved residues in the simplified pattern shown in Fig. 1. The structural similarity between the PHOSPHATE1 and PHOSPHATE2 regions, important for evolutionary arguments, is reflected at the sequence level (see conserved residues in the two motifs): a search with the PHOSPHATE1 motif detects the PHOSPHATE2 region of actin and hsp70-like proteins (including MreB, FtsA, and StbA). The PHOSPHATE2 region of hexokinase-like proteins is more diverged. Protein sequences are labeled by their Swiss-Prot data base identifiers [release 17, February 1991 (15)]. ACTS_HUMAN, actin; HS7C_BOVIN, hsp70 [HS70/BYV of beet yellow closterovirus heat shock-related protein (16)]; DNAK_ECOLI, prokaryotic heat shock-related protein; FTSA_ECOLI (FTSA/BACSU), FtsA (filamenting temperature sensitive) protein of *Escherichia coli* (*Bacillus subtilis*); MREB_ECOLI, MreB protein; STB1_ECOLI, plasmid-encoded StbA (plasmid stability) protein. HXKB_YEAST, hexokinase; HXKH_RAT, glucokinase; FUCK_ECOLI, fucokinase; GLPK_BACSU (GLPK_ECOLI), glycerokinase of *B. subtilis* (*E. coli*); GNTK_BACSU, gluconokinase of *B. subtilis*; XYLK_ECOLI, xylulokinase; KIRI_SALTY, ribulokinase of *Salmonella typhimurium*. Residue numbers were taken from the Swiss-Prot data bank entries except for actin, in which the numbering of Kabsch *et al.* (1) is used, with residue 1 corresponding to residue 3 of ACTS_HUMAN.

comparing the mismatch score of the worst-scoring known positive (8 mismatches), with the best-scoring rejected sequence [PEPC_MACFU, Swiss-Prot code (15), a sequence that is clearly homologous to the known structure of pepsinogen and scores with 18 mismatches]. The separation of correct and unexpected positives (up to 9 mismatches) from all other data base proteins (≥18 mismatches) is, in our experience, a strong indication of significance, stronger than some statistical tests based on random sequence shuffling.

RESULTS AND DISCUSSION

Comparison of Three-Dimensional Structures and Localization of Sequence Patterns. The general domain organization of these three structures is very similar, as described in detail by Flaherty *et al.* (4) for actin and Hsc70. The two domains containing the ATP binding site (domains I and II in Fig. 1) are divided into two subdomains (Ia and IIa; Ib and IIb). The two larger subdomains Ia and IIa have the same fold—i.e., the same topology of secondary structure elements and loop connections—and are related by a spatial symmetry transformation. The small subdomains Ib and IIb are different from one subfamily to another, most strongly when comparing actin/Hsc70 on the one hand and hexokinase on the other. In the case of actin, subdomains Ib and IIb are implicated in the monomer–monomer contact in filament formation. It may therefore be reasonable to propose that subdomains Ib and IIb correspond to specialized functions in each of the subfamilies.

Comparison of the structures in three dimensions reveals a clear cluster of equivalent residues in the vicinity of the ATP binding site (see Fig. 1). This common core includes three ATP binding loops [the two loops contacting the β- and γ-phosphate of ATP and the adenosine base binding site (1, 3, 4)] and a large part of the interface between subdomains Ia and IIa. Other than these regions, there is only one additional loop in the vicinity of the ATP binding site that is not part of the pattern. This loop follows β-4 in subdomain Ia. In this position, hexokinase has a long insertion, most probably related to sugar binding (17), which has no counterpart in actin/Hsc70.

From the many known homologous sequences and from the structural superpositions of the crystal structures, a pattern of conserved and invariant residues was derived. The pattern, consisting of five motifs (Figs. 1 and 2), covers most of the secondary structure elements: 7 of 10 β-strands, 3 of 5 α-helices in these domains, and all but 1 of the 6 loops lining the ATP binding site. Regions of subdomains Ia and IIa not included in the five motifs are primarily solvent-exposed loops and the most exterior secondary structure elements.

The structural role of the conserved residues in the five motifs can best be understood by mapping them onto the three-dimensional structure. We first discuss the common features apparent in the actin/hsp70 families and then point out where the hexokinase family differs.

The first set of invariant residues is involved in ATP binding. The metal ion complexed to ATP, usually Mg^{2+}, makes favorable interactions with D11 and D154 on the two ATP binding hairpins and with Q137 on connecting region CONNECT1 (Fig. 1). (Amino acids are in one-letter code, immediately followed by the residue number, and we use actin residue numbers unless otherwise stated.) Nearby, G13 and G156 are involved in the tight β-turn, such that addition of side chains would lead to steric clashes with the β- or γ-phosphate. The adenine part of ATP is in direct contact with the peptide unit following G302. The backbone NH of the preceding G301 forms a hydrogen bond to the α-phosphate.

The conserved segments connecting the two lobes of the structure are region CONNECT1, crossing from domain I to domain II, and region CONNECT2, crossing in the opposite direction (Fig. 1). The two regions make a close helix–helix contact, involving the invariant G342 (on CONNECT2) and Q137 as well as A144 (on CONNECT1). The helix–helix contact appears to have the properties of an interdomain hinge: it is present in both the closed (actin/hsp70) and open (hexokinase) crystal structures, in spite of considerable shift of nearby regions. It is plausible that the sequence requirements for such an invariant point of contact are restrictive.

The invariant residues, although scattered throughout the sequence (Figs. 1 and 3), all map to the same spatial region, within ≈8 Å (C_α positions) of the bound ATP (1), and are linked by a network of interactions. Perhaps the most remarkable of these is a set of mutual contacts linking Q137 on connecting region CONNECT1, D11 on the first ATP binding hairpin, T106 (not used in the pattern) on the adjacent β-strand, and V339 on CONNECT2.

The pattern of five motifs is essentially identical in the hexokinase crystal structure. However, in the second ATP binding hairpin, in place of the conserved 154-DxG-156 of actin/hsp70, hexokinase has 232-GT-233 (PHOSPHATE2; Figs. 1 and 2). The precise role of these residues in ATPase activity is not yet clear, as the currently available hexokinase structure does not contain a nucleotide and the hairpin appears to be one residue short compared to actin/hsc70. This difference in structural detail may be related to the higher efficiency of the hexokinase ATPase compared with actin/hsp70. Taking these differences into account, we have defined a sequence pattern separately for the actin/hsp70 family on the one hand and hexokinase on the other. The two patterns differ only in PHOSPHATE2.

Data Base Search. Many examples of proteins with essentially the same fold and very low sequence similarity are known. Ordinary pairwise sequence alignment cannot be used routinely to reveal very distant relationships because pairwise sequence alignment (without additional information) starts becoming unreliable below a level of ≈25% sequence identity [for alignments covering 80 or more residues (13)].

In the present case, there are several reasons for confidence in the pattern description of the actin/hexokinase/hsp70 ATPase fold. (*i*) Sequence similarity between any two of the three known structures (actin, hsc70, and hexokinase) is very low; yet, they have the same fold in subdomains Ia and IIa. (*ii*) The relative spacing of the five sequence motifs turned out to be remarkably similar in the new members (Fig. 3) compared with the known members. As neither the length of the gaps nor the order of the motifs were specified in the data base search, this represents independent supporting information. (*iii*) The pattern search scores of the known structures and of the new members are well separated from the rest of the data base: the known and new members have

Actin			377
Hsp70			640
MreB			347
FtsA			420
StbA			320
Hexok.			486
Xylulok.			484
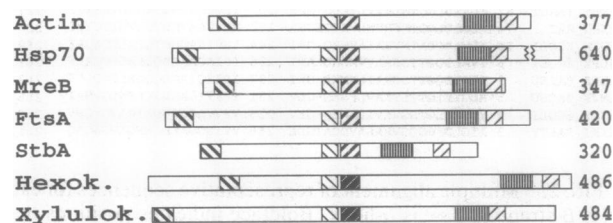
FIG. 3. Location of the five sequence motifs in each protein family: actin (muscle and cytoskeletal proteins), hsp70 (heat shock proteins), MreB, FtsA (prokaryotic cell cycle proteins), StbA (plasmid stability protein), hexokinase, xylulokinase (sugar kinases). Note that the relative position of the motifs is identical in all families and that the distances between the motifs are similar. The exception, StbA, may reflect the deletion of subdomain IIb. The motifs are hatched as in Fig. 2. Numbers on the right indicate total length of the proteins.

mismatch scores between 0 and 9, while the best rejected data base protein does not appear until a mismatch score of 18. (*iv*) The five-motif pattern covers not only active site or binding site residues but also "structural" residues—i.e., those with a probable crucial role in the structural core (e.g., hydrophobic residues in β-strands in the protein interior, small residues at the helix–helix contact point of the hinge region). For these reasons, there is little doubt that the sequence pattern is capable of reliably detecting new members in this ATPase domain class. It is therefore plausible to assume that the new members do not differ in aspects common to the known members: an ATPase domain with the same fold in subdomains Ia and IIa, a very similar ATP binding pocket down to the level of conserved hydrogen bonds with particular residues, a similar ATPase mechanism, and, finally, the capacity to react by hinged interdomain motion to the altered state of the substrate as a result of the catalytic reaction. These statements apply to the newly identified sugar kinases as well as to Mreb, Ftsa, and Stba, discussed in detail below.

Interestingly, recently developed methods of searching a sequence data base with profiles derived from known three-dimensional structures (see refs. 18 and 19) also were capable of detecting the relationship between actin and hsp70 (but not hexokinase) at the sequence level, where ordinary profile searches fail. The discriminatory power of the patterns derived here rests in part on the use of the superposition of three known crystallographic structures and in part on generalizing from observed amino acid frequencies in the corresponding multiple sequence alignments to amino acid properties required at particular positions (14).

Sugar Kinases. In addition to the known members of the actin/hsp70/hexokinase class, the pattern search unambiguously identifies several putative newly discovered members of this structural class, with mismatch scores well separated from the rest of the data base of >23,000 protein sequences. The first group of new members—fucokinase, glycerokinase, gluconokinase, xylulokinase, and ribulokinase—all are sugar kinases. The mutual relationship of fuco-, glycero-, glucono-, and xylulokinase is detectable by standard sequence comparison methods [e.g., program FASTA (20)], but their relationship to ribulokinase or to hexokinases is in the region of low sequence similarity, where, so far, only particular selective pattern searches are successful. In terms of the sequence pattern defined here (Fig. 2), these proteins are clearly more related to the eukaryotic hexokinases than to actin and hsp70. In particular, all these kinases have the invariant GT in PHOSPHATE2. In addition, they have other conserved regions that are not present in actin or hsp70 (data not shown). Together with hexokinases, they constitute a major structural subclass of prokaryotic and eukaryotic sugar kinases involved in the glycolytic pathway.

FtsA. The other proteins identified by the pattern search with a low mismatch score (Fig. 2) are reported to be prokaryotic cell cycle proteins. Protein FtsA is important for the later stages of cell division (21) and probably is present in the septum complex (22). *E. coli* FtsA was reported to be homologous to eukaryotic cell cycle kinases (23), but comparison with the more recently determined *B. subtilis* version of FtsA (24) (30% identical residues in a limited region compared to the *E. coli* sequence) makes this highly unlikely: regions aligned as conserved between FtsA and cdc kinases in that report (23) do not match the regions conserved between the two FtsA variants (data not shown). In contrast, our results lead to the hypothesis that FtsA has an ATPase domain similar to that of actin, hsc70, and hexokinase.

Additional evidence can be obtained from some of the known temperature-sensitive mutants of FtsA (25). One of these, FtsaA "allele 6," has D-217 mutated to N217. The equivalent amino acid in Hsc70 is D206, probably the most important residue in ATP hydrolysis (4).

MreB. A related candidate (Fig. 2), MreB protein, is reported to be involved in the positive control of cell elongation and/or negative control of cell division (26). The functions of MreB and FtsA appear to be related (27) and significant sequence similarity between them has been reported (26) in what we now know to be the phosphate (VVDIGGGT) and adenosine (VLTGG) binding motifs.

StbA. Little is known about the function of the third bacterial protein with a significant pattern match. The *E. coli* protein StbA appears to be involved in plasmid stability—i.e., deletion mutants suffer from unstable inheritance of certain low copy number plasmids (28). In the family described here, StbA is unusual in that it appears to lack subdomain IIb (Figs. 2 and 3). Our prediction is that expression of StbA changes along the cell cycle in analogy with FtsA (21).

Based on these findings, one may hypothesize that the ATPase fold found here for MreB and FtsA is a kinase domain that regulates the activity of enzymes specific for different prokaryotic cell cycle stages. Alternative hypotheses are suggested by the fact that MreB and FtsA are to be grouped with actin and hsp70, rather than with the sugar kinases, because of similarity in the sequence motif in the PHOSPHATE2 region (Fig. 2). In analogy to Hsc70, these proteins may have a chaperone function. In analogy to actin, MreB, FtsA, and/or StbA may have a direct structural role in determining cell shape. Consistent with such a role is the observation that FtsA participates in the structure of the septum (22).

Divergent or Convergent Evolution? A divergent evolutionary pathway is sketched in Fig. 4. Conceivably, an ancestral protein of ≈150 residues acquired the capacity to dimerize and bind ATP in an active site between the two subunits. Evidence for this comes from the overall structural symmetry
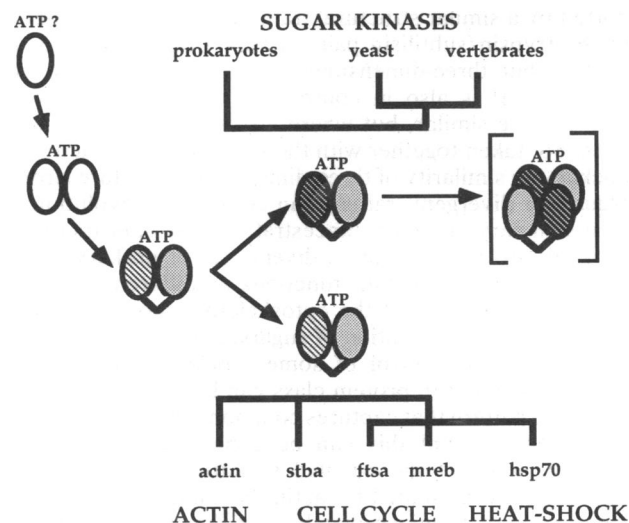
FIG. 4. Evolutionary tree of the ATPase domain present in widely divergent protein families, as detected by the sequence pattern. The patterns for the two main branches, actin/hsp70 and sugar kinases, differ primarily in the details of the PHOSPHATE2 motif (Figs. 1 and 2). The trees for each branch are qualitative and merely indicate grouping into subfamilies. The grouping is made by inspection of the multiple sequence alignment in the regions of the five motifs (Fig. 2). For example, in these regions FtsA and MreB are clearly more similar to hsp70 proteins than they are to actin. A possible evolutionary scenario includes an ancestral ATP binding homodimer and proceeds through gene duplication and subsequent divergence of the domains I (right oval) and II (left oval). In the sugar kinase branch, the difference between domains I and II is more pronounced than it is in the actin/hsp70 branch, indicated by heavier shading (left). A further gene duplication event for the high molecular weight mammalian hexokinases (29, 30) is indicated in brackets.

between the two domains of the ATPase as well as the symmetric arrangement of the two phosphate binding loops, which is also reflected at the sequence level (Fig. 2). Later, gene duplication and fusion led to the common ATPase fold of ≈300 or more residues characterized here. One major branch evolved into the ancestors of actin, the eukaryotic cell cycle proteins, and hsp70, in part by insertion of subdomains (Ib, top right; IIb, top left in Fig. 1). The other major branch evolved into sugar kinases. Since members of the Hsc70 family are present in eukaryotes and prokaryotes, the divergence between sugar kinases, actin, and Hsc70 has to be anterior to prokaryote–eukaryote branching, as is also mentioned in ref. 31.

The internal structural duplication in each of the three known structures (striking similarity of subdomains Ia and IIa after optimal superposition of backbone traces) is very likely the result of gene duplication. Comparing the subdomains of different members of the structural family, say, of actin and hexokinase, the three-dimensional similarity is as striking as that apparent in the internal duplication. So the most plausible explanation for both similarities is one of divergent evolution—otherwise, one would be trying to explain two comparable three-dimensional relationships by different mechanisms. As evolutionary pressure is significantly weaker away from sites of interaction with substrates or other entities, it appears very unlikely that the rather particular and complicated fold of this domain would arise twice as the result of selective evolutionary pressure. This particular argument is qualitative, as we have not performed a statistical analysis contrasting sequence similarity resulting from evolutionary divergence with that resulting from similar structural constraints, as was done for rhodanese (32).

However, there is a second reason in support of evolutionary homology. The ATPase active site pockets are very similar in actin, hsc70, and hexokinase (see above), supported by a similar structural framework. This is in contrast to the trypsin/subtilisin pair, in which the active sites are similar, but three-dimensional structures have no overall similarity. It is also in contrast to rhodanese, where the domains are similar, but where we do not have two active sites. So, taken together with the similarity of the active site pocket, the similarity of three-dimensional structure strongly argues for divergent, rather than convergent, evolution.

In summary, a dimeric ancestral ATPase appears to have evolved into what is now a diverse set of prokaryotic and eukaryotic enzymes with functions as different as muscle action, construction of the cytoskeleton, protein refolding, metabolic phosphorylation of sugars, and, possibly (indirect evidence only), control of some aspects of bacterial cell division. The entire protein class can be characterized by a sequence pattern that captures common functional and structural elements and this can be exploited to transfer, by analogy, some of the detailed structural and biochemical knowledge accumulated for actin, hexokinase, and hsp70 to the recently determined members of the class. This may be particularly useful in planning experiments to elucidate further the normal and defective function of the prokaryotic cell cycle proteins, such as FtsA, or of sugar kinases, such as glycerokinase.

1. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. (1990) *Nature (London)* **347**, 37–44.
2. Anderson, C. M., Stenkamp, R. E. & Steitz, T. A. (1978) *J. Mol. Biol.* **123**, 15–33.
3. Flaherty, K. M., Deluca-Flaherty, C. & McKay, D. B. (1990) *Nature (London)* **346**, 623–628.
4. Flaherty, K. M., McKay, D. B., Kabsch, W. & Holmes, K. C. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5041–5045.
5. Pollard, T. D. (1990) *Curr. Opin. Cell Biol.* **2**, 33–37.
6. Gething, M. J. & Sambrook, J. (1992) *Nature (London)* **355**, 33–42.
7. Middleton, R. J. (1990) *Biochem. Soc. Trans.* **18**, 180–183.
8. Schulz, G. E. (1992) *Curr. Opin. Struct. Biol.* **2**, 61–67.
9. Bennett, W. S. & Steitz, T. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4848–4852.
10. Steitz, T. A., Shoham, M. & Bennett, W. S. (1981) *Philos. Trans. R. Soc. London Ser. B* **293**, 43–52.
11. Vriend, G. & Sander, C. (1991) *Proteins* **11**, 52–58.
12. Holm, L. & Sander, C. (1991) *J. Mol. Biol.* **218**, 183–194.
13. Sander, C. & Schneider, R. (1991) *Proteins* **9**, 56–68.
14. Bork, P. & Grunwald, C. (1990) *Eur. J. Biochem.* **191**, 347–358.
15. Bairoch, A. & Boeckmann, B. (1991) *Nucleic Acids Res.* **19**, 2247–2250.
16. Agranovsky, A. A., Boyko, V. P., Karasev, A. V., Koonin, E. V. & Dolja, V. V. (1991) *J. Mol. Biol.* **217**, 603–610.
17. Arora, K. K., Filburn, C. R. & Pedersen, P. C. (1991) *J. Biol. Chem.* **266**, 5359–5362.
18. Bowie, J. U., Luethy, R. & Eisenberg, D. (1991) *Science* **253**, 164–170.
19. Scharf, M. (1989) Thesis (Univ. of Heidelberg, F.R.G.).
20. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3338–3342.
21. Donachie, W. D., Begg, K. J., Lutkenhaus, J. F., Salmond, G. P. C., Martinez-Salas, E. & Vicente, M. (1979) *J. Bacteriol.* **140**, 388–394.
22. Pla, J., Dopazo, A. & Vicente, M. (1990) *J. Bacteriol.* **172**, 5097–5102.
23. Robinson, A. C., Collins, J. F. & Donachie, W. D. (1987) *Nature (London)* **328**, 766.
24. Beall, B. & Lutkenhaus, J. (1989) *J. Bacteriol.* **171**, 6821–6834.
25. Robinson, A. C., Begg, K. J., Sweeney, J., Condie, A. & Donachie, W. D. (1988) *Mol. Microbiol.* **2**, 581–588.
26. Doi, M., Wachi, M., Ishino, F., Tomioka, S., Ito, M., Sakagami, Y., Suzuki, A. & Matsuhashi, M. (1988) *J. Bacteriol.* **170**, 4619–4624.
27. Wachi, M. & Matsuhashi, M. (1989) *J. Bacteriol.* **171**, 3123–3127.
28. Tabuchi, A., Min, Y. N., Kim, C. K., Fan, Y. L., Womble, D. D. & Rownd, R. H. (1988) *J. Mol. Biol.* **202**, 511–525.
29. Easterby, J. S. (1971) *FEBS Lett.* **18**, 23–26.
30. Schwab, D. & Wilson, J. E. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2563–2567.
31. Doolittle, R. F. (1992) *Protein Sci.* **1**, 191–200.
32. Keim, P., Heinrikson, R. L. & Fitch, W. M. (1981) *J. Mol. Biol.* **151**, 179–197.