simultaneously. And even seven interconnected reactions are capable of surprisingly sophisticated behaviour, such as spike generation, oscillations and chaotic fluctuations. So, in this respect, there are clear limits to reductionism for biochemists. Victims of our own success, we have probed so deeply into our surroundings that we are no longer able to see the wood for the trees.

What can we do about it? There are, I suggest, two possible paths to salvation. The first is to loosen up. Abandon slavish reliance on molecules and look for explanations at a higher level. This is of course a normal process, because our accounts of the natural world are frequently multi-level, hierarchical, 'onion skins' of explanation. Who would describe a cilium or a chromosome by listing all the molecules they contain together with the properties of these molecules and how they interact? Cilia and chromosomes are clearly delineated structures with defined location and function in the cell. We represent them mentally as distinct entities that operate as units within the cell. Surely, then, this is the way to tackle other parts of the cell, even those that presently seem impenetrable. We might, for example, try to resolve the plethora of interconnected reactions involved in cell signalling into discrete computational modules. Perhaps the signalling complexes that are being found increasingly in many systems will provide us with the elements we need to construct a satisfying reductionist explanation of intracellular communication?

The second road to salvation is through the keyboard of a computer. Although we poor mortals have difficulty manipulating seven things in our head at the same time, our silicon protégés do not suffer this limitation. Computers have a gargantuan appetite for numbers, digesting them at lightning speed and regurgitating them in any desired form – even one that we can understand. Suppose you are interested in the structure of a protein molecule and have been given the precise spatial coordinates of all of the atoms it contains. How much would you learn by studying this long list of numbers? I suggest very little. But now feed these numbers into a molecular graphics package running on a workstation and... voilà! Now you have a three-dimensional image of your protein that you can invert, rotate and survey for different residues. You can manipulate this graphical image in ways impossible with the real molecule it stands for, and in so doing achieve an understanding.

The dramatic success of this approach encourages me to ask: why not do the same for living cells? If we can use computer-based graphical elements to understand the world of protein structure, why should we not do the same for the universe of cells? The data are accumulating and the computers are humming. What we lack are the words, the grammar and the syntax of the new language.

Reference

1 *The Limits of Reductionism in Biology*, CIBA Symposium (Vol. 213) (in press)

DENNIS BRAY

Department of Zoology, Downing Street, Cambridge, UK CB2 3EJ.
Email d.bray@zoo.cam.ac.uk

Archaeal introns: splicing, intercellular mobility and evolution

Jens Lykke-Andersen, Claus Aagaard, Mikhail Semionenkov and Roger A. Garrett

Until recently, it appeared that archaeal introns were spliced by a process specific to the archaeal domain in which an endoribonuclease cuts a 'bulge-helix-bulge' motif that forms at exon–intron junctions. Recent results, however, have shown that the endoribonuclease involved in archaeal intron splicing is a homologue of two subunits of the enzyme complex that excises eukaryotic nuclear tRNA introns. Moreover, some archaeal introns encode homing enzymes that are also encoded by group I introns.

ARCHAEA do not appear to carry either group I introns, group II introns or nuclear mRNA-type introns that are found in eukaryotes and/or bacteria[1]. Instead, they carry introns in their tRNA and rRNA genes that are spliced by an apparently archaeal-specific mechanism. However, despite the distinctiveness of the archaeal introns, there are strong indications that they share some structural and functional characteristics with other intron types. Here, the structural characteristics of archaeal introns, their splicing mechanism and the mode of action of their encoded homing endonucleases are summarized together with their possible evolutionary relationships to other bacterial and eukaryotic introns.

J. Lykke-Andersen, C. Aagaard, M. Semionenkov and **R. A. Garrett** are at the RNA Regulation Centre, Institute of Molecular Biology, University of Copenhagen, Sølvgade 83H, DK-1307 Copenhagen K, Denmark.
Email: garrett@mermaid.molbio.ku.dk

Structure and splicing mechanism of archaeal introns

Archaeal introns have been detected in some tRNA genes of euryarchaeotes and in tRNA and rRNA genes of crenarchaeotes. In pre-tRNAs, the introns are generally located one nucleotide 3′ to the anticodon, at the same location as all eukaryotic nuclear tRNA introns. There are exceptions, however, where archaeal introns are positioned at other tRNA sites, including one within the variable arm[2,3] (Fig. 1a). Archaeal introns are also found at diverse sites within the large pre-rRNAs, all of which appear to correspond to important functional centres on the ribosomal surface[4]. Some of these rRNA introns contain open reading frames (ORFs) of about 600 nucleotides, which are low in G+C content relative to the co-transcribed rRNA and intron core[5–7].

All archaeal intron transcripts generate a 'bulge-helix-bulge' motif at the exon–intron junction[8–12] (Fig. 1b). The higher order structure of this splicing motif remains unknown, although the three-nucleotide 'bulges' exhibit limited accessibility to chemical and enzymatic probes in the free pre-RNA[12], and gel-mobility

 0968–0004/97/$17.00 PII: S0968-0004(97)01113-4

studies indicate that the 'bulges' produce bending of the motif (J. Z. Dalgaard and R. A. Garrett, unpublished). The 'bulge-helix-bulge' motif is recognized by a splicing endoribonuclease that cuts at symmetrical positions within the three-nucleotide bulges, producing 2′,3′-cyclic phosphates and 5′-OH ends[9,10,13] (Fig. 1c). Subsequently, the ends of the exons are ligated and the introns circularize (Fig. 1c), although the latter process has only been demonstrated for three of the larger introns that carry ORFs[6,13]. So far, no archaeal RNA ligase has been isolated.

The splicing endoribonuclease has been detected exclusively in archaeal cell extracts[9,13]. Moreover, it has a broad substrate specificity and it seems likely that the endoribonuclease of any archaeon can cleave any archaeal exon–intron junction[13]. It is also likely that this enzyme has a more general processing function, because the 'bulge-helix-bulge' motif is found in other archaeal cellular RNA structures, including the long processing stems of the large rRNAs[4]. Given the wide variation of the primary sequence of the substrates, it is probable that substrate recognition occurs mainly at a tertiary structural level. Insight into this process is minimal, however, because although the enzyme has been purified to different levels from cell extracts[9,14], only recently has a gene from the haloarchaeon *Haloferax volcanii* been cloned and sequenced[15].

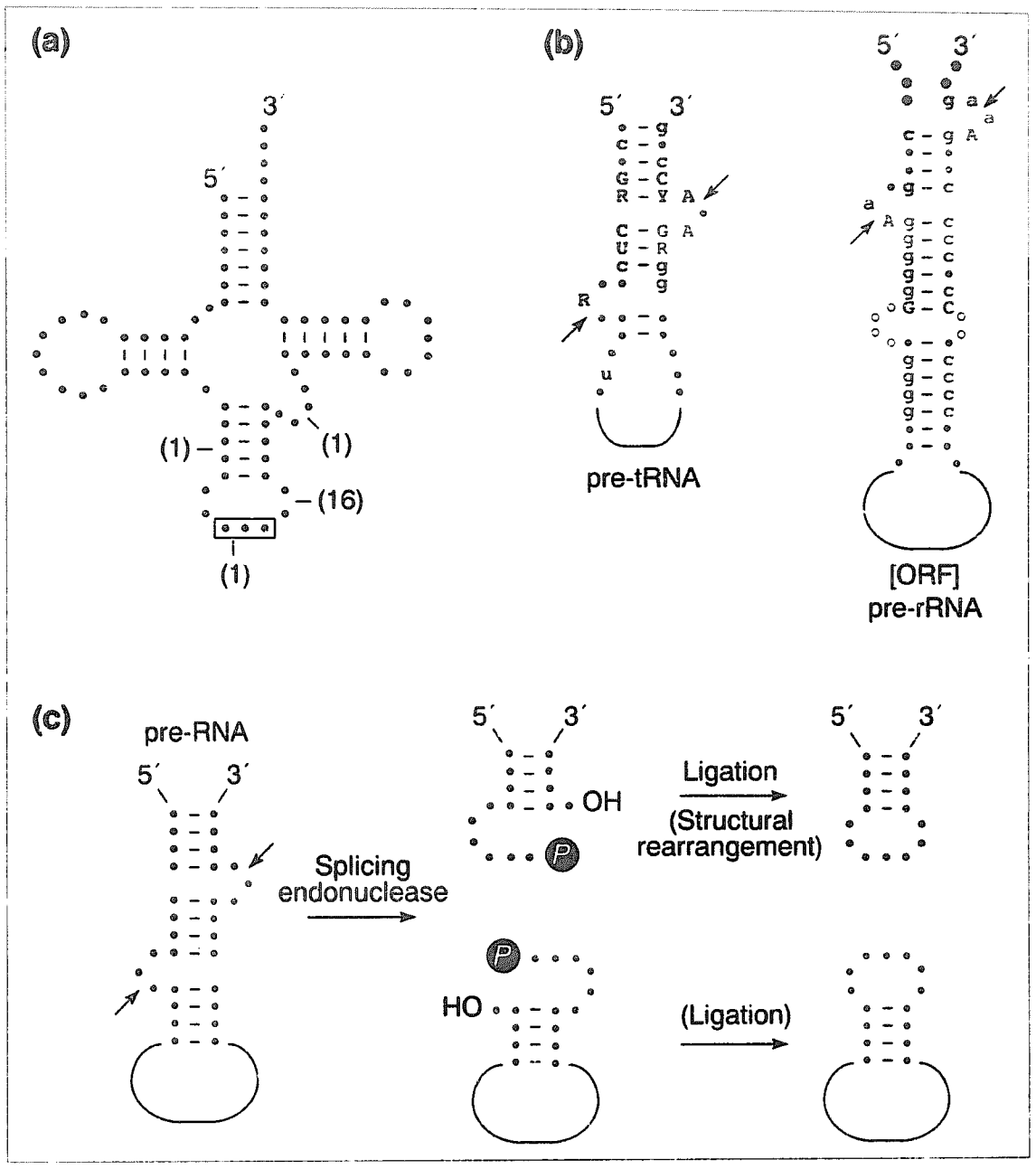

Figure 1

(a) Secondary structural model of tRNA where the locations and numbers (in brackets) of known archaeal introns are indicated; the anticodon is boxed. **(b)** Conserved 'bulge-helix-bulge' motifs at the exon–intron junctions of archaeal tRNA introns (left) and rRNA introns (right) where cleavage positions are indicated by arrowheads. Exon and intron nucleotides are shown in brown and red, respectively: upper-case letters indicate nucleotide conservation of ≥85%, and lower-case letters 60–85% conservation of the known sequences; less-conserved positions are presented as dots. R, purine and Y, pyrimidine. Base pairs, including G–U pairs, are indicated when present in ≥85% of the structures. The internal loop in the core structure of the rRNA intron, indicated by open circles, is variable in size. **(c)** Splicing mechanism of archaeal introns. The splicing endoribonuclease cleaves at the exon–intron junctions (indicated by arrows) and generates 5′-OH and 2′,3′-cyclic phosphates. Subsequently, the exons are ligated by an unknown mechanism, and sometimes a structural rearrangement occurs post-ligation[11]. The large rRNA introns circularize.

A common evolutionary origin

The following lines of evidence support a common origin for archaeal and eukaryotic nuclear tRNA introns. (1) Both splicing mechanisms require an endoribonuclease that generates a 2′-3′-cyclic phosphate and 5′-OH. However, whereas the archaeal enzyme recognizes and cuts a 'bulge-helix-bulge' motif (Fig. 1c), the nuclear enzyme primarily recognizes the tRNA structure and employs some kind of 'ruler' mechanism to recognize the intron–exon junction[16]. (2) Sequence-comparison studies indicate that the archaeal enzyme and the Sen2p and Sen34p subunits of the yeast heterotetrameric tRNA-splicing endoribonuclease are homologues[15,17,18]. (3) Most archaeal tRNA introns and all nuclear tRNA introns are located one nucleotide 3′ to the anticodon[11,16] (Fig. 1a).

Their putative similar origin raises the question as to how these introns and their splicing apparatus evolved. Searches within databases of bacterial genome sequences have failed to reveal homologues of the splicing endoribonuclease, which is consistent with it having evolved after the bacteria separated from the putative archaeal/eukaryotic branch. It is also likely that the more-complex splicing enzyme developed within the eukaryotic branch. Similarily, failure to detect archaeal- or nuclear-tRNA-type introns in bacterial genomes suggests that the introns also arose after this putative split. However, for the small tRNA introns and rRNA core introns, it remains likely that they derive from more ancient introns that may have been lost by at least some bacteria. Moreover, such a loss (or gain) of tRNA introns seems to be a fairly common event judging by their irregular distribution in the same tRNAs of different organisms.

Some archaeal rRNA introns encode homing endonucleases

Some of the rRNA introns contain ORFs positioned in the terminal loop of the intron core[5–7] (Fig. 1b). Such introns have only been found in the crenarchaeotal kingdom in the genera Desulfurococcus and Pyrobaculum, despite the sequencing of many archaeal 16S and 23S rRNA genes[4] and of the whole genome of *Methanococcus jannaschii*[19]. The encoded proteins carry two copies of a partly degenerate LAGLIDADG motif, which is

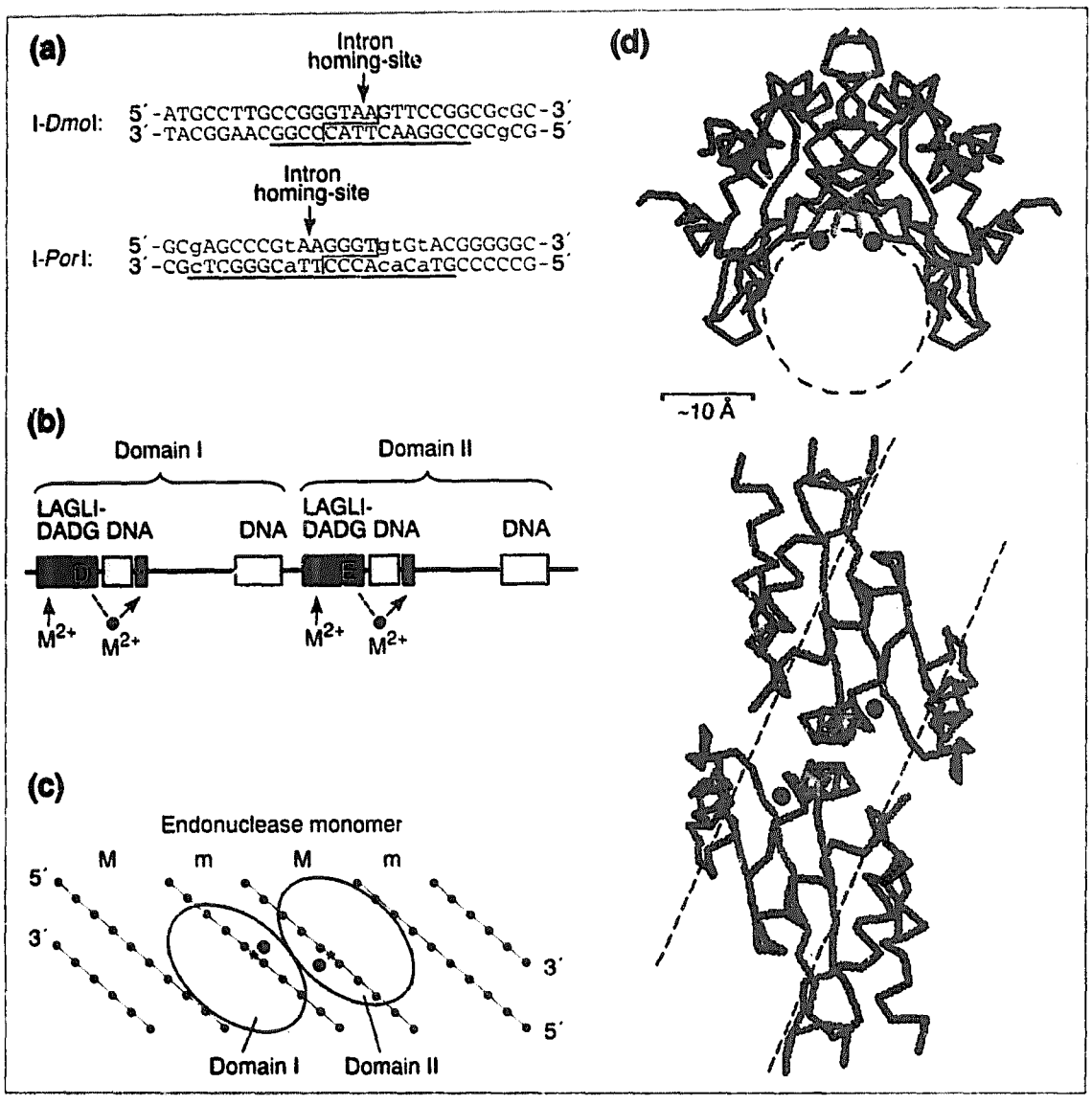

Figure 2

(a) Homing sites for introns encoding the archaeal endonucleases I-*Dmo*I and I-*Por*I are indicated by arrowheads and the cutting positions are marked by red lines. The approximate limits of the recognition site of the homing enzymes are underlined[21,25]. Base pairs conserved among all known crenarchaeotal rRNA genes are indicated by upper-case letters. **(b)** Diagram of archaeal homing endonucleases showing the linear organization of the two domains and the sites implicated in DNA and divalent metal ion binding (M^{2+}) that were deduced from protein footprinting experiments performed on I-*Dmo*I– and I-*Por*I–DNA complexes[22,26]. LAGLIDADG motifs and putative DNA-contact regions are represented as dark and light blue boxes, respectively. Regions of protein backbone positioned close to the divalent metal ions are indicated by orange boxes. The catalytic metal ions, coordinated by the conserved acidic residues (shown in yellow) of the LAGLIDADG boxes, are given as red circles[26]. **(c)** A two-dimensional projection of a homing endonuclease–DNA–metal ion complex, determined from the results of both DNA- and protein-footprinting experiments[22,25,26]. A probable distortion of the DNA double helix[25] is not illustrated. The sugar–phospate backbone is shown as lines, where dots indicate the positions of sugar residues. Base pairs are depicted as vertical lines and the major and minor grooves are labelled M and m, respectively. The scissile phosphates are indicated with asterisks. The two putative domains of the monomeric endonuclease are presented as yellow ellipses and the two catalytic metal ions are shown as red circles. **(d)** Crystal structure of I-*Cre*I, a homing endonuclease related to I-*Dmo*I and I-*Por*I. The DNA substrate is indicated by a dashed circle (upper) or dashed lines (lower). In the lower panel, the DNA is located above the protein. Protein-footprinting data on I-*Dmo*I and I-*Por*I are superimposed on the I-*Cre*I structure, using the same colour code as in (b). The putative positions of the catalytic metal ions (red circles) were deduced from Fe^{2+}–hydroxyl radical and mutagenesis experiments on the archaeal homing enzyme–DNA complexes[26].

also found in some proteins encoded by group I introns of mitochondria and chloroplasts, and occurs in inteins[1]. At least two of the archaeal intron-encoded proteins are homing endonucleases that specifically cleave intron$^-$ alleles of the rRNA genes that encode them[20,21] (Fig. 2a).

The DNA interactions of two of the archaeal homing enzymes, I-*Dmo*I and I-*Por*I, have been examined in some detail. Both recognize large regions of double-stranded DNA (15–20 bp) and, like all characterized LAGLIDADG endonucleases, they generate four nucleotide 3′-extensions and 5′-phosphates on cleavage[20,21] (Fig. 2a). Protein footprinting studies on I-*Dmo*I– and I-*Por*I–substrate complexes indicated that both endonucleases exhibit a two-domain structure with one conserved LAGLIDADG motif and two DNA-contacting sites within each domain[22] (Fig. 2b,c). Recent analyses of the crystal structures of monomeric PI-*Sce*I[23] and dimeric I-*Cre*I[24], two eukaryotic homing endonucleases, distantly related to monomeric I-*Dmo*I and I-*Por*I, confirmed these predictions. Each DNA-binding region, one within each protein half (Fig. 2b), consists of two β-strands that together form an antiparallel β-sheet. As inferred from DNA-footprinting data[25], the four-strand β-sheets of each protein domain contact the major groove on either side of cleavage sites that are positioned across the intervening minor groove (Fig. 2c,d). The DNA-footprinting results also indicate that I-*Dmo*I binding induces some distortion or bending of the DNA close to the cleavage sites and this may be a general property for other homing enzyme–DNA complexes[25].

The crystal structures reveal that the first part of the conserved LAGLIDADG motifs participate in domain–domain interactions (or monomer–monomer interactions for I-*Cre*I). Moreover, it was inferred from substrate-docking experiments that a fully conserved acidic residue at the penultimate position of both LAGLIDADG motifs is positioned close to each of the DNA-cleavage sites.

DNA cleavage requires divalent metal ions, and Fe^{2+}-hydroxyl radical and mutagenesis experiments showed that the catalytic divalent metal ions at each of the cleavage site phosphates are coordinated to one of the conserved acidic residues of the LAGLIDADG motifs and are close to the protein backbone lying carboxy-terminal to each motif[26] (Fig. 2b,d). Two other putative metal ions map within each of the LAGLIDADG motifs[26] (Fig. 2b,d).

Not all the archaeal proteins carrying LAGLIDADG motifs are homing endonucleases. For example, pPO2, encoded by the 23S rRNA gene of *Pyrobaculum organotrophum*, that also encodes I-*Por*I, does not cut an intron$^-$ allele[21]. Moreover, exceptionally for a LAGLIDADG protein, pPO2 has closely related homologues in other organisms, including a free-standing ORF in *M. jannaschii*[19] (R. Aravalli, unpublished); it remains to be seen whether it exhibits maturase activity as do some of the yeast mitochondrial proteins encoded in group I introns[1].

Intercellular mobility of archaeal introns

The encoding of homing-type endonucleases by crenarchaeotal rRNA introns raised the possibility that these introns were mobile. Such mobility has been demonstrated, intracellularly, for group I

and group II introns, and for an intein, between intron$^+$ and intron$^-$ alleles within eukaryotic cells[1]. However, most crenarchaeota contain only one set of rRNA genes per cell[1,4] and, therefore, have no obvious intracellular homing site. Experiments were performed to test for intercellular mobility of an archaeal intron between rRNA genes[27], a possibility that was enhanced by the observation that rDNA sequences flanking putative homing sites are often highly conserved (Fig. 2a). The rDNA intron from *Desulfurococcus mobilis* was transformed into intron$^-$ *Sulfolobus acidocaldarius* cells on a non-replicating and non-transcribing bacterial vector. Subsequently, a small number of introns recombined into the archaeal genome and the resultant intron$^+$ cells rapidly outgrew the intron$^-$ cells (Fig. 3a). It was demonstrated experimentally that this reflected both a selective advantage of intron$^+$ cells over intron$^-$ cells and intercellular mobility and homing of the intron[27] (Fig. 3b,c).

This result raised two important questions: (1) what is the basis of the intron-conferred selective advantage; and (2) what is the mechanism of intercellular intron mobility? At present, we have no clear insight into the molecular basis of either of these processes, although our favourite laboratory hypothesis for the selective advantage is that the highly stable, high-copy-number, circular intron species that encodes the homing endonuclease[12,13], or possibly the homing enzyme itself, is excreted from intron$^+$ cells and invades intron$^-$ cells where it acts as a viroid, virus-like particle, or antibiotic, by exposing intron$^-$ alleles to the homing endonuclease and, thereby, impairing or killing the intron$^-$ cells[12,13].

Intron mobility between archaeal cells may occur at a DNA level and this hypothesis is reinforced by experimental evidence for genetic exchange occurring by cell fusion in euryarchaeotes[28]. If cell fusion occurs, then the homing endonuclease could cleave the intron$^-$ allele during mating between intron$^+$ and intron$^-$ cells, and the intron could be copied and recombined into the intron$^-$ genome by a double-strand-break repair mechanism[1]. A low frequency of cell fusion in an *S. acidocaldarius* culture would provide a rationale for the relatively low level of intercellular mobility observed (Fig. 3b,c).

An alternative mechanism is that the circular RNA intron invades an intron$^-$ cell and undergoes reverse splicing and reverse transcription before recombining into the intron$^-$ allele. Clearly a number of difficult experiments are required in order to evaluate these models.

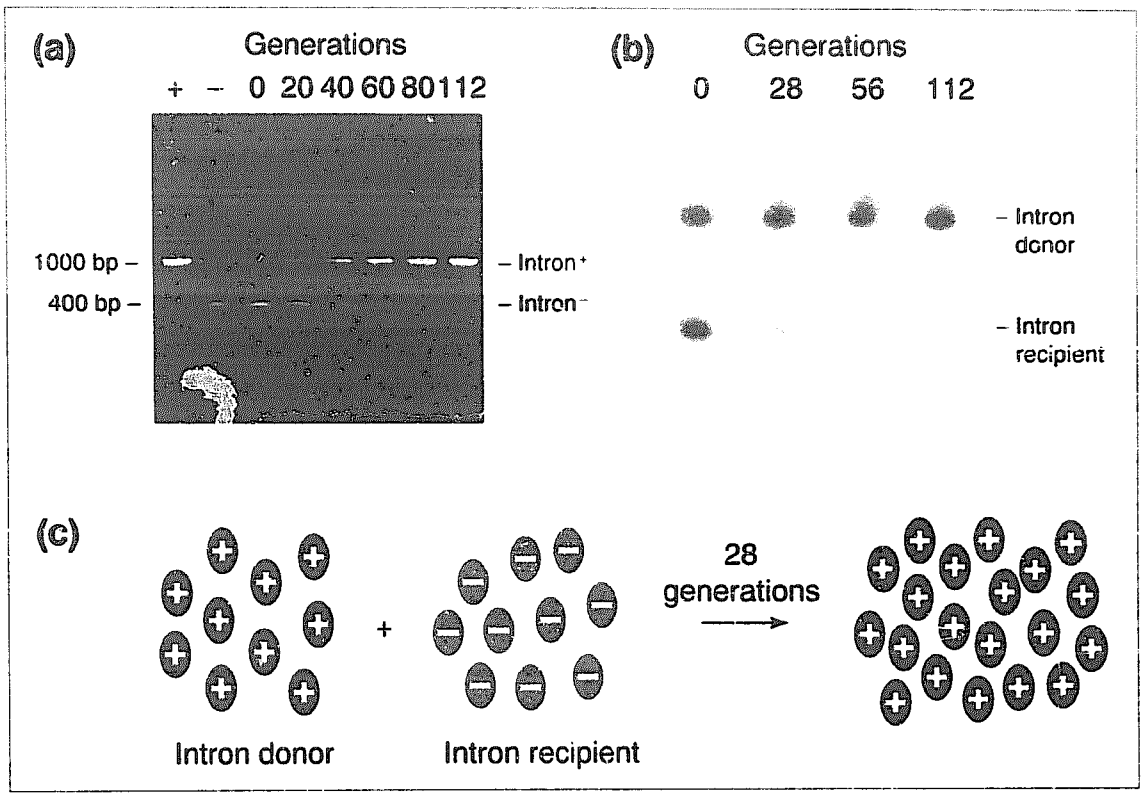

Figure 3
Intron homing and mobility between *Sulfolobus acidocaldarius* cells. **(a)** PCR reactions showing intron homing in *S. acidocaldarius*. A non-replicating plasmid containing the 23S rDNA intron of *Desulfurococcus mobilis* was transformed into *S. acidocaldarius* cells. PCR was performed on total genomic DNA isolated from cells after the indicated number of generations (doubling time about 6 hrs) using oligodeoxynucleotide primers flanking the putative intron homing site. Control PCR reactions on intron$^+$ DNA (+) and intron$^-$ DNA (–) are shown. **(b)** Selective advantage of intron$^+$ cells over intron$^-$ cells. Intron$^+$ cells (donor) were grown with an equal amount of intron$^-$ cells (recipient) and, after the indicated number of generations, primer extension reactions were performed on 23S rRNA that contained a genetic marker in the intron recipient cells and enabled the donor and recipient cells to be distinguished. (a) and (b) are modified from Ref. 27. **(c)** Diagram illustrating both the selective advantage of intron$^+$ *S. acidocaldarius* cells over intron$^-$ cells and the intercellular mobility of the intron. Intron donor and recipient cells are coloured dark and light green, respectively. The presence (+) or absence (–) of an intron is also indicated. After about 28 generations in the mixed culture, most cells are intron$^+$ donor cells, and the remaining intron recipient cells had all accrued the intron.

Given that natural environments are highly competitive for microorganisms, including hyperthermophiles, it is puzzling that the introns are not more widespread in archaea if they both confer a selective advantage on their host cells and are mobile intercellularly, especially because some homing sites constitute highly conserved sequence regions in the archaeal, and other, rRNAs (Fig. 2a).

A clue to the answer may be found in the genome of *M. jannaschii* where no intron-encoded ORFs were detected, but no less that 18 inteins occur, all of which carry double LAGLIDADG motifs, and some of which may be homing endonucleases[19]. If these enzymes are the cause of selective advantage, then it may require less energy for the cell to express these proteins together with low-copy-number enzymes, which are generally involved in DNA processing, than from the high-copy-number rRNA molecules.

Loss and gain of ORFs encoding LAGLIDADG proteins

The presence of ORFs in some group I intron cores, but not in others, led to the suggestion that the ORF itself could enter and leave the core structure of the intron[29]. In support of this hypothesis, highly divergent ORFs are found in group I intron cores that are closely related in sequence. Moreover, sequences flanking both the intron homing site and the putative ORF insertion site show close similarity for a few group I introns, such that, in principle, a homing endonuclease could cut a site within an ORF-less intron. This hypothesis can be extended to the archaeal rRNA introns where ORF-less intron core structures are also found, and sequence similarities also exist between the intron homing site and the putative ORF insertion site[11,12].

It has been proposed for group I introns that ORF insertion events occurred at a DNA level by illegitimate recombination after cleavage of the ORF-less intron core[30]. Another possibility is that

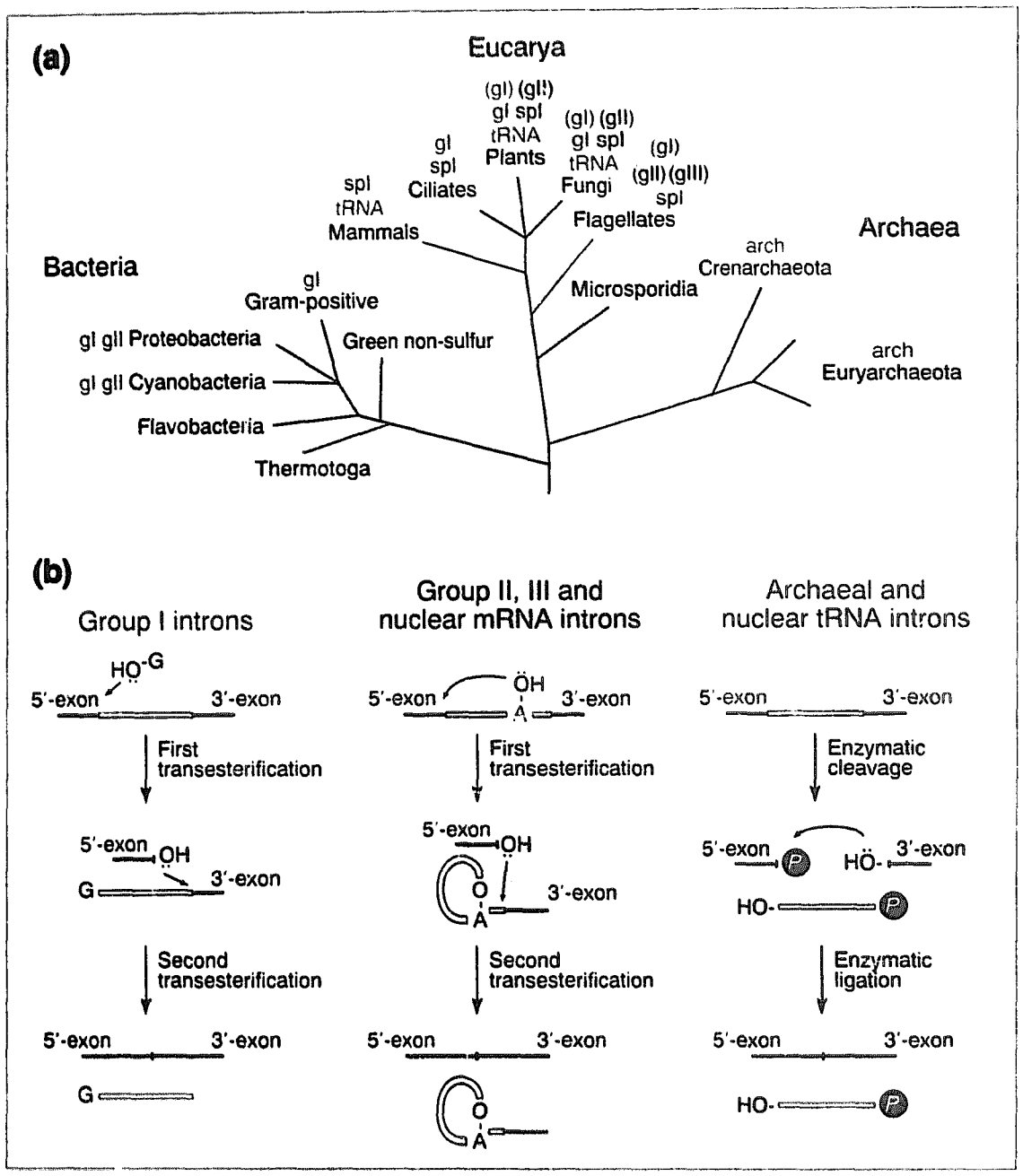

Figure 4

(a) Phylogenetic tree showing the known distribution of the different classes of introns that are colour-coded according to their splicing mechanism as shown in (b): archaeal, arch (red); nuclear tRNA, tRNA (red); group I, gI (green); group II, gII (blue); group III, gIII (blue); nuclear mRNA, spI (blue) (also called spliceosome introns). Mitochondrial and chloroplast introns are given in brackets. **(b)** The three mechanisms of intron removal. Group I introns (green) are removed by the two transesterification reactions that are illustrated. The subsequent circularization of some group I introns is not shown. Group II, group III and nuclear mRNA introns (blue) are also excised by two consecutive transesterifications, that are outlined, to produce ligated exons and an intron lariat. Archaeal and nuclear tRNA introns (red) are excised by a splicing endoribonuclease that generates 5'-OH and 2',3'-cyclic phosphates and then the exons are ligated. Circularization of archaeal rRNA introns is shown in Fig. 1a.

reverse splicing occurred at an RNA level, which is consistent with the coding RNA generally being located within accessible loops of the core structures of both group I introns and archaeal introns[12] (Fig. 1b). Moreover, the possibility that some features of the ORF insertion process are shared between these introns, is supported by the observation that archaeal, and many group I introns, encode proteins carrying double LAGLIDADG motifs which are presumed to be homologues. Furthermore, because these proteins occur throughout the three domains of life, where they are either encoded by introns or produced as inteins or free-standing proteins, they must pre-date the divergence of the three domains (Fig. 4a).

Are all three mechanisms of intron-excision ancient?

Introns occur widely within the three domains of life (Fig. 4a) and, although six different classes of introns have been identified, they are excised by three basic mechanisms that can be categorized as follows: (1) group I introns; (2) group II, group III and nuclear mRNA introns; and (3) archaeal and eukaryotic nuclear tRNA introns (Fig. 4b).

The first two mechanisms involve two consecutive transesterification steps, which differ only in the nucleophile used in the first step; whereas group I introns invoke the 3'-OH of an external guanosine[31], group II, group III and nuclear mRNA introns use the 2'-OH of an internal nucleotide, generally an adenosine[32,33]. Although the capacity of group I and group II introns to self-splice *in vitro* suggests that they existed during a putative period of RNA-based biology, they have evolved a dependence on protein factors for *in vivo* splicing, which probably helps to stabilize the intron core structures and renders the splicing reaction more efficient[34]. The nuclear mRNA introns have probably evolved from group II introns and their splicing depends on various *trans*-acting ribonucleoprotein complexes (snRNPs)[33]; this may also be true for group III introns[32].

The third mechanism of splicing, common to archaeal and nuclear tRNA introns (Fig. 4b), is dependent on an endoribonuclease, although experiments with human and plant nuclear tRNA introns have provided evidence for protein-independent cleavage at the intron–exon junctions *in vitro*[35]. The latter cleavage is likely to be induced by the specific positioning of divalent metal ions at the catalytic centre and, as such, would resemble self cleavage by 'hammerhead' ribozymes that also generate 2',3'-cyclic phosphates[36]. Thus, the third splicing mechanism for archaeal and nuclear tRNA introns may also derive from a ribozyme-based reaction.

Acknowledgements

We thank B. L. Stoddard for providing the crystal structure coordinates of the I-*Cre*I enzyme. The authors' research was supported by a Danish RNA Regulation Centre grant and European Union grants BIO2-CT93-0274 and BIO4-CT96-0270. J. L. was supported by Copenhagen University.

References

1 Belfort, M., Reaban, M. E., Coetzee, T. and Dalgaard, J. Z. (1995) *J. Bacteriol.* 177, 3897–3903
2 Kjems, J., Leffers, H., Olesen, T. and Garrett, R. A. (1989) *J. Biol. Chem.* 264, 17834–17837
3 Wich, G., Leinfelder, W. and Böck, A. (1987) *EMBO J.* 6, 523–528
4 Garrett, R. A. *et al.* (1991) *Trends Biochem. Sci.* 16, 22–26
5 Kjems, J. and Garrett, R. A. (1985) *Nature* 318, 675–677
6 Dalgaard, J. Z. and Garrett, R. A. (1992) *Gene* 121, 103–110
7 Burggraf, S., Larsen, N., Woese, C. R. and Stetter, K. O. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 2547–2550
8 Kjems, J., Jensen, J., Olesen, T. and Garrett, R. A. (1989) *Can. J. Microbiol.* 35, 210–214
9 Thompson, L. D. and Daniels, C. J. (1988) *J. Biol. Chem.* 263, 17951–17959
10 Thompson, L. D. and Daniels, C. J. (1990) *J. Biol. Chem.* 265, 18104–18111
11 Kjems, J. and Garrett, R. A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 439–443

12 Lykke-Andersen, J. and Garrett, R. A. (1994) *J. Mol. Biol.* 245, 846–855
13 Kjems, J. and Garrett, R. A. (1988) *Cell* 54, 693–703
14 Garrett, R. A. *et al.* (1994) *Syst. Appl. Microbiol.* 16, 680–691
15 Kleman-Leyer, K., Armbruster, D. W. and Daniels, C. J. (1997) *Cell* 89, 839–847
16 Phizicky, E. M. and Greer, C. L. (1993) *Trends Biochem. Sci.* 18, 31–34
17 Trotta, C. R. *et al.* (1997) *Cell* 89, 849–858
18 Lykke-Andersen, J. and Garrett, R. A. *EMBO J.* (in press)
19 Bult, C. J. *et al.* (1996) *Science* 273, 1058–1073
20 Dalgaard, J. Z., Garrett, R. A. and Belfort, M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 5414–5417
21 Lykke-Andersen, J., Thi-Ngoc, H. P. and Garrett, R. A. (1994) *Nucleic Acids Res.* 22, 4583–4590
22 Lykke-Andersen, J., Garrett, R. A. and Kjems, J. (1996) *Nucleic Acids Res.* 24, 3982–3989
23 Duan, X., Gimble, F. S. and Quiocho, F. A. (1997) *Cell* 89, 555–564
24 Heath, P. J., Stephens, K. M., Monnat, R. J., Jr and Stoddard, B. L. (1997) *Nat. Struct. Biol.* 4, 468–476
25 Aagaard, C., Awavez, M. J. and Garrett, R. A. (1997) *Nucleic Acids Res.* 25, 1523–1530
26 Lykke-Andersen, J., Garrett, R. A. and Kjems, J. (1997) *EMBO J.* 16, 3272–3281
27 Aagaard, C., Dalgaard, J. Z. and Garrett, R. A. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 12285–12289
28 Tchelet, R. and Mevarech, M. (1994) *Syst. Appl. Microbiol.* 16, 578–581
29 Perlman, P. S. and Butow, R. A. (1989) *Science* 246, 1106–1109
30 Loizos, N., Tillier, E. R. M. and Belfort, M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 11983–11987
31 Cech, T. R. (1990) *Annu. Rev. Biochem.* 59, 543–568
32 Michel, F. and Ferat, J. L. (1995) *Annu. Rev. Biochem.* 64, 435–461
33 Moore, M. J., Query, C. C. and Sharp, P. A. (1993) in *The RNA World* (Gesteland, R. F. and Atkins, J. F., eds), pp. 303–357, Cold Spring Harbor Laboratory Press
34 Caprara, M. G., Lehnert, V., Lambowitz, A. M. and Westhof, E. (1996) *Cell* 87, 1135–1145
35 Weber, U., Beier, H. and Gross, H. J. (1996) *Nucleic Acids Res.* 24, 2212–2219
36 Scott, W. G. *et al.* (1996) *Science* 274, 2065–2069

PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox

PAS domains, previously reported in proteins from mammals, insects, plants, fungi and cyanobacteria, are typically paired with a repeat domain and involved in protein–protein interactions[1–4]. They have been associated with light reception, light regulation and clock proteins[3,4]. We have identified highly conserved regions (S_1 and S_2 boxes) that are present in PAS domains and in a large family of sensor proteins of all kingdoms, including Archaea. This family includes many well-known sensory proteins that were not previously recognized as members of the PAS domain family. We also expand the range of signal transduction events that are associated with PAS domains. Of particular interest is the sensing of oxygen and redox potential.

The S-boxes were initially detected in bacterial sensors. Aer, a novel *Escherichia coli* signal transducer for behavior, responds to changes in the concentration of oxygen, redox carriers and carbon sources[5]. NifL of *Azotobacter vinelandii* is the redox sensor of a novel two-component regulatory system that controls expression of nitrogen fixation genes[6]. Identification of a conserved sequence that is shared by putative sensing domains of Aer and NifL[5] prompted a global search for other proteins that

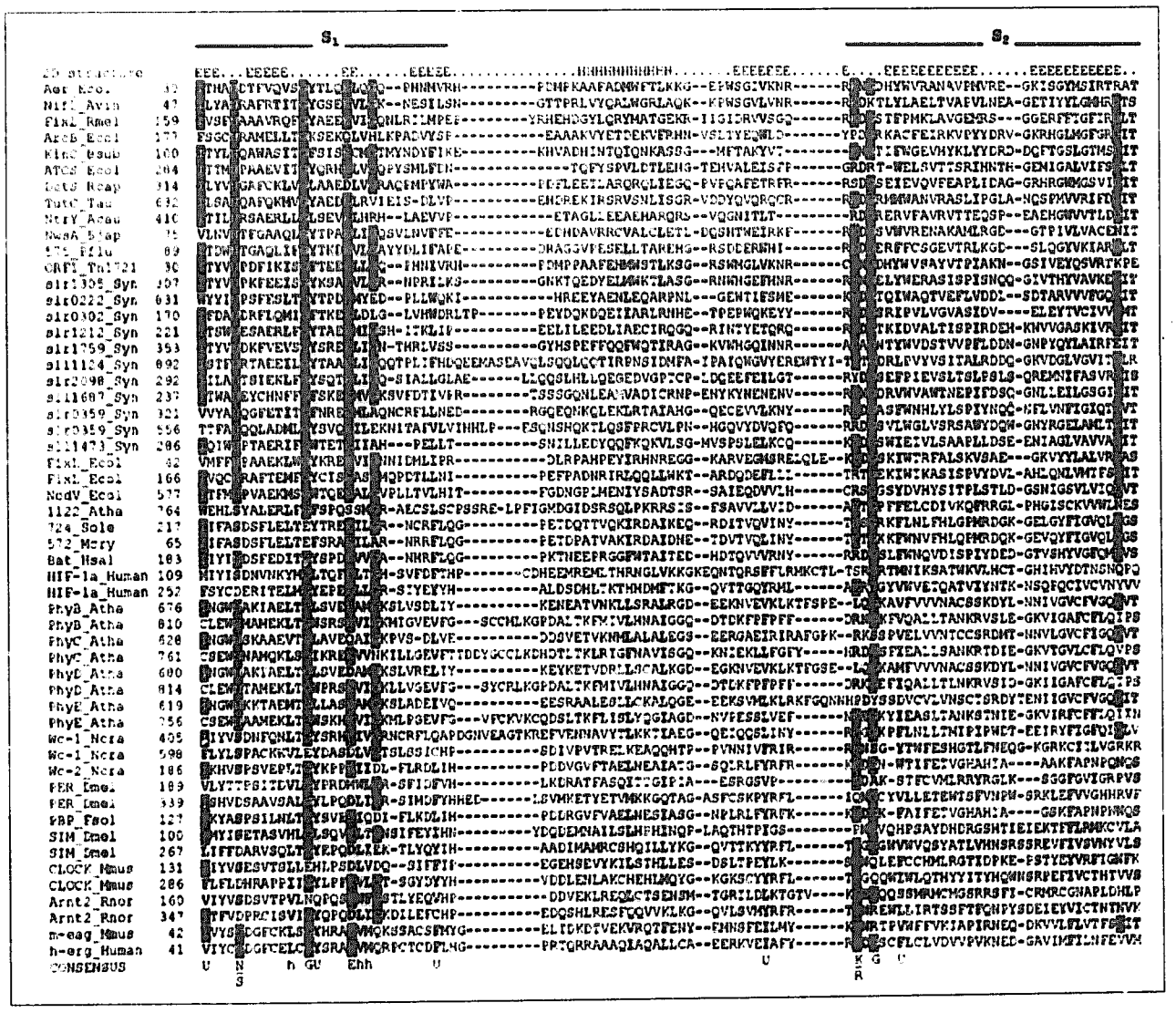

Figure 1
Multiple alignment of S-boxes. Sequence database searches were performed with the BLASTP program[7] using initially the putative sensing domain sequence of Aer and NifL as a query. Further searches with various representatives of the family gave significant cross-kingdom alignments. For example, queries using the archaebacterial Bat sequence gave P-values of 1.3×10^{-21} (scoring matrix blosum62) with the spinach protein kinase, 1.1×10^{-19} with the eubacterial *Synechocystis* slr0359 protein, 3.9×10^{-17} with the *Neurospora* Wc-1 protein and 3.4×10^{-8} with ion channel sequence from humans. The sequences were aligned using the programs CLUSTALW[18] and PILEUP[19], and manual alignment. Profiles were obtained using the GCG program PROFILEMAKE[20] and searching of the SWISS-PROT and PIR databases carried out with PROFILESEARCH[8]. Z-scores ranged from 3.7 to 14. Species abbreviations are indicated in Table I. Identical amino acids conserved in at least 60% of the sequences are highlighted in red; similar residues conserved in at least 70% of sequences are highlighted in yellow. The consensus line summarizes residue properties conserved in at least 80% of all sequences (U, bulky hydrophobic: FILMVWY; h, hydrophobic: ILMV). The secondary structure elements predicted using the PhD server[21] are shown above the alignment: H denotes an α-helix, and E represents a β-strand.

 0968–0004/97/$17.00 PII: S0968-0004(97)01110-9