could be brought back to life with disastrous effects? If microorganisms have been revived, what precautions have been taken to establish that the revived organism is indeed revived and not simply an environmental contaminant?

To begin grappling with some of these issues and to present a hitherto underused resource, we have set up a database of organisms which have been preserved for more than 100 years. This database is split into two sections. The first contains information about organisms that, although dead, have been preserved. DNA material may or may not be available from these organisms. The second contains a list of those organisms that have been revived after storage for long periods of time (see table).

The database is currently in its infancy, but it is hoped that others may contribute information about other long preserved microorganisms so that a central list can be kept as a potential resource. Researchers knowing of any other long-preserved organisms are encouraged to contact us directly.

Max J. Kennedy
Sarah L. Reader
The New Zealand Institute for Industrial Research and Development, PO Box 31-310, Lower Hutt, New Zealand
Fax: 64 4 569 0132

1. Karkhanis, S. N. *Nature* **261**, 406–407 (1976).
2. Kazmierczak, J. & Golubik, S. *Nature* **261**, 404–406 (1976).
3. Purdy, L. H. & Purdy, B. A. *Bot. Gaz.* **143** (4), 551–553 (1982).
4. Anon *Science* **253**, 1353 (1991).
5. Kendig, F. & Hutton, R. *Life-Spans or How Long Things Last* (Holt, Rinehart and Winston, New York, 1979).
6. Park, P. *New Scient.* p 12 (1 June 1991).
7. Seawa, M. R. D., Cross, T. & Uns, B. A. *Nature* **261**, 407–408 (1976).
8. Coghlan, A. *New Scient.* p 2 (14 September 1991).

Familiar strangers

SIR — All three protein structures[1–3] reported in a recent issue of *Nature* display folding motifs previously observed in other protein structures. In only one case did the authors concerned notice this similarity.

The amino-terminal domain in the staphylococcal enterotoxin B (SEB) structure[1] seems to be the fifth example of a novel common fold. This fold consists of a five-stranded β-barrel capped by an α-helix between the third and fourth strands. It has been observed in three other proteins secreted by bacterial pathogens: staphylococcal nuclease[4], the B-subunits of heat-labile enterotoxin[5] and verotoxin-1 (ref 6). It has also been found in the anticodon-binding domain of Asp-tRNA synthetase[7]. These four proteins, with very different amino-acid sequences, are very similar in secondary structure, even in minor details. They all bind oligonucleotides or oligosaccharides and their binding sites, where known, are similarly located on a particular side of the barrel[8]. These observations suggest that the corresponding site on SEB's amino-terminal domain may be essential for its toxic function. The carboxy-terminal domain of this protein seems to be a variation of the β-grasp motif[9] previously observed in ubiquitin, chloroplast-type ferredoxins and in the immunoglobulin G-binding domain of streptococcal protein G.

Musacchio *et al.*[2] found no protein fold closely related to the one they revealed in the structure of Src-homology (SH3) domain of spectrin. But this fold has been observed previously in the structure of R67 dihydrofolate reductase[10]. The fold, common to both structures, is composed of an antiparallel β-sheet of five strands coiled into a barrel-like structure. The common secondary structure also includes a turn of 3_{10} helix between the fourth and fifth strands. Despite having different amino-acid sequences, the close structural resemblance of SH3 and R67 proteins may be a clue to an as yet unidentified function of the SH3 domain.

The histidine-rich actin-binding protein hisactophilin was described in ref. 3 as having the same fold, the β-trefoil fold[11], as interleukin-1 and fibroblast growth factors. It is actually the eleventh example of this fold, which has also been observed in Kunitz trypsin inhibitors, ricin B chain[11] and in the interleukin-1 receptor antagonist protein[12]. The amino-acid sequences of proteins from different functional families show few significant similarities. However, in addition to their similar fold the proteins share another common feature: they all perform recognition functions.

Not only these three, but a substantial fraction of proteins whose structures have now been determined and which have no clear sequence similarity to previously known structures, seem to share already known folding motifs[8]. This information supports the suggestion that a few different protein folds perform a much larger number of biological functions.

Alexey G. Murzin
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

1. Swaminathan, S. *et al. Nature* **359**, 801–806 (1992).
2. Musacchio, A. *et al. Nature* **359**, 851–855 (1992).
3. Habazettl, J. *et al. Nature* **359**, 855–858 (1992).
4. Arnone, A. *et al. J. biol. Chem.* **246**, 2302–2316 (1971).
5. Sixma, T. K. *et al. Nature* **351**, 371–377 (1991).
6. Stein, P. E., Boodhoo, A., Tyrrell, G. J., Brunton, J. L. & Read, R. J. *Nature* **355**, 748–750 (1992).
7. Moras, D. *Trends biochem. Sci.* **17**, 159–164 (1992).
8. Murzin, A. G. & Chothia, C. *Curr. Opin. struct. Biol.* **2**, 895–903 (1992).
9. Overington, J. P. *Curr. Opin. struct. Biol.* **2**, 349–401 (1992).
10. Matthews, D. A. *et al. Biochemistry* **25**, 4194–4204 (1986).
11. Murzin, A. G., Lesk, A. M. & Chothia, C. *J. molec. Biol.* **223**, 531–543 (1992).
12. Stockman, B. J. *et al. Biochemistry* **31**, 5237–5245 (1992).

DNA correlations

SIR — Despite your editorial[1] saying that the long-range correlations in DNA reported by Peng *et al.*[2] and by Voss[3] had yet to be explained, we had already proposed one explanation of the phenomenon[4,5].

We concluded that non-coding DNA sequences (introns) tend to have longer-ranged correlation than coding DNA sequences (exons). Some, but not all, intron sequences exhibit even longer-ranged correlation and the resulting power spectra are $1/f^{\alpha}$ ($\alpha \approx 1$)[4,5]. Both conclusions appeared in refs 2 and 3, confirming our findings. We had proposed that the longer-ranged correlation is due to the abundance of repetitive patterns which are common in intron sequences.

We were led to this study when one of us was searching for long-range correlation and $1/f^{\alpha}$ spectra in cellular automata — a class of sequence manipulation rules which do not change the sequence length. It was soon realized that it is much easier to generate long-range correlation when the sequence length is allowed to increase.

This observation led to the study of a class of mathematical models called expansion–modification systems in which the sequence lengths become longer and longer[6,7]. These systems are extremely simple, and the only two basic operations contained are duplication (or amplification if more than two copies are made) and mutation. It was found that if the two operations are applied repeatedly, the elongated sequences can have $1/f^{\alpha}$ power spectra.

According to common belief, the lengths of present-day DNA sequences are longer than those of three billion years ago, when life started. One of the mechanisms for sequence elongation is gene or oligonucleotide duplication[8], during which copies of a segment of the sequence are made and these copies are inserted back into the original sequence. Similar to the case in expansion–modification systems, the gene or oligonucleotide duplication process is not perfect. If it occurred at different stages of evolution, the length of the repeated pattern could be different, leading to correlations at different length scales.

The fingerprints of oligonucleotide duplication are preserved better in intron sequences where, as we have indeed

1. Maddox, J. *Nature* **358**, 103 (1992).
2. Peng, C.-K. *et al. Nature* **356**, 168–170 (1992).
3. Voss, R. *Phys. Rev. Lett.* **68**, 3805–3808 (1992).
4. Li, W. & Kaneko, K. *Europhys. Lett.* **17**, 655–660 (1992).
5. Li, W. *Int. J. Bifurc. Chaos* **2**(1), 137–154 (1992).
6. Li, W. *Europhys. Lett.* **10**, 395–400 (1989).
7. Li, W. *Phys. Rev.* A**43**, 5240–5260 (1991).
8. Ohno, S. *Evolution by Gene Duplication* (Springer, Heidelberg, 1970).