

A new protein sequence data bank

SIR—A protein sequence data bank, called PGtrans, is available from our laboratory. This data bank is generated by automatic computer translation of the well-known nucleotide sequence library GenBank (established by Bolt, Beranek and Newman Inc. and Los Alamos National Laboratory). For this we designed a translator program which scans and interprets the specific information fields of each GenBank entry: DEFINITION, SOURCE and FEATURE table. The translator program can thus decide if the corresponding nucleotide sequence includes a coding region, and what is the proper genetic code to use to translate it into a putative protein sequence.

For each coding region successfully translated, the program creates a PGtrans entry including a text constituted of several fields: CODE, NAME, SOURCE, REFERENCE and COMMENT, followed by the amino-acid sequence. The translator program checks for unexpected 'stop' codons eventually found in coding regions. If a 'frame shift' allows the errored coding region to be rescued, the tentative amino-acid sequence is included in PGtrans. However, a warning is included into the COMMENT field to indicate a potential problem. The current PGtrans databank corresponding to GenBank (35.0) contains 3,107 entries for a total of 483,649 residues. Figure 1 shows the

entry-length distribution for PGtrans 29.0. A listing of the encounter anomalies is routinely communicated to the makers of GenBank.

The main purpose of PGtrans is to offer a direct access to all amino-acid sequences coded among GenBank nucleotide sequences and thus to be an efficient tool for protein homology searches. Its format is compatible with the rest of our sequence analysis software¹ and will be consistent with the recommendations of the CODATA task group on protein sequence data banks². Since we conserved the minimum of text information necessary to identify the protein, PGtrans is a compact data bank (1.4 Mbytes) which could fit for few more years into microcomputer environments. Among other efforts in gathering protein sequence data^{3,4}, the role of PGtrans is to provide a rapid update of new protein sequences determined by nucleic acid analysis. Magnetic tape copies (card image) of PGtrans will be distributed free of charge upon request to our laboratory.

JEAN-MICHEL CLAVERIE
ISABELLE SAUVAGET

Unité d'Informatique Scientifique,
Institut Pasteur,
75724 Paris Cedex 15,
France

1. Claverie, J.-M. *Nucleic Acids Res.* **12**, 397-407 (1984).
2. Lesk, A.-M. *Nature* **314**, 318-319 (1985).
3. Orcutt, B.C., George, D.G. & Dayoff, M.O. *A. Rev. Biophys. Bioeng.* **12**, 419-441 (1983).
4. Doolittle, R.F. *Science* **214**, 149-159 (1981).
5. Seto, Y. & Sakakibara, Protein Research Foundation Minoh, Osaka 562, Japan.

Three-dimensional electron lattice

SIR—*Nature* reported on 14 February (313, 527; 1985) the discovery of a three-dimensional electron lattice in *n*-HgCdTe by Rosenbaum *et al.* (*Phys. Rev. Lett.* **54**, 241; 1985). In fact the discovery was made in my laboratory in 1979. Our response to Rosenbaum *et al.* and relevant references can be found in *Physical Review Letters* for 22 July (55, 443; 1985).

G. NIMTZ

Universität zu Köln,
Physikalisches Institut,
Zulpicherstrasse 77,
5 Köln 41, FRG

Shadow boxing with Darwin

SIR—Dawkins's taunting review¹ of Eldredge's book *Time Frames: The rethinking of Darwinian evolution and the theory of punctuated equilibria* raised important issues of evolutionary theory that can only but invite comment.

There always have been two components to Eldredge and Gould's theory of punctuated equilibria, as the phrase suggests explicitly — long periods of stasis (equilibria) punctuated by relatively much shorter periods of change. To reiterate, yet again, that the periods of change potentially fall within traditional boundaries of gradual Darwinian evolution by means of natural selection, and to ignore the major periods of stasis (a device which allows Dawkins to dismiss the whole concept as an "interesting little theory . . . lying firmly within the neo-Darwinian synthesis") reduces scientific evaluation to a parlour game of one's own invention, which whilst entertaining and superficially coherent, does not offer proper scientific guidance for the general reader.

If we accept, for the moment, that stasis is a widespread observation in the geological record, and if too we accept that the forces behind selection are the only means for promoting biologically useful novelties throughout a population (setting aside contemporary knowledge that the molecular behaviour of the genes can effect similar changes), then it is surprising that little significant evolution occurs over periods of time that may span from hundreds of thousands to millions of generations, in a diverse range of species. The point is not that evolution is gradual "during the times that it is actually happening" (Dawkins' emphasis) but that it seemingly is not happening for most of the time. This was the message of the infamous Chicago meeting, which many of us took home, and which demands an explanation.

Attempts to bring stasis within the neo-Darwinian synthesis (see ref. 2 for the best of them) are not wholly satisfactory in that they plead for a general process called stabilizing selection, coupled to an

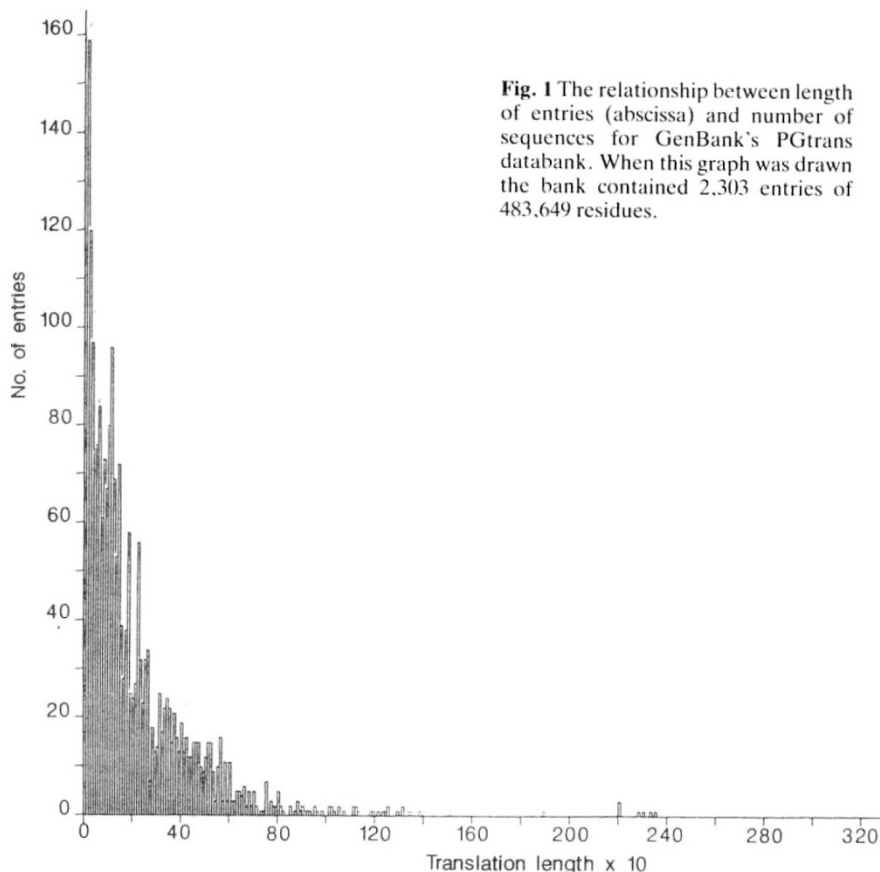


Fig. 1 The relationship between length of entries (abscissa) and number of sequences for GenBank's PGtrans databank. When this graph was drawn the bank contained 2,303 entries of 483,649 residues.