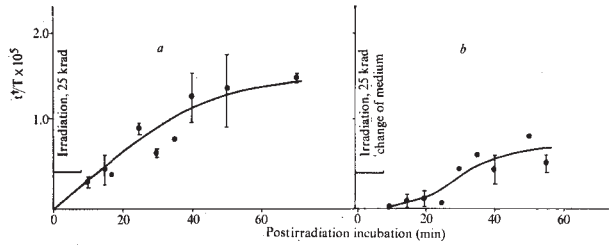


Fig. 3 Kinetics of the removal of damaged thymine residues (t†) from the DNA in CHO cells following exposure to 25 krad. The pooled and neutralised fractions of peaks 1 of the Dowex $50W \times 8(H^+)$ chromatograms of the type shown in Fig. 1 were analysed for their content in products of the 6-(hydroxy- or hydroperoxy)-5,6-dihydrothymine-type (t†) by a scaled up version of the reductive assay of Hariharan and Cerutti[8]. For reduction 0.1 ml of 0.5 M Ca_2CO_3 buffer, *p*H 10.3, and 20 mg of $NaBH_4$ dissolved in 0.1 ml of 0.01 N NaOH were added to 2 ml samples. Reduction was allowed to proceed for 20 h at 10° C in the dark. Excess reducing agent was then destroyed by the addition of 40 µl of glacial acetic acid and the samples were applied to columns containing an upper portion of Dowex $1 \times 10(OH^-)$ (1 × 9 cm) and, separated by a layer of sand, a lower portion of Dowex $50W \times 8(H^+)$ (1 × 8 cm). The columns were eluted with decarbonated water, and forty 0.5 ml fractions were collected directly into scintillation vials for counting. For the intracellular material (*a*), only the second of two radioactivity peaks represented t† (fraction 17–27). The first peak (fractions 11–16) contained some TTP which was not retained by the double column. Only a single radioactivity peak corresponding to t† was obtained from the extracellular material (*b*). The data from two experiments are given as ratios of the radioactivity obtained in the reductive assay over the total radioactivity originally contained in the culture medium plus the cells (T). The points with error bars were run in duplicates or triplicates. A correction has been made for the values for unirradiated samples otherwise carried through identical procedures, but not for the yield in the reductive assay[8].

grams were analysed for their content in products of the 6-hydroxy (or hydroperoxy)-5,6-dihydrothymine type (t†) by a modification of the reductive assay of Hariharan and Cerutti[8]. Radiation products of the t† type radioactively labelled with tritium (originating from the culture medium and cells) or with ^{14}C (originating from the γ-irradiated thymidine marker) were detectable only in the extracellular and intracellular peak I fractions. Figure 3 shows the time course of t† release. Our results demonstrate that γ-ray-damaged thymine residues are removed from the DNA within approximately 60 min postirradiation incubation. Approximately two times more t† accumulates inside the cell than in the culture medium. The removal of t† seems to be more rapid in CHO cells relative to that of photodimers in human cells. To judge from the appearance of unscheduled synthesis, the repair process started to level off only after 3–4 h postirradiation incubation in human skin fibroblasts[15] and in granulocytes in peripheral blood[16]. Whether the removal of radiation products from the DNA is the result of an ultraviolet type repair process or a facet of the controlled DNA degradation induced by γ rays in CHO cells cannot be decided at present.

We thank Dr R. B. Painter for stimulating discussions. This work was supported by a grant from the National Institutes of Health and a contract from the US Atomic Energy Commission. This work will be contained in the thesis submitted to Princeton University by M. R. M. in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

M. R. Mattern
P. V. Hariharan
B. E. Dunlap
P. A. Cerutti

Department of Biochemistry, University of Florida, Gainesville, Florida 32610

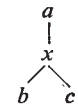
Received April 25; revised June 29, 1973.

1 Regan, J. D., Trosko, J. E., and Carrier, W. L., *Biophys. J.*, **8**, 319 (1968).
2 Cleaver, J. E., and Trosko, J. E., *Photochem. Photobiol.*, **11**, 547 (1970).
3 Klimek, M., *Neoplasm*, **12**, 559 (1965).
4 Klimek, M., *Photochem. Photobiol.*, **5**, 603 (1966).
5 Trosko, J. E., Chu, E. H., and Carrier, W. L., *Radiat. Res.*, **24**, 667 (1965).
6 Trosko, J. E., and Kasschau, M. R., *Photochem. Photobiol.*, **6**, 215 (1967).
7 Setlow, R. B., Regan, J. D., German, T., and Carrier, W. L., *Proc. natn. Acad. Sci. USA*, **64**, 1035 (1969).
8 Hariharan, P. V., and Cerutti, P. A., *J. molec. Biol.*, **66**, 65 (1972).
9 Van der Schans, G. P., Bleichrodt, J. F., and Blok, J., *Int. J. Radiat. Biol.*, **23**, 133 (1973).
10 Painter, R. B., and Cleaver, J. E., *Nature*, **216**, 369 (1967).
11 Carrier, W. L., and Setlow, R. B., *Abstr. WPM-F9, Seventh Annual Meeting, Biophysical Society*, Columbus, Ohio, 1973.
12 Brent, T., *Abstr. WPM-F8, Seventh Annual Meeting, Biophysical Society*, Columbus, Ohio, 1973.
13 Winans, L. F., Dewey, W. C., and Dettor, C. M., *Radiat. Res.*, **52**, 333 (1973).
14 Painter, R. B., in *Effects of Radiation on Cellular Proliferation and Differentiation*, 91 (International Atomic Energy Agency, Vienna, 1968).
15 Rasmussen, R. E., dissertation, Univ. California (1968).
16 Connor, W. G., and Norman, A., *Mutat. Res.*, **13**, 393 (1971).

Evolution of 5S RNA and the Non-randomness of Base Replacement

Sequences of 5S rRNA have been published for five widely divergent organisms: *Escherichia coli*, *Pseudomonas fluorescens*, yeast, human KB carcinoma, and *Xenopus laevis*[1–5]. The question arises whether sequences for the ancestors of these organisms, represented by X, Y, and Z in Fig. 1, can be reconstructed to any degree of confidence; were this possible, statistical analysis of mutation types would become feasible. Such reconstructions are well known in protein studies[6] and for tRNAs[7], but they are somewhat more difficult for 5S rRNA.

The main problem is to align the various sequences so that bases in corresponding position in different sequences are fairly certain to reflect a common term in the ancestral sequence. In protein studies, the fact that sequences are based on a twenty-symbol alphabet makes it unlikely that identical di- or tri-peptides in different sequences in approximately the same part of the molecule could have arisen by accident, and this is of great help in constructing alignments. With tRNA, the fixed position of certain modified bases in many sequences, plus base-pairing constraints in parts of the molecule, similarly ease the solution of the alignment problem. No such clues are available with 5S RNA. Alignment problems are primarily due to the confusing effects of base insertion or deletion mutations, which necessitate the use of gaps in alignment schemes. Since it is seldom clear in which sequences and between which positions gaps should be placed, the task of evaluating possible alignments becomes a combinatorial problem of great complexity. This problem has recently been solved[8] for the case where the evaluation is based on a minimal mutation criterion. The alignment algorithm developed for this solution is computationally impractical for large numbers of sequences, but for the case of three known sequences *a*, *b*, and *c*, and one unknown sequence *x*, in the configuration it is quite easy to implement a computer program for the method. (For example *a* = *E. coli*, *b* = *P. fluorescens*, *c* = *yeast*, *x* = unknown proto-prokaryote).



The algorithm is based on a recursion relationship which calculates, for three sequences, $a_1, a_2, \ldots, a_m$; $b_1, \ldots, b_n$; and $c_1, \ldots, c_p$, the minimal total number of mutations along

branches ax, bx, and cx necessary to relate the observed sequences to any hypothetical or reconstructed sequence x. Once this total is known, a minimising sequence $x_1, \ldots, x_q$ can be constructed.

Weighting insertions, deletions, and replacements of single bases equally, and letting $d(i, j, k)$ be the minimal total number of mutations required for the subsequences $a_1, \ldots, a_l$; $b_1, \ldots, b_j$; and $c_1, \ldots, c_k$, we have

$$d(i,j,k)=\min\begin{cases} d(i\text{–}1, j\text{–}1, k\text{–}1)\begin{cases}+0 \text{ if } a_l=b_j=c_k\\ +1 \text{ if } a_l=b_j\neq c_k,\ a_l=c_k\neq b_j \text{ or } b_j=c_k\neq a_l\\ +2 \text{ if } a_l\neq b_j\neq c_k\neq a_l\end{cases}\\ d(i\text{–}1, j\text{–}1, k)\begin{cases}+1 \text{ if } a_l=b_j\\ +2 \text{ if } a_l\neq b_j\end{cases}\\ d(i\text{–}1, j, k\text{–}1)\begin{cases}+1 \text{ if } a_l=c_k\\ +2 \text{ if } a_l\neq c_k\end{cases}\\ d(i, j\text{–}1, k\text{–}1)\begin{cases}+1 \text{ if } b_j=c_k\\ +2 \text{ if } b_j\neq c_k\end{cases}\\ d(i, j, k\text{–}1)\ +1\\ d(i, j\text{–}1, k)\ +1\\ d(i\text{–}1, j, k)\ +1\end{cases}$$

for $i=1, \ldots, m$; $j=1, \ldots, n$; and $k=1, \ldots, p$, where $d(r, 0, 0)=d(0, r, 0)=d(0, 0, r)=r$ for $r=1, 2, \ldots$ Once $d(m, n, p)$ has been calculated by systematically computing $d(1, 1, 1)$, $d(1, 1, 2)$, $d(1, 2, 1)$, . . . , an optimal alignment can be constructed by reversing the calculation, that is, by seeing which of a_i, b_j and/or c_k contributed to the minimal choice in calculating each $d(i, j, k)$, starting with $d(m, n, p)$. These procedures are described in more detail and proofs of optimality given elsewhere[8].

After finding the alignment, the unknown sequence can be largely reconstructed as follows. If $a_i=b_j=c_k$ are aligned, the three of these must be aligned with an x_h, where $x_h=a_i=b_j=c_k$. If $a_i=b_j$ or $b_j=c_k$ or $c_k=a_i$ (or two gaps) are aligned, then they are also aligned with an identical x_h (or a gap). All other cases lead to uncertainty in the x sequence (for example, if $a_i\neq b_j\neq c_k\neq a_l$ are aligned).

We used this procedure to establish a preliminary reconstruction for X in Fig. 1, by setting $a=$ *E. coli*, $b=$ *P. fluorescens*, $c=$ yeast, and $x=$ X. Similarly, a tentative sequence for Y was constructed by setting $a=$ KB, $b=$ *Xenopus*, $c=$ yeast, and $x=$ Y. In both cases, the uncertain positions were filled in by means of a random choice among the three different aligned bases (or a gap) at that position. An approximation to Z could then be derived by using $a=$ X, $b=$ Y, $c=$ yeast, and $x=$ Z.

A better version of X (according to the minimal mutation criterion over the whole tree of Fig. 1) could then be produced by setting $a=$ *E. coli*, $b=$ *P. fluorescens*, $c=$ Z and $x=$ X. Similar attempts to improve Y and Z produced no change. The five known sequences and the three reconstructed sequences are aligned in Fig. 2.

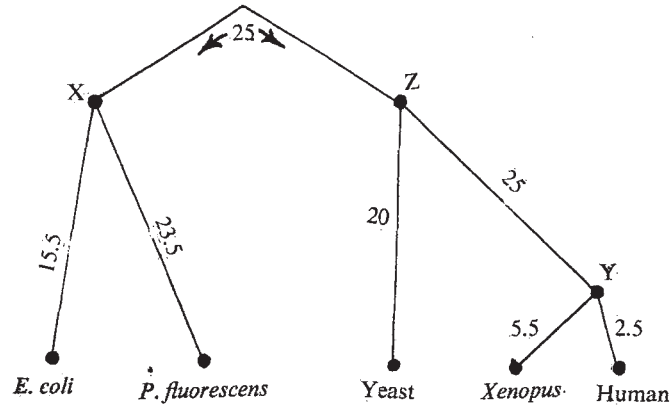


Fig. 1 Phylogenetic tree of known and ancestral organisms. X, Proto-prokaryote; Y, proto-vertebrate; Z, proto-eukaryote. Numbers represent estimated numbers of mutations along each branch.

```
EC UGC CUGGCGGCCG UAGC GCGGUGGUCCCACCUGACC CCAUGCC
PP UG? CU GCGGCCA UAGCAGC?UUGGAA?CACCUGAUC CCAU?CC
PF UGUUCU UUGACGAGUAGUAGCAUUGGAA CACCUGAUC CCAUCCC
PE  G? CU GCGGCCA UACCA?C UUGAAAGCACC?GAUCUCCGU CC
YT  GG UU GCGGCCA UACCAUC UAGAAAGCACC GUUCUCCGU CC
KB  GU CU ACGGCCA UACCACC CUGAACGCGCCCGAUCU CGU CU
PV  G? CU ACGGCCA UACCACC CUGAAAGCGCCCGAUCU CGU CU
XL  GC CU ACGGCCA CACCACC CUGAAAGUGCCCGAUCU CGU CU
                 |         |         |         |
                10        20        30        40

EC GA ACUCAGAAGUGAAACGC C GU AGCGCC GA UG GUA GUGU
PP GA ACUCAGAAGUGAAACGC U GU AGCGCC GA UG GUA GUGU
PF GA ACUCAGAGGUGAAACGA U GC AUCGCC GA UG GUA GUGU
PE GAUACUC?GAAGU UAA GC UGGU AG?GCCUGA UGAGUA GUGU
YT GAUAACCUGUAGU UAA GC UGGUAAGAGCCUGACCGAGUA GUGU
KB GAU CUCGGAAGC UAA GCAGGGU CGGGCCUGG UUAGUACUUGG
PV GAU CUCGGAAGC UAA GCAGGGU CGGGCCUGG UUAGUACUUGG
XL GAU CUCGGAAGC CAA GCAGGGU CGGGCCUGG UUAGUACUUGG
          |         |         |         |         |
         50        60        70        80        90

EC   GGGGUCUCCCCAUG CGAG A GUAGGGAAC U GCCAGGCAU
PP   GGGGU?UCCCCAUG CGAG A?CUA GGA?C U G CAGGCAU
PF   GGGGUUUCCCCAUGUCAAG AUCU  CGACCAUAG  A GCAU
PE A UGGGUGACC?CAUG CGAA ACCUA GGUGC U G CAGGC?U
YT AGUGGGUGACCAUACG CGAA ACCUA GGUGC U G CA AUCU
KB A UGGGAGACCGCCUG GGAAUACC G GGUGC U G UAGGCUU
PV A UGGGAGACCGCCUG GGAAUACC A GGUGC U G UAGGCUU
XL A UGGGAGACCGCCUG GGAAUACC A GGUGU C G UAGGCUU
          |         |         |         |
        100       110       120       130
```

Fig. 2 Alignment of known and reconstructed 5S RNA sequences. From top to bottom: EC, *E. coli*; PP, proto-prokaryote; PF, *P. fluorescens*; PE, proto-eukaryote; YT, yeast; KB, human KB carcinoma; PV, proto-vertebrate; XL, *Xenopus*. Reconstructed sequences are shaded; ?, undetermined base or gap.

Once the alignments and reconstructed ancestral sequences are available, it becomes possible to infer which mutations have occurred along each of the branches X→*E. coli*, X→*P. fluorescens*, Z→yeast, Z→Y, Y→KB, Y→*Xenopus*, and X→Z, although for the latter there is no way of knowing whether X or Z retains the ancestral base. Total mutations along each branch are indicated in Fig. 1. In Table 1, the 117 mutations reconstructed by our method are broken up into deletions, insertions and base replacements of various types. This shows that each transition type of mutation (pyrimidine-pyrimidine or purine-purine replacement) occurs more frequently than any transversion (pyr-pur or pur-pyr). In additon, transitions involving C and U are by far the most frequent type of mutation.

Through a systematic study of protein sequences, Vogel[9] and others have found that the evolution of the DNA coding for these proteins similarly shows evidence for a predominance of transitions, particularly pyr-pyr mutations. The frequency of fixed mutations from these and our own studies is thus: pyr-pyr > pur-pur > pur-pyr, pyr-pur. In addition, the transitions > transversions effect is also observable in tRNA data[10].

It could be argued that the protein data are explicable by some selective force controlling the type of amino acid substitution permitted. The resulting non-random protein mutations could then be related by the genetic code to non-random DNA mutations. One might also argue that the RNA data are the result of selection based on either function or three dimensional structure. That one finds identical patterns of non-random distributions in both protein and RNA data, however, indicates that the process responsible for this is independent of the particular type of molecule. Moreover, if selective forces are involved, this could only be at the level of DNA replication and/or RNA transcription. The basis of this selection could be structural, that is, the excess of transitions could imply that mispairing is permitted more often between a purine and a

Table 1 Estimated Number of Each Mutation Type Occurring in 5S RNA

Original base	New base A	C	G	U	None
A		5.0	8.0	6.3	3.3
C	3.7		6.2	11.3	4.3
G	7.3	4.0		6.3	4.0
U	3.0	14.0	4.2		3.0
None	6.3	4.3	6.0	6.3	

Bottom row represents insertions; extreme right column lists deletions. Transition mutation types underlined.

pyrimidine. This proposal is in agreement with the tautomerisation theory of mutations, in which G (enol form) pairs with T, and C (imino form) pairs with A. Another possibility is that known DNA repair enzymes correct pur-pur and pyr-pyr mismatches more often since they disrupt the DNA duplex more than pur-pyr mismatches. Neither tautomerisation nor selective repair, however, can explain the fact that pyr-pyr mutations are fixed more often that pur-pur.

Figure 2 suggests another general tendency in the evolutionary history of 5S RNA; that insertions and deletions have involved mainly single nucleotides. This contrasts with tRNA[7], and protein data[12], where simultaneous insertion or simultaneous deletion of several consecutive nucleotides has been shown.

This work was supported in part by le Ministère de l'Education du Québec.

DAVID SANKOFF
CRISTIANE MOREL
ROBERT J. CEDERGREN

Centre de Recherches Mathématiques and
Département de Biochimie,
Université de Montréal,
Case Postale 6128, *Montréal* 101

Received June 21, 1973.

1 Brownlee, G. G., Sanger, F., and Barrell, B. G., *J. molec. Biol.*, **34**, 379 (1968).
2 DuBuy, B., and Weissman, S. M., *J. biol. Chem.*, **246**, 747 (1971).
3 Brownlee, G. G., Cartwright, E., McShane, T., and Williamson, R., *FEBS Lett.*, **25**, 8 (1972).
4 Forget, B. G., and Weissman, S. M., *Science, N.Y.*, **158**, 1695 (1967).
5 Hindley, J., and Page, S. M., *FEBS Lett.*, **26**, 157 (1972).
6 Dayhoff, M. O., *Scient. Am.*, **221**, 86 (1969).
7 Cedergren, R. J., Cordeau, J. R., and Robillard, P., *J. theor. Biol.*, **37**, 209 (1972).
8 Sankoff, D., *Publ. Cent. Rech. Math., Technical report 262* (1973).
9 Vogel, F., *J. mol. Evol.*, **1**, 334 (1972).
10 Dayhoff, M. O., in *Atlas of Protein Sequence and Structure*, **5**, 113 (Nat. Biomed. Res. Found., Washington, DC, 1972).
11 Watson, J. D., and Crick, F. H. C., *Cold Spring Harb. Symp. quant. Biol.*, **18**, 193 (1953).
12 Dayhoff, M. O., in *Atlas of Protein Sequence and Structure*, **5**, 44 (Nat. Biomed. Res. Found., Washington, DC, 1972).

RNA in Nascent DNA from Cultured Human Lymphocytes

RECENT studies in microorganisms indicate the presence of RNA in nascent DNA and point to a role of the RNA as primer for DNA synthesis[1–8]. We report here evidence that RNA is also present in human cells in short chain intermediates of DNA replication, apparently serving a primer function as in microorganisms.

Early intermediates in lymphocyte DNA replication were isolated by their single-stranded properties as previously demonstrated for HeLa cells[9,10], Chinese hamster fibroblasts[11], Ehrlich ascites cells[12] and regenerating rat liver[13]. The single-stranded (nascent) fraction of pulse-labelled DNA, separated on hydroxylapatite[14], after very short labelling times contains a disproportionately large amount of the incorporated label (Table 1). This reaches a maximum by 60 s at 37° C whereas label in the double-stranded (bulk) DNA fraction increases progressively with time.

A second method used to separate nascent DNA involves simple ultracentrifugation of cell lysates to remove bulk DNA from low molecular weight single-stranded replicative intermediates, which remain in the supernatant (Table 2). Label in both the hydroxylapatite single-stranded fractions and the supernatant from ultracentrifugation can be chased into the bulk DNA, thus indicating their precursor nature (Tables 1 and 2).

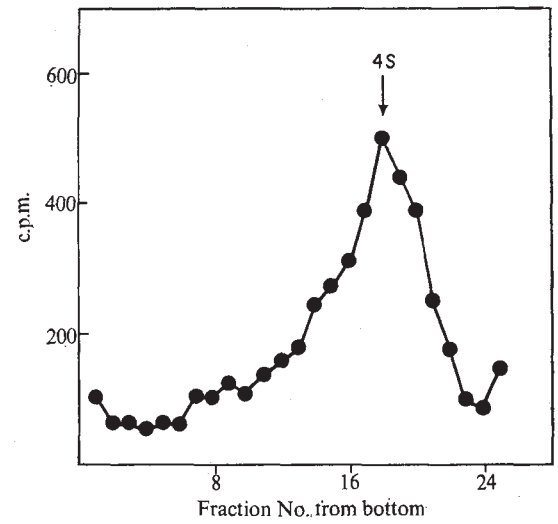


Fig. 1 Sedimentation analysis of ^{3}H-thymidine labelled DNA from cultured human lymphocytes isolated by hydroxylapatite chromatography. Human diploid lymphocytes (line "8866", established by Dr George Moore and given to us by Dr Richard Lerner) were maintained in the exponential phase of growth (2 to 8×10^5 cells ml^{-1}) in suspension culture at 37° C with modified Eagle's medium. Cells (4×10^7 ml^{-1}) suspended in 60 mM Tris-HCl, *p*H 7.4, 60 mM NaCl, 11 mM glucose, 2 mM $MgCl_2$ were incubated 5 min at 27° C, methyl ^{3}H-thymidine (Schwarz-Mann, 55 Ci mmol^{-1}) was added to a final concentration of 200 μCi ml^{-1} and incubation continued for 2 min. The incubation was terminated by addition of 4 volumes of 0.5% SDS, 0.05 M sodium phosphate, *p*H 6.8, 5 mM EDTA. The lysate (2×10^8 cells) was sheared by serial passage through 18 and 25 gauge needles, carrier DNA (100 μg) was added and the mixture was applied to 7 cm$^2\times10$ cm hydroxylapatite (Bio-Rad) column. A wash of 250 ml 0.05 M sodium phosphate, *p*H 6.8, was followed by an 800 ml linear gradient of 0.1 to 0.35 M sodium phosphate, *p*H 6.8, eluting denatured DNA at 0.19 M (36,000 c.p.m.) and native DNA at 0.26 M (120,000 c.p.m.). Pooled fractions of single-stranded nascent DNA were recovered by rotary evaporation and salts removed by passage over Sephadex G-50. Centrifugation was in neutral 5–20% sucrose gradients (1 M NaCl, 10 mM Tris-HCl, *p*H 7.5, 1 mM EDTA) for 6 h at 56,000 r.p.m., 20° C, in the Spinco SW 56 rotor. Marker *E. coli* tRNA was centrifuged in a separate tube. Fractions were collected from below, precipitated with acid, filtered on glass fibre disks, and counted in a liquid scintillation spectrometer.

Preliminary characterisation on neutral sucrose gradients of nascent DNA isolated by hydroxylapatite shows a broad size distribution with a peak at approximately 4S (Fig. 1). Nascent DNA isolated by the second method, ultracentrifugation of whole lysates, has a similar sucrose gradient pattern and also behaves as single-stranded DNA on hydroxylapatite (data not shown). Various sizes, including 4S, have been described in studies on mammalian cells[11,13,15–21].

The nascent single-stranded DNA from hydroxylapatite