24. H. E. Zimmerman, J. D. Robbins, R. D. McKelvey, C. J. Samuel, L. R. Sousa, *ibid.* **96**, 1974 and 4630 (1974).
25. H. E. Zimmerman, P. Baeckstrom, T. Johnson, D. W. Kurtz, *ibid.* **94**, 5504 (1972); *ibid.* **96**, 1459 (1974).
26. H. E. Zimmerman and J. A. Pincock, *ibid.* **94**, 6208 (1972); *ibid.* **95**, 2957 (1973).
27. H. E. Zimmerman, K. S. Kamm, D. P. Werthemann, *ibid.* **96**, 7821 (1974); *ibid.* **97**, 3718 (1975); H. E. Zimmerman and G. A. Epling, *ibid.* **92**, 1411 (1970); *ibid.* **94**, 8749 (1972); H. E. Zimmerman and G. E. Samuelson, *ibid.* **89**, 5971 (1967); *ibid.* **91**, 5307 (1969).
28. W. R. Ware, in *Creation and Detection of the Excited State*, A. A. Lamola, Ed. (Dekker, New York, 1971), vol. 1, part A; L. M. Bollinger and G. E. Thomas, *Rev. Sci. Instrum.* **32**, 1044 (1961); G. Laustriat, A. Choche, H. Lami, G. Pfeffer, *C. R. Acad. Sci.* **257**, 434 (1963); W. R. Ware, L. J. Doemeny, T. L. Nemzek, *J. Phys. Chem.* **77**, 2038 (1973).
29. A. E. W. Knight and B. K. Selinger, *Spectrochim. Acta Part A* **27**, 1223 (1971).
30. H. E. Zimmerman, D. P. Werthemann, K. S. Kamm, *J. Am. Chem. Soc.* **95**, 5094 (1973); *ibid.* **96**, 439 (1974).
31. H. E. Zimmerman and T. P. Cutler, *Chem. Commun.* (1975), p. 598.
32. H. E. Zimmerman and L. M. Tolbert, *J. Am. Chem. Soc.* **97**, 5497 (1975).
33. H. E. Zimmerman and C. J. Samuel, *ibid.*, pp. 448 and 4025.
34. H. E. Zimmerman and G. E. Keck, *ibid.*, p. 3527.
35. K. R. Kopecky and C. Mumford, *Can. J. Chem.* **47**, 709 (1972).
36. E. White, J. Wiecko, D. R. Roswell, *J. Am. Chem. Soc.* **92**, 2167 (1970).
37. T. R. Darling and C. S. Foote, *ibid.* **96**, 1625 (1974).
38. W. H. Richardson, F. C. Montgomery, M. B. Yelvington, *ibid.* **94**, 9277 (1974).
39. D. R. Kearns, *Chem. Rev.* **71**, 345 (1971).
40. E. M. Evleth and G. Feler, *Chem. Phys. Lett.* **22**, 499 (1973); D. R. Roberts, *J. Chem. Soc. Chem. Commun.* (1974), p. 683; J. J. S. Dewar and S. Kirschner, *J. Am. Chem. Soc.* **96**, 7578 (1974); N. J. Turro and A. Devaquet, *ibid.* **97**, 3859 (1975).
41. H. E. Zimmerman, R. W. Binkley, J. J. McCullough, G. A. Zimmerman, *J. Am. Chem. Soc.* **89**, 6589 (1967).
42. H. E. Zimmerman and A. A. Baum, *ibid.* **93**, 3646 (1971).
43. H. E. Zimmerman, R. J. Boettcher, W. Braig, *ibid.* **95**, 2155 (1973).
44. Supported by NSF, NIH grant GMO7487, and the U.S. Army Research Office.

Evolution of Repeated DNA Sequences by Unequal Crossover

DNA whose sequence is not maintained by selection will develop periodicities as a result of random crossover.

George P. Smith

A considerable portion of the DNA of some eukaryotes consists of sequences repeated very large numbers of times (*1*). These highly repetitious DNA's are often called satellites. The repeated unit is relatively homogeneous within each species, but major differences are observed between related repetitious DNA's in different species, even of the same genus (*2, 3*). Some satellites have been shown to consist of short, relatively homogeneous tandem repeats; the repeats in different satellites ranged in length from 2 to about 12 base pairs (*2, 4*). Other repetitious DNA's are more complex. For instance, partial sequence analysis of guinea pig alpha satellite (*5*) and mouse satellite (*6*) shows that neither is composed of a single very short repeating sequence. Instead, these DNA's appear to contain subrepeats of homologous but not identical sequences within a larger repeating unit.

Botchan (*7*) and Southern (*8*), among others, have investigated long-range periodicities in the more complex repetitious DNA's by digesting them with restriction enzymes, which cleave DNA at particular base pair sequences. This approach is illustrated by the work of Southern (*8*) on digestion of mouse satellite DNA with the restriction enzyme Eco RII. The major products are fragments whose lengths are small integral multiples of about 240 base pairs. Thus there seems to be an approximately 240 base pair periodicity in this DNA, with some of the repeats missing the Eco RII site, so that some fragments of higher multiples of 240 base pairs are released. In addition to these major fragments, small amounts of "fractional" fragments with lengths equal to 0.5, 1.5, 2.5, . . . times 240 base pairs are also obtained. The 120 base pair and 360 base pair fractional fragments are released in roughly equimolar amounts. Southern points out that these equimolar yields make it very unlikely that the fractional fragments arise exclusively by straightforward mutation somewhere near the middle of the 240 base pair repeat to produce new Eco RII sites; for in that case most such mutations would result in two 120 base pair restriction fragments, and very few would result in 360 base pair fragments. I will discuss the origin of fractional fragments later. When purified 240 base pair fragments were denatured and allowed to reassociate, a large proportion of the reassociated DNA was in high molecular weight complexes formed by reassociation of the complementary strands in a staggered register. This shows that the 240 base pair unit is composed of subrepeats, thus confirming the indirect conclusion from sequence analysis (*6*).

The author is assistant professor in the Division of Biological Sciences, University of Missouri, Columbia 65201.

Role of Unequal Crossover

I will argue in this article that repetitious DNA's with these characteristics will arise and evolve naturally as a result of random unequal crossover between sister chromatids—that is, between the two daughters produced by replication of a single DNA molecule. These unequal crossovers, which must occur in the germ line to be evolutionarily significant, might happen either at meiosis or at any one of the many germ line mitoses.

Repetitious DNA's might arise and evolve by many different mechanisms. I have singled out unequal crossover because there is good evidence that it actually occurs. Sister chromatid crossovers, which might be either equal or unequal, have been demonstrated to occur at a rate of several exchanges per cell per division in a variety of eukaryotic cells (*9, 10*). In many of these studies, exchange was detected with the aid of bromodeoxyuridine or [^{3}H]thymidine, which can artificially induce crossovers. Nevertheless it is very likely that there is an appreciable rate of crossover even in the absence of artificial induction, since exchanges occur at roughly comparable rates in ring chromosomes, where they can be detected without artificial agents by virtue of producing dicentric rings (*10*). I know of no direct evidence for unequal sister chromatid crossover. However, unequal nonsister chromatid crossover has been well known since the work of Bridges and Sturtevant on the *bar* locus of *Drosophila* (*11*), and there is indirect evidence for unequal sister chromatid exchange at the *bar* (*12*) and ribosomal RNA (*bobbed*) (*13*) loci of the same organism. I think that this evidence, taken together, strongly suggests that unequal sis-

ter chromatid crossover occurs at more or less the rate I will assume in this article. Hence, I am not advancing a speculative hypothesis in which a largely unprecedented process is invoked specially in order to explain the origin and evolution of repetitious DNA's. Rather, I argue that these phenomena are probable consequences of a process that is thought on completely different and rather convincing grounds to happen in chromosomes.

Crossover can lead to many complex patterns in the resulting two DNA molecules, and thus in the four recombinant molecules into which these two molecules segregate at the next replication (*14*). Figure 1 illustrates the process as I will assume it to occur and the terms I will use to describe it. As shown in Fig. 1, each of the four recombinant molecules resulting from a crossover (after segregation at the next replication) is assumed to be equivalent to one of those that would have been produced if the two daughter molecules taking part in the crossover had been aligned in a staggered fashion, cleaved at bonds that are aligned with each other, and rejoined crosswise. I will call the position of the cleaved bonds the points of crossover. I will also write of crossover as occurring "between" two points or regions in the parent molecule: the points referred to are those equivalent to the points of crossover in the daughter molecules, and the regions are the aligned sequences adjacent to those points. As can be seen in Fig. 1, each of the four recombinant molecules resulting (after replication) from a crossover harbors either a deletion or a tandem duplication of the stretch of base pairs lying between the points of crossover in the parent molecule. An unequal crossover can thus be thought of as deleting or tandemly duplicating a region of the DNA sequence, depending on which of the four recombinant molecules figures in the evolutionary lineage being discussed (*15*). Crossover is assumed to be initiated by local base-pairing between antiparallel strands from the two participating DNA molecules; this base-pairing would require at least a minimal degree of complementarity between the single strands, and consequently unequal crossover ought only to occur between sequences with at least a minimal degree of homology.

To understand how random unequal crossover generates repetitious DNA's, imagine that there is in the genome of some species a segment of DNA whose sequence is not maintained by natural selection. Other forms of selection may or may not be operative; for instance, the total length of the segment may or may not be maintained within certain limits. This segment may contain, and will continuously acquire as a result of mutation, short chance regions of homology at different points in the sequence, such as those indicated in Fig. 1. Random unequal crossover (either sister chromatid or nonsister chromatid) between these regions will generate variant segments in which sequences deriving from various parts of the original segment have been deleted or tandemly duplicated. Subsequent crossovers between tandem repeats aligned in register will either increase or decrease the number of

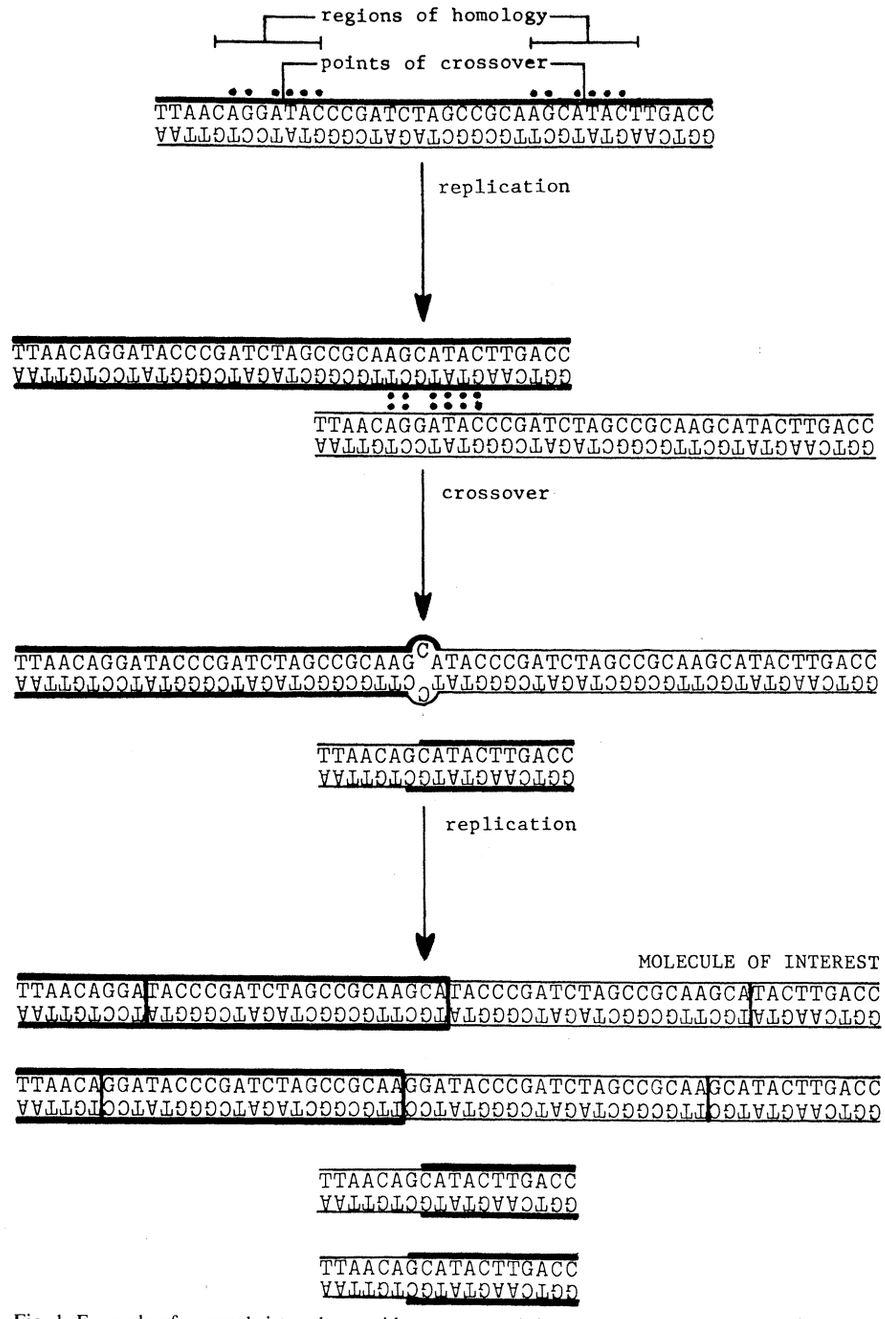

Fig. 1. Example of unequal sister chromatid crossover as it is assumed to occur in this article. The DNA single strands which derive from the upper (right side up) strand of the starting molecule are indicated by heavy lines, while those which derive from the lower (upside down) strand are indicated by light lines. The two molecules produced by the crossover are shown with heteroduplex regions, in which one strand derives from one of the parental molecules and the other strand from the other; mismatched base pairs, such as that shown in one of the recombinant molecules here, are possible in these heteroduplex regions. Despite this complexity of the two molecules produced by crossover, each of the four molecules resulting from subsequent replication has the pattern expected if it were produced by a single breakage and reunion at single points of crossover in the recombining molecules. The points of crossover corresponding to the final molecule of interest are indicated in the starting sequence; those corresponding to the other three final molecules would be different. The two longer final molecules carry tandem duplications of that part of the starting sequence lying between their respective points of crossover; the duplicated sequences are delineated by vertical bars.

tandem repeats; in these cases, I will write that the tandem array has been expanded or contracted, respectively.

The deletions and duplications I have just described would radically disrupt the sequence of the original segment and therefore would be rapidly eliminated from the population if selection were maintaining that sequence. Even in the absence of such selection, random genetic drift will result in the elimination of most deletions and duplications in a few generations (*16*). But random drift will also result in the fixation of an occasional deletion or duplication—that is, in the stochastic increase in frequency of chromosome segments harboring that deletion or duplication, until those chromosome segments supplant all homologous segments in the gene pool (*16*). The continual fixation of deletions and duplications will result in the continual accumulation of deletions and duplications in the chromosomes present in the organismal population. In the absence of selection for or against deletions and duplications, their rate of accumulation per chromosome per year will approximately equal their rate of occurrence per chromosome per year, and will not depend on other factors such as the size of the organismal population (*16*).

The accumulation of deletions will gradually eliminate from the population the descendant sequences of more and more of the original segment. Indeed, assuming that all sequences descending from the original segment do not disappear altogether, they will eventually be derived from a single base pair in some ancestral DNA molecule. Counteracting this loss of sequences descending from some parts of the original segment will be the accumulation of duplications of other parts of the original segment. These duplications will include expansions of tandem arrays, which will eventually give rise by chance to relatively long tandem arrays. Because long tandem arrays will harbor extensive regions of homology in each of many different alignments, crossover will occur much more frequently between the repeats in such arrays than between short chance regions of homology. Consequently, duplications will come increasingly to take the form of expansion of established arrays of tandem repeats. This process will continue until the whole segment consists of a single tandem array of repeats, unless mutations accumulate so frequently relative to crossovers that they obscure the similarities between repeats before the process can be completed.

Once a DNA segment has become repetitious, deletions and duplications will predominantly delete or duplicate integral numbers of repeats. As a result, very large numbers of deletions and duplications can accumulate without changing the fundamental repeat pattern. It has often been pointed out that repeats within an array will tend to remain homogeneous under these circumstances, even as mutations and other changes accumulate in the repeated sequence (*17, 18*). The reason is that any mutation (or other change) which arises in the array will either be eliminated from the population by the accumulation of deletions or be spread through the entire array in all the chromosomes in the population by the accumulation of duplications, all without changing the fundamental repeat pattern. In the latter case, I will write that the mutation (or other change) has undergone crossover fixation; and the average time interval between the occurrence of a mutation (or other change) destined to become fixed and its ultimate fixation I will call the crossover fixation time. If crossover is sufficiently frequent relative to mutation, the elimination and crossover fixation of new mutations (or other changes) will be so rapid that the repeats will not often be appreciably polymorphic for an appreciable number of mutations (or other changes) at any one time, and thus will remain quite homogeneous (*19*).

From these considerations I conclude that if unequal crossover is sufficiently frequent relative to mutation, and if crossover between repeats in long tandem arrays is sufficiently more probable than crossover between short chance regions of homology, any segment of DNA whose sequence is not maintained by natural selection will become and remain repetitious in the course of evolution.

Computer Simulation of Unequal Crossover

In order to illustrate the foregoing principles, I simulated random unequal crossover on a computer. It was necessary in these simulations to simplify the evolutionary model by assuming that nonsister chromatid crossover—which I imagine happening much less frequently than sister chromatid exchange (*20*)—does not happen at all. In this simplified model, once single strands have segregated, the DNA sequences that descend from them can never remix. Consequently, there will be only

```
RANDOM STARTING SEQUENCE

3033123233002101222000202320020003210101320302000331012321031122330203112020233301301321020100103200

0222320012210010111132312310323323023131312311103221330112133121120011210321320013211333002220200011

1321201023130323123322110313003102223312303003332131101131130031112331002003103020120310101300202223

3312223011201000232130033231133211211320223003322023010331002013310111221322111110003031221310222332

2332333223200001032212230322330133022022302133303100313001033032221231130231313231220122032012121213


FINAL SEQUENCE AFTER 200 CYCLES

2033310322103221032210322103221032210322103221032210322003221032210322103221032200322103221032210322

1032210322103221032200322103221032210322103221032210322103221032210322103221032210322103221032210322

1032210322003221032210322003221032210322103221032210322103221032210322103221032210322103221032200322

1032210322103221032210322003221032210322103220032210322103221032210322103221032200322103221032210322

103221032210322103221032210322103221032210322103222003012121 3
```

Fig. 2. Random starting sequence and final sequence after 200 cycles in one of the simulations summarized in Table 1 (fourth row). Each of the digits from 0 to 3 stands for a different one of the four possible base pairs. There is a small amount of anomalous sequence at each end of the array. This is because anomalous terminal sequences can only be lost by crossover at chance regions of homology within the terminal sequence; as the terminal sequences become small, such a crossover becomes highly improbable and the anomalous sequence persists.

a single evolutionary lineage leading to a given descendant DNA molecule. At any one time, of course, many DNA lineages for a given segment of the genome will coexist in the same species. From a long-term point of view, however, only one of these lineages will be important, for eventually all but one of them will become extinct, either because of random genetic drift or because of natural selection. For the purposes of the simulations, therefore, I could consider evolution as proceeding along a single DNA lineage, and each crossover in this lineage as producing a single recombinant molecule carrying either a deletion or a tandem duplication of some portion of the contemporary sequence. If I had not made this simplification, the program would have had to keep track of innumerable lineages simultaneously because of the possibility that any two of them could remix by nonsister chromatid exchange. It is most unlikely that my conclusions would have been altered significantly had it been possible somehow to dispense with this simplification, since the general arguments by which I justify them, both in the previous section and elsewhere in this article, apply equally to any form of unequal crossover.

In the simulations a DNA sequence, which at the outset was composed of 500 base pairs, was subjected to various numbers of evolutionary "cycles." Each cycle consisted of two steps. The first step was the introduction of a random base pair replacement at a single position of the contemporary sequence. The second step was a certain number of "attempted" unequal crossovers between two exact copies of the contemporary sequence. Each attempted crossover was generated by two random numbers. The first random number specified the alignment of the two copies of the sequence; all alignments for which the product of the crossover would be at least 450 base pairs but no more than 550 base pairs were equally likely. A second random number then specified the exact points of crossover. These could be at any position, starting from the point to the left of the leftmost of the overlapping base pairs, and extending to the point to the left of the rightmost m overlapping base pairs. In order to make crossover dependent on homology I introduced a certain criterion of homology which an attempted crossover had to meet in order to be actually executed. This requirement was that the m base pairs immediately to the right of the prospective points of crossover had to be identical in the two aligned copies of the sequence. The criterion m was an adjustable parameter.

I chose this criterion of local homology for ease of computation. There is no reason to suppose that it closely reflects the mode in which natural crossover depends on local homology; but neither is there the slightest reason to suppose that in comparison to other possible modes of dependence it is particularly favorable to the establishment or evolution of tandem repeats. It seems safe to extend my results qualitatively to any model of unequal crossover in which crossover frequency depends in some way on local homology.

Table 1. Summary of simulations starting with 500 base pair random sequence. Each cycle consisted of one random mutation and 500 attempted crossovers. For the attempted crossover to be executed, four residues to the immediate right of the prospective crossover points had to be identical.

Cycles	Crossovers executed	Predominant repeat length
100	924	39
100	1118	12
200	3057	15,16
200	5444	5
200	1916	29
200	3358	16
200	4069	11
200	3244	14,10
200	2503	18
200	8059	7
200	2232	21,37
200	4206	11
200	1700	29
200	1836	31
200	4524	12
200	4933	11
200	2410	17
200	1661	Heterogeneous
80	546	Heterogeneous

Generation of Repeats from Nonrepetitious DNA

Table 1 summarizes the results of several simulations in which the 500 base pair starting sequence was random. In these simulations there was one mutation and 500 attempted crossovers per cycle, and four residues were required to match for a prospective crossover to be executed (m = 4). Periodicities developed in all these simulations. Figure 2 shows the starting random sequence and the final sequence for one of these simulations; the 5 base pair periodicity in the latter is striking.

The demonstration that tandem repeats can develop readily from a nonrepetitious sequence supports my contention that random unequal crossover accounts for the origin of repetitious DNA's. It also suggests that these DNA's will have a strong tendency to remain repetitious. For even if a repetitious sequence occasionally becomes grossly heterogeneous despite the homogenizing effect of crossover fixation, it can readily redevelop repeats. For these reasons, I feel that repetitiousness is the usual state of DNA whose sequence is not maintained by natural selection.

In my simulations the total length of the sequence was kept within 10 percent of the starting length. These limits were imposed for convenience of calculation, and I do not mean to imply that very long tandem arrays necessarily arise from equally long nonrepetitious sequences. Indeed, I think it more likely that they evolve by expansion of small arrays arising within short nonrepetitious regions.

Evolutionary Instability of Long Repeats

Long repeats would probably be rather unstable in size. The reason is that mutation would be expected to generate chance points of homology within the repeat. An occasional crossover between such points will give rise to an abnormal sequence, which can be represented ... $ABABAAB$ ABA ..., where A and B represent different sequences (not necessarily the same size) that together constitute the parental repeat AB. The abnormal sequence contains a tandem array consisting of two repeats of the sequence A, which is shorter than the parental repeat AB. The tandem array AA can potentially be expanded to occupy the entire DNA segment, thus in effect decreasing the repeat length. Alternatively, the abnormal sequence produced by out-of-register crossover can be thought of as containing tandem repeats of an internally repetitious sequence that is longer than the parent repeat, the internally repetitious repeat being one of the sequences ABA, $ABABA$, and so forth. Such an array could also be expanded, thus in effect increasing the repeat length. This process is one way in which subrepeats of the sort present in mouse satellite could arise. An array of internally repetitious repeats would presumably be relatively unstable, however, because of the possibility of further out-of-register crossovers between homologous subrepeats. This instability could be resolved in two ways. On the one hand, the array will contain, or can readily develop by out-of-register crossover between homologous subrepeats, tandem repeats of the internally nonrepetitious sequences A or AB; expansion of these arrays can give rise to relatively stable arrays consisting of repeats that are less than or equal to the original parent repeat AB in length. On the other hand, an array of internally repetitious repeats might persist long enough for mutation to reduce the homology between subrepeats, thus reducing out-of-register crossover and stabilizing a repeat that is longer than the parent repeat. Thus crossover between chance points of

homology in a tandem array can ultimately lead to an increase or a decrease in repeat length, although it would seem more likely to lead to a decrease than to an increase.

The change of repeat length by out-of-register crossover between chance points of homology within the repeat is illustrated in the stages in the evolution of the repeat pattern shown in Fig. 2. The development of this pattern was followed at 20-cycle intervals, and the probable evolution of the final 5 base pair repeat from an original 22 base pair repeat could thereby be reconstructed. The scheme is shown in Fig. 3.

Short repeats would be much less likely than long repeats to develop internal points of homology and thus to shift in size in the manner described in this section. The reason is that the shorter the repeat the smaller the number of ways in which points of homology could develop within a single repeat. (Out-of-register points of homology could still theoretically develop in different repeats in the array, but this would require the coexistence in the same array of very different repeats despite the homogenizing effect of crossover fixation.) Because of this preferential occurrence of out-of-register crossovers in long repeats, and because, as mentioned above, the ultimate result of such a crossover is more likely to be a decrease than an increase in the repeat length, the process described in this section will, in the long run, tend to decrease the repeat length.

A special case of the evolutionary instability of long repeats is represented by the "spacer" portions of tandemly repeated functional genes. In several such tandem arrays, such as those for the 18*S* and 28*S* ribosomal RNA's and for the 5*S* ribosomal RNA's in *Xenopus* (*21*), each repeat contains an untranscribed segment, called a spacer. This spacer is somewhat variable within the species and highly variable between species. Although the spacer undoubtedly contains some sites (such as those for the binding of RNA polymerase) necessary for the functioning of the genes, its high variability suggests that much of its sequence is not maintained by natural selection. Under these circumstances, all the processes described in this section could occur, with the important proviso that natural selection would not permit extensive expansion of an array whose repeats do not contain intact functional regions. Consequently, the spacer would be expected to lengthen and shorten in evolution. For the reasons outlined in the previous paragraph, I would expect the length of the spacer to decrease in the long run in the absence of any selection against short spacers; conversely, the existence of a long spacer suggests that its length (but not necessarily its sequence) is being maintained by natural selection. Under these circumstances, any shortenings of the spacer would be more or less balanced by lengthenings. Moreover, lengthenings would be expected to take the form preferentially of expansions of arrays of internal subrepeats generated by out-of-register crossovers between chance points of homology. Thus long spacers ought often to be internally repetitious. Subrepeats have indeed been described within the spacers of the 5*S* repeats of *Xenopus laevis* by Brownlee *et al.* (*22*).

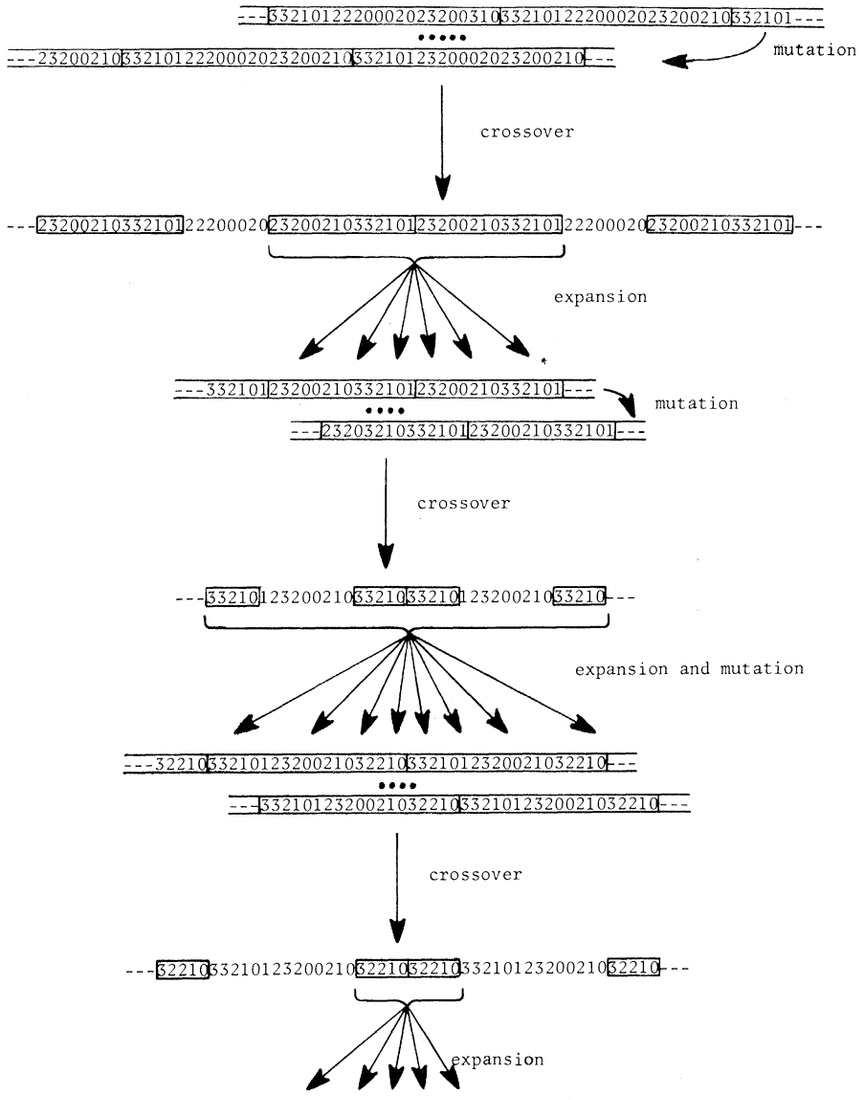

Fig. 3. Probable evolution of the final 5 base pair repeat pattern in Fig. 2. The proposed out-of-register crossovers occur at the short regions of homology indicated by dots. Each out-of-register crossover gives rise to a product that can be formulated . . . *ABABAABABA* . . ., where *A* and *B* represent distinct sequences; the sequence represented by *A* is boxed in the product of each of the three out-of-register crossovers here. Repeats in the molecules participating in the crossovers giving rise to those products are also boxed. Following each of the three crossovers shown, the two-repeat array indicated by the bracket—either *AA* or (in one case) *ABAABA*—was expanded by subsequent crossovers, as symbolized by multiple arrows emanating from the bracket. Only the last of these expansions proceeded to fill nearly the whole DNA segment. The expansions indicated here were accompanied by a process that might be called "dissemination." Crossover between points on either side of an array can duplicate the entire array to produce two separate arrays, thus potentially disseminating arrays throughout the DNA. Subsequent crossovers between different arrays will merge the arrays, resulting ultimately in a single continuous array, just as if expansion were not accompanied by dissemination.

Generation of Higher-Order Periodicities

Arrays of tandem repeats would be subject to instability of another sort suggested by Southern (*8*): the development of higher-order periodicities composed of an integral number of shorter sequences. This phenomenon is illustrated in several simulations in which the starting sequence, rather than being random, consisted entirely of a tandemly repeated 5 base pair

Table 2. Summary of simulations starting with 100 5 base pair repeats. Each cycle consisted of one random mutation and 100 attempted crossovers, in which the two copies of the sequence were aligned so that 5 base pair repeats were in register. For the attempted crossover to be executed, four or eight residues to the immediate right of the prospective crossover points had to be identical (m = 4 or 8). Each simulation comprised 100 cycles.

m	Number of simulations resulting in periodicities of indicated length					
	5	10	15	20	25	30
4	33	14	3			
8	3	5	3	2		1

sequence. In such a family, crossover in alignments in which the 5 base pair repeats are in register would greatly predominate over crossover in other alignments. In order to reduce computation time, therefore, I restricted the attempted crossovers to such alignments, reducing the number of attempted crossovers per cycle by a factor of 5 to keep the rate of crossover close to what it would have been if this restriction had not been introduced.

Table 2 shows the results of 64 100-cycle simulations. In the 50 simulations summarized in the upper row, the number of base pairs m that had to match in the aligned sequences in order for an attempted crossover to be executed was 4; in the 14 simulations summarized in the lower row, m was 8. Higher-order periodicities developed in a high proportion of these simulations. A typical resulting 10 base pair periodicity was the sequence

XACACACACACACAABABACACAB
ACACACABACBACABAABABACAC
ABACBAACABACABACBACACABA
CACABACBADACACACACACADA
CAABXX

where A represents the starting 5 base pair sequence 03212; B, C, and D represent, respectively, 03211, 23211, and 03311; and X represents one of three sequences that appear only once apiece. (In these number sequences, each of the digits from 0 to 3 represents one of the four possible base pairs.) The sequence is made up almost entirely of the 10 base pair units AB and AC. The generation of higher-order periodicities thus provides another way in which unequal crossover can generate a pattern of subrepeats within a larger repeating unit, as is seen in such repetitious DNA's as mouse satellite (*6*, *8*).

If we were to take the pentamer labeled A in the above sequence as constituting a restriction enzyme cleavage site, the pattern of fragments released would be as shown in the upper row of Table 3. The 10 base pair fragments predominate, but there are a few fragments of 0.5 and 1.5 times that length, released in about equimolar amounts. Such fractional fragments were a common but not invariable feature in my simulations. Thus unequal crossover provides a plausible explanation of the fractional fragments observed in mouse satellite (*8*), although in that case much longer fragments than 5 to 15 base pairs were involved. In one simulation, I followed at intervals of four cycles the pattern of fragments that would be released by cleavage at each occurrence of a certain 5 base pair sequence. Some representative patterns are shown in the lower four rows of Table 3. It can be seen that the proportion of fractional fragments fluctuated widely during the period in which a 10 base pair repeat pattern was developing from a 5 base pair repeat pattern. This fluctuation invalidates Southern's (*8*) estimate of the number of crossovers that have occurred in mouse satellite DNA, since his estimate assumes that the frequency of fractional repeats will increase in some constant proportion to the total number of crossovers.

The reason for the development of higher-order periodicities, as explained by Southern (*8*), is that the vagaries of random crossover and mutation will occasionally lead to an array whose repeats are appreciably heterogeneous. A crossover in such an array will lead to deletion or tandem duplication of a series of contiguous repeats. In the latter case, subsequent crossovers will occur preferentially between these tandemly duplicated series, for in such alignments the various heterogeneous repeats within the series are in register, whereas fewer identical repeats will, on the average, be in register in other alignments. As a result of this preferential crossover, there is a good chance that such an array of series—each series consisting of several of the original repeats—will expand and be maintained long enough for mutation to further reduce the homology between repeats within the series and thereby stabilize the higher-order periodicity.

Long periodicities generated in this manner, like those produced de novo from nonrepetitious DNA, will shift in length as random mutations generate new points of homology within them, in general tending to revert back to shorter periodicities. (This did not actually occur in my simulations because I limited crossover alignments to those in which 5 base pair segments were in register, thus greatly reducing the probability of such an event.) Hence, two countervailing processes—the shortening of long repeats by crossover at chance points of homology and the lengthening of repeats by development of higher-order periodicities—will affect the repeat length, which will consequently fluctuate in the course of evolution, and which will be heterogeneous in the transition periods. It is not surprising in this view, therefore, that some repeated DNA's are composed of simple periodicities while others are more complex. The probability of finding some contemporary DNA with a given pattern of repeats will, of course, depend critically on the exact rules governing crossover. For example, Table 2 shows that increasing m leads to more frequent development of higher-order periodicities, and to longer periodicities, under the conditions of my simulations.

Table 3. Pattern of fragments released by complete digestion of sequences by hypothetical restriction endonucleases, which are hypothesized to cleave at particular 5 base pair sequences. In the 200-cycle simulation referred to in the upper row, which is the same simulation whose final sequence is given in the text, the restriction enzyme is hypothesized to cleave at every occurrence of the 5 base pair sequence labeled A in the text. In the 420-cycle simulation summarized in the lower four rows, the enzyme is hypothesized to cleave at another 5 base pair sequence.

Cycle	Number of fragments released with indicated length						
	5	10	15	20	25	30	35
200	5	40	4				
388	92						
412		7	7	10	1	2	1
416		36					
420	3	43	4				

The particular algorithm I used may not have been able to reproduce a distribution of repeat patterns quantitatively similar to that in natural repetitious DNA's. For if the homology criterion m were increased in order to favor the approximately 100 to 1000 base pair repeats observed in some repetitious DNA's (*7*, *8*), the generation by chance of regions of sufficient homology to meet the criterion would become very improbable; as a result, processes that depend on crossover between such chance regions of homology might require unreasonable numbers of evolutionary cycles to complete. It would not, in any case, have been feasible to do simulations which could hope to reproduce quantitatively such long-range periodicities. In order to accommodate repeats 100 to 1000 base pairs long, the simulations would have had to encompass much more total DNA than the 450 to 550 base pairs actually involved. And if m had been substantially increased, thus decreasing the ratio of executed to attempted crossovers, the number of attempted crossovers per cycle would have had to be increased in order to keep the number of executed crossovers per cycle sufficiently high to maintain homogeneity

the repeats. These changes would have increased the number of calculations in the simulations far beyond the practical limits. I do not think that the quantitative implausibility of the particular algorithm I used seriously undermines my general thesis that unequal crossover can account for the origin and evolution of repetitious DNA's. Considering the room for speculation left by the large gaps in our understanding of crossover, it would be astonishing if there were not numerous plausible algorithms that would be perfectly consistent with the known facts and would lead under realistic conditions to patterns of periodicity similar in detail to the patterns observed in nature.

"Parahomologous" Deletions and Tandem Duplications

According to the crossover theory as I have so far presented it, the origin and evolution of tandem repeats depend on the generation of deletions and tandem duplications by crossover between short chance regions of homology. It is not necessary to the theory, however, that these deletions and duplications occur by crossing-over. All that the theory demands is that they not require extensive stretches of homology and that they be rare in comparison to unequal crossover between repeats in long tandem arrays. I will call all deletions and tandem duplications which do not require extensive homology parahomologous, so that I can refer to them without regard to their underlying mechanism.

Parahomologous deletions and duplications have been observed by Stewart and Sherman (*23*) to occur spontaneously (as well as after mutagenesis) in a 44 base pair segment of the *iso*-1-cytochrome c gene of yeast. Some of the deletions occurred between short regions of homology and thus may have been due to unequal crossover; but the spontaneous tandem duplications did not occur between regions of homology and therefore may well have been due to another mechanism. Although the rate of occurrence of parahomologous deletions and duplications was not measured, the fact that they could be detected at all within a single short segment of DNA implies that they occur at a rate quite high enough to accommodate the present theory of the evolution of repetitious DNA's.

Rate of Crossover

We now ask whether a rate of unequal sister chromatid crossover consistent with the observed overall rate of sister chromatid exchange would be sufficient to maintain intraspecies homogeneity in the very large amounts of DNA contained in some families of repeats—about 10^8 base pairs in the three major *Drosophila virilis* satellites (*24*) and about 2.5×10^8 base pairs in mouse satellite (*8*). By making some plausible although very insecure assumptions, a provisional answer can be obtained. During the simulated evolution of the family shown in Fig. 2, the approximately 100 5 base pair repeats remained quite homogeneous from cycle 160 to cycle 200, during which time 40 mutations per 500 base pairs occurred and 2291 crossovers were executed. Since 2000 crossovers correspond roughly to the number required for crossover fixation in 100 repeats allowed to vary between 90 and 110 (*17*), I conclude that a mutation rate as high as 40 mutations per 500 base pairs per crossover fixation time is not high enough to cause undue heterogeneity. If a plausible figure of 5×10^{-9} mutations per base pair per year is taken as the mutation rate (*25*)—my first assumption—then the above figure corresponds to a crossover fixation time of 1.6×10^7 years. My second assumption is that the average number of sister chromatid exchanges required for crossover fixation varies linearly with the number of repeats. This supposition has been confirmed approximately for numbers of repeats up to about 10^4 (*17, 26*), but extrapolation to numbers of the order of 10^7 is clearly very uncertain. If this assumption turns out to be correct, then the average number of crossovers required for fixation in the roughly 10^7 repeats in *D. virilis* satellites would be about 2×10^8. If this is so, a rate of sister chromatid exchange equal to about 2×10^8 crossovers during the fixation time of 1.6×10^7 years calculated above, or about 12.5 crossovers per year, would be sufficient to explain their homogeneity. Assuming three organismal generations per year and 20 germ line mitotic generations per organismal generation, this corresponds to 0.4 exchanges in the satellite regions (which constitute about 40 percent of the genome) per diploid cell per mitotic generation. The corresponding figure for the 10^6 240 base pair repeats in mouse satellite would be about 0.1 crossovers in the satellite regions (about 7 percent of the genome) per cell per mitotic generation, assuming one organismal generation and 20 germ line mitotic generations per year. Neither of these numbers seems inconsistent with the observed overall rates of sister chromatid exchange—several exchanges per cell per division (*9, 10, 27*).

The generation of higher-order periodicities is very similar to crossover fixation, except that it involves crossover fixation of a series of contiguous repeats rather than a single repeat. Provided that the rules governing crossover are sufficiently favorable to the fixation of series of repeats, higher-order periodicities ought to develop in times comparable to the crossover fixation time, as in fact they did in my simulations. Since it was argued earlier that the observed rate of sister chromatid exchange is not inconsistent with an evolutionarily reasonable crossover fixation time, we can conclude provisionally that higher-order periodicities could also be generated in evolutionarily reasonable times.

The two other processes described in this article—generation of repeats from nonrepetitious DNA and nonintegral changes of repeat length—also occurred in times comparable to the crossover fixation time in my simulations. This result argues that these two processes, which differ from crossover fixation in depending on parahomologous deletions and duplications, can take place in evolutionarily reasonable times, provided that parahomologous deletions and duplications occur at a sufficiently high rate. In the previous section I mentioned some evidence suggesting that parahomologous deletions and duplications do indeed occur at appreciable rates in eukaryotic chromosomes.

I emphasize that the foregoing calculations are very uncertain because of the uncertainty of the assumptions used to obtain them. I only intend them to show that the theory cannot at present be rejected on the ground that it requires implausibly high rates of unequal crossover or long evolutionary times.

Summary

It is often supposed that highly repetitious DNA's arise only as a result of unusual mechanisms or in response to selective pressure. My arguments and simulations suggest, by contrast, that a pattern of tandem repeats is the natural state of DNA whose sequence is not maintained by selection. The simulations show that periodicities can develop readily from nonrepetitious DNA as a result of the random accumulation of random mutations and random homology-dependent unequal crossovers. The lengths of these periodicities, and the patterns of subrepeats within them, would fluctuate in evolution, with the probability of a given pattern being dependent on the unknown exact nature of the crossover mechanism. Qualitatively, then, unequal crossover provides a reasonable and uncontrived explanation for the prevalence of highly repeated sequences in DNA and for the patterns of periodicity they evince.

References and Notes

1. R. J. Britten and D. Kohne, *Science* **161**, 529 (1968); K. D. Tartof, *Annu. Rev. Genet.*, in press.
2. J. G. Gall and D. D. Atherton, *J. Mol. Biol.* **85**, 633 (1974).
3. H. J. Cooke, *ibid.* **94**, 87 (1975).
4. M. N. Swartz, T. A. Trautner, A. Kornberg, *J. Biol. Chem.* **237**, 1961 (1962); D. M. Skinner, W. G. Beattie, F. R. Blattner, B. P. Stark, J. E. Dahlberg, *Biochemistry* **13**, 3930 (1974); S. A. Endow, M. L. Polan, J. G. Gall, *J. Mol. Biol.* **96**, 665 (1975).
5. E. M. Southern, *Nature (London)* **227**, 794 (1970).
6. P. A. Biro, A. Carr-Brown, E. M. Southern, P. M. B. Walker, *J. Mol. Biol.* **94**, 71 (1975).
7. M. R. Botchan, *Nature (London)* **251**, 288 (1974).
8. E. M. Southern, *J. Mol. Biol.* **94**, 51 (1975).
9. J. H. Taylor, P. S. Woods, W. L. Hughes, *Proc. Natl. Acad. Sci. U.S.A.* **43**, 122 (1957); A. S. Latt, *J. Histochem. Cytochem.* **22**, 478 (1974); H. Kato, *Nature (London)* **251**, 70 (1974); S. Wolff and P. Perry, *Chromosoma* **48**, 341 (1974); R. Tice, J. Chaillet, E. L. Schneider, *Nature (London)* **256**, 642 (1975); T. Ikushima and S. Wolff, *Exp. Cell Res.* **87**, 15 (1974); J. R. Korenberg and E. F. Freedlender, *Chromosoma* **48**, 355 (1974).
10. B. McClintock, *Cold Spring Harbor Symp. Quant. Biol.* **9**, 72 (1941); J. G. Brewen and W. J. Peacock, *Mutat. Res.* 7, 433 (1969).
11. A. H. Sturtevant and T. H. Morgan, *Science* **57**, 746 (1923); A. H. Sturtevant, *Genetics* **10**, 117 (1925); C. B. Bridges, *Science* **83**, 210 (1936).
12. H. M. Peterson and J. R. Laughnan [*Proc. Natl. Acad. Sci. U.S.A.* **50**, 126 (1963)] showed that changes in the number of tandem repeats could occur without recombination of outside markers.
13. A. Schalet [*Genetics* **63**, 133 (1969)] showed that changes in phenotypes presumed to be due to changes in the number of ribosomal RNA genes could occur without recombination of outside markers.
14. C. M. Radding, *Annu. Rev. Genet.* **7**, 87 (1973).
15. Crossover is often multiple, in that the recombinant DNA molecule it produces is equivalent to that which would be produced by two or more breakages and reunions in the same alignment. Such a recombinant molecule is the same as that which would be produced by two or more successive single unequal sister chromatid crossovers. In discussing the evolutionary consequences of unequal crossover, therefore, I will consider random multiple crossover as equivalent to two or more random single crossovers. Strictly speaking this will not be true, because in order to be equivalent to a multiple crossover, successive single crossovers must be restricted to certain alignments and cannot therefore be considered truly random. But I doubt that this restriction would have a significant effect on my results or conclusions.
16. J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper & Row, New York, 1970).
17. G. P. Smith, *Cold Spring Harbor Symp. Quant. Biol.* **38**, 507 (1973).
18. K. D. Tartof, *ibid.*, p. 491; G. M. Edelman and J. A. Gally, in *The Neurosciences: Second Study Program*, F. O. Schmitt, Ed. (Rockefeller Univ. Press, New York, 1970), p. 962.
19. This process alone will not account for the relative homogeneity of repetitious DNA's like mouse satellite, which are found at many separate positions in the genome. For if there were not in addition exchange of sequences between separated tandem arrays, the arrays would rapidly diverge as if they were in separate species. Indeed, separation of an originally continuous array into different positions in the genome is a plausible explanation for groups of related but not identical satellites in some species, such as *Drosophila virilis* (*2*). But while separation would certainly favor divergence, the divergence would not necessarily always occur rapidly or even at all. I assume that only an occasional exchange of sequences between separated arrays—perhaps by crossovers between them in which the chromosomal arms lying to either side of the arrays are not recombined—would be sufficient to keep the arrays similar. I make this assumption by analogy with ordinary population genetics theory, which shows that very occasional cross migration of individuals is sufficient to keep otherwise isolated populations of organisms evolving together (*16*, p. 268).
20. There are many more mitoses than meioses in the germ line. I assume that nonsister chromatid exchange occurs only very infrequently in the mitoses, in which homologous chromosomes are not paired. Because the number of chiasmas observed in meioses roughly equals the number of sister chromatid exchanges observed in mitoses, I conclude tentatively that the average number of sister chromatid exchanges per organismal generation occurring in the lineage leading to a given germ line chromosome greatly exceeds the number of nonsister chromatid exchanges.
21. D. D. Brown and K. Sugimoto, *Cold Spring Harbor Symp. Quant. Biol.* **38**, 501 (1973).
22. G. G. Brownlee, E. M. Cartwright, D. D. Brown, *J. Mol. Biol.* **89**, 703 (1974).
23. J. W. Stewart and F. Sherman, in *Molecular and Environmental Aspects of Mutagenesis*, L. Prakash, F. Sherman, M. W. Miller, C. W. Lawrence, H. W. Taber, Eds. (Thomas, Springfield, Ill., 1975), p. 102.
24. M. Blumenfeld, *Cold Spring Harbor Symp. Quant. Biol.* **38**, 423 (1973).
25. M. Kimura, *ibid.*, p. 515.
26. K. D. Tartof and S. Litwin, personal communication.
27. These calculations were derived from simulations in which nonsister chromatid exchange did not occur at all. Intuitively, I think it unlikely that crossover fixation would be any slower if sister chromatid exchange occurred at a certain rate and were accompanied by nonsister chromatid crossover than if sister chromatid exchange occurred at the same rate and were not accompanied by nonsister chromatid crossover, but I do not know how to confirm or deny this.
28. I did part of this work under postdoctoral supervision of Oliver Smithies, Departments of Genetics and Medical Genetics, University of Wisconsin. I thank Drs. Edward H. Coe and Donald L. Riddle for discussions and Elizabeth Hollis for the figures. This work was supported by National Institutes of Health grants GM20069 to O. Smithies and GM22686 to G.P.S.

Boom Towns May Hinder Energy Resource Development

Isolated rural communities cannot handle sudden industrialization and growth without help.

John S. Gilmore

The energy boom town in western United States is apt to be a bad place to live. It's apt to be a bad place to do business.

This is a problem for more than the people in the boom town. It also affects federal agencies seeking to increase energy resource production in the Rocky Mountain West, as well as the firms building and operating energy resource extraction and conversion facilities. The situation can be frustrating for local and state governments charged with protecting the health, safety, and welfare of their populations. The problems result from the traditional, business-as-usual boom in which unmanaged growth is the cumulative result of many different corporate, governmental, and individual decisions, mostly made in total isolation from each other. "Business-as-usual" is a characterization applicable at all levels of government, as well as to industry.

The results of such unmanaged growth are probably the leading source of upsets and conflicts that can be seen or anticipated in the process of western energy resource development. The boom town is a major source of social tension in an area or a region, provoking both litigation and legislation. It is a major contributor toward the potential confrontation between state and federal governments about who shall make which decisions affecting western energy resource development. Besides fostering conflict, this sort of boom growth almost inevitably generates a situation that causes overruns in both the time and the money required to get projects built and operating.

Pistol Shot, U.S.A.

The best way to explain these effects is to describe the typical business-as-usual boom town. Therefore, let us consider the very real situation in the imaginary town of Pistol Shot in some state in the West. Pistol Shot's problems are typical of those encountered by a small, isolated western community that is being impacted or is about to be impacted by the development of coal, oil shale, uranium, or even geothermal resources.

The author is Senior Research Economist at the University of Denver Research Institute, Denver, Colorado 80210, and professor in the university's College of Law. The article is adapted from a presentation at Vail Symposium V, Vail, Colorado, 12 August 1975.