4146–4151 *Nucleic Acids Research, 1996, Vol. 24, No. 21*

Statistical evidence for a biochemical pathway of natural, sequence-targeted G/C to C/G transversion mutagenesis in *Haemophilus influenzae* Rd

Rainer Merkl and Hans-Joachim Fritz*

Institut für Molekulare Genetik, Georg-August-Universität Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany

Received August 9, 1996; Revised and Accepted September 10, 1996

ABSTRACT

Markov chain analysis of the *Haemophilus influenzae* Rd genome reveals striking under-representation of three palindromic tetranucleotide strings (CCGG, GGCC and CATG), accompanied by over-representation of six tetranucleotide strings that are derived from the former by exchanging strand location of the two residues making up a G/C nucleotide pair at the terminal palindrome position. Constraints are outlined for a molecular model able to explain the phenomenon as the result of sequence-targeted, enzyme-driven G/C to C/G transversion mutagenesis. Possible participation in the process by components of known DNA mismatch repair or restriction/modification systems (in particular, cytosine methylation) is discussed. The effect widens the spectrum of enzyme-driven, specific mutagenesis beyond the formerly described C/G to T/A transition (VSP repair of *Escherichia coli*). Potential evolutionary benefits of enzymatic pathways of specific mutagenesis can be envisioned.

INTRODUCTION

The integrity of information stored in DNA is threatened both by replication errors (reviewed in ref. 1) and by chemical reactions (e.g. hydrolytic, oxidative, radiation-induced) that change the structure and, as a result, the coding properties of nucleotide residues within the polymer (reviewed in ref. 2). A plethora of elaborate mechanisms have evolved to keep the rate of DNA sequence alterations low (3); metabolic costs, however, set finite limits to the achievable accuracy. In addition, spontaneous mutations are not merely accidents but are the origin of genetic variation and hence the necessary raw material of evolution. The overall picture of DNA being handed down from one cell generation to the next is one of extreme chemical conservatism; the very narrow margin left to flexibility is nonetheless essential.

It follows that mutation rates by themselves are subject to evolutionary optimization and it is not surprising that they vary considerably from one species to another as a function of genome size and special selection pressures (as imposed, for instance, upon a parasite by the host immune system). The error rate of ~10^{-10} per nucleotide pair per generation as observed with *Escherichia coli* (4) may be taken as typical for a genome in the size range of several 10^6 nucleotide pairs. The most obvious targets of evolutionary fine tuning of mutation rates are the accuracy of DNA polymerases and the efficiency of various DNA repair processes. It is conceivable, however, that enzymatic pathways, dedicated to mutagenesis, also contribute to generating sequence variation, possibly in special ways not accessible by mere imperfections in the fidelity of DNA replication and repair.

A candidate for such a pathway has been identified in the form of very short patch (VSP) DNA mismatch repair of *E.coli* K-12. This mechanism has dual effects; it counteracts the mutagenic effect of hydrolytic deamination of DNA 5-meC residues (for a recent review ref. 5). On the other hand, and significantly for the point made here, VSP repair actively promotes T/A to C/G transition mutagenesis in a sequence-targeted manner (6–8; cf. Fig. 1). At present it is unclear whether this driven mutagenesis constitutes a useful function by itself, as has been suggested (8,9) or, as the alternative view holds (5), whether it is just an inescapable side-effect of efforts to avoid otherwise even stronger mutagenesis through 5-meC deamination.

The notion of enzymatically driven mutagenesis pathways possessing their own functional significance would be fostered if additional mechanisms could be identified, especially if these were not linked to mutation avoidance. Since statistical sequence analysis had already been successful in the discovery of the mutagenic effect of VSP repair (6–8), we followed the same strategy to search the genome of *Haemophilus influenzae* Rd for conspicuous over- and under-representations of short nucleotide strings.

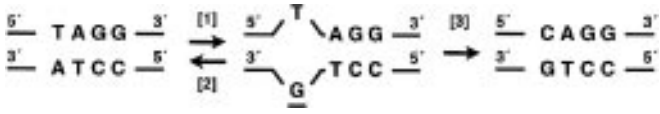

Figure 1. Generation of T/A to C/G transition mutations by VSP DNA mismatch repair. [1] Replicational misincorporation of a guanosine residue (underlined) opposite a thymine residue of the template strand. [2] DNA mismatch repair by the MutHLS system to restore the starting sequence. [3] DNA mismatch repair by the VSP system leading to fixation of a T/A to C/G mutation. The figure gives an example of one particular tetranucleotide sequence out of an entire family defined as follows. The prototype sequence of VSP repair is $C\underline{T}^{A}/_{T}GG$ (mismatched T residue underlined); either the first or the last nucleotide can deviate from the prototype sequence. The key step of VSP repair is endonucleolytic incision by Vsr endonuclease on the 5′-side of the mismatched thymine residue leaving behind a 5′-phosphate and a 3′-OH group.

*To whom correspondence should be addressed. Tel: +49 551 39 3801; Fax: +49 551 39 3805; Email: hfritz@uni-molgen.gwdg.de

MATERIALS AND METHODS

The database

The file *GHI.1con* (version 1.0, see ref. 10) contains in one contig 1 830 140 nucleotides (119 not unambiguously identified) of the *Haemophilus influenzae* Rd genome.

The statistical evaluation algorithm

Frequencies of occurrence $f(a_1a_2)$ $f(a_1a_2a_3)$ and $f(a_1a_2a_3a_4)$ ($a_i \in$ {A,C,G,T}) of dimer, trimer and tetramer sequences were determined by sliding windows of length 2–4 along the contig. Expected frequencies f_{exp} $(a_1a_2a_3a_4)$ of tetramer sequences were calculated as a second-order Markov chain:

$$f_{exp}(a_1a_2a_3a_4) = f_{M4,2}(a_1a_2a_3a_4) = \frac{f(a_1a_2a_3)f(a_2a_3a_4)}{f(a_2a_3)} \quad \mathbf{1}$$

Deviation of observed frequency $f(a_1a_2a_3a_4)$ from expected frequency $f_{exp}(a_1a_2a_3a_4)$ is quantitatively expressed by representation bias factor $r_{M4,2}$ as

$$r_{M4,2}(a_1a_2a_3a_4) = \frac{f(a_1a_2a_3a_4)}{f_{exp}(a_1a_2a_3a_4)} \quad \mathbf{2}$$

The test of statistical significance

Statistical significance of the deviation of observed numbers of occurrence of tetranucleotides from expectation was determined by the χ^2 test (11). χ^2 values were computed and compared with χ^2 distributions calculated with 81 ($3 \times 3 \times 3 \times 3$) degrees of freedom for the global dataset of all 256 tetranucleotide strings and with one degree of freedom for individual tetranucleotides.

RESULTS

The genome of *H.influenzae* Rd (10) was searched for over- and under-representation of tetranucleotide sequences by second-order Markov-chain analysis. Results are displayed in Figure 2. The χ^2 test (11) applied to the global set of 256 tetranucleotides resulted in a value of 2.5×10^4 which indicates non-random distribution with very high significance. As a measure of over- or under-representation of a tetranucleotide string $a_1a_2a_3a_4$ we define its representation bias factor $r_{M4,2}$ $(a_1a_2a_3a_4)$ as the observed frequency $f(a_1a_2a_3a_4)$ of that string divided by its expected frequency calculated as a second-order Markov chain. Thus, a bias factor greater than one indicates over-representation, one smaller than unity indicates under-representation. The abscissa in Figure 2 gives absolute frequencies, the ordinate the bias factors, so that strings that are rare in absolute terms and also under-represented relative to statistical expectation are located in the lower left corner of the graph. Figure 2 demonstrates a striking under-representation of three tetranucleotide strings (CCGG, GGCC and CATG); along the ordinate, all three are separated from the rest by a sizable gap. A common structural feature is that all three sequences are palindromes and have G/C base pairs at their ends.

The most straightforward explanation for such a finding would be that CCGG, GGCC and CATG are for some reason (such as, for instance, selective pressure exerted by restriction enzymes with the correponding target sequences) counterselected in *H.influenzae* Rd. Most simply, this would mean that any sequence alteration destroying such a site should have the same selective

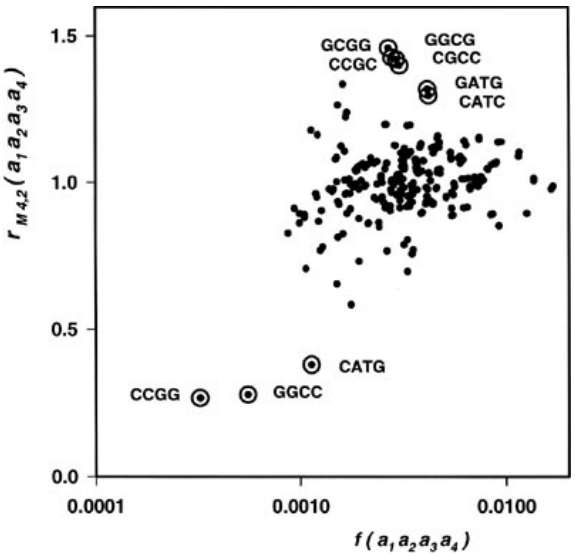

Figure 2. Two-dimensional plot of frequencies and representation bias factors of all 256 tetranucleotide strings in the genome of *Haemophilus influenzae* Rd. Frequencies of occurrence $f(a_1a_2a_3a_4)$ and representation bias factors $r_{M4,2}(a_1a_2a_3a_4)$ were calculated for all 256 tetranucleotide strings $a_1a_2a_3a_4$ (derived from the DNA strand deposited in the data base) as described under Materials and Methods. Each string is represented in the graph as a filled circle. Values of $r_{M4.2}$ greater than one indicate over-representation of a string relative to statistical expectation; $r_{M4.2}$ values smaller than one indicate under-representation.

benefit and the three tetranucleotides would have been thinned out in the past mainly by random point mutations. This process should have lead, in each case, to a concomitant and evenly distributed over-representation of 12 tetranucleotide strings (*tetra_H1*$_{a1a2a3a4}$) which differ from the starting sequence in a single position, provided the 12 different sequence changes have equal propensity to occur and the frequency distribution of the 12 product strings is not subsequently distorted by biological selection. In addition, the over-representation of each individual product string would be small (in absolute terms one twelfth of the under-representation of the corresponding starting string) and could easily be obscured by any less-than-perfect validity of the idealized premises mentioned. Results of comparing the occurrence of each of the three under-represented strings with those of the elements of its corresponding set *tetra_H1*$_{a1a2a3a4}$ are compiled in Figure 3.

Two points emerge from Figure 3 that are pertinent to the argument made above. First, there is no general over-representation of the three sets of twelve tetranucleotide sequences. Second, two tetranucleotide strings in each case are located at the extreme upper end of the representation scale, well-separated from the rest. Strikingly, these six most prominently over-represented sequences stand to their respective reference tetranucleotide strings in a consistent structural relationship; in every case they are derived from the under-represented sequence by exchanging a guanosine for a cytosine residue (or vice versa) at position one or four of the palindrome. Considered for double-stranded DNA, this corresponds to strand-swapping between a guanosine and its juxtaposed cytosine residue and, due to the two-fold symmetry, the operation is the same at the two equivalent sequence positions (see Fig. 4). Since in the process the two-fold symmetry of the sequence is broken, it links to each palindrome two non-symmetrical tetranucleotide strings and defines three families,

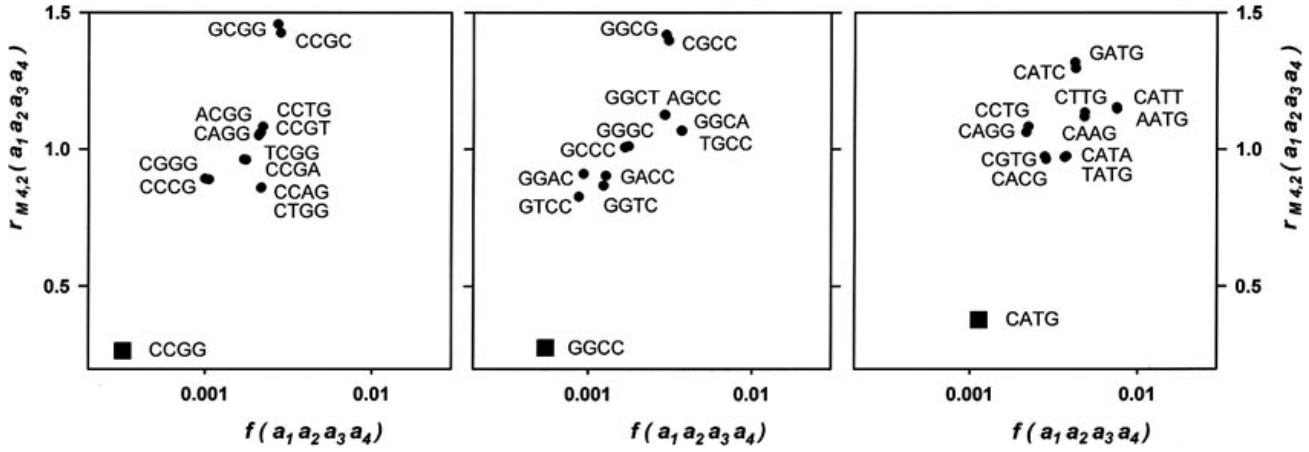

Figure 3. Occurrence of tetranucleotide strings having Hamming distance 1 to reference sequences CCGG, GGCC and CATG. Each panel shows the occurrence of one of the three under-represented tetranucleotide strings (filled square) and, for comparison, those of all twelve tetranucleotide sequences that differ from each respective reference sequence by a single nucleotide (filled circles). Meaning of symbols and mode of plotting as in Figure 2.

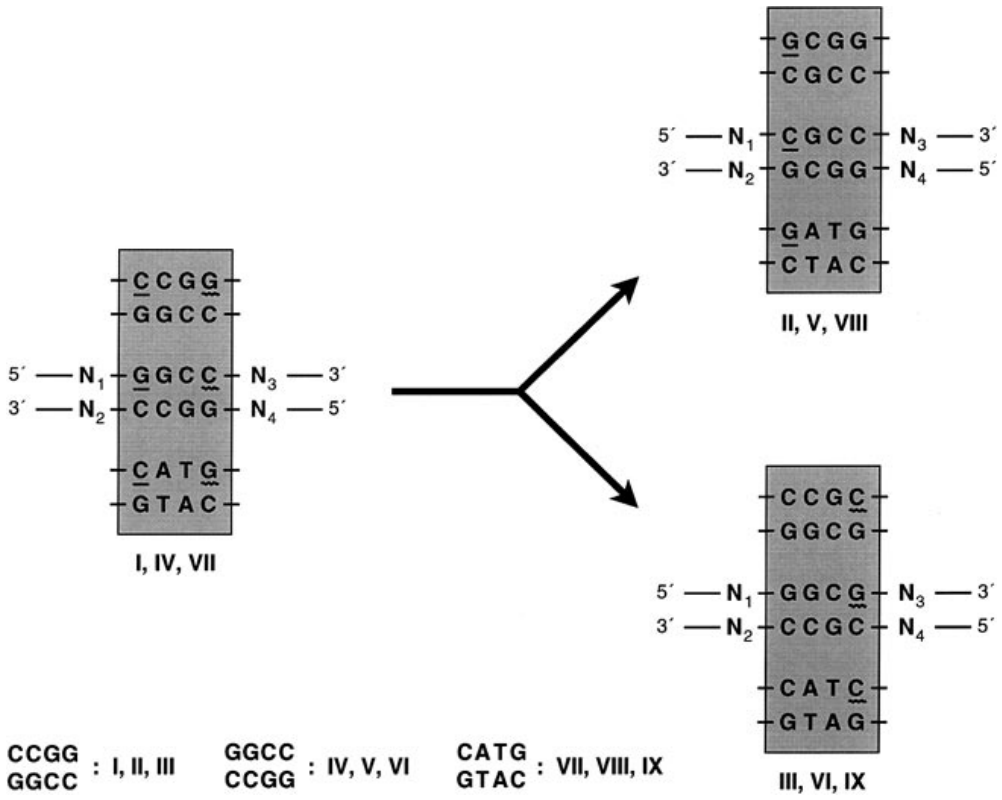

Figure 4. Postulated process of G/C for C/G substitutions, formal description. The process depicted explains the statistical findings illustrated in Figure 2. Replacing a G/C base pair for C/G at the extreme end of any of the three palindromic tetranucleotide sequences I, IV or VII leads to the corresponding non-symmetrical tetranucleotide sequences II, V, VIII, III, VI and IX. I is converted to II or III, IV to V or VI and VII to VIII or IX. If the process is unidirectional as indicated, under-representation of I, IV and VII and corresponding over-representation of the other six sequences is the necessary consequence.

each consisting of three strings, one under-represented and two over-represented ones. Thus, CCGG is linked to GCGG and CCGC (family 1); GGCC is linked to GGCG and CGCC (family 2) and CATG is linked to GATG and CATC (family 3). These findings strongly suggest that indeed the pool of one class of tetranucleotide sequences (the three palindromes) is depleted to the benefit of another but that this, in contrast to the starting assumption of random point mutations, happens in a single specific fashion that is consistent for all three families.

Going back to Figure 2, one finds the six over-represented sequences at the extreme upper end of the representation scale of all tetranucleotide strings. These are, in descending order of over-representation, GCGG, CCGC, GGCG, CGCC, GATG and CATC. Interestingly, family 1 contains the strings forming the extreme

ends of the distribution on either side, family 2 the second-most under-represented tetranucleotide and strings #3 and #4 in the over-representation ranking and, finally, family 3 the third-most under-represented tetranucleotide and strings #6 and #7 at the other end. The only string that interrupts this matching scheme is the fifth-most over-represented tetranucleotide (GTCG), the high abundance of which has evolved for reasons apparently unrelated to the effect discussed here.

The χ^2-test was applied to all 256 individual tetranucleotide strings and strings were ranked in order of decreasing significance of their deviation from the expected number of occurrences in the *H.influenzae* Rd genome (Table 1). At the high significance end of the scale, the three under-represented strings occupy ranking positions 1–3, the six over-represented ones positions 6–11. Taken together, the findings suggest that linking the two non-symmetrical members of each sequence family to the third one by a G/C to C/G base pair exchange is more than a formal description but that the over-represented strings are generated at the expense of the under-represented ones by a process of natural transversion mutagenesis targeted to the terminal positions of the three palindromic sequences CCGG, GGCC and CATG.

Table 1. Ranking of individual tetranucleotide strings by significance of deviation from expected number of occurrence

$a_1a_2a_3a_4$	χ^2	rank
CATG	2113	1
GGCC	1905	2
CCGG	1196	3
GCGC	966	4
TATA	816	5
GCGG	715	6
GGCG	675	7
CCGC	656	8
CGCC	637	9
GATG	587	10
CATC	523	11

First 11 positions of a ranking order by decreasing value (underlined strings are not part of the set of nine tetranucleotides considered).

DISCUSSION

Drift or selection for a function?

In the case of T/A to C/G transition mutagenesis driven by VSP repair, the known quantitative grading of sequence context preference of Vsr endonuclease (the key enzyme of the process) is reflected in the grading of statistical under- or over-representation of the corresponding tetranucleotide strings, a finding that indicates free drift with selection superimposed only in second approximation (8). In the absence of conflicting evidence, free drift is also the simplest assumption for the mutagenesis process considered here—with one notable exception. To some extent, selection for a known biological function contributes to the high abundance of GCGG and CCGC. These strings are part of the DNA uptake signal sequence (USS) of *H.influenzae* Rd. The two orientations of USS (plus: AAGTGCGGT, minus: ACCGCACTT) occur in the genome 734 and 731 times, respectively (12). This comprises a significant fraction of the corresponding total occurrences of 4982 and 5161. That this fraction is not dominating, however, is illustrated by the following. GCGG and CCGC have a mean

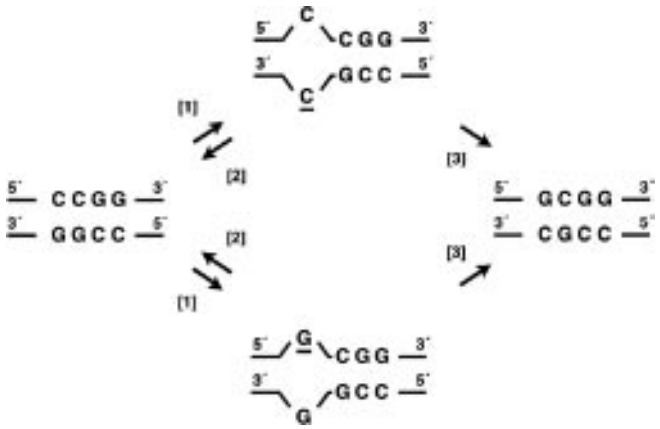

Figure 5. Transversion mutagenesis pathway in *Haemophilus influenzae* Rd, possible mechanism. [1] Replicational misincorporation of a cytosine residue opposite another cytosine residue (top) or, alternatively, a guanosine residue opposite another guanosine residue (bottom). Misincorporated residues are underlined. [2] Post-replicative DNA mismatch repair to restore starting sequence. [3] Postulated sequence-specific DNA mismatch repair leading to fixation of a G/C to C/G transversion mutation. Note that in [3] the newly synthesized DNA strand containing the erroneously incorporated nucleotide serves as the synthesis template.

representation bias factor of 1.44 (cf. Fig. 2), whereas the other 10 tetranucleotide strings that make up the USS (both orientations) have a mean bias factor of only 1.10. Thus, the presence of GCGG and CCGC in the USS contributes only moderately to their over-representation. Intergenic dyad sequences (IDS) are another class of statistically prominent nucleotide strings within the *H.influenzae* Rd genome (13); they do not contain any of the nine tetranucleotide sequences discussed here.

Constraints for modelling a molecular mechanism

No chemical reaction is known that could plausibly occur in a living cell and modify either DNA guanine or DNA cytosine residues in such a way that it would trigger the transversion of a G/C nucleotide pair to C/G. This makes a G/G or a C/C mismatch resulting from a replication error (Fig. 5) the likely source of mutagenesis and one would have to conclude that either or both of these mismatches, in the special sequence context provided by the three strongly under-represented tetranucleotide strings, are poor substrates of post-replicative DNA mismatch repair or are actively corrected ‘the wrong way’, i.e. with the newly synthesized DNA strand providing the template for repair synthesis.

It seems difficult to explain the sequence specificity of the mutagenesis process without implying the participation of a sequence-recognizing protein. A specific gap in the substrate spectrum of post-replicative mismatch repair comprising G/G and/or C/C in the three particular tetranucleotide sequence contexts is a possibility that is consistent with the demands of formal logic but is not likely to be true, given what we know about DNA/protein interaction in general and the properties of DNA mismatch repair in particular. Positive recognition of the mismatch(es) in the relevant sequence contexts is more plausible and could bring about the effect in either of two different ways. A protein could bind to the mismatch and make it inaccessible to repair by (e.g.) preventing its binding to a MutS homolog. Alternatively, the site could be recognized by an enzyme that initiates repair by incising at or near the mismatched residue in the parental DNA strand.

Possible connections to known components of macromolecular DNA metabolism

Palindromic tetranucleotide sequences are typical targets of type II restriction/modification (R/M) systems. Specifically, the three sequences considered here serve this function in various *H.influenzae* strains and other *Haemophilus* species, CCGG in *H.influenzae* RFL2 and *H.influenzae* RFL5, GGCC in *H.aegypticus* and *H.haemoglobinophilus* and CATG in *H.influenzae* RFL1 and *H.influenzae* RFL8 (selection taken from ref. 14). Even though *Hin*dII (GTYRAC) and *Hin*dIII (AAGCTT) are the only R/M systems described to date for *H.influenzae* Rd, the coincidence is suggestive enough to invite thoughts on how components of an R/M system could participate in the postulated mutagenesis pathway.

The substrate recognition properties of either the methyltransferase or the restriction endonuclease of an R/M system could be relaxed to the extent that the enzyme binds to one or both of the mismatched intermediates shown in Figure 5 and protects it from being processed by post-replicative DNA mismatch repair. Binding of DNA cytosine-C^5 methyltransferase to a mismatched target site has been documented for M.*Hha*I and M.*Hpa*II (15), however not with a G/G or C/C mismatch. Beyond binding and passively blocking correction of a replication error, a restriction endonuclease could actively initiate repair with fixation of the mutation by setting an endonucleolytic cut in the parental DNA strand; corresponding activities have been described for restriction enzymes *Eco*RV (16) and, with the mismatch-equivalent 2-aminopurine/thymine opposition, for *Sso*II (17).

The fact that three different palindromes are affected raises the obvious question of whether three different sequence-specific proteins must be invoked. This is not necessarily the case since multispecific DNA cytosine-C^5 methyltransferases which recognize up to five different target regions (18) have been described. Specifically, the (CCGG, GGCC, CATG) set has a two out of three overlap with the substrate sequences of the trispecific DNA cytosine-C^5 methyltransferase of *Bacillus subtilis* phage SPR (CCGG, GGCC, $CC^A/_TGG$) (19). Hence, a single multifunctional protein as the cause of the effect is not a far-fetched idea, if this protein is to be a DNA cytosine-C^5 methyltransferase.

Enzymes of type II R/M systems are not the only known candidates for driving the process under consideration. Components of DNA mismatch repair pathways also fulfill the requirements if they are specific with respect to both the nature of the mismatch and its sequence context. Two mismatch repair enzymes already known offer reasonable analogies to what has to be postulated for sequence-targeted G/C to C/G transversion mutagenesis. These are Vsr endonuclease of *E.coli* K-12 (8,9) and DNA mismatch glycosylase Mig.*Mth* of *Methanobacterium thermoautotrophicum* (20). Vsr endonuclease acts on a well-defined, small family of substrate sequences, as required here, but it processes T/G mismatches exclusively and hence drives A/T to C/G mutagenesis (6–8). Mig.*Mth*, on the other hand, while being most active on T/G mismatches also processes G/G (20), a suitable intermediate for G/C to C/G mutagenesis (Fig. 5). Its sequence recognition selectivity, however, seems to be rather relaxed (20).

The genome of *H.influenzae* Rd contains several candidate genes for R/M systems and for members of the Endo III/MutY/Mig.*Mth* family of enzymes but no reading frame related to *vsr* (10; R. Merkl unpublished). The relevant protein, however, need not necessarily be encoded in the bacterial chromosome; a plasmid or phage genome is a viable alternative.

CONCLUSIONS

This study rests entirely on statistical evidence. The discovered distortion of the tetranucleotide composition of the *H.influenzae* Rd genome, however, is strikingly prominent and the assumption of a single mutagenic process suffices to explain the entire set of phenomena in a consistent manner. This lends credibility to the postulate of a mutagenic process that is simple in its formal description and can plausibly be sustained by known components of macromolecular DNA metabolism (which is not to say that participation of a hitherto unknown type of enzyme is ruled out). With VSP repair the process shares the property of introducing specific mutations into a small, well-defined family of substrate sequences. In contrast to VSP-driven mutagenesis, however, the enzymatic activity required is not functionally linked in an obvious way to the processing of a common chemical DNA damage. The postulated new pathway expands the spectrum of enzymatically driven mutagenesis from transitions (VSP repair) to tranversions.

As to the evolutionary significance of the phenomenon, interpretations at different levels of complexity are possible. Most simply, it may just reflect free drift, i.e. be the fortuitous and largely fitness-indifferent result of side-reactions catalyzed by enzymes with entirely different tasks in macromolecular DNA metabolism. Alternatively, it may resemble the footprints of restriction enzymes (possibly encoded by ‘visiting’ phages or plasmids) that exert selective pressure against presence in the bacterial genome of their respective target sequences; the open question with that model would be why one particular type of point mutation should be prefered so prominently over all others. Finally, and most interestingly, the enzymatically driven mutagenesis may be a device of creating sequence diversity within larger populations that become useful in connection with a pathway of DNA uptake and short-patch repair. This could constitute a recombination mechanism that transfers genetic information in short blocks of DNA sequence, in close analogy to what has been suggested for VSP repair (6,8,9).

The case put forward above may provide a model for how one can use the large data sets furnished by genome sequencing projects to discover patterns in DNA sequences that significantly deviate from statistical expectation and how experimentally falsifiable hypotheses as to the molecular origins of these patterns can be derived starting from statistics alone. Experiments to test the ideas laid out in the preceding paragraphs have recently been initiated in this laboratory.

NOTE ADDED DURING REVISION

After submission of this manuscript, the DNA sequence of the entire *Methanococcus jannaschii* genome was published [C.J. Bult *et al*. (1996) *Science* **273**, 1058–1073]. A first statistical survey reveals the same phenomenon to prevail in this archaeal genome, albeit with different tetranucleotide strings involved; GATC (under-represented) is coupled to CATC and GATG (over-represented); likewise, CTAG and/or GTAC (under-represented) is/are coupled to CTAC and GTAG (over-represented).

ACKNOWLEDGEMENT

This work was supported by Fonds der Chemischen Industrie.

REFERENCES

1 Umar, A. and Kunkel, T. (1996) *Eur. J. Biochem.* **238**, 297–307.
2 Lindahl, T. (1993) *Nature* **362**, 709–715.
3 Friedberg, E.C., Walker, G.C. and Siede, W. (1995) *DNA Repair and Mutagenesis*. ASM Press, Washington, D.C.
4 Schaaper, R.M. (1993) *J. Biol. Chem.* **268**, 23762–23765.
5 Lieb, M. and Bhagwat, A.S. (1996) *Mol. Microbiol.* **20**, 467–473.
6 Merkl, R., Kröger, M., Rice, P. and Fritz, H.-J. (1992) *Nucleic Acids Res.* **20**, 1657–1662.
7 Bhagwat, A. S. and McClelland, M. (1992) *Nucleic Acids Res.* **20**, 1663–1668.
8 Gläsner, W., Merkl, R., Schellenberger, V. and Fritz, H.-J. (1995) *J. Mol. Biol.* **245**, 1–7.
9 Hennecke, F., Kolmar, H., Bründl, K. and Fritz, H.-J. (1991) *Nature* **353**, 776–778.
10 Fleischmann, R.D. *et al.* (40 authors) (1995) *Science* **269**, 496–512.
11 Sachs, L. (1982) *Applied Statistics. A Handbook of Techniques.* Springer Verlag, New York, Heidelberg, Berlin.
12 Smith, H.O., Tomb, J.-F., Dougherty, B.A., Fleischmann, R.D. and Venter, J.C. (1995) *Science* **269**, 538–540.
13 Mrázek, J. and Karlin, S. (1996) *Trends Biochem. Sci.* **21**, 201–202.
14 Kessler, C. and Manta, Y. (1990) *Gene* **92**, 1–248.
15 Yang, A.S., Shen, J.-C., Zingg, J.-M., Mi, S. and Jones, P.A. (1995) *Nucleic Acids Res.* **23**, 1380–1387.
16 Alves, J., Selent, U. and Wolfes, H. (1995) *Biochemistry* **34**, 11191–11197.
17 Petrauskene, O.V., Schmidt, S., Karyagina, A.S., Nikolskaya, I.I., Gromova, E.S. and Cech, D. (1995) *Nucleic Acids Res.* **23**, 2192–2197.
18 Schumann, J., Willert, J., Wild, C., Walter, J. and Trautner, T.A. (1995) *Gene* **157**, 103–104.
19 Günthert, U., Lauster, R. and Reiners, L. (1986) *Eur. J. Biochem.* **159**, 485–492.
20 Horst, J.-P. and Fritz, H.-J. (1996) *EMBO J.* **15**, 5459–5469.