MECHANISMS OF MUTAGENESIS AND THE ROLE OF LOCAL DNA SEQUENCE COMPLEXITY

Chuzhanova N.A., Cooper D.N.

1 Department of Computer Science, Cardiff University, PO Box 916, Cardiff CF24 3XF, UK
2 Institute of Mathematics, RAS, 630090, Novosibirsk, Russia
3 Institute of Medical Genetics, University of Wales College of Medicine, Cardiff CF14 4XN, UK

e-mail: nadia.chuzhanova@cs.cardiff.ac.uk

Key words: sequence complexity, micro-deletion, micro-insertion, indel, mutagenesis

Resume

Motivation: A detailed study of deletion and insertion mutagenesis could improve our understanding of the molecular mechanisms underlying micro-insertions, micro-deletions and indels and could be an invaluable aid to the optimisation of mutation search strategies in molecular diagnostic medicine.

Results: In the present study we examine the local DNA sequence complexity and its role in previously postulated mechanisms for deletion and insertion mutagenesis such as slipped mispairing and strand switching, secondary loop excision and quasi-palindrome correction, Moebius loop resolution and excision. A novel mutational mechanism mediated by the insertions of inverted repeats is proposed. It was found that the mechanism of mutation and the type of repeat involved into micro-insertions and micro-deletions depend upon the change in a certain complexity measure. Data from the Human Gene Mutation Database (Krawczak et al., 2000a) is used to compare and contrast 3767 micro-deletions, 213 different indels and a small amount of insertions in order to propose mechanistic processes that could account for their genesis.

Introduction

The most common types of mutation causing human genetic disease are single base-pair substitutions and micro-deletions (Antonarakis et al., 2001). The remainder comprises an assortment of larger deletions, insertions, inversions, expansions and complex rearrangements. One relatively uncommon type of mutation is indel, a combined micro-deletion/ micro-insertions that result in the apparent replacement of one or more base-pairs by others, not necessarily the same number. Although it is likely that the component micro-insertion and micro-deletion events occur contemporaneously (i.e. as part of the same complex mutational event), this need not necessarily be so. The study of mutational lesions causing human genetic disease has revealed that, irrespective of their type, the nature, frequency and location of mutations are invariably non-random (Cooper, Krawczak, 1993; Antonarakis et al., 2001), being strongly influenced by the complexity of the local DNA sequence environment (Krawczak et al., 2000a). Complexity analysis was used to examine the change in complexity and the subsequent involvement of different types of repeat in deletion/insertion event.

Methods and Algorithms

Complexity analysis, as devised by Gusev et al. (1999), was used to examine the potential contribution of the local DNA sequence complexity to the several postulated mechanisms of deletion/insertion mutagenesis and also to the two-step process of indel formation.

Complexity analysis is based on a definition of sequence complexity by taking into account different types of repetitive element including direct and inverted repeats and inversions thereof occurring in a given sequence. One can conceive of the sequence as being decomposed into "words", where each word is the longest among all possible words for which a direct or inverted repeat, or an inversion thereof, occurs somewhere upstream of the current position. An overlap between two repeat copies is also permitted. The length of the first fragment is always 1. It is apparent that this decomposition, H(S), contains the minimal number of words. The number of words in this minimal decomposition, H(S), is called the complexity of S. We shall denote respectively by C1(S), C3(S), C4(S), C5(S) and C2(S) the complexities or number of words in decompositions H1(S) computed with respect to direct repeats, H3(S) with inverted repeats, H4(S) with symmetric elements, H5(S) with inversions of the inverted repeats and H2(S) computed using a combination of all the above mentioned types of repeat.

Let us consider as an illustrative example a DNA fragment S from the RET gene with reported indel delCCinsGG (the precise location of the indel and the deleted nucleotides are indicated by lower case letters). Decompositions of S into words and the corresponding complexities are:

* Corresponding author.
Some of the repeated words are underlined. Different types of repeats are marked by arrows in order to indicate their orientation. Parameters C1–C5 represent suitable measures of complexity (regularity) of a sequence S, since any abundance in S of direct and inverted repeats and inversions thereof serves to reduce the corresponding complexity. As can be seen from the above example, the minimum complexity among C1, C3–C5 is achieved with measure C4. The second measure in ascending order is measure C5. Fragment S is thus comparatively rich in symmetric repeats and inversions of the inverted repeats that together make a major contribution to measure C2. To be able to compare the complexities of two or more sequences that differ in length, one can use the complexity per base, c=C/N, where N is the length of a given sequence in base-pairs.

Intuitively, complexity should decrease as a consequence of deletion and increase as a result of an insertion but this is not invariably so. Whether the complexity is increased or decreased can be strongly influenced by the appearance or disappearance of prominent repeats through either the insertion or deletion of bases. We show later that the mechanism of mutation and the type of repeat involved into micro-insertions and micro-deletions depend upon the change in a certain complexity measure.

Implementation and Results

Direct repeats and slipped mispairing models. One mechanism frequently implicated in the generation of micro-deletions and micro-insertions is slipped mispairing that involves the misalignment of short direct repeats. During DNA replication, the template strand can slip forward, producing a single-stranded loop that can subsequently be excised and repaired thereby fixing a micro-deletion. This event serves to decrease the C1 complexity of a sequence. Thus, for example, a short micro-deletion in the APC gene gives rise to a slight decrease in complexity C1 from 13 to 12. The decompositions are as follows (two direct repeats causing micro-deletion are underlined):

H1 before deletion: A-AAA-G-AA-T-AGa-tag-T-C-T-TC-CTT-TA, C1=13, c1=0.54;
H1 after deletion: A-AAA-G-AA-T-AG-T-C-T-TC-CTT-TA, C1=12, c1=0.60.

Conversely, one of the nascent strands may slip backwards thereby templating a micro-insertion. In this case, the complexity of a sequence bearing an insertion must either remain the same or can increase slightly owing to the presence of imperfect direct repeats. For example, an insertion of gcg into the AAP gene does not change the complexity of the fragment. The decompositions are as follows:

H1 before insertion: G-G-A-GG-C-GGCGGCGGC-C-A-CCA, C1=9, c1=0.45;
H1 after insertion: G-G-A-GG-C-GGCGgcgGCGGC-C-A-CCA, C1=9, c1=0.39.

A modified slipped mispairing model was proposed by Krawczak and Cooper (1991) to account for deletions not readily explicable by the standard model. If the DNA sequence flanking the deleted bases also occurs as a contiguous sequence in the immediate vicinity, the intervening non-homologous bases may loop out thereby potentiating the formation of a second direct repeat copy. Transient misalignment of the two repeats may then allow the deletion of the intervening bases before strand alignment is restored. The juxtaposition of two repeat copies as a consequence of the deletion serves to decrease the complexity of the sequence. For example, in the CFTR gene, a short deletion [ATTCTGTTCTcaGTTTTCCTGG] leads to the creation of a second copy of TTCTGTT with a concomitant decrease in complexity from 12 to 9 (complexity per base also decreases from 0.55 to 0.45).

Inverted repeats and secondary structure. Inverted repeats have also been implicated in the generation of micro-deletions and micro-insertions (Cooper, Krawczak 1993). An inverted repeat or palindrome comprises a series of bases that are complementary to another contiguous sequence upstream on the same DNA strand. By definition, therefore, an inverted repeat allows hairpin loop formation; excision repair of such a loop may yield a micro-deletion. Inverted repeats may also promote slipped mispairing of the nascent strand and subsequent duplication of downstream sequence. In the case of a micro-deletion in the KIT gene this led to a decrease in complexity:

H3 before deletion: T-C-T-GA-A-C-TCA-a-agGAT-TCC-CAGAGT-T, C3=12, c3=0.55;
H3 after deletion: T-C-T-GA-A-C-TCA-GTC-TCC-CAGAGT-T, C3=9, c3=0.45.

In the case of a micro-insertion in the F9 gene this led to an increase in complexity:

H3 before insertion: T-ATA-C-C-A-A-GGTAT-CC-CGG-TAT-TG-T, C3=12, c3=0.5;
H3 after insertion: T-ATA-C-C-A-A-GGTAT-CC-C-a-a-gtta-cc-a-a-GGTAT-TG-T, C3=20, c3=0.47.

A number of mechanisms that can account for both micro-deletions and micro-insertions in the vicinity of inverted repeats involve quasi-palindromic sequences (imperfect inverted repeats). Quasi-palindromes are thought to promote "strand-switching", or aberrant templating in the formation of the nascent strand.

131
Symmetric elements and Moebius loops. Symmetric elements, sequences that possess an axis of internal symmetry or which are symmetrical to another contiguous sequence on the same DNA strand upstream, have also been implicated in the generation of deletions and insertions, being thought to facilitate the formation of secondary structure intermediates. Krawczak and Cooper (1991) proposed that such an intermediate could be a Moebius loop-like structure formed after strand separation, twisting and re-annealing to the opposite strand in reverse orientation. Mismatched bases could loop out in such a structure thereby facilitating their own excision. An example of this process is provided by the 1bp micro-deletion in the APC gene responsible for the appearance of an 8bp self-symmetric fragment that decreases C4 complexity:


Alternatively, such a Moebius loop may partially resolve if one of the DNA strands disconnects and breaks. The repair of this region by a DNA polymerase would effectively result in the duplication of a sequence from the end of the symmetric element that initially broke off.

Inversions of inverted repeats and a novel mutational mechanism. Inversions of inverted repeats were found to be over-represented in the genomes of various organisms (Tchurikov, 1992). It was suggested that these elements may facilitate the formation of secondary structures, henceforth to be termed knots. Such pseudo-palindromic intermediates are structurally similar to hairpin-loops and may be implicated in the generation of micro-deletions and micro-insertions. Thus excision repair of such a loop may yield a micro-deletion whilst slipped mispairing of the nascent strand could lead to duplication of downstream sequence. For example, in a micro-deletion in the BRCA1 gene, a mismatched base that looped out of the knot could have facilitated its own excision:

H5 before deletion: AAAATATTTGgGAAAACCTAT, C5=12, c5=0.57;
H5 after deletion: AAAATATTTGAAAACCTAT, C5=10, c5=0.50.

The existence of this knot structure may help to resolve the ambiguity in deletion position that occurs as a consequence of the repetitive nature of the deletion-prone site. It is therefore likely that it was the third G that was deleted.

Repeats involved in insertion/deletion events and the possible path of mutation. Each indel may be regarded as having been the result of a two-step insertion/deletion process; the first step transforming the wild-type sequence to an intermediate, the second step transforming the intermediate to the final mutated sequence. There are essentially three possibilities. As it was shown above, the insertions increase the complexity while the deletions decrease it. This means that the complexity of a fragment remains more or less the same before and after the insertion/deletion event, i.e. if during the first step, a decrease in complexity is observed, then the second step must reverse this process leading to an increase in complexity and vice versa.

Discussion

The analysis demonstrated that changes in local DNA sequence complexity can be accounted for the involvement of a certain postulated mutational mechanism in micro-insertion and micro-deletion events and in indel formation; it can predict both the number and identity of the bases deleted and/or inserted. Proposed approach could also be applied to the analysis of gross rearrangements, which have so far been refractory to analysis. This is the short summary of a new approach; the detailed results will be reported elsewhere.

References