Tandem DNA Repeats: Generation and Propagation in the Microgene Polymerization Reaction and in vivo

Mark Itsko\textsuperscript{1,4}, Eitan Ben-Dov\textsuperscript{2}, Avinoam Rabinovitch\textsuperscript{3} and Arieh Zaritsky\textsuperscript{1}

Departments of \textsuperscript{1}Life Sciences and \textsuperscript{3}Physics, Ben-Gurion University of the Negev, PO Box 653, Be’er-Sheva 84105, Israel, \textsuperscript{2}Achva Academic College, MP Shikmim, 79800, Israel
\textsuperscript{4}Present address: Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA

1. Introduction

Short and long-motif periodicities are omnipresent in genomes of both eukaryotes (Britten & Kohne 1968) and prokaryotes (Hofnung & Shapiro, 1999). Repetitive DNA can amount to more than half of higher eukaryotic genomes, specifically in \textit{Homo sapiens} (International Human Genome Consortium, 2001) and \textit{Zea mays} (Meyers et al., 2001). Even in prokaryotes it can account for about 6% of the total genome, specifically in \textit{Mycoplasma pneumoniae} (Ruland et al., 1990) and \textit{Neisseria meningitides} (Parkhill et al., 2000). Two major types of repetitive DNA exist (Dogget, 2000): tandem (head-to-tail contiguous) and interspersed (non-contiguous) with a specific pattern scattered all over the genome and lengths varying up to several hundred nucleotides (nt).

Emergence and propagation of interspersed repeats such as REP (Repetitive Extra-genic Palindrome) found in bacteria and \textit{Alu} repeats that are abundant in the human genome are putatively attributed to reproduction processes mediated by transposons (Gilson et al., 1984) or retroviruses (Ullu & Tschudi, 1984). These are multi-step complex enzymatic processes and will not be considered here. This chapter deals with tandem repeats and their generation \textit{in vitro} by the Microgene Polymerization Reaction (MPR) (Itsko et al., 2008a; 2008b; 2009), and proposes a mechanism for their \textit{in vivo} generation as well.

Most tandem repetitive DNA sequences in higher eukaryotes are located near the chromosomal telomers or centromers where they play important roles in maintaining genome integrity (Blackburn, 1991) and segregation (Catasti et al., 1994). In addition to this un-transcribed but evidently functional repetitive DNA, the human genome contains many apparently non-functional repetitive DNA sequences in the forms of micro-satellites (a variety of di-, tri-, tetra-, and penta-nucleotide tandem repeats) and mini-satellites (30-35 bp long, with variable sequences and conserved cores of 10-15 bp) (Dogget, 2000). The number of repeats is prone to expand during replication because its constituent strands can slide over each other between the multiple complementary regions (Wells, 1996). Thermodynamically unfavorable structures bulging out from DNA duplex that accompany the strand sliding
process can be stabilized by the inner base-pairs which facilitate the expansion (Kang et al., 1995). Expansion of DNA repeats is associated with a variety of human hereditary diseases (Mirkin, 2007).

The conceptual views on repetitive DNA were changed from considering it as parasitic DNA (Orgel & Crick, 1980) up to necessary organizer of genomic information (Shapiro & von Sternberg, 2005). Furthermore, numerous periodicities encrypted in encoded proteins may reflect the evolution of modern coding sequences based on primordial oligomeric repeats (Ohno, 1987). Unused preexisted long repetitive sequences may yield coding frames expressing unique enzymes even currently (Ohno, 1984).

Different in vitro systems are used to study mechanisms underlying genomic repeat expansion and possible evolutionary aspects of this process. These systems include double- or single-stranded DNA with relatively short (4-8 nt) repetitive units and thermophilic DNA polymerases operating under isothermal conditions. Staggered re-annealing of constituent strands is putatively involved in expansion of repetitive DNA duplexes (Tuntiwechapikul & Salazar, 2002; Fig. 1A) whereas hairpin elongation was proposed for extension of single-stranded DNA with palindromes (Ogata & Miura, 2000; Ogata & Morino, 2000; Fig. 1B). Applying temperatures near the melting point of the hairpin-coil transition in the former or in the starting duplex in the latter profoundly facilitate such expansions. Total repetitive DNA synthesis is greatly accelerated if hairpin elongation is combined with endonucleolytic digestion of obtained long repetitive products (Liang et al., 2004; Fig. 1C).

The MPR, which was initially developed to produce artificial proteins containing repetitive motifs (Shiba et al., 1997), can be used to study not only DNA repeat propagation but their enigmatic generation as well. In this system, a medium-sized (40-50 nt) non-repetitive homoduplex DNA (HD) evolves into multiple head-to-tail repetitive products during heat-cool cycles in the PCR.

In this chapter we consider this reaction, in strict terms of physical and polymer chemistry, to decipher the steps composing it and the mechanisms underlying their consecutive development. Several unorthodox views and consistent experimental results are described that obey basic thermodynamic rules, and analogous in vivo reactions are discussed in light of repeated motifs observed among subspecies of an entomopathogenic bacterial species.

Fig. 1. in vitro systems for repeat expansion of a homo-duplex (A), single stranded (ss) DNA, (B) and ss-DNA with administered endonuclease (C). Black arrows, forward repeating units; grey arrows, reverse repeating units
Tandem DNA Repeats: Generation and Propagation in the Microgene Polymerization Reaction and in vivo

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequencea</th>
<th>GC (%)</th>
<th>Tm b (°C)</th>
<th>∆Gos c (kcal mol⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOMUL</td>
<td>5'-GGAAATAGAAGAAGATCTTTAGATAGTTACACGC-3'</td>
<td>28.6</td>
<td>67 (68)</td>
<td>-43</td>
</tr>
<tr>
<td></td>
<td>3'-CCCTTATCTTCTGGAATTAGAAATATCTCTCGTGAGG-5'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOMU</td>
<td>5'-GGTGATAGAAGAACTTAAATCTTTATTAGAAATAATCTCTGC-3'</td>
<td>33.3</td>
<td>68 (68)</td>
<td>-45</td>
</tr>
<tr>
<td></td>
<td>3'-CCACTATCTTCTTGGAGAATTTAGAAATAATCCTCTGGACG-5'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVNA</td>
<td>5'-GGTGATAGAAGTAATTAAATCTTTATTAGAAATAATCTCTGC-3'</td>
<td>38.1</td>
<td>71 (72)</td>
<td>-48</td>
</tr>
<tr>
<td></td>
<td>3'-CCACTATCTTCTTGGAGAATTTAGAAATAATCCTCTGGACG-5'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVNAH</td>
<td>5'-GGTGATAGAAGTAATTAAATCTTTATTAGAAATAATCTCTGC-3'</td>
<td>42.9</td>
<td>73 (74)</td>
<td>-50</td>
</tr>
<tr>
<td></td>
<td>3'-CCACTATCTTCTTGGAGAATTTAGAAATAATCCTCTGGACG-5'</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Bold type letters indicate differences in composition between EVNA and NOMU. Underlined bold type letters indicate differences in composition between NOMUL and NOMU and between EVNA and EVNAH.
b Experimental and UNAFold-predicted (in brackets) melting temperature of the corresponding OS types.
c Association free energy (Visual OMP6-predicted) of complementary strands into OS at 37°C.

Reprinted from Itsko et al., 2008b with permission from Elsevier.

Table 1. Original homo-duplexes, called Original Singlets (OS) used in MPR research

Fig. 2. Demonstration of MPR with EVNA OS (Table 1) and its concentration-dependence on the reactants. Change of product length with concentrations of EVNA with Vent (A, B) and of dNTP (C with Taq, D with Vent) on 0.8% (A, C) and 2.5% (B, D) agarose gels. (A, B) Lanes 1-7, 10.4, 5.2, 2.6, 1.3, 0.64, 0.32, 0.16 μM of EVNA, respectively. Lanes 8 or M, DNA ladder. (C, D) The following concentrations (in μM) of each dNTP were used: lane 1, 100; lane 2, 200; lane 3, 300; lane 4, 400; lane 5, 500; lane 6, 600; lane 7, 700; lane 8, 800; lane 9, 900; lane 10, 1,000. Concentration of the EVNA homo-duplex, 5.6 μM. Lanes M, DNA ladder. Reprinted from Itsko et al., 2008a with permission from Elsevier
2. Demonstration of the MPR

MPR can easily be demonstrated by applying long (up to 64 cycles) end-point-detection PCR heat/cool cycling conditions on reaction mixtures containing DNA in the form of homo-duplex (HD) of length 40-50 bp, the sequence of which does not include any repetitive motif (e.g., Table 1). The PCR product is composed of long DNA stretches of heterogeneous length that is visualized as a smear on loose agarose gels and resolvable into discrete bands of HD multiples on dense gels (Fig. 2).

3. Overall scheme of MPR

The MPR is kinetically divided into three stages: initiation, amplification and propagation (Fig. 3), each is subdivided into a number of steps that are considered later in the chapter. Here they are formulated in a simplified way with the following variables and parameters:

- \( HD_i \) - homo-duplexes of DNA containing a variable number \( i \) repeats. Correspondingly, \( HD_1 \) is original non-repetitive homo-duplex called also \( OS \) (original singlet) with which the MPR starts, and \( HD_2 \) is doublet of homo-duplexes designated as \( D \), including initial doublet (ID) that triggers propagation.
- \( k_i \) is the constant rate of MPR initiation.
- \( k_{Ampl} \) is the constant rate of ID amplification by \( OS \).
- \( k_{Pr} \) is the constant rate of MPR repeat propagation per PCR cycle, assumed to be independent of polymer length (measured in repeat units \( n \)).

The symbols embraced in square brackets designate concentrations of corresponding DNA species.

**Initiation**

The minimal repetitive unit that is prone to expand by staggered re-annealing and replication of overhangs (Fig. 3B) is \( D \). Propagation is therefore initiated by generation of a so-called initial doublet (ID). The equation formulating the simplified process of initiation (Fig. 3 A) is:

\[
2OS \overset{k_i}{\longrightarrow} ID .
\]

The mechanism of this reaction and its molecularity will be discussed in sections 5 and 6.

**Amplification**

The initiator (ID) can be amplified by the \( OS \) (Fig. 3 B) according to:

\[
OS + ID \overset{k_{Ampl}}{\longrightarrow} 2D .
\]

Since \( ID \) and \( D \) are the same molecular species generated by different mechanisms, the reaction describes an autocatalytic process, in which the mass concentration of the initiator is rapidly brought to that of \( OS \) (see section 7 below). If \( k_{Ampl} = k_{Pr} \), the amplification stage is kinetically included in the following propagation stage.
Propagation
After generation of doublets, the number of repeats in DNA is expanding according to

\[ HD_i + HD_j \xrightarrow{k_p} 2HD_n, \]  

where \( n \leq i + j - 1 \), since at least one repeat is always hidden in the overlap paired region (Fig. 3C).

Propagation is also an autocatalytic process resulting in an exponential growth of the number of repeats per one polymer molecule \( \langle n \rangle_N \). This is justified as follows. In the extension reaction (eq. 3), two molecules of lengths \( n_1 \) and \( n_2 \) (\( n_1 \leq n_2 \)) yield two molecules with lengths in the range between \( n_1 + 1 \) and \( n_1 + n_2 - 1 \). Assuming that the population of such repetitive products is uniformly distributed with a common difference of one repeat unit,
the average length of the product $\langle n \rangle = \frac{(n_1 + 1) + (n_1 + n_2 - 1)}{2} = \frac{n_1 + n_2}{2}$. Averaging over all possible reactions of this sort yields

$$\langle n \rangle_{N+1} = \langle n_1 \rangle_N + \langle n_2 \rangle_N / 2 .$$

Since $\langle n_1 \rangle_N = \langle n_2 \rangle_N = \langle n \rangle_N$ (belonging to the same distribution), the average length of the polymer increases by a factor of 1.5 at each cycle:

$$\langle n \rangle_{N+1} = 3\langle n \rangle_N / 2 \text{ or } \langle n \rangle_N = \langle n \rangle_0 \times 1.5^N .$$

Propagation is finished when nucleotides (dNTPs) are depleted.

4. Extent of polymerization in MPR

The final polymer length in chain-growth polymerization reactions is determined by the “kinetic chain length” defined as the number of monomer units consumed in the propagation stage per active center produced in the initiation stage (“Flory, 1953; Atkins, 1994) as visualized in Fig. 4. Accordingly, the mean MPR product size increases with decreasing OS and increasing dNTP concentrations (Itsko et al., 2008a). The nucleotide concentration [dNTP] presented in the reaction mixture determines the extent of propagation, whereas that of the [ID] (equal to [OS]₀ due to amplification (eq. 2)) represents active centers. Thus, in MPR-produced multiple repetitive DNA, the final length

$$\langle n \rangle = \frac{1}{m} \times \frac{[dNTP]}{[OS]_0}$$

where $m$ is the number of nucleotides composing one OS. Eq. 6 is concordant with the experimental results (Fig. 2).

![Fig. 4. Extent of polymerization in the MPR. Illustrative demonstration of its dependence on concentrations of dNTP (blue spots) and ID (red stars), at given concentrations of the reactants (A), at twice the concentration of ID (B) and at twice the concentrations of both (C). I, initial reactants; F, final products](www.intechopen.com)
5. Molecularity of the initiation stage

Progress of MPR can be followed by Real-Time PCR (Fig. 5). The initial concentration of OS \([OS_0]\) determines the cycle number (the operative threshold cycle) \(N_{th}\) (in cycle units \(N\)) at which the signal (the microgene expansion process) is first detected. A 2-fold decrease in \([OS_0]\) delayed \(N_{th}\) by approximately 4 cycles (i.e., \(\Delta N_{th} = 4\); Fig. 5).

Following the initiation process, total MPR products (\(Tot\)) are propagated exponentially formulated by

\[
[Tot]_N = [ID] \times (1 + E)^N
\]  

where \(E\) is the amplification efficiency per cycle. Average \(E\) determined experimentally is about 0.7 (between 0.8 and 0.57), larger than that calculated from general considerations (eq. 5). The generation-rate of the initial doublet \(ID\) is assumed to be proportional to the power \(m\) (molecularity) of the concentration of OS \(d[IDS]/dN \propto [OS]^m\) (Fig. 3, A, B). The rate of amplification that has exponential nature (eq. 2) is much higher than that of the initiation, and hence the consumption of OS molecules in the initiation process is negligible compared to that during the amplification stage. The initiation process would therefore not affect \([OS_0]\) hence \([ID]\) would be \(\propto N \times [OS]^m_0\). The MPR initiation is reasonably assumed (Itsko et al., 2008a) to occur in all \([OS_0]\) at the first cycle \(N = 1\) (though the low sensitivity of RT-PCR detects the products at a later stage, i.e., at the \(N_{th}\)). The experiments were performed with a series of 2-fold-\([OS_0]\) values (Fig. 5). The predicted ratio between \([ID]\) of two successive dilutions is thus

\[
\frac{[ID_2]}{[ID_1]} = \left(\frac{2 \times [OS_0]^m}{[OS]^m_0}\right) = 2^m .
\]  

The exponential propagation of \(Tot\) following generation of \(ID\) depends on the number of cycles and amplification efficiency (eq. 7). If \(E\) (derived from the slope in each line of Fig. 5) remains constant, the ratio between successive \([ID]\) values in the RT-PCR experiments with two-fold different initial concentrations of OS may be derived at points with equal amount of \([Tot]_{N_{th}}\) as for example the threshold points (Fig. 5) according to

\[
\frac{[ID_2]}{[ID_1]} = (1 + E)^{N_{th1} - N_{th2}} ,
\]  

and the molecularity \(m\) of the reaction can be derived by equating eq. 8 with eq. 9:

\[
m = (N_{th1} - N_{th2}) \times \frac{\ln(1 + E)}{\ln 2} .
\]  

The observed values of \(E\) however, do vary slightly (0.8–0.6) between successive lines (Fig. 5). A molecularity of 3.1 (ranging between 2.6 and 3.4) for the initiation process was estimated, using the average \(E\) value (0.71). It can thus be concluded that three OS must somehow interact to initiate doublet formation in the MPR.
Fig. 5. Relative total MPR products generated by RT-PCR with the following initial EVNA homo-duplex concentrations (in μM): closed circles, 0.32; open circles, 0.16; closed squares, 0.08; open squares, 0.04. Reprinted from Itsko et al., 2008a with permission from Elsevier

6. Kinetics and thermodynamics of initiation

Third-order kinetics of initiation leads to a simple mechanism for the generation of ID: a rare and reversible association between three OS generates a nucleation complex (NC), (Fig. 6), which converts to ID according to:

\[ 3OS \rightleftharpoons NC \rightarrow \text{ID} + 2OS \]  
(11),

where I and II denote first (non-enzymatic fast equilibrium) and second (enzymatic rate-limiting) stages in the MPR initiation process.

One of these three OS (labeled in grey) aligns and bridges the other two, fixing them in the required proximity for the DNA polymerase to skip the inter-template gap while displacing the confronting non-template strand of the adjacent OS. This bridging occurs putatively through occasional Watson-Crick bonds between aligning and aligned homo-duplexes. Such putative bridging complexes are not covalently bonded, very unstable and have not been demonstrated directly. Experimental system for their revelation has still not been elaborated, but they can be predicted using the following software packages:
1. Visual OMP6 visualizes different 2nd-rank structures that can be potentially formed from constituent strands of OS (Fig. 7).
2. UNAfold is not so illustrative but much more comprehensive in providing schematic landscape of all hybridization patterns possible between complementary strands of OS and their temperature dependence (Fig. 8).

Fig. 6. A model for the MPR initiation. OS, original singlet HDs (a pair of complementary primers); NC, nucleation complex (an arrangement of three OS); grey strands correspond to the OS bridging the gap; GZ, gap zipper (conformation composed of half NC (hNC)) and a DNA polymerase-generated (by gap skipping) single strand of ID. Adapted from Itsko et al., 2008b with permission from Elsevier

UNAfold probability dot plots demonstrate that increasing temperature increases probabilities of generation of overall 2nd-rank structures and those that can be involved in NC (Fig. 8) among them. The van’t Hoff equation that formulates this tendency is:

$$\frac{d \ln K_{1 \rightarrow 2}}{d(1/T)} = -\frac{\Delta H_{1 \rightarrow 2}}{R},$$

where $K_{1 \rightarrow 2}$ and $\Delta H_{1 \rightarrow 2}$ are, respectively, the equilibrium constant and the standard enthalpy for the transition between more stable 1st-rank (OS) and less stable 2nd-rank (NC) structures in I part of eq. 11. $\Delta H_{1 \rightarrow 2} > 0$ because the stability of 2nd-rank structures is lower. Therefore, as the temperature increases ($1/T$ decreases), $\ln K_{1 \rightarrow 2}$ and hence $K_{1 \rightarrow 2}$ must increase, reflecting a rise in the probability of the 2nd-rank conformations.

Generation of ID from NC (Fig. 6, part II of eq. 11) requires keeping inside of NC 1st-rank...
structure of OS for DNA polymerase to skip the inter-template gap by extending the pre-
hybridized OS constituent strands. On the other hand, the equilibrium constant of the
formation of this 1st-rank structure from separated complementary DNA strands will
decrease with temperature according to the same van’t Hoff equation (eq. 12) in which \( \Delta H < 0 \) due to exothermic property of the hybridization. Decrease in this constant results in
impeding this polymerization step and overall amplification and propagation stages.
Thus temperature rise increases the rate of initiation but decreases the rates of
amplification/propagation.

<table>
<thead>
<tr>
<th>Structure</th>
<th>( \Delta G_{hNC} ) 67ºC (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOMUL</td>
<td>1.80</td>
</tr>
<tr>
<td>NOMU</td>
<td>2.78</td>
</tr>
<tr>
<td>EVNA</td>
<td>-1.34</td>
</tr>
<tr>
<td>EVNAH</td>
<td>-0.36</td>
</tr>
</tbody>
</table>

Fig. 7. Suggested structures of hNC (half NC) of OS types with approximate \( \Delta G_{hNC} \) of their
formation from single strands (at corresponding \( T_m \)), predicted by Visual OMP6.
Conventional (|) and G:T pairings. Reprinted from Itsko et al., 2008b with permission from Elsevier.

Following MPR kinetics by RT-PCR at different temperatures can test the above
reasoning. Backward extrapolation of the amplification curves with different types of OS
(next paragraph) to the first cycle retrieves the ratio \([ID]/[OS]_0\) and the value of \( \ln(1+E) \)
designated as amplification rate (A). The variability in changes of A (Fig. 9A) and
\([ID]/[OS]_0\) (data not shown) with temperature is exceedingly high and not amenable for
analysis. However these two parameters are strictly related in an exponential mode (Fig.
9B).
Fig. 8. Hybridization-probability dot-plot of EVNAH (A) and NOMUL (B). Each symbol designates pairing between strands’ nucleotides. Filled-in squares designate pairs existing in OS at all tested temperatures with probability close to 1. Open circles (enclosing dots), 2nd rank pairs at 68°C for NOMUL and 74°C for EVNAH. Dots, pairs existing at 71°C for NOMUL and 77°C for EVNAH, some of which (enclosed in open squares) form bridging structures (hNC) leading to NC (Fig. 7). Crosses, conformations existing at 74°C for NOMUL and at 80°C for EVNAH. Reprinted from Itsko et al., 2008b with permission from Elsevier.

As \( A \) rose, the calculated ratios \([ID]/[OS]\)_0 exponentially declined over seven decades. Increased \( A \) corresponds to a decrease in the applied temperature (Fig. 9A); hence, formation of \( ID \) is accelerated as temperature rises (Fig. 9B), consistent with the above reasoning. The high variability in the above parameters when they are plotted against temperature (Fig. 9A) seems to result from limited accuracy of the temperature maintenance by the heat block of the RT-PCR apparatus for the duration of each cycle (repeatedly launched by rapid cooling from 95°C to the desired temperature). The actual temperature that acts on each multiple sample may therefore differ from the registered one. The temperature-dependent amplification rate \( A \) (Fig. 9B) was used alternatively as a more stable indicator because it reflects the average temperature during the cycle. At each calculated \( A \), the ratio \([ID]/[OS]\)_0 for NOMUL was lower than that for EVNAH, consistent with lower stability of the putative bridging structure involved in NC formation of NOMUL than that for EVNAH (Fig. 7). In addition, at \( A=0.46 \) for NOMUL and 0.51 for EVNAH, the ratio lines were drastically bent (Fig. 9B), justified by lowering the fraction of hNC structures (Fig. 7) in overwhelmed number of emerging alternative pairings between constituent OS single strands at relatively high temperatures (Fig. 8, crosses) that correspond to lower amplification rates.
Fig. 9. A, Arrhenius plots of MPR amplification rate, with NOMUL (open circles) and EVNAH (filled-in circles). B, Dependence of calculated $[ID]/[OS]_0$ on MPR amplification rate, with NOMUL (open circles) and EVNAH (filled-in circles). Reprinted from Itsko et al., 2008b with permission from Elsevier.
7. Kinetics of amplification

Kinetics of the MPR amplification stage demonstrates biphasic behavior for three of the four types of OS used (Fig. 10). Single-burst kinetics was seen with NOMU at all tested temperatures and with the rest of OSs at temperatures exceeding by 2-4°C the melting point of corresponding OS. The biphasic kinetics is explained by two parallel processes (with

Fig. 10. MPR kinetics with four OS types: NOMUL (open circles, 68°C; filled-in squares, 69°C; open triangles, 70°C; filled-in circles, 71°C; filled-in triangles, 72°C). NOMU (open circles, 68°C; open squares, 69°C; open triangles, 70°C; filled-in circles, 71°C; filled-in squares, 72°C; filled-in triangles, 73°C). EVNA (open circles, 71°C; open squares, 72°C; filled-in circles, 73°C; open diamonds, 74°C; filled-in triangles, 75°C). EVNAH (open circles, 71°C; open squares, 72°C; filled-in circles, 73°C; filled-in squares, 74°C; filled-in triangles, 75°C). Arrowheads point at biphasic (grey) and single-burst (black) kinetics. Reprinted from Itsko et al., 2008b with permission from Elsevier
distinct kinetic parameters) involved in the MPR expansion, amplification of the ID by OS and propagation of doublets (D) to triplets (T) and higher-level multiples through staggered conformations (SD) (Fig. 3C) Existence of an amplification stage was suggested for the first burst in the progressively increased fluorescence intensity during MPR (Fig. 10).

The amplification (Fig. 11) can be formulated by the chemical equation

\[ \text{OS} + \text{D} \xleftrightarrow{I} 2\text{H} \xrightarrow{II} 2\text{D}. \]  

where I and II denote first (fast equilibrium) and second (rate-limiting) steps of the process.

The rate of amplification (II) can be expressed as

\[ \frac{d[D]}{dt} = k[H]. \]

where \( k \) is the rate constant of the fast-equilibrium step and \( H \) is the hetero-duplex composed of hybridized OS and D (Fig. 11). At temperatures higher than \( T_m \) of OS (and obviously of \( H \)) and lower than that of D, step I involves melting of OS, fraying the ends of D and hybridization of the single strands of the former with the latter.

Assuming that this step occurs near its multi-state equilibrium with \( K_H = [H]/[\text{OS}][D] \), the overall process is expressed by

\[ d[D]/dt = kK_H[\text{OS}][D]. \]

\( k \) rises with temperature, but \( K_H \) decreases due to the dissociation of \( H \) into D and single stranded OS. In the reaction described by eq. 13, \( k \) rises less than the drop in \( K_{H} \), so that the overall value \( kK_H \) decreases with temperature. That is reflected in the highly negative value (around -62 kcal mole\(^{-1}\)) of \( \Delta G^\ddagger_{\text{Ampl}} \) derived from the Arrhenius plots for NOMUL and EVNAH (Fig. 9A). In other chemical processes as well, an exothermic fast-equilibrium stage leads to negative \( \Delta G^\ddagger \) of the overall process. Solving eq. 13 (for detailed derivation see Itsko et al., 2008b) yields an expression for total fluorescence (\( \text{Flu}_{\text{Tot}} \)):

\[ \text{Flu}_{\text{Tot}} = \alpha[\text{OS}]_0 \left( 1 + (\gamma - 1) \frac{(e^{AN} - 1)}{R + e^{AN}} \right), \]

where \( A = kK_H e([\text{OS}]_0 + [D]_0), \) ([OS]\(_0\) and [D]\(_0\) are initial concentrations of OS and D correspondingly where [D]\(_0\) is actually [ID], \( N \) is the number of cycles and \( \varepsilon \) is the cycle period, \( R = [\text{OS}]_0/[ID], \) \( \alpha \) is the arbitrary coefficient expressing the fluorescence brightness of OS; \( \gamma \) is the ratio between the values of the fluorescence brightness of D and OS. Eq. 15 was used to approximate the first burst in fluorescence intensity of MPR curves (Fig. 10) and retrieve \([\text{ID}]/[\text{OS}]_0\) ratio and amplification rate \( A \) from them.

### 8. Thermodynamics of transition from amplification to propagation

Denatured single strands of doublet (ssD) can be hybridized in fully aligned manner generating doublet homo-duplex (D, right in Fig. 12A) or in staggered manner generating staggered doublet (SD, left). Transition from amplification to propagation is mediated by the appearance of SD. Obviously SD has lower stability than D due to half number of hydrogen bonds (\( \Delta G_{SD\rightarrowD} < 0 \)). SD cannot readily return to D due to the energetic barrier (\( \Delta G^\ddagger_{SD\rightarrowD} \)) between these conformations, the magnitude of which is determined by the...
stability ($\Delta G_{WSD}$) of the additional 2nd-rank intermediate structure called weak staggered doublet (WSD) (Fig. 12B). The more stable it is, the more easily SD switches to D due to smaller $\Delta G_{SD \rightarrow D}$. SD switching to D prevents the propagation process and results in the biphasic mode of expansion (Fig. 10). Disappearance of the biphasic kinetics (Fig. 10) at temperatures higher than 71°C for NOMUL, 73°C for EVNA and 76°C for EVNAH (Table 1) is explained by increased probability of SD with temperature rise. UNAFold also predicts formation of SD at temperatures higher than: 73°C for NOMUL, 78°C for EVNA and 80°C for EVNAH (Fig. 13A). Formation of SD as a structure that is less stable than fully aligned hybridized D is stimulated by rising temperature according to the van't Hoff equation (eq. 12).

Fig. 11. The model for amplification: OS, Original Singlet HD (a pair of complementary primers), D, Doublet. $H_1$ and $H_2$, hybridized denatured D and OS. Adapted from Itsko et al., 2008b with permission from Elsevier

Melting the doublets at high temperature and their quick-cooling afterwards can demonstrate SD conformations. Quick cooling prevents the D constituent strands from finding the final most stable fully-aligned conformation when they encounter each other but entrap them in first available conformation, mostly a staggered one. Such procedure was accomplished on MPR end products generated from all four OS types, with concentrations of dNTP and OS that limit the extent of expansion (Eq. 6 and Fig. 4), and revealed that out of the four, the OS of NOMU was the only one to diminish (bottom band, lane 5 in Fig. 13B), most likely due to its pairings with overhangs formed by staggered structures of multiple-repeated DNA. Thus, among all used OSs staggering of D of NOMU is the most facilitated process. It is also consistent with the observation that only this OS displayed single burst kinetics at all tested temperatures in RT-PCR (Fig. 10).

The facilitation of staggering and of transition from amplification to propagation stage in the NOMU case is reasoned by the least stability (highest $\Delta G_{hNC}$; Fig. 7) of its bridging structure among all used OSs, that would reflect also the least stability of its WSD (Fig. 12B) because the latter is just a staggered tandem OS.
Fig. 12. Schematic free energy diagram of the various conformations that participate in the suggested transition from amplification to propagation. Dashed lines denote energy levels. \( G_{ssD} \) and \( \Delta G_{D \rightarrow T} \) are, respectively, the free Gibbs energy of single stranded \( D \) and of the difference between \( T \) and \( D \). B, WSD conformations (tandem staggered OS) of the four \( D \) types (related to Fig. 7). (Conventional (|) and G:T pairings). Adapted from Itsko et al., 2008b with permission from Elsevier.
Fig. 13. Revealing SD structures *in silico* (A) and *in vitro* (B). A, Probability dot plot of hybridization of EVNAH’s doublet. Each symbol designates pairing between strands’ nucleotides. Filled-in squares designate pairs existing in \( D \) at all tested temperatures with probability close to 1. Open circles (enclosing dots or not, respectively) designate 2nd-rank pairs at 80ºC or 81ºC, with probabilities of approximately \( 10^{-6} \). Arrows point at two lines corresponding to two SD, effective for propagation (black) and not (grey). B, MPR-generated products after 65 PCR cycles from 4 OS types, separated on 2.5% agarose gel. Lanes 1 and 2, with EVNAH; 3 and 4, EVNA; 5 and 6, NOMU; 7 and 8, NOMUL. Odd numbers designate samples denatured (10 min at 95ºC) and then cooled rapidly. Reprinted from Itsko et al., 2008b with permission from Elsevier.
9. Kinetics of propagation

The overall process of MPR propagation includes the following recurring set of three steps:

1. Aligning \( A_i \) forward and \( B_j \) reverse complementary DNA strands containing \( i \) and \( j \) repeats in staggered mode according to the equation:

\[
A_i + B_j \xleftarrow{k_1,k_{-1}} A_iB_j, \tag{16}
\]

where \( k_1 \) and \( k_{-1} \) are rate constants of alignment and of melting, respectively.

2. Association of DNA polymerase with above staggered homo/hetero-duplex:

\[
A_iB_j + \text{Enz} \xleftarrow{k_2,k_{-2}} A_iB_j\text{Enz}, \tag{17}
\]

where \( k_2 \) and \( k_{-2} \) are rate constants of association and of dissociation, respectively.

3. DNA-polymerase mediated filling-in overhangs according to:

\[
A_iB_j\text{Enz} \xrightarrow{k_3} A_n + B_n + \text{Enz}, \tag{18}
\]

where \( k_3 \) is the rate constant of the enzyme-driven polymerization (turnover number) for filling in \((n - i)\) repeats.

The mathematical descriptions of these steps are:

1. \[
\frac{d[A_iB_j]}{dt} = k_1[A_i][B_j] - k_{-1}[A_iB_j] - k_2[A_iB_j][\text{Enz}] + k_{-2}[A_iB_j\text{Enz}], \tag{19}
\]

2. \[
\frac{d[A_iB_j\text{Enz}]}{dt} = k_2[A_iB_j][\text{Enz}] - k_{-2}[A_iB_j\text{Enz}] - k_3[A_iB_j\text{Enz}], \tag{20}
\]

3. \[
\frac{d[A_n(i,j)\text{Enz}]}{dt} = k_3[A_iB_j\text{Enz}], \tag{21}
\]

where square brackets designate concentrations of corresponding species and \( A_{n(i,j)} \) is \( A_n \) generated from given \( i \) and \( j \).

The rate of generation of intermediates such as \([A_iB_j]\) and \([A_iB_j\text{Enz}]\) is taken close to zero according to the assumption of steady-state kinetics for them (Atkins, 1994). The final expression for the above chemical reaction is:

\[
\frac{dA_{n(i,j)}}{dt} = \frac{k_3[\text{Enz}_{\text{tot}}][A_i][B_j]}{\text{Enz}_{\text{tot}}k_3/k_1 + K_D^{-1}K_M + [A_i][B_j]}, \tag{22}
\]

(for derivation see Itsko et al., 2009), where \( K_M^{\text{app}} = K_M^{-1}K_M + K_D^{-1}K_M \), \( K_M^{\text{app}} = \text{Enz}_{\text{tot}}k_3/k_1 \), \( \text{Enz}_{\text{tot}} \) is the total concentration of enzyme, \( K_D = k_1/k_{-1} \) is the equilibrium constant for duplex formation, \( K_M = (k_{-2} + k_3)/k_2 \) is the Michaelis-Menten constant.

Assuming \( t = N\epsilon \), where \( N \) is the number of cycles and \( \epsilon \) is the cycle period, yields

\[
\Delta A_{n(i,j)} = \frac{k_{\text{pol}}[\text{Enz}_{\text{tot}}][A_i][B_j]}{\text{Enz}_{\text{tot}}k_3/k_1 + K_D^{-1}K_M + [A_i][B_j]}, \tag{23}
\]
where \( k_{\text{Pol}} = k_3 \varepsilon \) and \( \varepsilon = 240 \text{ sec cycle}^{-1} \) and \( \Delta A_n \) \( [= A_n(N+1) - A_n(N)] \) is increments in \( A_n(N) \) after one cycle.

Fig. 14. Experimental versus model distributions. (A) Consecutive samples run on 0.8% agarose gel: M, DNA size markers; lane 1, cycle 16; lane 2, cycle 17; lane 3, cycle 18; lane 4, cycle 19; lane 5, cycle 20. (B) Scanned and digitized samples of cycles 17-20 (from A). (C) Matlab simulations. Final distributions of total MPR product at cycles 14-23. Model parameters: \( k_{\text{Pol}} = 0.09 \text{ repeat/cycle} \) \( (k_3 = 3.75 \times 10^{-4} \text{ repeat/sec}) \), \( K_{M}^{\text{app}} = 1 \text{nM}^2 \) (Eq. 22); \( k_i = 10^{-24} \text{nM}^{-3} \text{sec}^{-1} \) (Eq. 1); \( E = 0.735 \) (Eq. 7). (D) Juxtaposition of the model (full lines for cycles 14-23) with four (cycles 17-20) experimental distributions (crosses, open circles, open squares and open diamonds, respectively), each integrally normalized. Reprinted from Itsko et al., 2009 with permission from Elsevier

Calculation of the overall change \( A_n \) during a PCR cycle requires consideration of inflow to the \( A_n \) category from all the possible combinations of \( A_i \) and \( B_j \) that together allow the generation of such product and outflow of \( A_n \) to a group of longer lengths:

\[
\Delta A_n = k_{\text{Pol}} \left( \sum_{i=1}^{n-1} \sum_{j=n-i+1}^{\infty} \Delta A_{n(i,j)} + \sum_{j=2}^{\infty} \frac{(j-1)\Delta A_{n(n,j)}}{n+j-1} \right)
\]  

(24)
A similar equation may be formulated for $\Delta B_n$. Since the initial concentrations of the complementary DNA strands dealt with here are identical, $A_1(0) = B_1(0)$, the propagation kinetics of $A_n$ and $B_n$ are identical, hence $B_i = A_i$.

The above set of difference equations (limited to 3,000) describing kinetics of MPR was solved numerically using Matlab 7 (MathWorks, Natrick, MA (http://www.mathworks.com)). The experimental results were compared with the model (Fig. 14) and satisfactory resemblance was found between them. A discrepancy was revealed between the skewness of the experimental and modeled distributions of final MPR length. The experimental distributions are skewed to longer lengths whereas the model predicts negative skewness. This means that for every $(A_i, B_i)$ involved in the extension step (see eqs. 3 and 18) the generated product is much shorter than the expected maximum ($(i+j-1)$). This is consistent with lower processivity of the Vent polymerase (Kong et al., 1993).

10. MPR as an evolutionary process

MPR can be viewed as an evolutionary process to develop biological polymers such as DNA. The repeat generation aspect included in the MPR brings this process even closer to a common scenario of Darwinian evolution, namely, spontaneous appearance of the physico-chemical trait (ID) that is selected (amplified) since it directly contributes to its own reproduction in the form of repeats propagation. Theoretically, two variants of ID can be generated in the reaction mixture from a given OS with comparable probabilities, cis and trans (Fig. 15). Only the cis form is prone to propagation; the trans form is "infertile" since it cannot be aligned in a staggered manner. MPR polymers indeed contain only direct repeats and never alternating inverse repeats.

![Diagram of MPR initiation](https://example.com/diagram.png)

Fig. 15. Cis (A) and Trans (B) species of ID putatively generated in MPR initiation, and their potency in the subsequent propagation. Black and grey regions are complementary to each other. Arrowheads denote 3’

The thermodynamic stabilities of second-rank structures of different OSs involved in NC formation vary depending on sequence hence they have different propensities for propagation. In MPR conducted with a mixture of different OSs, the final polymers would contain predominantly the repetitive motif corresponding to OS with the most stable NC structure. This prediction is confirmed by the finding (Saito et al., 2007) that the most abundant among 3 microgenic units in the combinatorial polymers generated from an equimolar mixture is the one with the lowest free energy of formation of the NC-promoting 2nd-rank structure (analysis not shown), thus lending strong support to the model proposed and
described here. This feature can be exploited to design special types of proteins with desired repeated motifs in vitro and perhaps in vivo as well.

11. MPR and in vivo expansion

Repeating motifs in natural proteins are widely documented and described (http://www.ncl-india.org/trips/), and their possible activity mechanisms discussed (Katti et al., 2000). The fact that they derive of encoding genes is sometimes not sufficiently considered. During studies with the entomopathogenic Bacillus thuringiensis subspecies, we encountered several such cases, where DNA repeats exist, reflecting their encoded protein motifs. Specifically, helper Orf2 proteins of some cry operons include different numbers of repetitive motifs. For example, Orf2cry2Aa, encoded by orf2 of the cry2Aa operon (Widner & Whiteley, 1989), contains 11 tandem repeats of a 15-amino-acids motif and seems to be involved in Cry2Aa crystallization (Ge et al., 1998). This motif is encoded by a 45-nt repeat in its gene orf2cry2Aa (Fig. 16A) that includes only 6 transitions and a single transversion among the 450 nt composing the first 10 repeats. All 5 additional changes are located at the 3’ end of the last (11th) repeat. orf2cry2Aa3 (Sasaki et al., 1997; Fig. 16B) includes 13 repeats with the same basic unit as orf2cry2Aa. In this gene, the first 9.5 repeats are separated from the rest by a 12 nt long ‘linker’, i.e., a non-homologous insertion. A second type of repeat unit among the known orf2 genes is found in orf2cry2Ca (Wu et al., 1991) and in orf2cry2Ac (Accession # AY007687). It differs from the repeat of orf2cry2Aa by two triplets that surround a conserved T(A/G)C found in both kinds of repeats and by an extra adjacent triplet. A string of 991 nt in orf2cry2Ca encoding almost 21 repeats of a 16-mer motif in Cry2Ca contains only 35 transitions (Fig. 16C), whereas the shorter (7 repeats) string in orf2cry2Ac is much less conserved and contains many deletions and insertions (Fig. 16D).

A. orf2cry2Aa (Access. # M23723)

```
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAT CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
```

B. orf2cry2Aa3 (Access. # D86064)

```
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
ACG TAT AAC CAA AGT CAG AAT GTA TGC CCA CAA GAT TTA GTT GAT
```

www.intechopen.com
C. orf2 cry2Ca Access. # X57252

ACG TAT AAC CAA AGC CAA AAT GCCG TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAG AGC CAA AAT GGT GCC TGC GGC CAA GAT TTA AGT GAT
ACG TAT AAC CAA AGT CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT

D. orf2 cry2Ac (Access. # AY007687)

ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT

E. cry11Bb2 (Access. # HM068615)

ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA GTG GAT
ACG TAT AAC CAA AGC CAA AAT GGT GCC TAC GGC CAA GAT TTA AGT GAT

Fig. 16. Individually aligned repeated motifs found in different B. thuringiensis (Bt) subspecies. A. orf2 cry2Aa from Bt subsp. kurstaki HD-1; B. orf2 cry2Aa3 from Bt subsp. sotto. C. orf2 cry2Ca from Bt S1; D. orf2 cry2Ac from Bt; E. cry11Bb2 from Bt strain K34. Fonts: red, pyrimidine transitions; blue, purine transitions; underlined, transversions. Highlights: green, the conserved triplet in the middle of the motif; yellow, region of varied length (modulus 3); grey, sequence homology between cry11Bb2 and orf2; red, non-homologous insertions.

Repetitive motifs are encountered in certain cry genes themselves. The newly discovered Cry11Bb2 (Melnikov et al., 2010) contains 7 tandem repeats of a 16-amino acids motif, the role of which is still to be determined. Seven out 16 triplets constituting the repetitive motif of Cry11Bb2 (highlighted grey in Fig. 16E) are identical/homologous in composition and location to those found in previously described orf2 genes (Fig. 16, A-D). Repeated blocks have been found in other mosquito larvicidal δ-endotoxins (de Maagd et al., 2003): Cry11Ba1 (four repeats), Cry11Bb1 (5 repeats), Cry20Aa (eight), Cry27Aa (8 repeats).
The spectrum of replacements in the repeats of orf2cry2Aa and orf2cry2Ca is reciprocal on two sides of the conserved T(A/G)C triplet: in its 5’ part there is a bias to pyrimidine transitions whereas purine transitions are exclusively observed at its 3’ part. Intriguingly, such pattern of changes may point to occurrence of double strand break (DSB) in the middle of the repeat. Repair of DSB is mediated by trans-resection of opposite strands surrounding it in a way that transiently exposes generated ssDNA stretches to increased mutagenesis with characteristic pattern of transitions encompassing the DSB (Yang, 2008). Moreover, much evidences has accumulated in eukaryotic cells that point to blunt dsDNA ends as key intermediates in the process leading to gene amplification (Messer & Arndt, 2007; Pace et al., 2009; Mondello et al., 2010).

To survive DSB, cells exploit two major processes: homologous recombination (HR) and non-homologous end-joining (NHEJ). HR retrieves lost genetic information in error-free way from undamaged homologous template using Rad50/Mre11 complex in yeasts (Wyman & Kanaar, 2006) or RecBCD in Escherichia coli (Dillingham & Kowalczykowski, 2008). In contrast, NHEJ break repair is a ligation-like non-template-directed process that occurs between the non-homologous termini of a DSB and is therefore considered to be more error-prone. In the course of NHEJ, a DSB is repaired by attracting the Ku70/80 hetero-dimer, which recruits the ligase IV complex (comprised of ligase IV, XRCC4 and XLF) to seal the DNA ends (Mahaney et al., 2009). NHEJ, regarded as exclusive prerogative of eukaryotic cells, has recently been found in prokaryotes by identifying bacterial homologues of Ku protein (Doherti et al., 2001; Brissett & Doherty, 2009). Moreover, a functional NHEJ repair pathway is essential for spore viability in Bacillus subtilis under conditions that yield DSBs (Weller et al., 2002; Wang et al., 2006; Moeller et al., 2007). It is noteworthy that many of the bacteria that contain the Ku ligase system are capable of sporulation (B. subtilis, Streptomyces coelicolor, Sinorhizobium loti) (Weller et al., 2002). The sporulating B. thuringiensis may contain NHEJ repair pathway as well. Exogenous oligonucleotides complementary to the broken ends can efficiently target DSB for repair in yeasts (Storici et al., 2006). We propose that partial complementarity between sequences flanking the broken ends may assist their sealing and contribute to repeat generation via MPR-like slipped structures (Fig. 17, panels A-H). An example involving putative 2nd rank complementarity that can participate in the process to seal the broken ends (Fig. 17I) was revealed in Cry2Bb2 (Melnikov et al., 2010).

Once the initial doublet has been generated inside the chromosome, the repeats may propagate during DNA replication since the constituent strands may slide over each other between the multiple complementary regions. Second rank complementary structures similar to WSD in MPR (Fig. 12B) may stabilize such slipped structures. The alternative mechanism for the propagation is possible recurrence of DSBs in the middle of the generated repeats. Bridging the broken ends by occasional Watson-Crick bonds due to partial complementarity (Fig. 17D) can bring about nucleotides insertions/deletions during fill-in by DNA polymerase similar to those occurring during conversion of the NC to the ID in MPR (Fig. 6 and Itsko et al., 2008a). Selection for functionality of a given coding region can restore the lost frame by generating an additional triplet or dropping an existing one. This can result in the two/three triplets varied region (highlighted yellow in Fig. 16) surrounding the conserved T(A/G)C triplet in the 4 orf2 versions and in cry11Bb2 listed here (highlighted green in Fig. 16). The inserted linker in that region as in orf2cry2Aa3 (highlighted red in Fig. 16B) is also consistent with this explanation.
Fig. 17. Generation of sequence duplication due to DSB repair via NHEJ-like repair pathway. Chromosome region (A) suffers a DSB (B). C. Trans resection of DNA strands flanking the break by a dedicated exonuclease. D. Bridging the broken ends in staggered mode by occasional Watson-Crick bonds due to partial complementarity. E. Filling-in single stranded regions by DNA polymerase. F. Back slippage of the broken ends under repair that is mediated by a weak 2nd rank complementarity. G. Filling-in single stranded regions by DNA polymerase. H. Restored integrity of the chromosome with the generated duplicated region. Dumbbells designate schematic location of the sequence undergoing duplication. Newly synthesized DNA is designated in grey. I. Fully-aligned and two slipped structures existing between constituent strands of cry2Bb2 repeat motif that may assist to bridge the DSB ends. Bold and underlined regions designate two sets of complementarity.
It is intriguing that the AGT triplet in orf2cry2Ca (Fig. 16C) is always adjacent to CAG whereas its 5'-replaced AGC is adjacent to CAA, both combinations encode the same amino acids, Serine by AGT/C and Glutamine by CAG/A. This coincidence cannot be explained by selection for suppression of one transition by the other. Such replacements do affect stability of slipped structures presumably emerging during replication of this region. Without the replacements, almost triplet repetition AAT CAA AGT CAA AAT (Fig. 16C) would be explicitly observed that could putatively propagate in uncontrolled way through slipped structures, bringing about genetic instability. Strong selection for silent transitions that save triplet coding but prevent deleterious expansion by disrupting repetition pattern would thus be anticipated to stabilize the genome.

12. Concluding remarks

The MPR was dissected to sub-reactions and their thermodynamics and kinetics were analyzed. Different propensities of various HDs to expand into multiple repeating units were justified in terms of different stabilities of NCs engaged in MPR initiation. The proposed models, thermodynamic for initiation and kinetic for propagation, agree satisfactorily with experimental results.

The MPR with non-repetitive HD presents an optional chemical evolutionary system in which the thermodynamic advantage of very weak interactions results in biased proliferation of a certain reaction product. The learned approach to optimize MPR is necessary in protein engineering. The importance of studying this phenomenon lies far beyond applied interest; it may reflect primordial molecular evolution of primitive DNA sequences into complex genomes.

Molecular participants in the repeat expansion process and reaction conditions accompanying it in vivo obviously differ from those existing in MPR mixtures. The facilitating factor in the in vitro repeat expansion, temperatures manipulations above the Tm of expanding oligonucleotide HD, is impossible in vivo, where an ensemble of various enzymes is operating. DNA/protein complexes can select for generation of specific functionality-relevant hybridized species of DNA molecules out of competing structures. This principal difference affects only the rate of the processes and their efficiencies but not their thermodynamic feasibilities. Even strongly unfavorable energetic interactions (high ΔH values) between DNA species (putatively bringing to repeat generation and their propagation) can occur in both systems when overcome by entropic component of their multiplicity and redundancy (TΔS) (Harvey, 1997). Thus, the basic thermodynamic consideration about stabilities of 2nd rank structures underlying repeat propagation propensity can be common to the repeat expansion phenomenon in both arenas.

13. Acknowledgments

This investigation was partially supported by a grant from the United States-Israel Binational Science Foundation (BSF, number 2007-037), Jerusalem, Israel (to A.Z.), and a Levi Eshkol scholarship (to M.I.) from the Israeli Ministry of Science, Culture and Sports.
14. References


Melnikov, O.; Baranes, N.; Einav, M.; Ben-Dov, E.; Manasherob, R.; Itsko, M. & Zaritsky, A. (2010). Tandem repeats in a new toxin gene from a field isolate of *Bacillus thuringiensis* and in other cry11-like genes. submitted


Progress of thermodynamics has been stimulated by the findings of a variety of fields of science and technology. The principles of thermodynamics are so general that the application is widespread to such fields as solid state physics, chemistry, biology, astronomical science, materials science, and chemical engineering. The contents of this book should be of help to many scientists and engineers.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
