

# Computational Studies of the Mechanical Stability for Single-Strand Break DNA

Paweł Krupa,<sup>1,2,\*</sup> David J. Wales<sup>3</sup> and Adam K. Sieradzan<sup>2</sup>

<sup>1</sup>Institute of Physics, Polish Academy of Sciences, Aleja Lotnikow 32/46, PL-02668 Warsaw, Poland, <sup>2</sup>Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland,

<sup>3</sup>Department of Chemistry, Cambridge University, Lensfield Road, Cambridge CB2 1EW, U.K.

---

\*Corresponding author; phone: +48 22 116 35 45; e-mail: [pkrupa@ifpan.edu.pl](mailto:pkrupa@ifpan.edu.pl)

## Abstract

The stability of DNA is crucial for the existence of most living organisms. Even a single DNA break can lead to serious problems, and even cell death. In this work the position-specificity of single strand breaks (SSB) and the stability of short DNA fragments of various lengths and sequence repetitions ( $d(AT)_{30}$ ,  $d(ATGC)_{15}$ ,  $d(GC)_{30}$ ,  $d(TTAGG)_{12}$ ,  $d(TTAGGG)_{10}$ , and  $d(TTTAGGG)_{9}$  with SSBs and  $d(GC)$  with 2-60 repetitions without SSBs) were examined, by performing a series of steered molecular dynamics simulations using the coarse-grained NARES-2P force field. Our results show that the stability of DNA with a SSB strongly depends on the position of the break, and that the minimum length of DNA required for stability is sequence dependent.  $d(GC)_{30}$  with SSB in position  $x$  was found to be less resistant to stretching than  $d(GC)_x$  without SSB, where  $x$  is the number of  $d(GC)$  repetitions. DNA sequences with longer repeated fragments (such as telomeres) exhibit greater stability in the presence of breaks positioned at the beginning of the chain, which could constitute a cellular defence mechanism against DNA damage.

# 1 Introduction

The study of basic mechanisms involved in cell function is currently one of the most active areas in molecular science. DNA is a principal cellular component: it is responsible for carrying the genetic information of all proteins, and determines the frequency of transcription or replication. However, DNA is very prone to damage, which can lead to cancer<sup>1</sup> or neurodegenerative disease.<sup>2</sup> Moreover, DNA damage is also associated with autoimmune conditions<sup>3</sup> and diabetes mellitus type 2,<sup>4</sup> while aberrations of telomeres can lead to cancer.<sup>5</sup> Radiation,<sup>6</sup> reactive oxygen species,<sup>7</sup> and chemical reactants<sup>8</sup> are among the agents that damage DNA. In adults over 1 kg of superoxide radical  $O_2^{\bullet-}$  is produced annually.<sup>9</sup> There are several different examples of DNA damage: single strand breaks (SSBs),<sup>10</sup> the appearance of 8-hydroxydeoxyguanosine residues,<sup>7</sup> creation of polycyclic aromatic hydrocarbon adducts,<sup>11</sup> and breaks in double strand helices (DSB).<sup>12</sup> Among the various examples of DNA damage, SSBs are the most common, arising at a frequency of tens of thousands per cell per day.<sup>13</sup> Unless quickly repaired the most likely consequence of SSBs is collapse or blockage of replication forks during the S phase of the cell cycle, leading to DSB, and finally to overall DNA instability.<sup>14,15</sup> In most cases, SSBs are rapidly reversed and do not lead to DSB,<sup>16,17</sup> but otherwise they can lead to serious damage to neuronal cells and even to heart failure.<sup>18</sup>

The first attempts to study the influence of SSBs on the stability and structure of the DNA started in the 1990s, when only 1 ns simulations could be performed for very small systems.<sup>19</sup> With advances in computational power and tools, longer simulations for larger systems with higher accuracy became possible. However, with all-atom methods, we are still limited to relatively small systems. A popular size for DNA studies is often the dodecamer on a sub-microsecond timescale.<sup>20</sup> Recently, larger systems, e.g. 31 bp DNA, have become accessible, however, with run lengths still limited to 100 ns and with a low number of replicas.<sup>21</sup> For pulling velocities of 1-12 cm/s<sup>21</sup> significantly higher force peak values were observed than in experiment.

In the present contribution we consider the mechanical stability of undamaged DNA of

various lengths compared to SSB-DNA. We apply steered molecular dynamics with the coarse-grained force field NARES-2P to study the effect of the SSB position on the mechanical resistance of damaged DNA using slow pulling speeds of 0.4 mm/s and 0.02 mm/s with at least 64 trajectories per system. We study different sequences: three non-telomeric and three telomeric, which differ in their mechanical properties.<sup>22</sup> We determine if the break position has a similar effect on mechanical stability to shortening of the DNA. The advantages of using an efficient coarse-grained force field facilitate: (i) using relatively low pulling speeds of 0.4 mm/s and 0.02 mm/s, (ii) performing 64 to 120 trajectories for each system, providing 7,080 trajectories in total, (iii) using larger systems, which can exhibit stability during simulations, each consisting of at least 60 bp. Our results show that stability of the single-break DNA is both position- and sequence-dependent, and that shorter DNA fragments without defects are more stable than long DNA with SSB in one of the chains in an analogous position.

## 2 Methods

The influence of the length of broken fragments on the stability of DNA during pulling was determined using canonical SMD simulations for the NARES-2P force field, implemented in recent work.<sup>22</sup> Each simulation consisted of 64 trajectories, each of approximately 150  $\mu$ s (30,000,000 steps, each of 4.89 fs, taking into account the approximate 3 to 4 order-of-magnitude speed-up associated with the NARES-2P force field), which were continued to full dissociation of the chains. A pulling speed of approximately 0.4 mm/s and a force constant of 2 kcal/(mol  $\text{\AA}^2$ ) for the harmonic spring were used in all simulations. Additionally, a series of runs with a 0.02 mm/s pulling speed (2.45 ms, 500,000,000 steps) were run to analyse the influence of the pulling speed on the results. In previous work<sup>22</sup> we adjusted pulling speed by trial and error and found that in some cases even with a pulling speed of 0.4 mm/s system may not have enough time to equilibrate during simulations, however, most of the regrabbing phenomenon could still be observed. In all simulations, restraints (anchors for pulling) were placed on the second residue from the 5' position of each chain, so pulling was performed parallel to the long

axis of the DNA.

NARES-2P<sup>23,24</sup> is a coarse-grained model for nucleic acids derived analogically to the physics-based coarse-grained model for proteins UNited RESidue (UNRES), which has been developed for over 20 years and used mostly to predict protein structure<sup>25</sup> and dynamics.<sup>26</sup> In NARES-2P only two interaction sites per nucleic acid are employed: a united sugar-base and a united phosphate group. In both force fields the energy function is derived from a Restricted Free Energy (RFE) or Potential of Mean Force (PMF) for all-atom systems, including solvent and counter-ions. In the simplified representation, the all-atom energy function is averaged out over the remaining degrees of freedom. All energy terms in NARES-2P are analogous to UNRES, including the temperature dependence ( $f_2(T)$ ),<sup>23</sup> while the effective energy function of the NARES-2P model is expressed by eq. (1).

$$\begin{aligned}
 U_{nucl}^{NARES-2P} = & w_{BB}^{GB} \sum_i \sum_{i<j} U_{B_i B_j}^{GB} + w_{BB}^{dip} f_2(T) \sum_i \sum_{i<j} U_{B_i B_j}^{dip} \\
 & + w_{PP} \sum_i \sum_{i<j} U_{P_i P_j} + w_{BP} \sum_i \sum_j U_{P_i B_j} + w_{bond} \sum_i U_{bond}(d_i) \\
 & + w_{bend} \sum_i U_{ang}(\theta_i) + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i) + w_{rot} \sum_i U_{rot}(\alpha_i, \beta_i) \quad (1)
 \end{aligned}$$

where  $U_{B_i B_j}^{GB}$  denotes the anisotropic Gay-Berne potential of the sugar-base sites and  $U_{B_i B_j}^{dip}$  denotes the mean field dipole-dipole interaction potential describing van der Waals and electrostatic-interaction energies for sugar-base groups, respectively.  $U_{P_i P_j}$  and  $U_{P_i B_j}$  denote the phosphate-phosphate and phosphate – sugar-base interaction potentials, respectively,  $U_{bond}$ ,  $U_{ang}$ ,  $U_{tor}$  and  $U_{rot}$  denote virtual-bond-stretching, virtual-bond-angle-bending, torsional, and sugar-base-rotamer potentials, respectively. Each term is multiplied by the appropriate weight ( $w_x$ ). For more details and description of the energy function, including van der Waals and base-base electrostatic-interaction components, please refer to equations 1-14 in ref.<sup>27</sup>

Starting structures were obtained using Nucleic Acid Builder, which is a part of the AMBER package,<sup>28,29</sup> to generate ideal all-atom models of the B-DNA structures. These structures were subsequently energy minimised for the NARES-2P force field,<sup>23,24</sup> as in previous work.<sup>22</sup> Each

nucleic chain contains approximately 60 nucleotides, which gives approximately  $205 \text{ \AA}$  (6 full twists) chain length. This length was found to be sufficient to observe distinct properties for the DNA sequences.

Three non-telomeric ( $d(AT)_{30}$ ,  $d(ATGC)_{15}$ , and  $d(GC)_{30}$ ) and three telomeric sequences ( $d(TTAGG)_{12}$ ,  $d(TTAGGG)_{10}$ , and  $d(TTTAGGG)_9$ ) were used to analyse the influence of single chain breaks in B-DNA. For each of six selected sequences with a single-chain break, 59 breaking positions were studied with 64 SMD trajectories each, giving 3,776 trajectories per sequence. Additionally, stretching of the GC30 was run in two different modes: with a lower pulling speed (0.02 instead of 0.4 mm/s) and a lower pulling force constant (0.1 instead of 2 kcal/(mol  $\text{\AA}^2$ )), and  $d(TTAGGG)_{10}$  was also run with a lower pulling speed (0.02 mm/s). In total,  $9 \times 59 \times 4 = 33,984$  independent trajectories, each of approximately 150  $\mu\text{s}$ , were run and subsequently analysed.

Additionally, a separate series of different length  $d(GC)_X$  chains, where  $X$  is the number of GC repetitions in the range of 4 to 120, were tested at two different pulling speeds (0.4 and 0.02 mm/s) and  $d(TTAGGG)_{10}$  at 0.4 mm/s, to investigate the relation between chain length and maximum stretching peak value, and verify if single strand breaks have purely a shortening effect. Each simulation consisted of 120 trajectories to improve sampling, giving in total  $2 \times 22 \times 120 + 15 \times 120 = 7,080$  trajectories from 150 to 2450  $\mu\text{s}$ .

The values and positions of the main force peaks during pulling were fitted to the analytical function<sup>30</sup>

$$f_c(N) = \frac{N\Delta G_{bp}}{N\delta - \delta_0} - \frac{\Delta G_{obs} + \Delta G_0}{N\delta - \delta_0}, \quad (2)$$

where  $f$  is the shearing force,  $N$  is the number of base pairs,  $\Delta G_{bp}$  is the free energy of disruption of a single base pair,  $\delta$  is a distance, and  $\Delta G_{obs}$  is the observed  $N$ -dependent free energy for shearing. For simplicity, the assumption that the transition state is a single base-pair state was applied, as in ref. 30:  $\delta = \delta_0$ , and  $G_{bp} = G_0$ .

## 3 Results and Discussion

### 3.1 SMD simulations of DNA chains with single break

Analysis of the force needed to pull two chains of B-DNA with single breaks shows that mechanical resistance to stretching is strictly dependent on the sequence. Example plots for  $d(GC)_{30}$  are shown in Figure 1, while additional plots can be found in SI Figures S1-S9.  $d(GC)_{30}$  exhibits one main force peak, whose height grows with the number of GC repeats, while the position is similar for breaking at positions 25 to 59 with lengths of 370-390 Å. No other significant peaks are observed. For  $d(ATGC)_{15}$  there is one main force peak, which for breaks at position 25 to 59 appears around 350 to 360 Å, with a secondary peak at lengths between 400 and 500 Å, observed mainly for the longer fragments (e.g. 59).  $d(AT)_{30}$  is least resistant to stretching, and the maximum force peak value barely exceeds 5 pN. Short fragments of AT are completely unstable, while the pulling pattern for longer fragments (over 23 nucleotides) indicates gradual untangling of the DNA chains without clear secondary peaks. For telomeric sequences a single main peak is present, plus at least one secondary peak, which is related to the untangling of the first repetition. This peak is clearly visible for  $d(TTAGGG)_{10}$  at a slow pulling speed of 0.02 mm/s (Figure S8). If the DNA chain has enough time during stretching it untangles by single repetition and jumping to another chain, manifested as the sawtooth line in the Figure S8. **For the faster pulling speed this effect is diminished and not so clearly visible.** Such behaviour is not observed for  $d(GC)_{30}$  pulled under the same conditions (Figure S4).

As indicated in Figures 1 and S1-S9, there are distinctive lengths for the broken chains that cause a rapid increase of the force needed to separate them and in the peak force values: 21-25 for  $d(AT)_{30}$ , 8-17 for  $d(ATGC)_{15}$ , 12-18 for  $d(GC)_{30}$ , 9-13 for  $d(TTAGG)_{12}$ , 7-13 for  $d(TTAGGG)_{10}$ , and 9-14 for  $d(TTTAGGG)_{9}$ . Hence for each sequence a different number of nucleotides is needed to provide a stable interacting fragment of the chain. Surprisingly, the minimal length that exhibits resistant force is smallest for both telomeric sequences and  $d(ATGC)_{15}$ , not for  $d(GC)_{30}$ , which involves the largest number of hydrogen bonds stabilising the DNA. This trend appears for fragments containing up to 12 nucleotides, for which all

other sequences (except AT) require a higher force to detach the broken fragment, and are therefore more stable (Figures 2 and 3 and Table 1). Such behaviour indicates that the length of the repetitive fragment, not the number of hydrogen bonds, is the main factor governing the stability of B-DNA with single breaks.

In general, it can be seen from Figures 2 and 3 that very short DNA chains (up to 5 nucleotides) are not stable for all of the analysed sequences, and the positions of the main force peaks appear random.  $d(AT)_{30}$  is the most unstable sequence for any position of breaking. When damaged DNA is pulled  $d(AT)_{30}$  requires a much longer sequence to be stable. Surprisingly, the second most unstable sequence for short pulled fragments is damaged  $d(GC)_{30}$ , which is consistent with the diagrams shown in Figures S1-S9. On increasing the length of the pulled DNA chain,  $d(GC)_{30}$  starts to be the most stable sequence. However, for smaller fragments telomeric sequences and  $d(ATGC)_{15}$  are more stable. A similar picture can be seen from the total work needed to dissociate the DNA chains (Figure S15). Analysis of the diagram in which main force peak values are scaled by the average number of hydrogen bonds (Figure S12), suggests that this factor is significant, as shown by the much smaller differences between the plots. However, the order of stability remains the same.

The mechanism for DNA untangling consists of several stages (Figure 4): (i) stretching of DNA chains up to about 165-170% of the initial length (e.g. for  $d(TTTAGGG)_9$  stretching occurs from 207 Å to about 346 Å) maintaining most of the hydrogen bonds (in line with previous studies<sup>31-33</sup>); (ii) hydrogen bonds start to break and DNA chains slide, reforming hydrogen bonds with other bases; (iii) the 3' ends of the DNA fail to reform regular hydrogen bonds during chain sliding, and intermediate triplex structures are formed; (iv) full dissociation of the DNA chains. If SSB results in a short fragment (up to about 10 bp) at the 3' end, then during stretching it dissociates before step (i).

Fitting of the analytical function (eq. 1) to the main force peak values (Table 2 and Figure S11) produces good alignment with the data. The observed  $\Delta G_{\tau_{obs}}/\Delta G_{bp}$  values for our standard pulling speed agree with previous results,<sup>30</sup> in the range 6 to 7, compared to 4.95 to 7.98



for all sequences, except  $d(AT)_{30}$ , for which  $\Delta G_{\tau_{obs}}/\Delta G_{bp}=14.35$ . The analytical function also fits well to the position of the main force peaks (Figure S10). For all cases  $\Delta G_{\tau_{obs}}/\Delta G_{bp}$  reaches a plateau for a given length, indicating that DNA is quickly reaching maximum resistance to stretching. The lower  $\Delta G_{\tau_{obs}}/\Delta G_{bp}$ , the higher is the resistance for shorter fragments, because they become more resistant faster, and the  $\Delta G_{\tau_{obs}}/\Delta G_{bp}$  values are lowest for the three telomeric sequences. For  $d(AT)_{30}$   $\Delta G_{\tau_{obs}}/\Delta G_{bp}$  is the largest, which is also manifested by instability for very short sequences.

### 3.2 SMD simulations of DNA chains with different lengths

Analysis of the simulations shows that very short chains of  $d(GC)_2$  and  $d(GC)_4$  are not stable and can dissociate for minimal forces. Increasing the number of GC repetitions from 12 to 32, the position of the main force peak increases linearly from 4.2 to 10.1 pN. The increase of force then slows down significantly, reaching a maximum for  $d(GC)_{30}$ , and decreasing slightly for longer sequences. The decrease for the long chains is caused by the extended total simulation time, which gives the system more time to relax, and subsequently decreases the main force peak value. To check this hypothesis, an additional series of simulations was run for  $d(GC)_X$  with a pulling rate of 0.02 mm/s, i.e. 20 times slower (Figure S13). We find that the maximum value of the force peak decreases from 13.2 to 7.3 pN for  $d(GC)_{30}$ , and that the maximum force peak value shifts to longer chains (Figure S14). DNA also maintains stability for much longer stretching when a slower pulling speed is used, e.g. for  $d(GC)_{60}$  with 0.4 mm/s pulling speed most trajectories dissociate at approximately 760 Å, and only a few of them are still bound up to 910 Å, while at 0.02 mm/s, most of the trajectories are bound until reaching approximately 1100 Å (Figure 5A and S13). However, despite the differences caused by pulling speed, with increasing chain length, the maximum force peak value increases only to a limiting value and reaches a plateau (Figure 5A-B), which is consistent with experiment.<sup>30,34</sup> Differences between long-range resistance between  $d(GC)_{60}$  and telomeric sequence of  $d(TTAGGG10)_{20}$  were explained in detail in previous work.<sup>22</sup>

In general, the behaviour of  $d(GC)$  sequences with different repetition lengths is similar to

$d(GC)_{30}$  with different positions for the SSB, but not identical. The shortest length of DNA providing stability is  $d(GC)_6$  compared to position 12 in SSB  $d(GC)_{30}$ , which is analogous. However, for more  $d(GC)$  repetitions the position of the main force peak increases much faster and almost linearly, compared to  $d(GC)_{30}$  with different SSB positions (Figure 2 and Figure S15). The fit of  $\Delta G_{\tau_{obs}}/\Delta G_{bp}$  is also much better for  $d(GC)_X$  than for chains with SSB. For  $d(GC)_{60}$  with 0.02 mm/s pulling speed a secondary force peak can be observed (Figure S13) at around 1080 Å, which is the result of averaging over 120 trajectories, in which release and recapture of DNA chains occurred in several cases. It should be noted that such behaviour can be observed only for sequences of sufficient length and at sufficiently slow pulling speed.

Interesting behaviour can be seen in the total work needed to dissociate the DNA chains (Figure S17). For the short sequences (up to 26  $d(GC)$  repetitions) more work is required to dissociate the chains at faster pulling speeds, while for longer sequences (more than 28  $d(GC)$  repetitions), more work is needed for slower pulling speeds. These results may be due to the lower stability of short DNA sequences, which start to untangle during longer simulations, and by the ability of longer DNA sequences to stabilise chains by forming intermediate triplex and by reformation of duplex structures with loose ends, as observed in previous work.<sup>22</sup> The presence of various high order DNA structures was recently confirmed by experimental studies.<sup>35</sup>

These observations are confirmed by Figure S17B, showing the total work needed to dissociate the DNA chains divided by the number of  $d(GC)$  repetitions. For our standard pulling speed, the work per repetition ( $w_r$ ) increases rapidly up to 18  $d(GC)$  repetitions, then stabilises up to 22  $d(GC)$  repetitions, and slowly decreases for longer chains. Pulling speed has a significant effect on the stability of the DNA chains, and even using a relatively slow value of 0.4 mm/s, differences in equilibration time for shorter and longer chains have a significant impact on the work needed to dissociate them. For simulations with a pulling speed of 0.02 mm/s the shortest stable DNA chain is  $d(GC)_6$ , then  $w_r$  slowly increases up to 12  $d(GC)$  repetitions, and from 14 to 28  $d(GC)$  repetitions a rapid increase of  $w_r$  is observed, which is associated with the formation of marginally stable intermediate structures. Above 28  $d(GC)$  repetitions only

a minor increase of  $w_r$  is observed, which suggests that these structures gained full mechanical stability for stretching.

## 4 Conclusions

A series of MD simulations were run for DNA breaks in different positions and GC sequences with a variable number of repetitions, to examine the influence of DNA length and the position of the SSB on the stability of the chains. Our results show that breaks at the end of the DNA chain (less than 6 nucleotides from the end) cause significant instability, producing minimal resistance to pulling. Fragments longer than 6 nucleotides are relatively stable, depending on both the sequence and the length of the fragment. Although  $d(GC)_{30}$  forms the most hydrogen bonds, it is the most stable only for fragments longer than 12 nucleotides. For DNA chains damaged in positions 6 to 12, sequences with longer repetitive fragments (such as  $d(ATGC)_{15}$  and telomeric sequences) were found to be the most stable under stretching, which may explain why nature chose these fragments to protect the termini of DNA. Our simulations show that if the SSB position or length of the DNA chain is long enough, stretching up to 170% can be supported without breaking hydrogen bonds, in agreement with previous studies.<sup>31</sup> Further stretching causes disruption of the hydrogen bonds and sliding of the chains past each other. The main force peaks are related to the first part of stretching (up to 170%), and scale well with the total number of hydrogen bonds stabilising the helices. The highest peaks are observed for  $d(GC)$  sequences, and the lowest for  $d(AT)$ . Further stretching depends more on the sequence repetitions than on the number of hydrogen bonds. In addition to the previously observed influence of chemical modification of the DNA,<sup>36</sup> and simulation details, such as pulling speed<sup>37,38</sup> or initial pulling angle,<sup>39</sup> our study confirms that pulling speed plays important role. However, the most important factor is DNA length, which affords greater mechanical resistance to pulling, largely due to formation of higher order structures.

Telomeric sequences are most stable for SSBs present in short fragments, and their stability is achieved very quickly (the lowest  $\Delta G_{\tau_{obs}}/\Delta G_{bp}$ ). Moreover, because telomeric sequence rep-

etitions are longer, a sawtooth force profile is observed (clearly visible for e.g.  $d(TTAGGG)_{10}$ ), which additionally stabilises the DNA chains by producing multiple barriers to dissociation. We also observed the tailing effect of telomeric sequences seen in previous work,<sup>22</sup> which seems to be connected both to the length of the repetitive fragment and the flexibility of the DNA chains. This effect increases with AT content, and is manifested by resistance in highly stretched chains (longer than 520 Å), caused by formation of intermediate triplex and quadruplex structures. Overall, telomeric sequences appear to be optimal for protecting DNA chains from SSB, even very close to the end of a DNA chain.

## 5 Acknowledgements

The work was supported by the National Science Center (Poland) Sonata UMO-2015/17/D/ST4/00509. DJW gratefully acknowledges support from the EPSRC (EP/N035003/1). Computational resources were provided by (a) the supercomputer resources at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdansk, (b) computational resources at Interdisciplinary Center for Mathematical and Computer Modeling in Warsaw (ICM), grant GA71-23, and (c) our 682-processor Beowulf cluster at the Faculty of Chemistry, University of Gdansk.

### 5.1 Conflict of Interest statement

The authors declare that there are no conflicts of interest.

## ASSOCIATED CONTENT

Supporting Information Available:

Additional diagrams of the force needed to stretch DNA with SSB for a given distance for different repetitive fragments and conditions (Figures S1-S9), diagrams of main-force-peak values and their positions (Figures S10-S12), analogical plots for  $d(GC)_x$  without SSBs (Figure S13-S15), and plots of work (Figures S16-S17).

## REFERENCES

- [1] Totter, J. R. Spontaneous cancer and its possible relationship to oxygen metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77*, 1763–1767.
- [2] El-Khamisy, S. F.; Saifi, G. M.; Weinfeld, M.; Johansson, F.; Helleday, T.; Lupski, J. R.; Caldecott, K. W. Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature* **2005**, *434*, 108–113.
- [3] Bashir, S.; Harris, G.; Denman, M. A.; Blake, D. R.; Winyard, P. G. Oxidative DNA damage and cellular sensitivity to oxidative stress in human autoimmune diseases. *Ann. Rheum. Dis.* **1993**, *52*, 659–666.
- [4] Jackson, S. P.; Bartek, J. The DNA-damage response in human biology and disease. *Nature* **2009**, *461*, 1071–1078.
- [5] Graham, M. K.; Meeker, A. Telomeres and telomerase in prostate cancer development and therapy. *Nat. Rev. Urol.* **2017**, *14*, 607–619.
- [6] Rak, J.; Chomicz, L.; Wiczak, J.; Westphal, K.; Zdrowowicz, M.; Wityk, P.; Zyndul, M.; Makurat, S.; Golon, L. Mechanisms of damage to DNA labeled with electrophilic nucleobases induced by ionizing or UV radiation. *J. Phys. Chem. B* **2015**, *119*, 8227–8238.
- [7] Breen, A. P.; Murphy, J. A. Reactions of oxyl radicals with DNA. *Free Radical Biol. Med.* **1995**, *18*, 1033 – 1077.
- [8] Cariello, N. F.; Keohavong, P.; Sanderson, B.; Thilly, W. G. DNA damage produced by ethidium bromide staining and exposure to ultraviolet light. *Nucleic Acids Res.* **1988**, *16*, 4157.
- [9] Halliwell, B. Free radicals and antioxidants: a personal view. *Nutr. Rev.* **1994**, *52*, 253.
- [10] Lai, H. Single- and double-strand DNA breaks in rat brain cells after acute exposure to radiofrequency electromagnetic radiation. *Int. J. Radiat. Biol.* **1996**, *69*, 513–521.

- [11] Pratt, M. M.; John, K.; MacLean, A. B.; Afework, S.; Phillips, D. H.; Poirier, M. C. Polycyclic aromatic hydrocarbon (PAH) exposure and DNA adduct semi-quantitation in archived human tissues. *Int. J. Environ. Res. Public Health* **2011**, *8*, 2675–2691.
- [12] Rogakou, E. P.; Boon, C.; Redon, C.; Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J. Cell Biol.* **1999**, *146*, 905–916.
- [13] Caldecott, K. W. Single-strand break repair and genetic disease. *Nat. Rev. Genet.* **2008**, *9*, 619–631.
- [14] Caldecott, K. W. Mammalian single-strand break repair: Mechanisms and links with chromatin. *DNA Repair* **2007**, *6*, 443 – 453.
- [15] Dilley, R. L.; Verma, P.; Cho, N. W.; Winters, H. D.; Wondisford, A. R.; Greenberg, R. A. Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature* **2016**, *539*, 54–58.
- [16] Satoh, M. S.; Lindahl, T. Role of poly (ADP-ribose) formation in DNA repair. *Nature* **1992**, *356*, 356.
- [17] Negritto, M. R. Repairing Double-Strand DNA Breaks. *Nature Education* **2010**, *3*, 26.
- [18] Higo, T.; Naito, A. T.; Sumida, T.; Shibamoto, M.; Okada, K.; Nomura, S.; Nakagawa, A.; Yamaguchi, T.; Sakai, T.; Hashimoto, A.; et al., DNA single-strand break-induced DNA damage response causes heart failure. *Nat. Commun.* **2017**, *8*, 15104.
- [19] Yamaguchi, H.; Siebers, J. G.; Furukawa, A.; Otagiri, N.; Osman, R. Molecular dynamics simulation of a DNA containing a single strand break. *Radiation protection dosimetry* **2002**, *99*, 103–8.
- [20] Chakraborty, K.; Mantha, S.; Bandyopadhyay, S. Molecular dynamics simulation of a single-stranded DNA with heterogeneous distribution of nucleobases in aqueous medium. *J. Chem. Phys.* **2013**, *139*, 075103.

- [21] Landuzzi, F.; Palla, P. L.; Cleri, F. Stability of radiation-damaged DNA after multiple strand breaks. *Phys. Chem. Chem. Phys.* **2017**, *19*, 14641–14651.
- [22] Sieradzan, A. K.; Krupa, P.; Wales, D. J. What makes telomeres unique? *J. Phys. Chem. B* **2017**, *121*, 2207–2219.
- [23] He, Y.; Mozolewska, M. A.; Krupa, P.; Sieradzan, A. K.; Wirecki, T. K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagieła, D.; Ślusarz, R.; et al., Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 14936–14941.
- [24] Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; et al., A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* **2014**, *20*, 2306.
- [25] Krupa, P.; Mozolewska, M. A.; Wiśniewska, M.; Yin, Y.; He, Y.; Sieradzan, A. K.; Ganzynkiewicz, R.; Lipska, A. G.; Karczyńska, A.; Ślusarz, M.; et al., Performance of protein-structure predictions with the physics-based UNRES force field in CASP11. *Bioinformatics* **2016**, *32*, 3270–3278.
- [26] Mozolewska, M.; Krupa, P.; Scheraga, H. A.; Liwo, A. Molecular modeling of the binding modes of the iron-sulfur protein to the Jac1 co-chaperone from *Saccharomyces cerevisiae* by all-atom and coarse-grained approaches. *Proteins: Struct. Funct. Bioinf.* **2015**, *83*, 1414–1426.
- [27] He, Y.; Liwo, A.; Scheraga, H. A. Optimization of a nucleic acids united-residue 2-point model (NARES-2P) with a maximum-likelihood approach. *J. Chem. Phys.* **2015**, *24*, 243111.
- [28] Case, D.; Cerutti, D.; Cheatham, I. T.; Darden, T.; Duke, R.; Giese, T.; H, G.; Goetz, A.; Greene, D.; Homeyer, N.; et al., University of California, San Francisco.

- [29] SalomonFerrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198–210.
- [30] Mosayebi, M.; Louis, A. A.; Doye, J. P.; Ouldridge, T. E. Force-induced rupture of a DNA duplex: from fundamentals to force sensors. *ACS Nano* **2015**, *9*, 11993–12003.
- [31] Punkkinen, O.; Hansen, P. L.; Miao, L.; Vattulainen, I. DNA overstretching transition: Ionic strength effects. *Biophys. J.* **2005**, *89*, 967–978.
- [32] Smith, S.; Cui, Y.; C, B. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **1996**, *271*, 795–799.
- [33] Bianco, P.; Bongini, L.; Melli, L.; Lombardi, V. PicoNewton-millisecond force steps reveal the transition kinetics and mechanism of the double-stranded DNA elongation. *Biophys. J.* **2011**, *101*, 866–874.
- [34] Hatch, K.; Danilowicz, C.; Coljee, V.; Prentiss, M. Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs. *Phys. Rev. E* **2008**, *78*, 011920.
- [35] Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **2018**, *10*, 631–637.
- [36] Severin, P. M. D.; Zou, X.; Gaub, H. E.; Schulten, K. Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.* **2011**, *39*, 8740–8751.
- [37] Lee, N.-K.; Thirumalai, D. Pulling-speed-dependent force-extension profiles for semiflexible chains. *Biophys. J.* **2004**, *86*, 2641–9.
- [38] Qiu, H.; Guo, W. Detecting ssDNA at single-nucleotide resolution by sub-2-nanometer pore in monoatomic graphene: a molecular dynamics study. *Appl. Phys. Lett.* **2012**, *100*, 083106.



- [39] Naserian-Nik, A. M.; Tahani, M.; Karttunen, M. Pulling of double-stranded DNA by atomic force microscopy: a simulation in atomistic details. *RSC Adv.* **2013**, *3*, 10516.

## 6 Figure legends

Figure 1. Diagrams of the force needed to stretch DNA for a given distance in GC30. Solid thin lines indicate averaged values over 64 trajectories for each possibility, fainter thick lines indicate the standard error for each average, and numbers in the legend indicate after which residue the break was introduced. Diagrams for other sequences, pulling speeds, and force constants can be found in the SI.

Figure 2. Position of the main force peak needed to separate chains with single breaks as a function of the break position.

Figure 3. Values of the main force peak needed to separate chains with single breaks, as a function of the break position.

Figure 4. Example mechanism of SSB stretching in trajectory number 1 (out of 64) of  $d(TTTAGGG)_9$ . Panel A shows an initial structure, and the other panels show the structures after a given simulation time for a pulling speed of 0.4 mm/s: B) 16  $\mu$ s, C) 35  $\mu$ s, D) 38  $\mu$ s, E) 44  $\mu$ s, F) 45  $\mu$ s, G) 63  $\mu$ s, H) 64  $\mu$ s. The full unbroken chain is shown by the yellow line, while the broken chain is coloured in blue (stretched fragment) and red (remaining part).

Figure 5. The force needed to stretch A) d(GC), B) d(TTAGGG) sequences as a function of the number of repetitions for a given distance.

Figure 6. Values of the main force peak needed to separate GC chains as a function of the number of GC repetitions (dots) and fitting of eq. 1 with parameters showed in Table 2 to this data (lines).

Table 1: Minimum number of nucleic acids in a single-break fragment necessary for maintaining stability, manifested as a main force peak value higher than 30, 60 or 120 pN.

	$> 30pN$	$> 60pN$	$> 120pN$
$d(AT)_{30}$	16	21	23
$d(ATGC)_{15}$	7	8	8
$d(GC)_{30}$	9	11	12
$d(TTAGG)_{12}$	7	9	10
$d(TTAGGG)_{10}$	6	6	8
$d(TTTAGGG)_9$	6	9	9

Table 2: Fitting of the obtained distance and force values from the simulations to eq. 1.

	$\Delta G_{bp}/\delta$	error	$\Delta G_{\tau_{obs}}/\delta$	error	$\Delta G_{\tau_{obs}}/\Delta G_{bp}$	error
$d(GC)_{30}$	801.81	10.39	6395.8	164.8	7.977	0.178
$d(GC)_{30} - slow$	703.13	10.58	7484.3	220.9	10.644	0.270
$d(GC)_{30} - weak$	843.06	13.39	6889.7	251.5	8.172	0.269
$d(ATGC)_{15}$	556.93	6.87	3176.6	93.7	5.704	0.153
$d(AT)_{30}$	499.54	7.75	7167.5	162.4	14.348	0.237
$d(TTTAGGG)_9$	524.50	7.12	3095.1	97.1	5.901	0.167
$d(TTAGGG)_{10}$	537.33	11.07	2660.0	135.9	4.950	0.232
$d(TTAGGG)_{10} - slow$	516.36	14.76	3851.2	234.5	7.458	0.401
$d(TTAGG)_{12}$	516.38	5.54	3090.2	82.5	5.984	0.146
$d(GC)_x$	995.48	24.42	8557.9	582.1	8.597	0.545
$d(GC)_x - slow$	606.20	19.02	7130.6	446.3	11.763	0.637

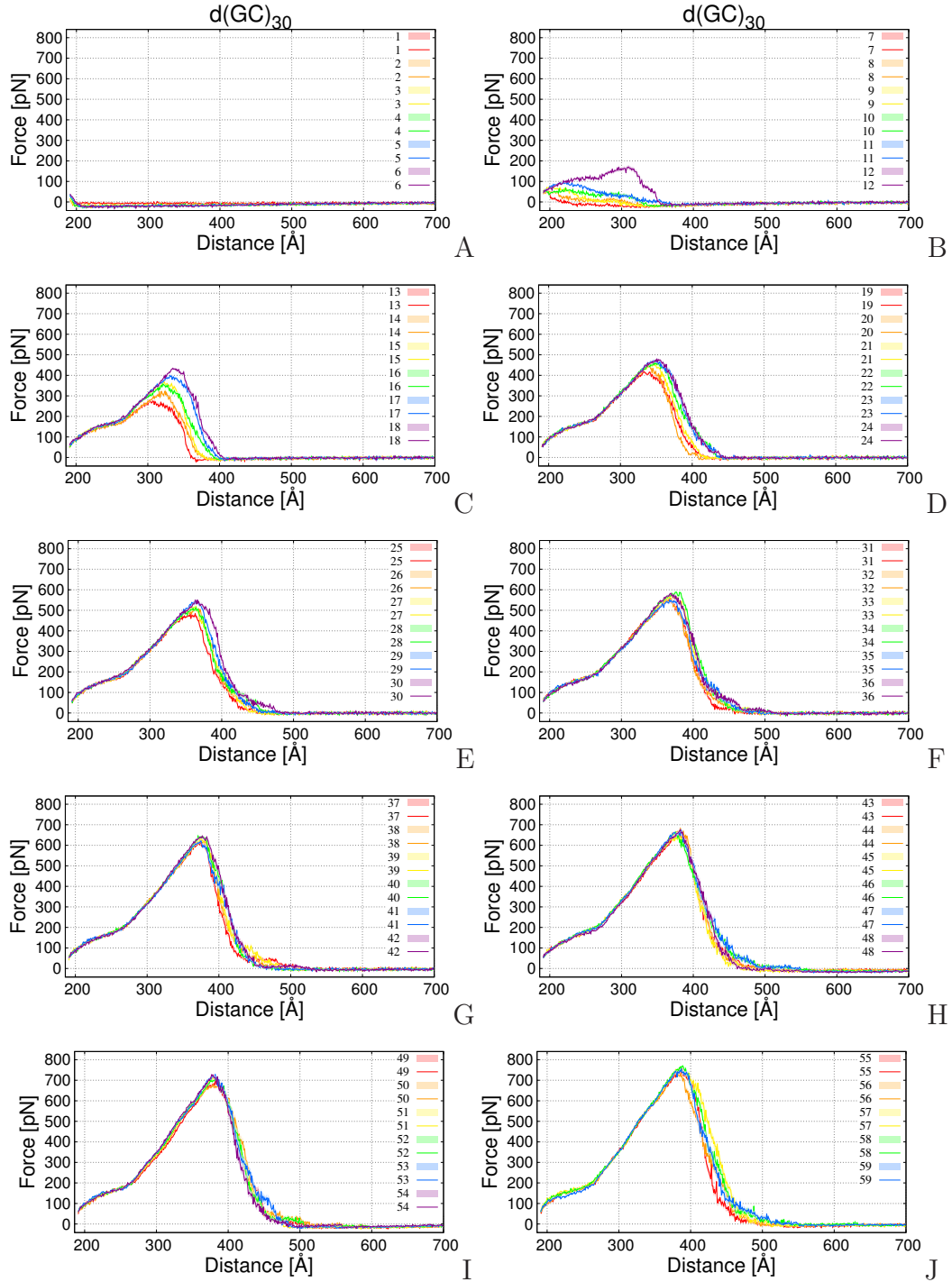


Figure 1

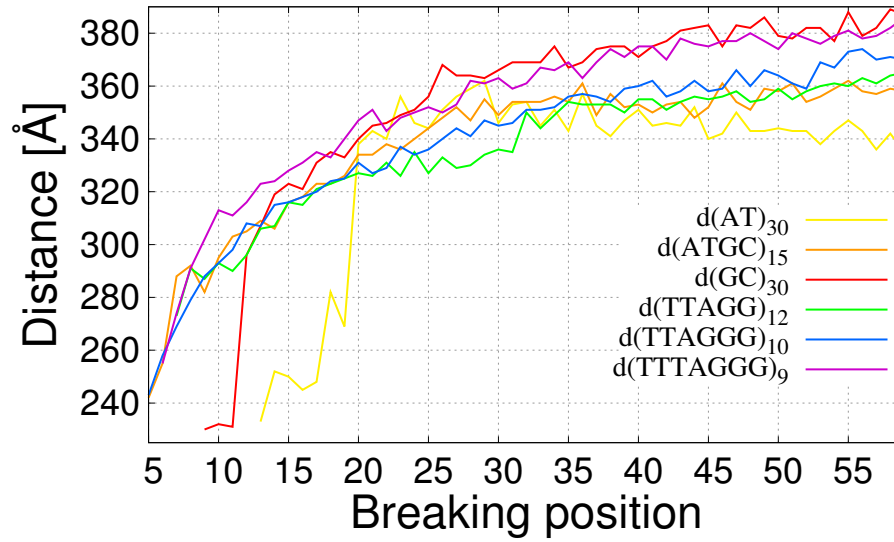


Figure 2

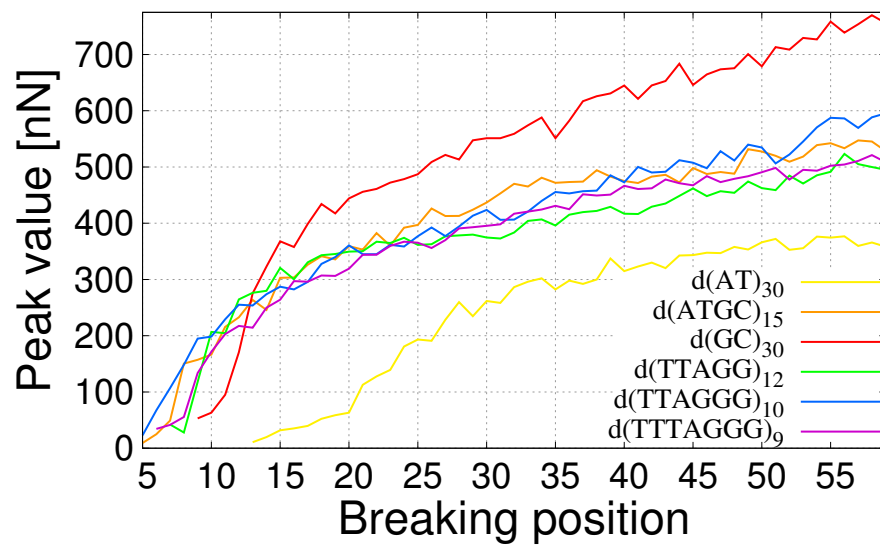


Figure 3

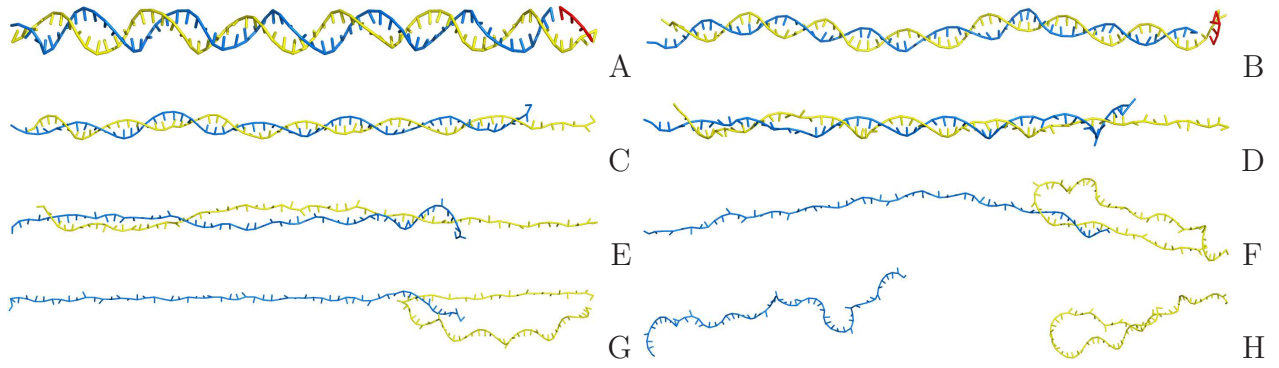


Figure 4

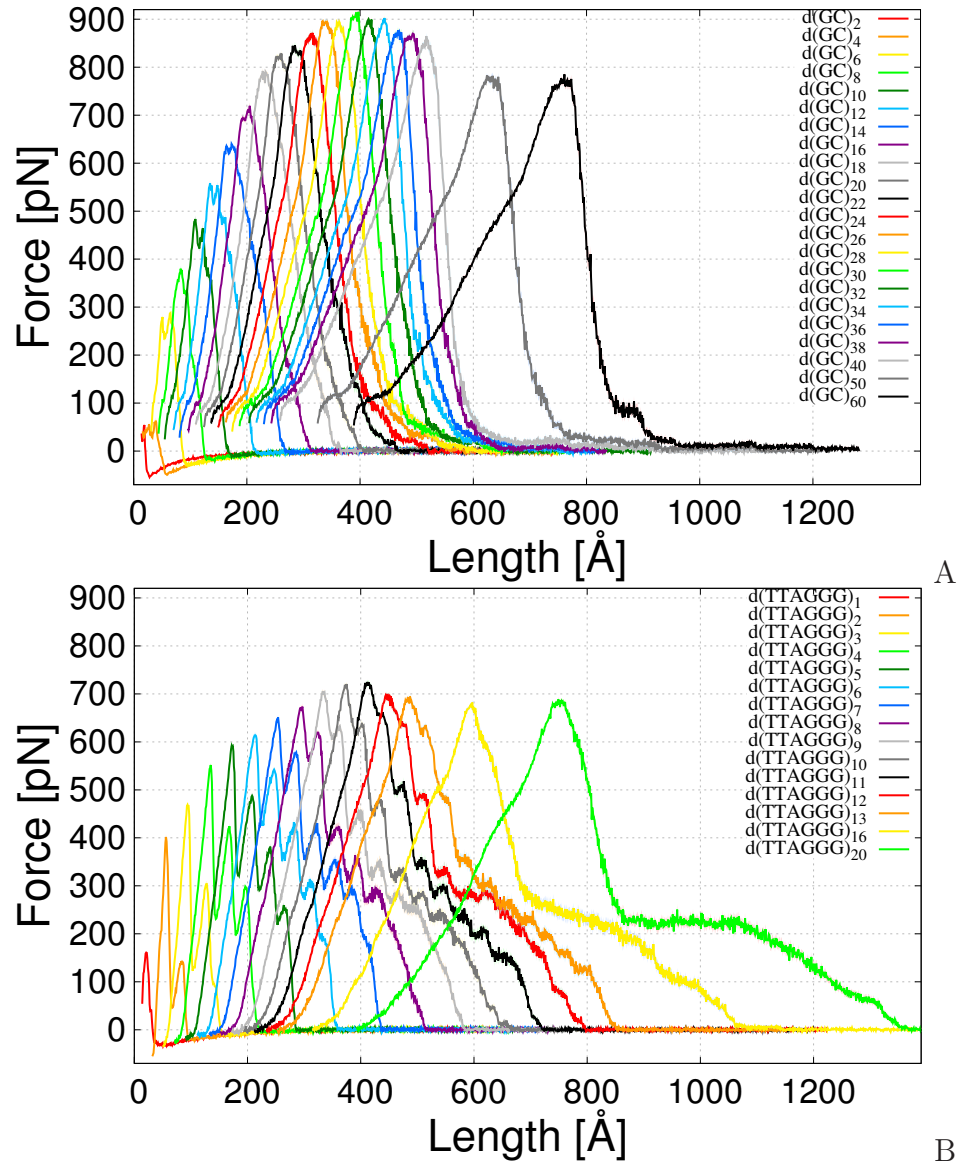


Figure 5

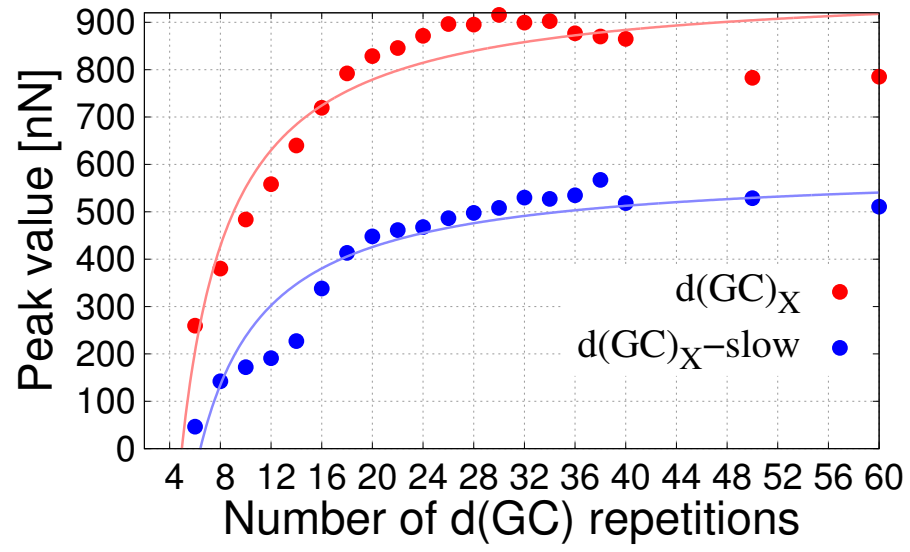


Figure 6



## TOC Graphic

