Reactive Molecular Dynamics Simulation on DNA Double Strand Breaks Induced by Hydrogen Elimination

Hiroaki Nakamura[1,2*], Kento Ishiguro[2], Ayako Nakata[3], Shunsuke Usami[1,4], Seiki Saito[5] and Susumu Fujiwara[6]

[1]*National Institute for Fusion Science, 322-6 Oroshi-cho, Toki, Gifu 509-5292, Japan*

[2]*Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan*

[3]*National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan*

[4]*The University of Tokyo, Tokyo 113-8654, Japan*

[5]*Yamagata University, 4-3-16 Jonan, Yonezawa, Yamagata 992-8510, Japan*

[6]*Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan*

We propose a scar model to simulate double-strand breaks (DSBs) in telomeric DNA induced by the β-decay of tritium into helium. The two hydrogens bonded to the 5′ carbon connecting the pentasaccharides and phosphate are removed in this scar model. Molecular dynamics simulations using a reactive force field were conducted across 10 cases involving telomeric DNA composed of 16 base pairs (32 nucleotides). Double-strand breaks (DSBs) are observed in structures with more than 24 scars, whereas only single-strand breaks (SSBs) occur in the case with 16 scars, Moreover, in the case of {16,0} and {0,16}, where only one of the strands had scars, SSB occurs only in the scarred strand. Secondly, in the {16,8} and {8, 16} cases, DSBs occurs. Thus, we conclude that the following conditions are necessary for DSBs：(i) Scars must be present on both the left (L) and right (R) strands. (ii) A substantial number of scars (24 or more) must be in close proximity.

1. Introduction

Brenner initially developed[1)] molecular dynamics simulations using reactive force fields to study carbon nanotubes, which have since been adapted for various materials, including DNA. This force field can handle the bonding and breaking of multiple atoms for a limited range of elements such as carbon, hydrogen, and oxygen. It also defines a measure of bond order and can distinguish the order of covalent bonds. This reactive force field has been improved, and now a reactive force field that can handle more elements, including biopolymers and metals, has been proposed. Our group has performed calculations using Brenner's reactive force field to simulate the behavior of the wall of a fusion reactor.[2–5)] Recently, we have attempted to treat the behavior of DNA[6–9)] by using this experience.

*E-mail: hnakamura@nifs.ac.jp

Several years ago, we started using molecular dynamics (MD) simulations[10,11] to assess tritium-induced DNA damage,[12–14] including transmutation effects.[6,7] Simultaneously with the simulations, we also started DNA damage experiments[7,15] that were complementary to the simulations.

In our initial study of DNA,[6] we focused on the structure of telomeres at the end of DNA and performed MD simulations, assuming that tritium (T) was absorbed into the cell, replacing hydrogen (H) in the guanine bases in telomeres and eventually undergoing β-decay into helium (He). Instead of the reactive force field, we employed the conventional force field CHARMM36,[16] which calculates the optimal DNA structure in a solvent environment but cannot handle covalent bond cleavage or formation. This was because, at the time, we could not find a reactive force field that could handle DNA. Thus, using the CHARMM36 force field, we were able to simulate the collapse of the double helix structure as the hydrogen in the guanine of the telomere underwent β-decay into helium, weakening the hydrogen bonds between the DNA double strands. As a result, we quantitatively evaluated the rate of double-strand unwinding caused by the β-decay of tritium in guanine to helium. However, the CHARMM36 force field used in this study is unable to model the most intriguing phenomenon, namely double-strand breaks (DSBs) of DNA.

With advancements since our earlier simulations, we were able to use reactive force fields capable of modeling DNA, enabling us to pursue the challenge of simulating double-strand breaks, which has been the focus of our research for some time. This analysis requires the evaluation of the electronic state through quantum chemical calculations to accurately assess DNA bond breakage, as indicated in related studies.[17] However, the amount of calculation is considerable, and it is still difficult to investigate the outline of various situations in practice due to the limitations of computer resources. Therefore, we will attempt to gain an understanding of the outline using classical molecular dynamics simulations with a reactive force field (ReaxFF). Specifically, we modeled the desorption of 5′ hydrogens that connect the pentasaccharides and phosphate, as these are considered the most likely sites for tritium substitution within the telomere. This paper describes the molecular dynamics (MD) simulations, in this case, on double-strand breaks (DSBs) performed using ReaxFF.

2. Molecular Dynamics Simulation Model

2.1 Telomeric DNA

As in our previous work,[6] we focused on telomeres, structures at the ends of eukaryotic chromosomes that consist of repeated base sequences (telomeric DNA).

The reason for choosing telomeric DNA as the simulation model is the same as in the previous study, as follows: The shortening of telomere replication in human cells plays a crucial role in cellular senescence.[18] Therefore, we consider that the human cell damage by tritium depends on the damage of the telomeric DNA.

We derived the telomeric DNA structure by removing the TRF2 protein from the TRF2-Dbd-DNA complex (PDB ID: 3SJM).[19] The telomeric DNA has 17 base pairs, d(TCTAGGGTTAGGGTTAG), which consists of 1,078 atoms as shown in Fig. 1. Here, the four types of bases are adenine A, guanine G, cytosine C, and thymine T. To distinguish between the two strands, the side with the sequence {TCTAGGGTTAGGGGTTAG} is denoted as L-strand, and the side with {AGATCCCAATCCCAATC} as R-strand. In addition, the orange squares in Fig.1 indicate the phosphorus atoms in the nucleotides with 16 phosphorus atoms per strand.

Since the telomeric DNA used in this simulation is initially negatively charged ($-32e$, where the elementary charge $e = 1.602176634\times10^{-19}$ C), sodium ions are added to neutralize the entire system. Furthermore, the salt concentration must be set to the concentration of the human body. Considering the above, 120 sodium ions (Na^+) and 88 chlorine ions (Cl^-) are added randomly and placed in the system. The initial volume V of the simulation box is 100 Å$\times$ 100 Å$\times$ 100 Å. We also add 30,773 water molecules into the simulation box.

First, we make the steady state of the telomeric DNA with the solvent at 310K. We employed the following three-step process to achieve a steady-state:

(1) $0 \le t \le 0.5$ ns and T increases from 0 K to 310K (NVT ensemble, CHARMM force field): At $t = 0$, we perform the energy minimization of the total system. Then using a canonical ensemble (NVT ensemble), we linearly increase the reference temperature T from 0 K to 310 K over 0.5 ns while keeping the DNA coordinates fixed. The volume V of the simulation box was maintained at 100 Å$\times$ 100 Å$\times$ 100 Å, with the periodic boundary conditions applied in the x, y, and z directions. In this first process, CHARMM force field is adopted[16,20] using Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) code[21,22] with Langevin thermostat algorithm[23] to control the temperature of the system. The time step of the MD simulation is set to 1 fs. During the first process, the coordinates of all atoms that make up telomeric DNA remain fixed. The solvent molecules, on the other hand, evolve in time. Eventually, the Na^+ and Cl^- ions in the solvent, which interact with the DNA's negative charge, reorganize themselves, allowing the solvent to reach a steady state. We comment here on the difference between the CHARMM force field and the reactive force field[24,25] (ReaxFF). The CHARMM force field was designed

to calculate structural changes in biomolecules, such as telomeric DNA, in a solvent, but it is unable to simulate covalent bond cleavage or formation. Therefore, it can be faster than calculations using the ReaxFF, which can handle covalent bond cleavage and formation.

(2) 0.5 ns $\leq t \leq$10.5 ns, T is fixed at 310 K and P is fixed at 1 bar (NPT ensemble, CHARMM force field): Once the reference temperature T reaches 310 K in the first process, T is maintained at 310 K. The constraints on the telomeric DNA are then gradually relaxed. Moreover, to examine the behavior of DNA under an NPT ensemble at 1 bar, the simulation conditions are transferred from NVT ensemble to NPT ensemble with sequential changes in the constraints of the telomeric DNA. The details of the simulation conditions are explained below. After t = 0.5 ns, the simulation condition is changed from NVT ensemble to NPT ensemble, where the pressure P is fixed at 1 bar. The total number of particles N is kept in the first process state. Conversely, the volume V of the simulation box is no longer treated as fixed but variable. From 0.5 ns to 1.5 ns, the positions of all atoms within the telomeric DNA remain fixed, as in the first process. From 1.5 ns to 3.5 ns, the main strand's atomic coordinates are fixed, while the side strands' atoms are allowed to move freely. From 3.5 ns to 10.5 ns, all telomeric DNA atoms are unfixed, allowing the DNA to change its shape freely. Thus, we achieved a stable structure of the telomeric DNA and the solvent using the CHARMM force field at 310 K and 1 bar.

(3) 10.5 ns $\leq t \leq$ 11.4 ns, T is fixed at 310 K and P is fixed at 1 bar (NPT ensemble, ReaxFF): We then conducted MD simulations using ReaxFF, capable of modeling covalent bond cleavage and formation. However, switching the force field from CHARMM to ReaxFF, based on the molecular structure derived from the previous process, destabilized the system, causing the collapse of the DNA structure from its ends. Therefore, the four hydroxyl (OH) groups at the ends of each DNA strand are fixed to prevent the DNA from unintentionally disintegrating its structure from the ends (See Fig. 2). Under these constraints at the four ends, MD simulations using the ReaxFF are performed under NPT ensemble (T = 310 K, P =1 bar) conditions. The time step is set to 0.1 fs, which is shorter than the one used in the CHARMM simulations. This approach ensures the stable configuration of the telomeric DNA in the solvent. Additionally, the time evolution of the root mean square deviation (RMSD)[6] is shown in Fig. 3. The RMSD remains nearly constant over the range from 10.7 ns to 11.4 ns, indicating the stability of the constructed telomeric DNA structure.

2.2 Scar Model

2.2.1 Definition of scar model In this simulation, DNA damages are treated as "the removal of two hydrogens binding to the 5′ carbon between the pentasaccharides and the phosphate group" as shown in Fig.4. This state, in which the hydrogen atoms are removed, will henceforth be referred to as a 'scar'. This scar model is based on the fact that molecular dynamics simulation by Tsuchida et al.[27] shows that atoms in the solvent collide with these hydrogens with the highest frequency. In this case, if tritiums are present in the solvent, the probability of hydrogen substitution between tritium in the solvent and hydrogen in the DNA is expected to be highest for this 5′ bound hydrogen. When this substitution occurs, tritium bound to the 5′ carbon will proceed to β-decay, decaying to helium with a half-life of about 12.3 years.[12] Unlike hydrogen isotopes, the helium produced in this reaction is not covalently bound to the 5′ carbon, so it eventually diffuses into the solvent, leaving the DNA. This phenomenon is assumed in the scar model. This model offers the advantage of applicability not only to the disintegration effects of tritium β-decay, as initially intended, but also to scenarios involving hydrogen desorption through direct or indirect action.

The removal of two hydrogens attached to the 5′ carbon in the scar model results in the charge distribution around the affected region. It is, therefore, necessary to calculate the charge distribution. The density functional theory (DFT) calculations using Gaussian09[28] are used to calculate the charge distribution. Since DFT calculations demand greater computational resources than the MD simulation, calculations are performed with the smallest possible structure, taking into account the effects of the environment around the scar position. For example (see Fig. 5), we consider the case of introducing a scar at $p_{\mathrm{L}} = 10$ in Fig. 1. Since the 5′ carbon, from which two hydrogens are removed, spans nucleotides, it is reasonable to consider the two-nucleotide structure as the smallest unit to be calculated by DFT. In the DFT simulations, we use B3LYP exchange-correlation functional[29,30] and cc-pVDZ bases-set.[31]

2.2.2 Configuration of scars in both strands Ten patterns of simulations were performed to investigate the effect of scars on the telomeric DNA, varying the number and location of scars. To distinguish between them, the number of scratches in the left or right strands is defined as n_{L} or n_{R}, respectively. We consider the arrangements $\{n_{\mathrm{L}}, n_{\mathrm{R}}\} = \{0,0\}, \{1,1\}, \{4,4\}, \{8,8\}, \{12,12\}, \{16,16\}, \{16,0\}, \{0,16\}, \{16,8\}$ and $\{8,16\}$ as shown in Figs. 6 and 7.

3. Molecular Dynamics Simulation Results

MD simulations were conducted using the 10 different scar models introduced in the previous section (Figs. 6 and 7) to investigate their dependence on the number and location of scars. The state diagrams of strand breaks obtained from the MD simulation for these 10 scar models are illustrated in Fig. 8. The evaluation method for the strand breaks used here is explained below.

Since a broken strand may reconnect after being severed, it is essential to assess the strand's state over time to understand its progression. In this simulation, the assessment of DNA strand breaks is conducted as follows.

First, the bond order used in the ReaxFF calculation is used to determine whether the DNA strand has been cut. The bond order $\mathrm{BO}_{ij}(t)$ between atoms i and j at time t is a quantity that expresses the degree of covalent bond (σ bond, π bond, or $\pi\pi$ bond) and is calculated directly from the interatomic distance using the following empirical formula.

$$\begin{aligned}\mathrm{BO}_{ij}(t) &= \mathrm{BO}^{\sigma}_{ij}(t) + \mathrm{BO}^{\pi}_{ij}(t) + \mathrm{BO}^{\pi\pi}_{ij}(t) \\ &:= \exp\left\{p_1\left[\frac{r_{ij}(t)}{r^{\sigma}_0}\right]^{p_2}\right\} + \exp\left\{p_3\left[\frac{r_{ij}(t)}{r^{\pi}_0}\right]^{p_4}\right\} + \exp\left\{p_5\left[\frac{r_{ij}(t)}{r^{\pi\pi}_0}\right]^{p_6}\right\}, \quad (1)\end{aligned}$$

where $r_{ij}(t)$ is the interatomic distance between atoms i and j at time t, r^{α}_0 term is equilibrium bond length for each bond $\alpha = \sigma, \pi$, or $\pi\pi$, and $p_{1,2,\cdots,6}$ terms are empirical parameters.[32] Moreover, the bond order $\mathrm{BO}^{\alpha}_{ij}(t)$ of each bond α is equal to zero in the absence of a bond, increases following the strength of the bond, and eventually reaches one in the case of a perfectly bond. Therefore, it takes in the following range:

$$0 \leq \mathrm{BO}^{\alpha}_{ij}(t) \leq 1 \quad \text{for } \alpha = \sigma, \pi, \text{or } \pi\pi, \quad (2)$$

$$0 \leq \mathrm{BO}_{ij}(t) \leq 3. \quad (3)$$

Next, we calculate the time average $\overline{\mathrm{BO}}_{ij}(t)$ from t to $t + (N_{\mathrm{BO}} - 1)\Delta t$ follows:

$$\overline{\mathrm{BO}}_{ij}(t) := \frac{1}{N_{\mathrm{BO}}} \sum_{l=0}^{N_{\mathrm{BO}}-1} \mathrm{BO}_{ij}(t + l\Delta t), \quad (4)$$

where we use 100, 000 as the step number of the MD simulation N_{BO}, and Δt is the time step of MD simulation with ReaxFF, *i.e.*, 0.1 fs. In other words, the bond order is averaged over the time interval of $N_{\mathrm{BO}}\Delta t = 0.01$ ns. Here, we comment on this time interval of 0.01 ns. In this simulation, the time interval that can be analyzed is (11.4 - 10.7) ns = 0.7 ns, as shown in Fig. 3, which shows the time evolution of the RMSD with a resolution of 1 ps, and the fluctuations can be seen. Therefore, we decided to suppress the fluctuations by taking the time

resolution of the time average of the bond order to 0.01 ns, which is ten times higher than the time resolution of the RMSD. On the other hand, the stability of the strand-breaking state over a longer time, 0.7 ns, can be identified in Figs. 10 – 19.

If the value of $\overline{\mathrm{BO}}_{ij}(t)$ calculated in this way is less than a threshold $\varepsilon_{\mathrm{BO}} = 0.03$, it is deemed that no bond exists between atoms i and j.[33] In particular, if atoms i and j are adjacent in the DNA strand, we express that a 'gap' has occurred between the pair of i and j. The number of these gaps was recorded for each chain on the left and right, respectively, and their temporal evolution is illustrated in Figs. 10–19. Moreover, gaps occurring in only one strand are classified as single strand breaks (SSBs), while gaps in both strands are classified as double strand breaks (DSBs).

Until $t = 11.4$ ns, DSBs occur in the four cases {12, 12}, {16, 16}, {16, 8} and {8, 16}, and SSB occurs in the three cases {8, 8}, {16, 0} and {0, 16}. In particular, for {16, 0}, and {0, 16}, gaps appear exclusively in the scarred strand. No gaps were observed during the simulation period for {1, 1} and {4, 4}. Obviously, in the no scarred 'original' telomeric DNA case, *i. e.*, {0, 0}, no gaps appear in the DNA, and the double-strand structure is preserved. The results can be categorized based on the total number of scars ($n_{\mathrm{L}} + n_{\mathrm{R}}$) as follows:

(1) $n_{\mathrm{L}} + n_{\mathrm{R}} = 2$ and 8: No gaps appear.

(2) $n_{\mathrm{L}} + n_{\mathrm{R}} = 16$: Single-strand break (SSB) occurs.

(3) $n_{\mathrm{L}} + n_{\mathrm{R}} = 24$ and 32: Double-strand breaks (DSBs) occur.

Thus, it can be demonstrated that the total number of scars ($n_{\mathrm{L}} + n_{\mathrm{R}}$) serves as a reliable indicator of telomeric DNA strand breaks. We also summarize, in Table I, the time of occurrence of SSB and DSBs, and the total number of gaps generated in the final state ($t = 11.4$ ns) for each $\{n_{\mathrm{L}}, n_{\mathrm{R}}\}$ pair. In the subsequent Subsections 3.1, 3.2, 3.3, and 3.4, the following four physical quantities are presented and discussed for each $\{n_{\mathrm{L}}, n_{\mathrm{R}}\}$ pair:

(i) The final molecular structure at $t = 11.4$ ns.

(ii) The location of $p_{\mathrm{R,\,L}}$ in the telomeric DNA where gaps appear at $t = 11.4$ ns.

(iii) The time evolution of the gap number from 10.7 ns to 11.4 ns.

(iv) The spatial ($i_{\mathrm{R,\,L}}$) distributions of the bond order at $t = 10.7$ ns and 11.4 ns. Here, the spatial position $i_{\mathrm{R,\,L}}$ is defined in Fig. 9. Moreover, bond order is a formal measure of the multiplicity of covalent bonds between two atoms.

3.1 Original Strand: O case

First, the time evolution of the telomeric DNA structure is reported for no scars cases, *i. e.*, $n_{\mathrm{L}} + n_{\mathrm{R}} = 0$ (O case in Table. I). This behavior is the basis for discussing the results in the

Table I. Times of single-strand break (SSB) and double-strand breaks (DSBs) occurrence, and the number of gaps at $t = 11.4$ ns.

Scars in Backbone			Occurrence Time [ns]		Num. of Gaps at $t = 11.4$ ns
Case	Pair $\{n_L, n_R\}$	Total Num. $n_L + n_R$	SSB	DSBs	
O	$\{0, 0\}$	0	11.357	null	2
D1	$\{16, 16\}$	32	10.774	11.242	8
D2	$\{16, 8\}$	24	10.904	11.105	6
D3	$\{8, 16\}$	24	11.039	11.041	7
D4	$\{12, 12\}$	24	10.884	11.160	4
S1	$\{16, 0\}$	16	10.904	null	9
S2	$\{0, 16\}$	16	10.993	null	3
S3	$\{8, 8\}$	16	10.945	null	1
N1	$\{4, 4\}$	8	null	null	null
N2	$\{1, 1\}$	2	null	null	null

scarred cases.

From Fig. 10, it is shown that one gap appears in the L-strand (green chain), the position of which is represented by black squares in the right figure of (O-a) with $p_L = 7$ and 8. In addition, the time evolution of the number of gaps in the left-hand side of Fig. 10 shows how the gaps are repeatedly created and extinguished until finally there are only one or two gaps, which eventually remain. This time evolution demonstrates that thermal fluctuations can induce gaps in the telomeric DNA by the simulation over the 0.7 ns period spanning $t_0 = 10.7$ ns to $t_1 = 11.4$ ns, even when the telomeres were initially unscarred. The reason for the gaps is as follows. The structure of the telomeric DNA used in this simulation is based on the initial structure of the TRF2-Dbd-DNA complex, from which the TRF2 protein has been removed, as explained in Subsection 2.1. While the original complex showed structural stability, it is not known whether the telomeric DNA structure alone becomes stable. Therefore, because the initial structure is unstable, it is believed that the structure changes due to thermal fluctuations, and gaps intermittently appear. This can be seen from the fact that the RMSD in Fig. 3 constantly fluctuates. However, it should also be noted that this gap may result in a single-strand break but not double-strand breaks. This must be taken into account when analyzing other cases of scars.

3.2 Double Strand Breaks (DSBs): D1, D2, D3 and D4 cases

In these cases, *i. e.,* $n_L + n_R = 24$ and 32, the double-strand breaks occur as shown in Figs. 11, 12, 13 and 14. The observation that once a gap is made it does not heal, as obtained from all time evolution of the number of scars on the strands in Figs. 11(D1-b), 12(D2-b), 13(D3-b) and 14(D4-b), shows that the gaps are expanding. This phenomenon occurs in both strands, leading to the DSBs. This behavior contrasts sharply with the repeated formation and recovery of gaps due to thermal fluctuations observed in the O case in Section 3.1, where the gaps do not expand.

3.3 Single Strand Breaks (SSB): S1, S2 and S3 cases

In these cases, *i. e.,* $n_L + n_R = 16$, the single-strand break occurs as shown in Figs. 15, 16 and 17. Among the three cases, S1, S2 and S3, the number of gaps increases in the order of S3 < S2 < S1. In each case, only one single-strand break is observed.

3.4 No Gaps: N1 and N2 cases

In these cases, where $n_L + n_R = 2$ and 8, no gaps are observed, as depicted in Figs. 18 and 19. In both cases, the DNA structure remains as stable as that of the original strands (Fig. 10).

4. Discussions

DSBs were observed exclusively in calculations with more than 24 scars, specifically in the cases of {16,16}, {16,8}, {8,16} and {12,12}. Furthermore, SSB occurred for simulations with 16 scars in the cases {8,8}, {16,0}, and {0,16}, and no gaps occurred for simulations with less than 16 scars. Thus, a significant correlation between the number of scars and the susceptibility of strand breakage was observed,

From Table I and Figs. 11(D1-b), 12(D2-b), 13(D3-b) and 14(D4-b), the number of gaps in DSBs are 8 for D1, 6 for D2, 7 for D3, and 4 for D4, respectively. Table I also shows that SSB, for all three cases (S1, S2, and S3), occurred at $t \sim 10.9$ ns, which is 0.4 ns earlier than in the original case (O case). This fact indicates that the installation of scars into the DNA strands increases the fragility of the DNA backbone. Furthermore, a detailed comparison of the occurrence times from Table I reveals that, among the four cases in which DSBs occur, the {16,16} case is the slowest with $t = 11.242$ ns. Conversely, the {8,16} case produced DSBs earliest, at 11.041 ns. It can, therefore, be concluded that the number of scars alone does not determine the fragility of the strands. It is suggested that the formation of DSBs in DNA is influenced not only by the number and distribution of scars, but also by the effects of ions and water molecules surrounding the DNA.

For the cases {16,0} and {0,16}, a gap occurs in the scarred strand, and the other strand

does not exhibit any effects resulting from the initial scars. In addition, Figs. 15(S1-b) and 16(S2-b) show that the final number of gaps in the {16,0} case is 9 and that in the {0,16} case is 3, indicating a significant difference in the number of gaps for the same SSB cases. In contrast, while DSBs occur in both the {16,8} and {8,16} cases, the {16,8} case shows a greater number of gaps in the strand with fewer scars. In a somewhat unexpected turn of events, in the {8,16} case, the gaps occur at $p_{\mathrm{L}} = 15$ and 16 on the L strand without any scars. This indicates that the correlation between the positions of the scars and gaps is not significant.

The distribution of gap occurrence at covalent bonds on the backbone of nucleotides (see Fig. 20) is examined to investigate the occurrence of gaps further. The positions of the scars observed across all 10 simulations as shown in Figs. 6 and 7 are reduced to the 'unit' nucleotide. A summary of these is provided in Table II. This table indicates that the covalent bond between P-O, representing the phosphodiester group, is extremely fragile, whereas the covalent bonds between O-C and C-C surrounding the pentasaccharide residue are durable and resistant to gaps. This fragility arises because the bond order between P-O ranges from only 0.4 to 0.6, even in the absence of a gap.

Additionally, a potential cause of the gap formation in the covalent bond between P-O is the interaction between the phosphate group and surrounding ions. Figure 21 visualizes the ions around the telomeric DNA backbone in the O case {0,0}. The left-hand figure shows that Na^+ and Cl^- are approaching the phosphate group before the gap occurs at $t = 11.12$ ns. Then they interact with P or O in the backbone, weakening the original P-O covalent bond. Such a phenomenon, in which multiple Na^+ or Cl^- approach the gap in the backbone, was also observed in other cases.

From the above points, we can infer that the process leading to gap formation occurs as follows:

(1) Excess Na^+ ions are attracted to negatively charged sites (phosphate groups) by hydrogen desorption or accidental approach by random water movement.
(2) Na^+ and phosphate groups make bonds.
(3) The covalent bonds in the backbone are weakened, and the phosphate groups are pulled out from the DNA.

Thus, it can be said that the more hydrogen desorption and the change of initial charge configuration due to the decomposition effect increase, the more likely the above processes are to occur, and the more the gap increases accordingly. This can be inferred from the work of Hishinuma *et al.*,[34] where they investigated the process of DNA breaking under irradiation-

induced thermal energy using MD simulations. According to them, it was reported that when counterions and water molecules are present around the DNA, the phosphate group and Na^+ form the intermediate Na^+-O-P-O, after which strand breakage is induced at the phosphate group. This report supports the notion that the processes (1)–(3) outlined in our study play a role in the occurrence of gaps within the DNA.

Table II. Distribution of covalent bonds on the backbone of the unit nucleotides of telomeric DNA. In the scar mode, two hydrogens are removed in C5′.

Position in Backbone	Num. of Gaps		
	in L-chain	in R-chain	Total num.
O3′ – P	5	10	15
P – O5′	12	5	17
O5′ – C5′	0	3	3
C5′ – C4′	0	3	3
C4′ – C3′	2	0	2
C3′ – O3′	1	0	1

5. Conclusion

We proposed a scar model to simulate the occurrence of double-strand breaks (DSBs) in telomeric DNA, induced by the effect of the β-decay of tritium into helium. The two hydrogens bonded to the 5′ carbon connecting the pentasaccharides and phosphate are removed in this scar model. Molecular dynamics simulations using ReaxFF were conducted across 10 cases, varying the total number and arrangement of the scars on the strands of the telomeric DNA. The simulation resulted in double-strand breaks being observed for structures with more than 24 scars for the telomeric DNA consisting of 16 base pairs (32 nucleotides). In the case of 16 scars, only single-strand breaks were observed.

Detailed observations of case with asymmetrical scar distributions on the L and R strands reveal that in the case of {16,0} and {0,16}, where scars are present only on one strand, single-strand breaks (SSBs) occurred solely in the scarred strand without developing into double-strand breaks (DSBs). Secondly, double-strand breaks occurred in the {16,8} and {8, 16} cases and were more pronounced in the strand with fewer scars.

Based on these results, the following conditions are essential for the formation of double-strand breaks:

(1) The “scars” must occur on both L and R strands.

(2) A significant number of scars (more than 24) must be in close proximity.

Acknowledgment

Computations were performed using the Research Center for Computational Science, Okazaki, Japan (Project: 24-IMS-C099) and the Plasma Simulator of NIFS. The research was supported by KAKENHI (Nos. 21H04456, 22H05131, 23H04609, 22K18272, 23K03362), by the NINS program of Promoting Research by Networking among Institutions (01422301) by the NIFS Collaborative Research Programs (NIFS22KIIP003, NIFS24KIIT009, NIFS24KIPT013, NIFS22KIGS002, NIFS22KISS021) and by the ExCELLS Special Collaboration Program of Exploratory Research Center on Life and Living Systems(24-S5).

References

1) D. W. Brenner,O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni and S. B. Sinnott, J. Phys. Condens. Matter **14**, 783 (2002).
2) A.Ito, H. Nakamura, J. Plasma Phys. **72**, 805 (2006).
3) A. M. Ito, A. Takayama, Y. Oda, T. Tamura, R. Kobayashi, T. Hattori, S. Ogata, N. Ohno, S. Kajita, M. Yajima, Y. Noiri, Y. Yoshimoto, S. Saito, S. Takamura, T. Murashima, M. Miyamoto and H. Nakamura, J. Nucl. Mat. **463**, 109 (2015).
4) H. Nakamura, A.M. Ito, S. Saito, A. Takayama, Y. Tmura, N. Ohno, and S. Kajita, Jpn. J. Appl. Phys. **50**, 01AB04 (2011).
5) H. Nakamura, K. Takasan, M. Yajima, S. Saito, J. Adv. Simulat. Sci. Eng. **10**, 132 (2023).
6) H. Nakamura, H. Miyanishi, T. Yasunaga, S. Fujiwara, T. Mizuguchi, A. Nakata, T. Miyazaki, T. Otsuka, T. Kenmotsu, Y. Hatano and S. Saito, Jpn. J. Appl. Phys. **59**, SAAE01(2020).
7) Y. Hatano, H.Nakamura, S. Fujiwara, S. Saito, and T. Kenmotsu, *Damages of DNA in tritiated water*, The Enzymes, Academic Press **51**, 131(2022). 10.1016/bs.enz.2022.08.009.
8) S. Fujiwara, H. Nakamura, H. Li, H. Miyanishi, T. Mizuguchi, T. Yasunaga, T. Otsuka, Y. Hatano and S. Saito, J. Adv. Simulat. Sci. Eng. **6**, 94 (2019).
9) H. Li, S. Fujiwara, H. Nakamura, T. Mizuguchi, T. Yasunaga, A. Nakata, T. Miyazaki, T. Otsuka, T. Kenmotsu, Y. Hatano, S.Saito, Plasma Fus. Res. **14**, 3401106 (2019).
10) M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1991).
11) D. Frenkel and B. Smit, *Understanding Molecular Simulations: From Algorithms to Applications* (Academic, San Diego, 2002).
12) T. Tanabe (ed.), *Tritium: Fuel of Fusion Reactors* (Springer Japan, Japan, 2017).
13) United Nations Scientific Committee on the Effects of Atomic Radiation, *Sources, Effects, and Risks of Ionizing Radiation Annex C Biological Effects of Selected Internal Emitters-Tritium* (United Nations Publication 2016).
14) L. Mullenders, M. Atkinson, H. Paretzke, L. Sabatier, S. Bouffler, Nat. Rev. Cancer **9**, 596 (2009).
15) Y. Hatano, Y. Konaka, H. Shimoyachi, T. Kenmotsu, Y. Oya, H. Nakamura, Fusion Eng. Des., **146**, 1200 (2018).

16) K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, A. D. MacKerell Jr., J. Chem. Theory Comput., **8**, 348 (2012).

17) T. Otsuka, M. Taiji, D.R. Bowler, T. Miyazaki, Jpn. J. Appl. Phys. **55**, 1102B1 (2016).

18) C. W. Greider, Cell **97**, 419 (1999).

19) S.K. Nair, S.K. Sliverman, J.H. Chen, and Y. Xiao, Crystal structure analysis of TRF2-Dbd-DNA complex. 2012, `http://dx.doi.org/10.2210/pdb3SJM/pdb`.

20) B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartles, S. Boresch, A. Caflish, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, J. Comp. Chem. **30**, 1545 (2009).

21) A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, S. J. Plimpton, Comp. Phys. Comm. **271**, 10817 (2022).

22) https://www.lammps.org/index.html.

23) T. Schneider and E. Stoll, Phy. Rev. B. **17**, 1302 (1978).

24) S. Monti, A. Corozzi, P. Fristrup, K. L. Joshi, Y. K. Shin, P. Oelschlaeger, A. C. T. van Duin and V. Barone, Phys. Chem. Chem. Phys. **51**, 15062 (2013).

25) H. M. Aktulga, J. C. Fogarty, S. A. Pandit, and A. Y. Grama, Parallel Computing, **38**, 245 (2012).

26) A. Vashisth, C.Ashraf, W. Zhang, C.E. Bakis, and A.C. T. van Duin, J. Phys. Chem. A, **122**, 6633 (2018).

27) Y. Tsuchida, S, Saito, H. Nakamura, Y. Yonetani, S. Fujiwara, Nihon Simulation Gakkai Ronbunshi, **13**, 32 (2021) in Japanese.

28) Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K.

Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

29) A. D. Becke, J. Chem. Phys., **98**, 5648(1993).

30) C. Lee, W. Yang, R. G. Parr, Phys. Rev. B **37**, 785(1988).

31) T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007(1989).

32) T.P. Senftle, S.Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A.Grama and A.C.T. van Duin, npj, Comput. Mater. **2**, 15011(2016).

33) J. R. Gissinger, S. R. Zavada, J. G. Smith, J. Kemppainen, I. Gallegos, G. M. Odegard, E. J. Siochi, K. E. Wise, Carbon **202**, 336 (2023).

34) N. Hishinuma, K. Oikawa, T. Okada, M. Kanno, H. Yamazaki, W. C. Chung, A. Saito, H. Kohno, SENAC, **50**, 3 (2017) in Japanese.

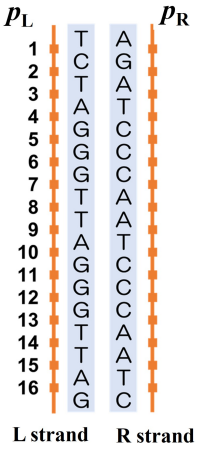

Fig. 1. The base sequence of the DNA. The four types of bases are adenine A, guanine G, cytosine C, and thymine T. The orange squares per strand denote the phosphorus atoms numbered on the L-strand with p_L and on the R-strand with p_R. The side with the sequence {TCTAGGGTTAGGGGTTAG} is denoted as L-strand, and the side with {AGATCCCAATCCCAAT} as R-strand.

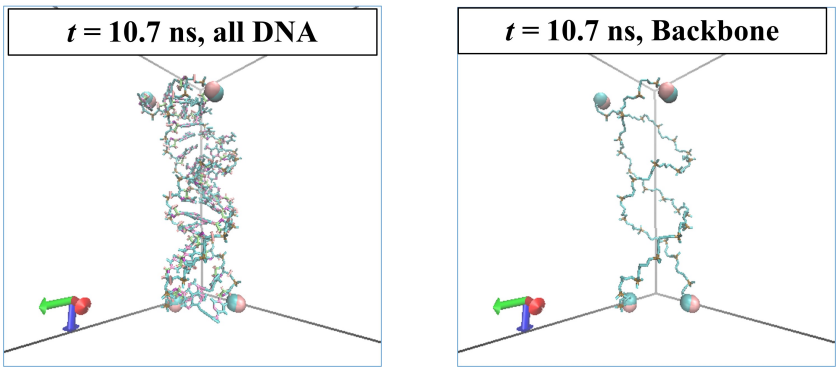

Fig. 2. The left figure shows snapshots of the structure of telomeric DNA at t = 10.7 ns. All the atoms that make up the telomeric DNA are drawn. In the right figure, only the main strands are picked up. In the third process, the four hydroxyl (OH) groups at the ends of the DNA strand (large balls) are fixed to prevent the DNA from collapsing from the ends.

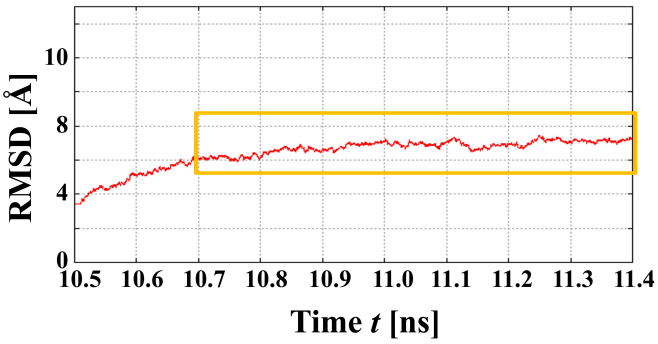

Fig. 3. Time evolution of the root mean square deviation (RMSD) of the atoms composing the telomeric DNA molecule. The RMSD remains nearby constant between 10.7 ns and 11.4 ns.

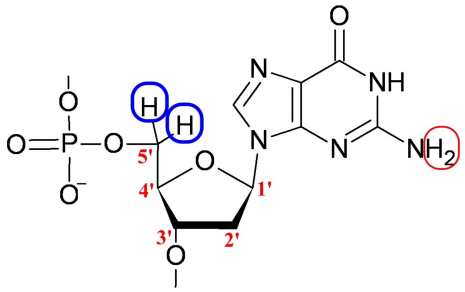

Fig. 4. Molecular structural formula of "scar model." This nucleotide is depicted as a base for guanine. The two hydrogens bonded to the 5′ carbon between the pentasaccharide and the phosphate group are highlighted by blue circles. In the scar model, these two hydrogens are removed. Incidentally, our previous work[6] employed the model in which the two hydrogens highlighted in red in the guanine were removed.

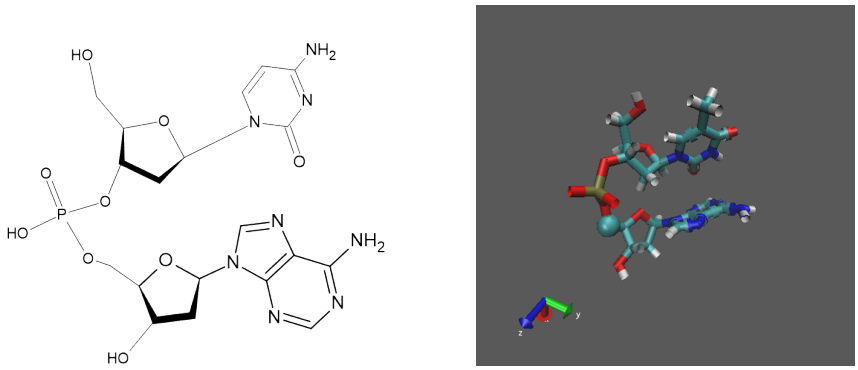

Fig. 5. The example of the smallest unit of the density functional theory (DFT) simulation for $p_{\mathrm{L}} = 10$ in Fig. 1. The two hydrogens are removed from the 5′ carbon between the nucleotides A and G. B3LYP exchange-correlation functional[29,30] and cc-pVDZ bases set[31] are used in the DFL simulation with Gaussian09.[28]

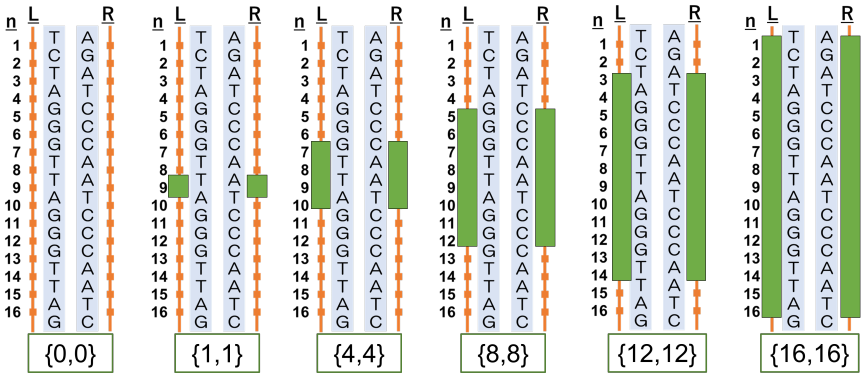

Fig. 6. Schematic diagrams of the scar model with varying numbers of scar pairs for $\{n_{\mathrm{L}}, n_{\mathrm{R}}\} = \{0,0\}, \{1,1\}, \{4,4\}, \{8,8\}, \{12,12\}$, and $\{16,16\}$. Each green squares represent scars.

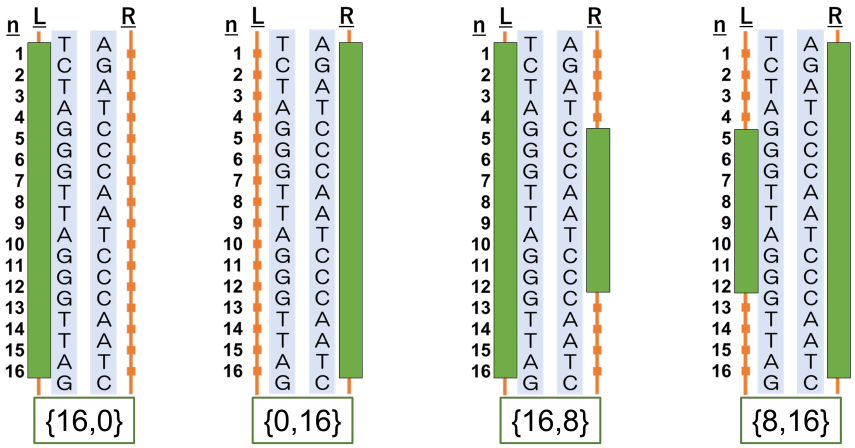

Fig. 7. Schematic diagrams of the scar model with varying numbers of scar pairs for $\{n_{\mathrm{L}}, n_{\mathrm{R}}\} = \{16,0\}, \{0,16\}, \{16,8\}$ and $\{8,16\}$. Each green squares represent scars, as in Fig. 6.

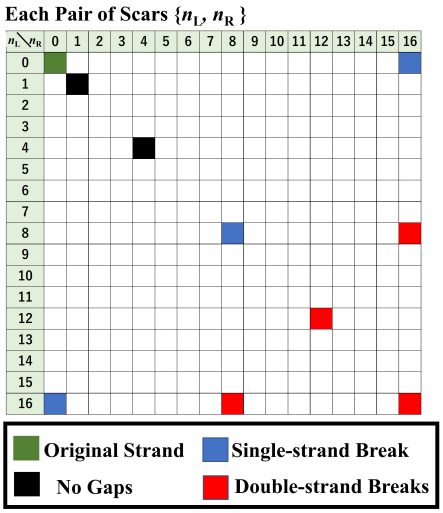

Fig. 8. Summary of the MD simulation results for $\{n_L, n_R\} = \{0, 0\}$, $\{1, 1\}$, $\{4, 4\}$, $\{8, 8\}$, $\{12, 12\}$, $\{16, 16\}$, $\{16, 0\}$, $\{0, 16\}$, $\{16, 8\}$ and $\{8, 16\}$ as shown in Figs. 6 and 7. The green square represents telomeric DNA in the absence of scars. At the end of the simulation ($t = 11.4$ ns), the black squares indicate no gaps in both DNA strands, the blue ones indicate single-strand breaks (SSB), and the red ones indicate double-strand breaks (DSBs).

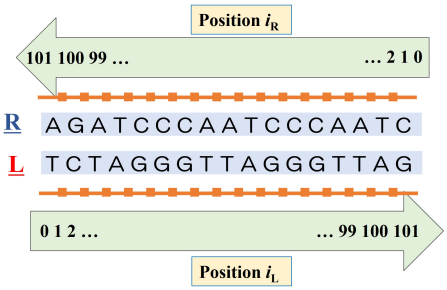

Fig. 9. Definition of the position $i_{R,L}$ on each R or L strand. The $i_{R,L}$ means atoms in the backbone of the R or L strand. On the other hand, the position $p_{R,L}$ in Fig. 1 means the positions of only phosphorous atoms in the backbone of the R or L strand.

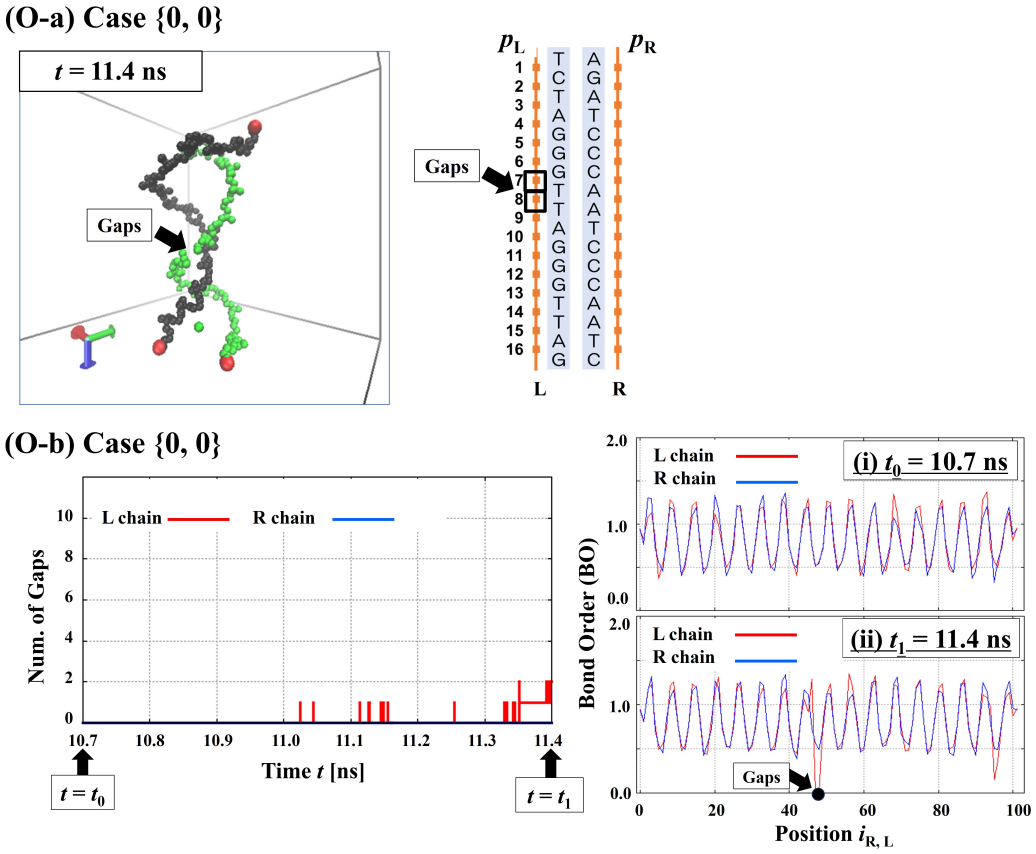

Fig. 10. For O case {0, 0}, the final (t = 11.4 ns) molecular structure is depicted on the left side of (O-a). There, one gap appears in the L-strand (green chain), the position of which is represented by black squares in the right figure of (O-a) with p_L = 7 and 8. The left side of (O-b) shows the time evolution of the number of gaps in each L (red) or R (blue) strand from t_0 =10.7 ns to t_1 =11.4 ns. The spatial distribution of the bond order for each strand is drawn on the right side of (O-b) at $t = t_0$ (upside) and t_1 (downside). Spatial distribution here means distribution with respect to $i_{R,L}$, which means the position of atoms in the backbone of the R or L strand as shown in Fig. 9. This figure shows that there is a gap (black circle) in the L strand at t_1 (downside).

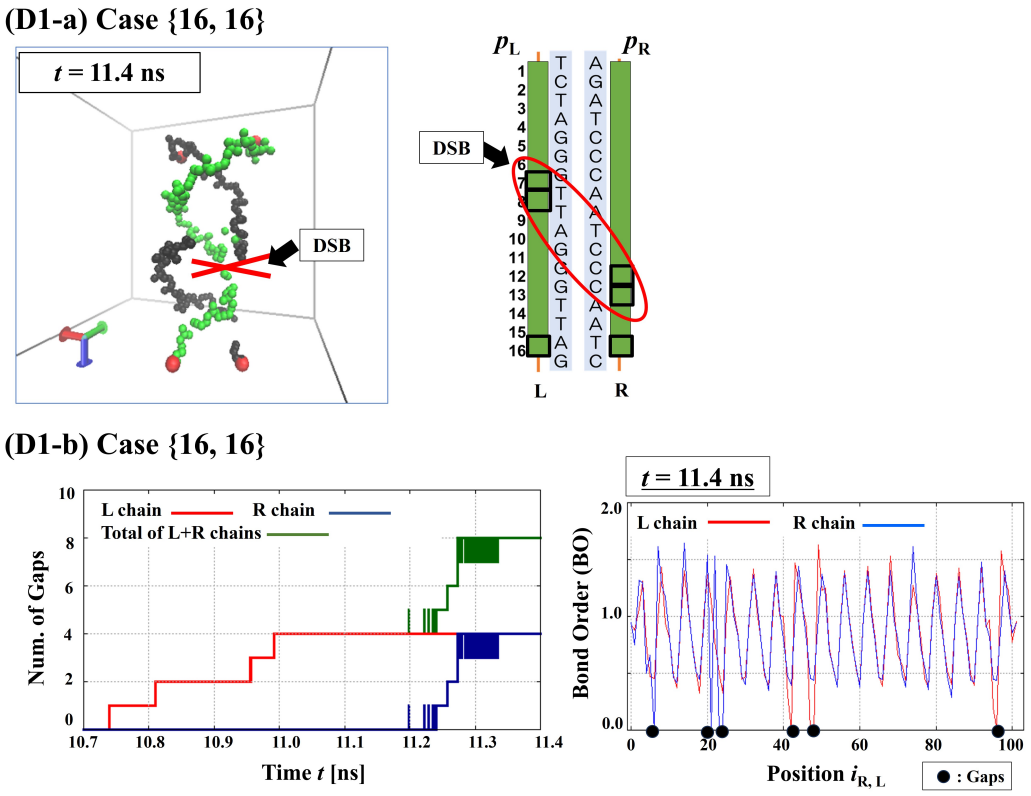

Fig. 11. In D1 case {16, 16}, there are 16 scars each on the R and L chains (32 scars in total), represented by green squares on the right-hand side of (D1-a). The red oval means the double-strand breaks. Moreover, double-strand breaks, which are depicted as the red line, appear in the final molecular structure shown on the left-hand side of (D1-a). Figure (D1-b) shows the time evolution of the number of gaps and the spatial distribution of the bond order.

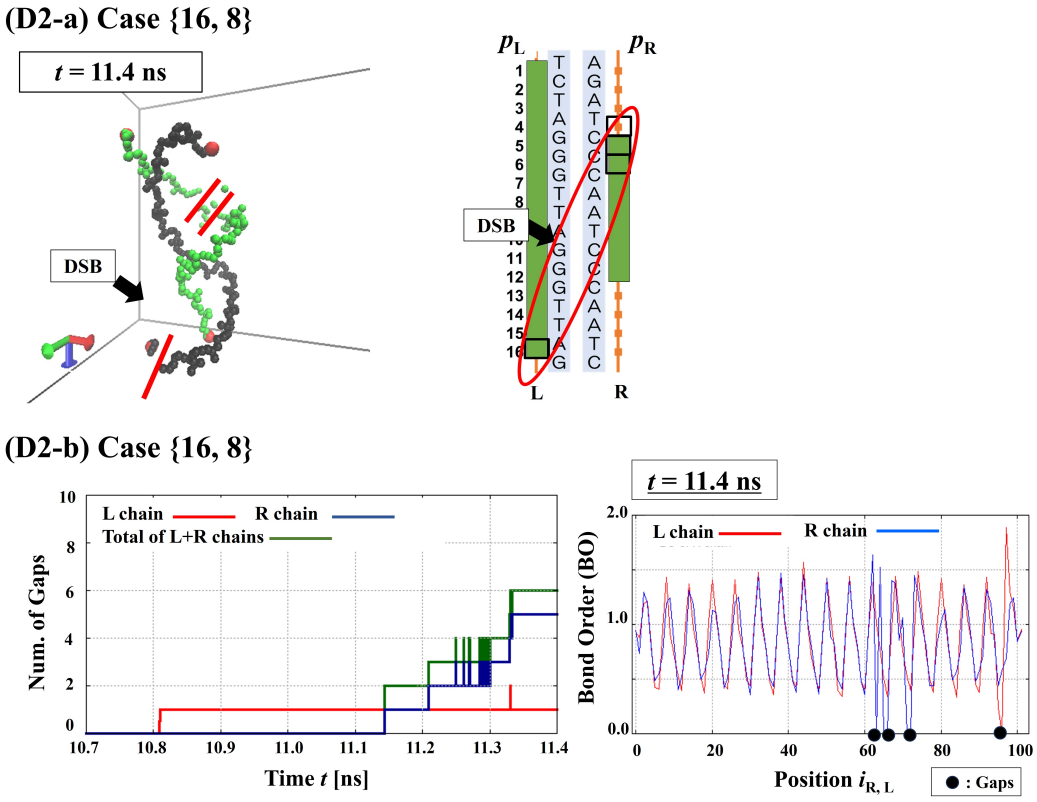

Fig. 12. In D2 case {16, 8}, there are 16 scars on the R chain and 8 scars on the L chain (24 scars in total), represented by green squares on the right-hand side of (D2-a). The red oval means the double-strand breaks. Moreover, double-strand breaks, which are illustrated by the red line, appear in the final molecular structure on the left-hand side of (D2-a). Figure (D2-b) shows the time evolution of the number of gaps and the spatial distribution of the bond order.

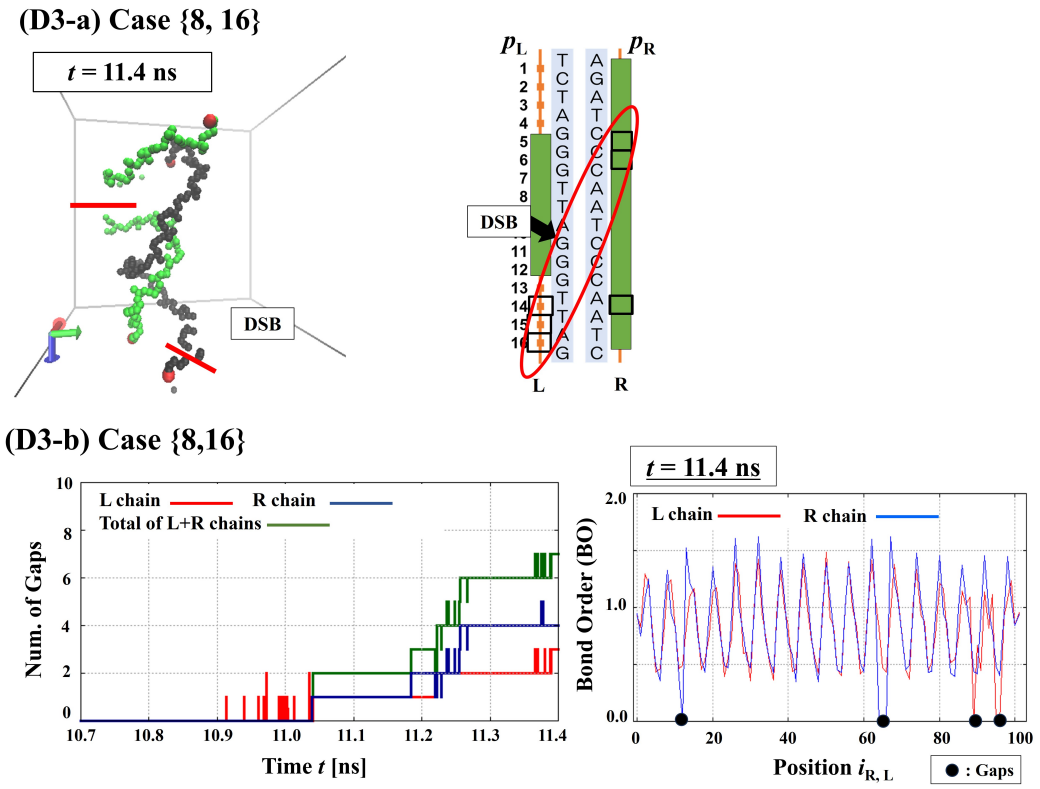

Fig. 13. In D3 case {8, 16}, there are 8 scars on the R chain and 16 scars on the L chain (24 scars in total), represented by green squares on the right-hand side of (D3-a). In this case, the left and right sides of case D2 are reversed. Moreover, double-strand breaks, which are illustrated by the red line, appear in the final molecular structure on the left-hand side of (D3-a). Figure (D3-b) shows the time evolution of the number of gaps and the spatial distribution of the bond order.

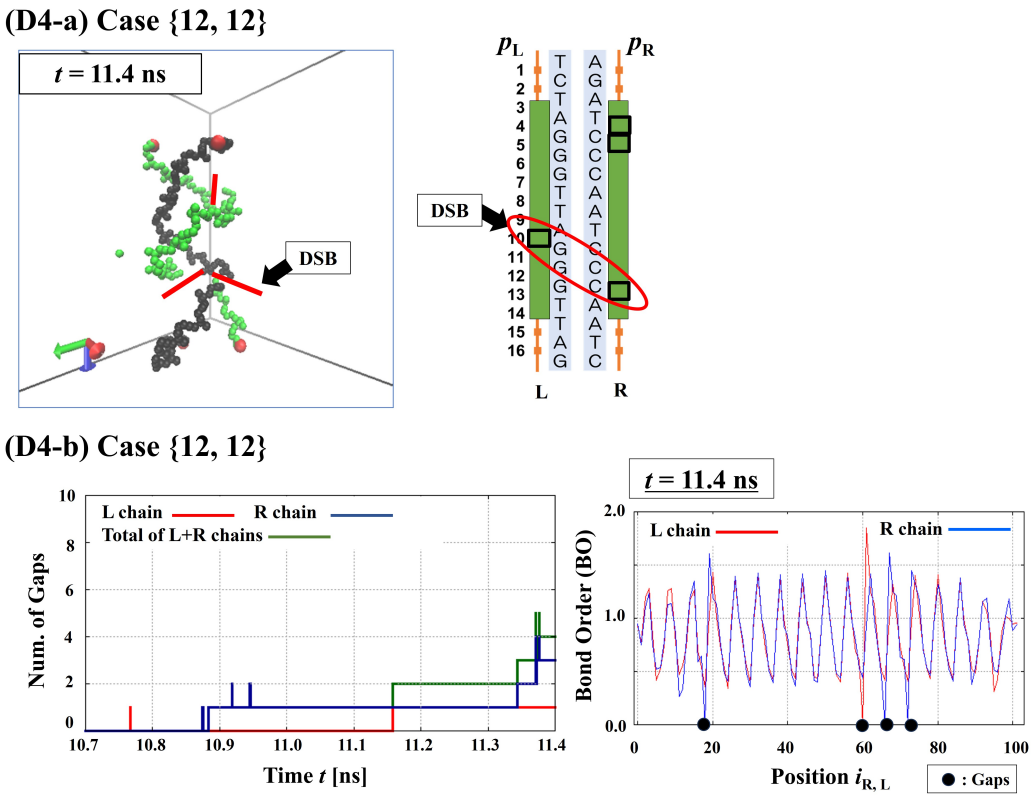

Fig. 14. In D4 case {12, 12}, there are 12 scars each on the R and L chains (24 scars in total). Although the total number of scars is less than in other cases D1, D2, and D3, DSBs are still occurring.

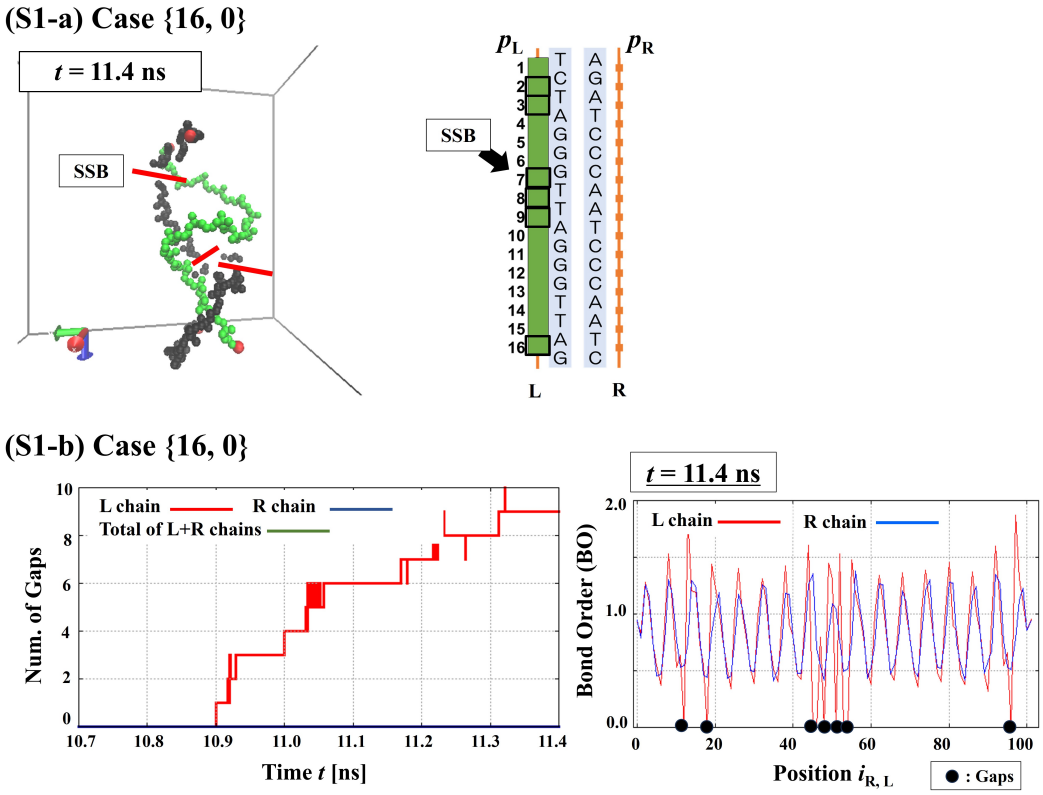

Fig. 15. In S1 case {16, 0}, there are 16 scars on the L chain and no scars on the R chain (16 scars in total), represented by green squares on the left-hand side of (S1-a). The gaps in the L chain increase as time goes by. In the R chain, no gaps occur.

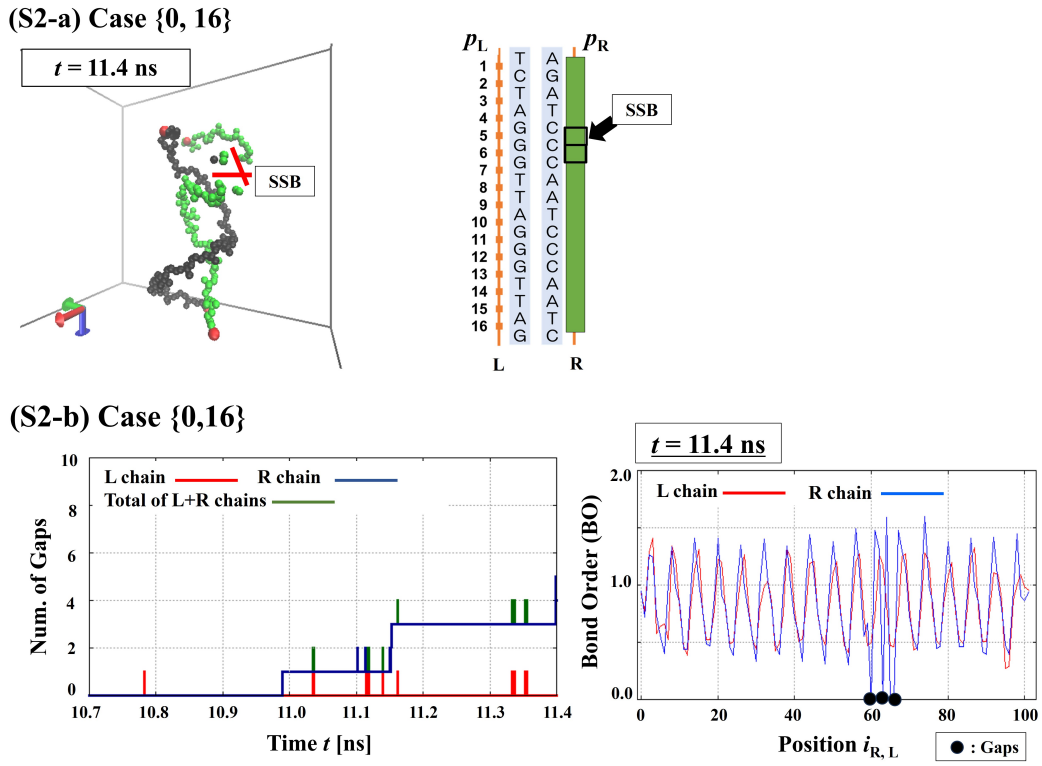

Fig. 16. In S2 case {0, 16}, there are 16 scars on the R chain and no scars on the L chain (16 scars in total) which is contrary to the S3 case, represented by green squares on the right-hand side of (S2-a). The gaps in the R chain increase as time goes by. However, gaps occur momentarily in the L chain but are quickly recovered.

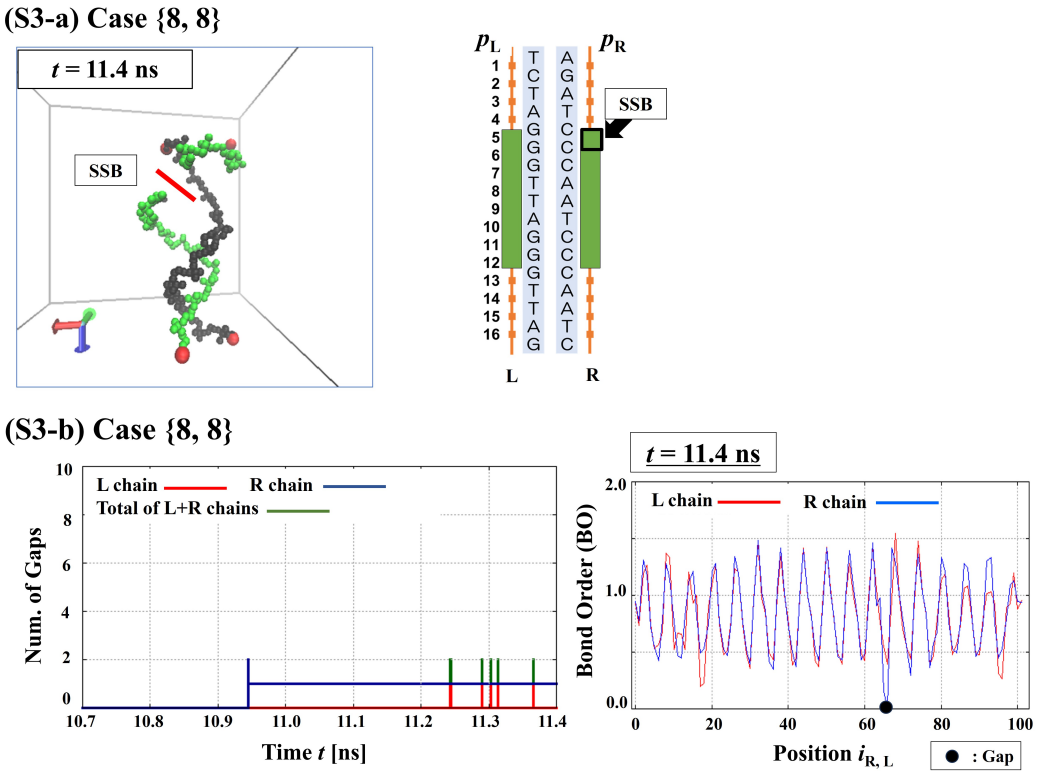

Fig. 17. In S3 case {8, 8}, there are 8 scars each on the R and L chains (16 scars in total). Gaps occur in the R chain, with one failing to recover and persisting. In contrast, once a gap emerges in the L chain, it is capable of recovery. This phenomenon results in the formation of an SSB.

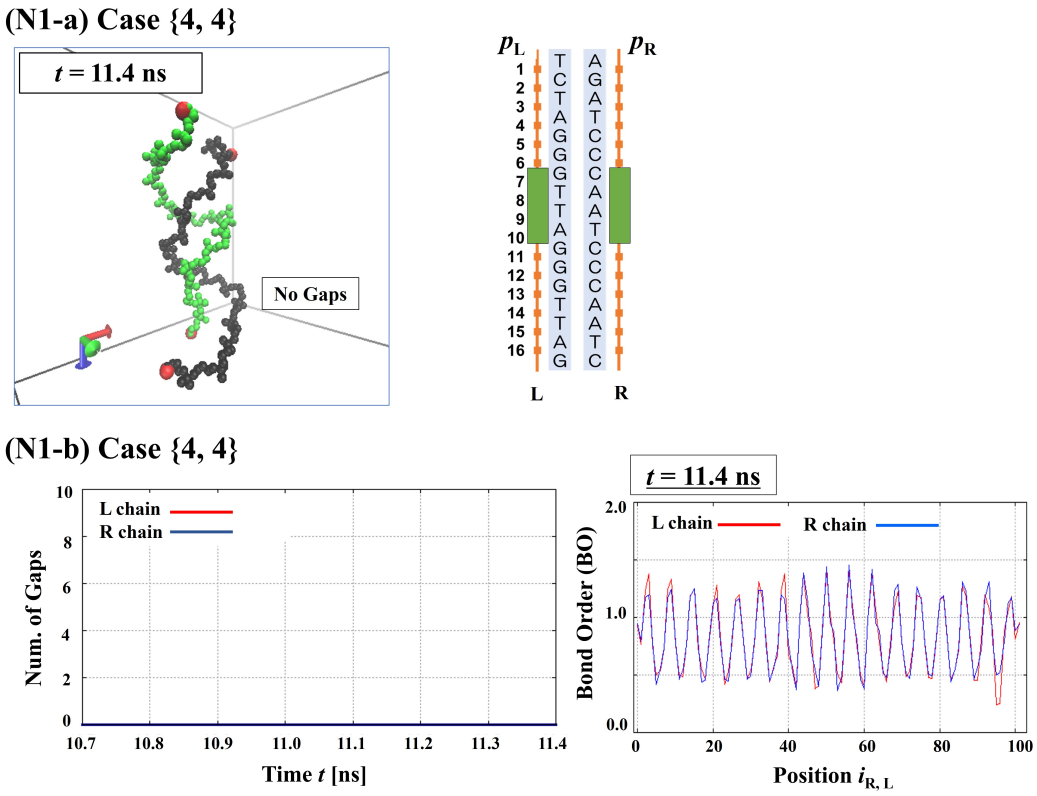

Fig. 18. In N1 case {4, 4}, there are 4 scars each on the R and L chains (8 scars in total). No gaps occur in the R and L chains.

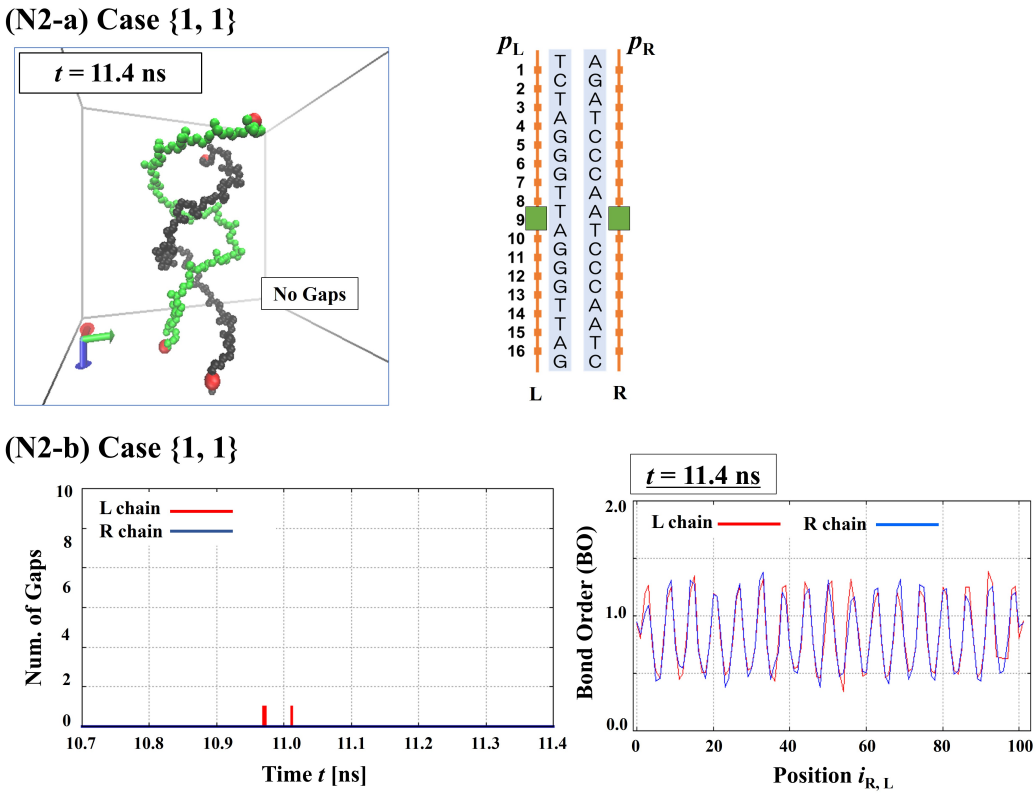

Fig. 19. In N2 case {1, 1}, there is one scar each on the R and L chains (2 scars in total). No gaps occur in the R and L chains.

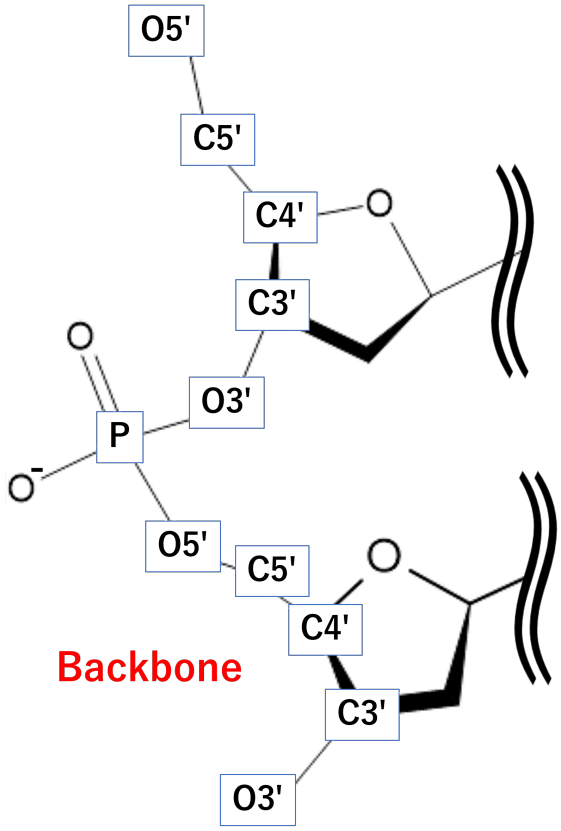

Fig. 20. The unit structure of telomeric DNA backbones

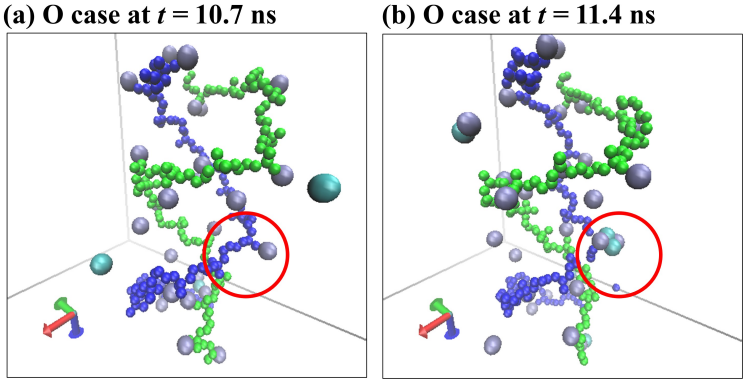

Fig. 21. Sodium (Na^+) and chloride (Cl^-) ions aggregate at the gap locations in O case {0,0}. A gap occurs at p_L = 7 and 8 circled in red. The R or L strand is drawn by green or blue balls. Light blue or purple ball means Na^+ or Cl^-, respectively. The time t = 10.7 ns is in the left-hand figure (a), and 11.12 ns is in the right-hand one (b).