Structural and energetic consequences of oxidation of d(ApGpGpGpTpT) telomere repeat unit in complex with TRF1 protein

Piotr Cysewski, Przemyslaw Czeleń

► To cite this version:

Piotr Cysewski, Przemyslaw Czeleń. Structural and energetic consequences of oxidation of d(ApGpGpGpTpT) telomere repeat unit in complex with TRF1 protein. *Journal of Molecular Modeling*, 2010, 16 (11), pp.1797-1807. 10.1007/s00894-010-0730-8 . hal-00592595

HAL Id: hal-00592595

<https://hal.science/hal-00592595>

Submitted on 13 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Editorial Manager(tm) for Journal of Molecular Modeling
Manuscript Draft

Manuscript Number: JMM01154R1

Title: Structural and energetic consequences of oxidation of d(ApGpGpGpTpT) telomere repeat unit in complex with TRF1 protein

Article Type: Special Issue Prague

Keywords: stacking, telomere, oxidative damage, 8-oxoguanine, TRF1, many-body contributions

Corresponding Author: Prof. Piotr Cysewski, Ph.D.

Corresponding Author's Institution: Collegium Medicum, Nicolaus Copernicus University

First Author: Piotr Cysewski, Ph.D.

Order of Authors: Piotr Cysewski, Ph.D.; Przemysław Czeleń, PhD student

Manuscript Region of Origin:

Abstract: The configuration hyperspace of canonical and oxidized 14-mers of B-DNA comprising telomere repeat units d(ApGpGpGpTpT) was sampled over 40ns via molecular dynamic (MD) simulations. The energetic and structural consequences of TRF1 binding to telomere B-DNA were compared with non-complexed systems. Energetic properties of analysed pairs, di- and tri-nucleotide steps occurring in central telomere repeat unit were estimated by means of advanced quantum chemistry computations including not only BSSE corrections, electron correlation contributions but also non-negligible many-body terms. These data along with bases pair and base step parameters distributions allow for quantization of consequences of oxidation and/or TRF1 binding to telomere repeat units. Occurrence of 8-oxoguanine in central telomeric triad (CTT) is the source of high stiffness if compared to non-modified oligomer. The origin of this property comes from significantly alteration of intermolecular interactions introduced by 8-oxoguanine. The increased stability observed for base-base interactions are accumulated and characterizes also di- and tri-nucleotides. The observed changes in the intermolecular interactions originate from structural alterations imposed by TRF1 binding to canonical and oxidized telomere B-DNA. First and most direct consequence of TRF1 binding to oxidized telomere repeat unit is alteration of shift-slide correlations if compared to canonical system. This in turn leads to large differences in purine-purine overlapping in oxidized structures. Thus, oxidized telomere B-DNA double strands are sensitive to interactions with protein ligands and numerous structural and energetic changes are imposed on base pairs forming CTT.

Response to Reviewers: Reviewer #1:

1. In supporting Figure S1 and Figure S2, is it median values or mean values of the parameters distributions?

In all cases median values were used and misleading captions were corrected.

2. The format and the sequence number of the references need to be checked (Such as Ref. 14). Thank you, the text was modified accordingly.

Reviewer #2:

1. Authors used a slightly different sequence for isolated and complexed DNA systems - the initial structure of the complex is from NMR, but no information is about an origin of the free oligomer. Was this structure prepared directly in AMBER or is it based on any experimental structure? Why authors did not use the same structure?

Indeed oligomers denoted as S1 and S2 (free B-DNA) had not exactly the same sequence as S3 and S4 (TRF1 in complex with B-DNA). Reason of this was strictly technical. Before finding NMR structures deposited in NDB database we have modeled non-complexed structures comprising canonical and oxidized telomere repeat unit in water solution. We did not expect that conclusions are sensitive to oligomer sequence especially that such differences are in the region outside of our interest.

Nevertheless all computations on both molecular dynamics and quantum chemistry levels were repeated for exactly the same sequences. Thus, solution structures of B-DNA bound to human TRF1 that are provided by 1IV6.pdb file (20 conformations) were used as the source of all four sets of data (S1, S2, S3 and S4). In the previous version of our paper just one structure was chosen as a source of initial conformation of S3 chain. This time we have performed independent molecular dynamics runs for all 20 available conformations. Based on data provided in 1IV6.pdb file new initial structures were prepared for S1 (after removing all atoms belonging to protein) and S2 (by replacement of guanine with 8-oxo-guanine in the central telomeric triad of S1) in each of 20 conformations. Thus, the whole paper covers 40ns dynamics for each of studied oligomer (S1-S4). However, independent 2.5ns runs were performed for every initial structure. This provide broad range of conformations that were starting from different experimental structures. Furthermore, in order to check adequateness of such short MD runs the time autocorrelation functions were estimated according to specification in NIST document (www.itl.nist.gov/div898/handbook). Below there are provide typical plots:

2. There is no information about a total number of simulations. There are four types of studied systems (free, complex, free+oxoG, complex+oxoG) but it seems to me that there is only one simulation for each type - is it true? ... The most critical point is the length of simulations which is - in my opinion - insufficient. Today, a common length of MD simulations is in tens of nanoseconds, frequently exceeding 50-100 ns. I am rather skeptic if the system (mainly in case of the complex, which contains ~1800 atoms and water box) is fully equilibrated after 0.5 ns and is able to show characteristic dynamical behavior in a few nanoseconds. To make qualified decisions from 2 ns of MD simulations is not possible and is necessary to extend simulations (minimally to 10 ns) and (if it is possible) to carry out multiple simulations to improve sampling. Only in this case it is possible to make meaningful conclusions.

Previous paper utilized just one short simulation. Now we performed 20x2.5ns molecular dynamics and as long as 40nps trajectories were used for averaging and most representative structures extraction. Fortunately after extension of trajectories the main concussions were almost not affected by the length of simulation. This is due to fortunate selection of initial structures. Below, it is presented demonstration of this "lucky" selection of starting points for MD runs. The correlation between median values of shift and slide for 20 distinct trajectories are plotted for S3 and S4 systems. Red symbol denotes structure that was used in previous version of our paper. Thus it is not surprising that essential structural and dynamic features were included in previous MD simulations.

I would like also to add one more comment. Previously we have selected every 20ps snapshot for ab initio computations. This resulted in 100 structures and was quite feasible for our computer cluster. After extension form 2000 points up to 40000 structures it is completely impractical such way of selection and we have proposed alternative approach described in the methodology part. In addition I would like to present here figure that is not included in the paper, but probably will be interesting for Referee. The workflow of proposed algorithm is presented in supplementary material within footnote to Fig. S1 so I will not to repeat it here. The main criterion for most representative structures selection

is similarity of IIE histograms obtained for probing set and one that was averaged over all structures used for all probing sets. In figure presented below all gray lines represent IIE distributions obtained within MM framework for different probing sets. Each probing set comprised 36 conformations of dinucleotide steps representing the most frequently occurring ranges of base step parameters. Smoothed histogram presented as red line represents averaging over all 360 structures coming from 10 distinct probing sets. Dotted line corresponds to histogram estimated based on 2500 dinucleotide conformations. It is clearly visible that averaging over 10 probing sets is as good as over thousand of dinucleotide conformations.

3. Do authors know about the parmbsc0 version of the Amber force field which is today commonly used for MD simulations of DNA molecules? This force field has modified backbone torsion profiles to avoid irreversible backbone changes (α/γ torsion angle switches) observed in long MD simulations. Although this is not this case I recommend authors to use this force field version in their future simulations.

Thank you for this comment. We are aware of Barcelona corrections to the Amber forcefield (Pérez et al. Biophys. J. 2007, 92(11): 3817-3829) and we use this set of parameters in all of our simulations. Just for ensuring this the citation is added in the reference list.

4. There is no figure showing DNA-TRF1 complex - it would be useful to see where the mutation is taking place in the complex.

I am sorry not to providing such figure but I am ready to fulfill this requirement if Referee will insist. Since in Ref.8 there were presented molecular details showing contacts between TRF1 and B-DNA, I do not think it is really important. Paper is quite overloaded with figures and data, so please forgive me not providing such figure.

5. The list of references is rather limited - I am missing more oxoG and oxoG/MD papers. Thank you more references and comments are provided now in the introductory section.

Structural and energetic consequences of oxidation of d(ApGpGpGpTpT) telomere repeat unit in complex with TRF1 protein

Received: 14.12.2010 / Accepted: 21.04.2010

Piotr Cysewski^{1,2,✉} and Przemysław Czeleń¹

¹Physical Chemistry Department, Collegium Medicum, Nicolaus Copernicus University, Kurpińskiego 5, 85-950 Bydgoszcz, Poland, email: piotr.cysewski@cm.umk.pl

²General Chemistry Department, University of Technology and Life Sciences in Bydgoszcz, Faculty of Chemical Technology and Engineering, Seminaryjna 3, 85-326 Bydgoszcz, Poland

✉Email: piotr.cysewski@cm.umk.pl

Abstract

The configuration hyperspace of canonical and oxidized 14-mers of B-DNA comprising telomere repeat units d(ApGpGpGpTpT) was sampled over 40ns *via* molecular dynamic (MD) simulations. The energetic and structural consequences of TRF1 binding to telomere B-DNA were compared with non-complexed systems. Energetic properties of analyzed pairs, di- and tri-nucleotide steps occurring in central telomere repeat unit were estimated by means of advanced quantum chemistry computations including not only BSSE corrections, electron correlation contributions but also non-negligible many-body terms. These data along with bases pair and base step parameters distributions allow for quantization of consequences of oxidation and/or TRF1 binding to telomere repeat units. Occurrence of 8-oxoguanine in central telomeric triad (CTT) is the source of high stiffness if compared to non-modified oligomer. The origin of this property comes from significantly alteration of intermolecular interactions introduced by 8-oxoguanine. The increased stability observed for base-base interactions are accumulated and characterizes also di- and tri-nucleotides. The observed changes in the intermolecular interactions originate from structural alterations imposed by TRF1 binding to canonical and oxidized telomere B-DNA. First and most direct consequence of TRF1 binding to oxidized telomere repeat unit is alteration of *shift-slide* correlations if

compared to canonical system. This in turn leads to large differences in purine-purine overlapping in oxidized structures. Thus, oxidized telomere B-DNA double strands are sensitive to interactions with protein ligands and numerous structural and energetic changes are imposed on base pairs forming CTT.

Keywords Stacking · Nucleobases interactions · Telomere · Oxidative damage · 8-oxoguanine · TRF1 · Many-body contributions

Introduction

Telomeres are important biological complexes playing crucial role in preserving genome stability [1]. Their main function is stabilization of chromosome ends and protecting them from being recognized as B-DNA double strand breaks [2]. In humans all telomeric B-DNA fragments consist of several kilobases of six nucleotides sequences (AGGGTT) constituting so-called telomere repeat unit (TRU) [3]. These chromosomes tails are shortened with every cell division by about 20-200 base pairs, therefore often they are called “mitotic clocks” [5-7]. There are specific protein ligands called TRF1 [3] and TRF2 [4] that can attach to telomere B-DNA in cells. Both these proteins interact with telomeric AGGGTT sequences by binding to major groove *via* strong contacts of different types [8]. Among many stabilization contributions there are produced distinctive hydrogen bonds between amino acids residues and G-C base pairs in GGG triplet [8]. Also direct interactions with ribose phosphate backbone on both sides of the major groove and N-terminal arm [9] are important. Stability of this complexes is highly sensitive to mild oxidative stress-induced by hyperoxia, mitochondrial dysfunction, presence of arsenic or exposure to UV irradiation [10, 11]. Occurrence of oxidative lesions in the telomeric fragment significantly affects process of DNA-protein complexes formation. Presence of one 8-oxoguanine molecule in the central telomeric triad (CTT) reduces the affinity of protein binding by factor 2 compared with standard non-damaged DNA [12]. Such profound reduction of complex instability can be addressed to significant changes in structural characteristics affecting dynamic properties of oxidized CTT [13]. Besides, observed alterations of electrostatic properties imposed by oxidation is an important factor contributing to TRF binding. In our previous works [13, 14] there were discussed some important consequences of telomere repeat unit oxidation. The application of MD/QM quenching method revealed significant stiffness of CTT as a results of CTT oxidation. This was related to skewness decrease of *roll* and *shift* distributions, increase of skewness and kurtosis of *opening* distributions and increase of kurtosis of *roll* distributions characterizing oxidized CTT with respect of canonical one. These structural changes imposed the increase of intermolecular interaction energies (IIE) characterizing stacking interactions of 8-oxoguanine with neighboring guanine molecules and hydrogen bonding of 8-oxoguanine with cytosine. The aim of this paper is to analyze consequences of TRF1 binding to canonical and oxidized telomere B-DNA oligomers on structure and energetics of telomere repeat unit. Although there were published important studies on dynamic and structural properties of oxidized B-DNA to our best knowledge consequences of TRF1 complexation and/or

oxidation were not described so far. For example nowadays Naômé et. al [15] characterized hydration and ion binding patterns around double-stranded B-DNA containing 8-oxoguanine supporting existence of early recognition modes utilized by hOGG1 enzyme. Molecular dynamics simulation of clustered DNA damage sites that contain 8-oxoguanine were studied by Fuhimoto et. al. [16]. Dodson and Lloyd [17] studied mechanism of substrate recognition for BER glycosylases *via* base flipping of DNA sequences containing 8-oxoguanine. The generated trajectories were used for identification of such properties of oligonucleotides in the vicinity of the altered site that are responsible for recognition. Dynamic behavior of DNA oligomers containing 8-oxoguanine was also studied by Cheng et.al. [18]. Besides, short 1ns molecular dynamics simulations were performed for elucidation of the effect of guanine lesion presence in B-DNA [19]. Much larger MD simulation were carried out by Fujimoto *et al.* combined with biomolecular visualization software [20]. However, most of theoretical studies on 8-oxoguanine in B-DNA are related to explanation [21, 22] of activities of human oxoguanine glycosylase (hOGG1) [23-26] and TRF1 binding was analyzed with much lower intensity [27-30]. This paper concentrates on energetic stabilities of CTT characterized on post-SCF quantum chemistry level including not only BSSE corrections [31, 32] and necessary electron correlation contributions [33-35] but also many-body terms [35] for all of analyzed systems. Applied here methodology was successfully tested in our previous investigations [36-39] for polymorphism-related heterogeneities of guanine stacking in B- and A-DNA forms [37], characteristics of *inter*- and *intra*-strand stacking interactions in d(CpG) and d(GpC) steps found in B-DNA, A-DNA and Z-DNA crystals [38], description of energetic heterogeneities in canonical and oxidized central guanine triad of B-DNA telomeric fragments [13,14] and quantification of all possible *intra*-strand stacking interactions between nucleobases [39]. The essential feature of applied here approach is characteristics of statistically significant probe of conformations and consequently obtaining of intermolecular interaction energies (IIE) distributions coming from quite long (40ns) molecular dynamics sampling of analyzed oligomers. The averaged IIE values characterizing intermolecular interactions within CTT can then be supplied with corresponding standard deviations giving an impression of magnitude of structural and energetic fluctuations.

Methods

The meaningful conformations of any molecular system require adequate sampling over configurational hyperspace using one of many available theoretical approaches [40-44]. Among them molecular dynamics (MD) is widely accepted as an accurate tool for probing dynamics of many models of diverse biological systems within picoseconds or even microseconds time scales [41-44]. Along with increase of computer power time of simulations also continues to increase providing large amounts of data. This ensures that sampling provide adequate insight into the dynamics of interactions. However, this is accomplished with interpretation problems since selection of the most representative structures (MRS) used for characteristics of analyzed system is not a trivial task. One of commonly applied approach of extraction such representative probe is clustering of molecular configurations according to their structural and energetic similarities. Unfortunately, there is no one ideal data-mining technique [45] providing univocal way of MRS selection and different algorithms lead to different descriptions [45-48]. As it was demonstrated by Shao *et al.* [45] some algorithms tend to produce homogeneously sized clusters, whereas others have a tendency to produce singleton clusters. For these reasons in this paper we adopted another strategy for selection of MRS characterizing structural and energetic heterogeneities of canonical and oxidized B-DNA CTT present in free oligonucleotides and complexed with TRF1 protein in water solution. The central telomeric triad comprising d(G₂pZ₃pG₄) sequence seems to be crucial from the perspective of TRF1 binding and this fragment was analyzed in details. Since conformations of dinucleotide steps are univocally characterized by set of six values of helical parameters [49, 50] the distributions of these quantities instead of actual atomic coordinates were used for MRS selection. Using trajectories generated during 40ns sampling with 1ps step there were extracted *shift*, *slide*, *rise*, *roll*, *tilt* and *twist* values using X3DNA package [51]. Then after normalization of all distributions there were prepared histograms by division of the whole span into six equal intervals. Based on derived histograms the weighting coefficients were obtained for each of base pair step parameter. Then there were randomly selected structures in number corresponding to estimated weights for each interval. This led to 36 structures for each probing distribution of each analyzed dinucleotide step. Typically this procedure was repeated ten times for generation of distinct probing distribution sets. The intermolecular interaction energies (IIE) were estimated using Amber forcefield [52, 53] for all dinucleotide step belonging to each probing distributions. For this purpose from selected pdb files corresponding to each MD trajectory snapshot there

were cut dinucleotide steps for which sugar phosphate backbones was replaced with hydrogen atoms. Position of N-glycosidic hydrogen atoms were optimized according to geometric criterions. Obtained values of dinucleotide stabilization energies were used for histograms plotting. There were also prepared histograms corresponding to all analyzed structures for given dinucleotide step (360 values). This lead to series of histograms of different shapes. For actual quantum chemistry computations the most representative set was selected by comparison to averaged histogram. The RMSD of histograms points were computed and the smallest value indicated the most representative set of conformations (MRS). The whole procedure was automatized by program written by authors. Initially different length of intervals in histograms, different number of selected structures for each interval and different number of probing sets were considered but the resulting IIE distributions were hardly dependent on these parameters. More details please find in supplementary materials (Fig. S1 and footnote). Only most representative sets of structures were used for precise intermolecular interaction energies (IIE) estimation at DF-MP2/aug-cc-pvDZ (aDZ) level of theory using MolPro package [54]. Since dinucleotide steps comprising stacked guanine molecules can exhibit significant non-negligible many-body contributions [35,36] all IIE values were corrected using linear relationships proposed elsewhere [56]. In all cases where quantum chemistry calculations were applied the simplified model was used in which sugar-phosphate backbones were replaced with hydrogen atoms. However, the conformations of nucleobases exactly matching to ones occurring in MD runs were prepared according to procedure applied in our previous investigations [13, 14, 36-39]. All IIE values were also corrected for BSSE error according to the counterpoise scheme [31, 32].

As it was mentioned beforehand the subject of our interest is the telomere repeat unit (AGGGTT). The water solution structures of d(GpTpTpA₁p**G₂pG₃pG₄**T₅pT₆pApGpGpG) oligomer in B-DNA form complexed with human TRF1 are available in NDB database [50] denoted as 1IV6 [55]. It comprises 20 distinct conformation of this protein-DNA tandem. These conformations were used as the source of initial structures of four studied systems (S1, S2, S3 and S4). Symbol S1 stands for free non-modified B-DNA oligomer, S2 represents oxidized analogue d(GpTpTpA₁p**G₂pZ₃pG₄**T₅pT₆pApGpGpG), S3 denotes non-oxidized telomere oligomer in complex with TRF1 and S4 refers to oxidized telomeric fragment bound to TRF1. Symbol Z stands for guanine (in S1 or S3) or 8-oxoguanine (in S2 or S4). Central telomeric triad was marked using bold style. The available 20 conformations coming from NMR experiment were then the source of all S3 initial conformations. Preliminary structures

of S1 were prepared by removing of all atoms belonging to protein from each of S3 conformation. Then S2 and S4 were build by replacement of G₃ with 8-oxoguanine in S1 and S3 respectively. Independent 2.5ns molecular dynamics runs were performed for all 20 available conformations. Last 2ns of all trajectory were used as the source of structural parameters collected every 1ps. The time autocorrelation functions [57,58] were plotted for ensuring that such short trajectories can still provide an adequate sampling of B-DNA dynamical structure. Since relaxation times of helicoidal parameters are proven to be shorten then 0.5ns [58] their distributions appear to be well stabilized within applied time scale. Additionally different starting points ensure proper sampling of dynamic features of analyzed systems. The X3DNA package [51] was used for analysis of B-DNA conformation. Thus, trajectories corresponding to 40ns dynamics for each of studied oligomer (S1-S4) were obtained. Each oligomer was immersed in TIP3P water box and neutralized by 26 sodium cations. The periodic boundary conditions were imposed. All molecular dynamic simulations were done at room temperature with an aid of AMBER package [53] with Barcelona corrected forcefield [59].

Results and discussion

Trajectories generated during molecular dynamics simulations allow for quantitative characteristics of structural diversities imposed by oxidation and/or TRF1 complexation of analyzed telomere repeat units. Besides, statistically significant probe of conformations taken for quantum chemistry calculations offers unique insight into heterogeneity of intermolecular interaction imposed by these two important factors. Due to commonly known inaccuracies of any available force-field parameterizations the reliable energetic analysis cannot be achieved using MM level. This is especially important for stacking interactions for which attractions are granted from dispersion forces and inclusion of electron correlation is a must [32,34]. To our best knowledge this is the first such comprehensive attempt to characterize structural and energetic changes imposed by protein-DNA complexation using quantum chemistry approach.

Structural diversities of telomere repeat unit

Description of oligonucleotide chain flexibility and its sensitivity to external factors poses some challenge on the way of presentation. In order not to be obscured by very details coming from time evolution of so many parameters, the analysis of the whole distributions was done in terms of three statistical measures. Obviously the median value is the first choice representing the averages of particular distribution. The comparison of medians supplemented with values of standard deviation can directly reveal structural consequences of particular factor acting on the oligomer. Since not all distributions are of normal type one can find such distributions which are characterized by almost the same median values (along with similar standard deviations) and still differences between them can be statistically significant. In such cases two additional statistical measures, namely skewness and kurtosis, are very useful since departures from normality are displayed by combinations of their values. The first quantity (skewness) simply evaluates the population asymmetry. Positive values of skewness indicate that right part of distribution is more extended than the left one. Consequently existence of left tail on distribution plots is indicated by negative values of this parameter. Thus, the non-zero values of skewness reveal changes in preferences of particular conformations adopted by analyzed oligomers. The second quantity used here is kurtosis, which offers very convenient way for characterizing shape of distributions in sense of the degree of peakedness. In case of normal distribution kurtosis is equal to zero and represent so-called mesokurtic distribution. A pure leptokurtic distribution has a higher peak than the normal distribution and has heavier tails. On the contrary a platykurtic distribution has a lower peak than a normal distribution and lighter tails. There is simple and direct structural interpretation of deviation from mesokurtic distributions, namely leptokurtic character indicates stiffness of analyzed oligomer structure with respect of particular parameter. On the other hand flexibility with respect of particular parameter is correlated with platykurticity and negative values of kurtosis. The distributions of all base step and base pair parameters characterizing flexibility of the whole telomeric repeat unit $d(A_1pG_2pZ_3pG_4pT_5pT_6)$ were analyzed. Each of dinucleotide steps is to be characterized by 18 parameters distributions for each of analyzed chains. In supplementary materials detailed characteristics of obtained distributions are collected in Tables S1 and S2 along with graphical presentation (Fig. S2 and S3) of median values variations for all telomeric repeat units belonging to four analyzed oligomers. Additionally in Fig. 1 there are provided the most representative plots.

<Figure 1 here>

Although there are observed some alterations of median values, the majority of base pair distributions differ by less than standard deviations. This suggests that only modest changes are enforced on such base pair parameters as *shear*, *stretch*, *stagger*, *propeller* and *opening* by either of analyzed factors (see Table S1, S2 and Fig. S1, S2). The only exceptions are visible for *buckle*. In Fig. 1 it is demonstrated that G₄ guanine (third one in CTT) is characterized by higher values of *buckle* in case of S4 oligomer. Also T₅ and T₆ pairs are affected on the same way by TRF1 binding to both canonical and oxidized telomere fragments. Values of *buckle* represent bending of hydrogen bonded pairs along shorter molecular axis (perpendicular to DNA chain in major-minor groove direction). Besides, the median value of *propeller* describing mutual rotation of molecular planes around shorter molecular axis is lower for bases inside CTT of S1 and S2 systems. However, the difference between non-complexed and complexed dinucleotide chains is of order of standard deviations. This suggest that rotation of H-bonded bases in pairs constituting CTT is restricted in some extend after TRF binding. Reduction of this parameter suggest better coplanarity of nucleobases belonging to opposite chains and is the highest in case of S3. Interestingly, kurtosis and skewness are rather small and almost identical for all distributions of all base pair parameters except of *stretch* and *opening*. In both cases skewness exceeds unity only for T₅ suggesting that there are significant portion of structures adopting higher values than average of this parameter. The higher values of *stretch* the longer distance between hydrogen bonded pairs, while higher values of *opening* indicate asymmetrical trends of hydrogen bond breathing. Very high values of kurtosis were noticed in case of Z₃ *opening* for S2 system and T₅ *stretching* for S3 and S4 oligomers. Thus, oxidized guanine in CTT is the source of high stiffness if compared to non-modified oligomer. On the other hand binding of TRF1 to both oxidized and non-oxidized chains reduces this kind of rigidity to the degree typical for free non-altered oligomer. Furthermore, TRF1 complex formation imposes strong stiffness of T₅ *stretching* in S3 and S4 systems. Thus, these few mentioned alterations imposed on base pair parameters are the main structural consequences of by oxidation or/and TFR1 binding in analyzed systems. In addition some interesting alterations are induced also on distributions of base pair parameters. In Fig. 1 there are presented three such spectacular instances. For example *shift* and *slide* are quite sensitive parameters affected by interactions of oxidized oligomer with analyzed protein. The median values of *shifting* of two base pairs along shorter molecular axis are reduced for Z₃/G₄ and increased for G₂/Z₃ pairs in S4 oligomer. This is accomplished by the significant reduction of skewness of *shift* populations of G₂/Z₃. This suggests that S4 system exhibits strong tendency toward reducing *shift* values to ones typical for S1 oligomer. Also increased

sliding of Z_3/G_4 and G_2/Z_3 pairs found for S4 system suggests significant movement along longer molecular axis of these two pairs imposed by TRF1. Furthermore, significant reduction of base pairs *twisting* is observed for S2 and S4 oligonucleotides indicating much smaller rotation of base pairs within $d(G_2/pZ_3)$ dinucleotide steps. Also, interactions with protein change also *roll* values characterizing S4 structure. *Rolling* of base step pairs describes mutual rotation of base pairs around shorter molecular axis. Negative *roll* values of Z_3/G_4 indicate increasing distance between stacked base pairs from major groove side. These changes are also accomplished with changes in *tilt* values describing Z_3/G_4 base pair of S4 structures suggesting tendency of increasing of distance between stacked bases on right side of major groove. Besides, there is observed significant *twisting* of Z_3/G_4 base step pair of S3 and S4 that is much larger if compared with S2 and S3 oligomers. Above statements confirms that binding of TRF1 to canonical or oxidized telomere repeat unit is important factor contributing to structure and dynamics of analyzed oligomers. The most important alterations imposed on geometry of CTT's that were discussed above are summarized in Fig. 2. Length of vectors and directions of angles schematically represents magnitude of differences.

<Figure 2 here>

Purines rings overlapping

In the central telomeric triad (CTT) three purine rings can overlap forming two pairs of *intra-strand* stacking. It is reasonable to expect that any changes imposed on structure and dynamics of CTT can lead to alterations of base-over-base displacement. This feature can be expressed qualitative by the analysis of fluctuations of overlapping area (in $\text{\AA}^2$) between polygons defined by atoms on successive bases within the same strand. Polygons are projected on the mean plane of the analyzed base-pair step. Such values are the measure of base ring overlaying excluding exocyclic atoms. Observed fluctuations of overlapping areas for two successive purines belonging to the first strand were plotted in Fig. 3 for all analyzed oligomers. Very interesting observation can be made from both presented plots and provided medians. First of all, native B-DNA telomere is very flexible structure and overlapping area of the first two guanine molecules in CTT changes in wide range (from 0 to $6 \text{ \AA}^2$). Supplied median values as well as histogram plots in Fig. 3 support quite flexible nature of GGG triad in native telomere repeat. The comparable values of medians and standard deviations of G_2/G_3 and G_3/G_4 suggest symmetrical movements of both G-C pairs. Replacement of G_3 with 8-oxoguanine has profound consequence on the area of overlapping. Momentous increase of

G₂/8oxoG₃ stacking contacts and slight decrease of 5'-Z₃/G₄-3' overlaying is observed in case of S2 system. Hence, symmetrical GGG triad become much more asymmetrical after oxidation of central guanine. Interestingly, similar consequences on overlapping areas are observed as a result of TRF1 binding to S1 structure. In the case of S3 oligomer the contacts of G₂/G₃ pair are significantly increased with unchanged G₃/G₄ overlapping. Oxidation of guanine in CTT in S4 complex even enhance this effect. Additionally reduction of overlapping areas in d(Z₃pG₄) dinucleotide steps is to be noticed. One can expect that the *intra*- strand interactions should be significantly affected by such changes of CTT structure. This aspect will be discussed in the next section. Thus, both oxidation and TRF1 binding can have significant influence on the overlapping of aromatic rings in central telomeric guanine triad.

<Figure 3 here>

Energetic consequence of oxidation and TRF1 binding

Structural alterations imposed by oxidation and complexation with TRF1 must affect intermolecular interactions between bases constituting central telomeric triad. Three distinct types of contacts determine IIE within central telomeric triad, namely hydrogen bonding, *intra*- and *inter*-strands stacking. They are systematically studied below for population comprising most representative structures extracted according to procures described above. As it was previously demonstrated [34-36] interactions in d(GpG) dinucleotide steps can exhibit significant non-additive character reaching several kcal mol⁻¹. Furthermore, many-body contributions to stabilization of dinucleotide steps comprising 8-oxoguanine are also non-negligible [56]. Fortunately, sum of all many body terms scale linearly with non-additive part and simple correction is possible. Regression analysis presented in Fig. 4 demonstrate that many-body contributions are not very high and in average are equal to -1.8 kcal mol⁻¹ for non-oxidized dinucleotide steps and -2.4 and -2.8 kcal mol⁻¹ for d(GpZ) and d(ZpG), respectively. Since many body contributions can change depending of nucleobases conformations, all additive IIE values were corrected according to regression formula presented in Fig. 4.

<Figure 4 here>

Base-base interactions in CTT

Results of IIE calculations were presented in Fig. 5 and Table 1. In case of purine/purine *intra*-strand stacking width of IIE distribution is about 2 kcal mol⁻¹ in terms of half of the most often occurring IIE values (differences between bottom and top 25% of values). In case of G₂/G₃ in S1 oligomers the range of observed *intra*-stacking starts from -5.6 and -0.7 kcal mol⁻¹. The *inter*-strand stacking of complementary cytosine molecules ranges between -6.1 and +1.3 kcal mol⁻¹. This give impression of the influence of structural fluctuations generated during MD run on the heterogeneities of actual values of intermolecular interactions. Due to non-normal distributions of intermolecular interaction energies instead of mean or median values the most frequently occurring energies (IIE^{MFO}) are used as the main energetic characteristics of intermolecular interactions. It simply corresponds to localization of maximum on smoothed histograms. Plots characterizing distributions of *intra*-strand stacking of two purines or two pyrimidines are exemplified in Fig. 5. From these data interesting trends can be concluded. First of all *intra*-strand stacking involving 8-oxoguanine in G₂/Z₃ pairs (of S2 and S4) is in average 3.5 kcal mol⁻¹ stronger than in G₂/G₃ pairs (of S1 and S3). This significant increase of stacking of oxidized guanine is the main consequence of observed structural differences between modified and non modified chains of telomere repeat unit. Interestingly, irrespectively of the presence or absence of TRF1 protein the *intra*-strand stacking of 8-oxoguanine with guanine is significant context-dependence since IIE^{MFO} of G₂/Z₃ and Z₃/G₄ sequences differ by more than values of standard deviations. Purines involved in the former pair interact much stronger than the latter. This rationale the observed discrepancies in overlapping areas. The *intra*-strand stacking of C₂/C₃ cytosine molecules are sensitive to TRF1 interactions with oxidized telomere repeat unit. Statistically significant decrease of IIE^{MFO} by about 2 kcal mol⁻¹ is noticed in case of C₂/C₃ stacking. For the rest of *intra*-strand stacking pairs of two cytosine molecules there is no observed significant influence of guanine oxidation and TRF1 binding. Besides, *inter*-strand stacking occurring in CTT is sequence dependent. For example 5'-Z₃|C₄-5' (in d(ZpG)) is about 30% weaker than 3'-Z₃|C₂-3' (in d(GpZ)) in all oligomers except S4. In this case both these *inter*-strand interactions are almost identical. Presence of the oxidized guanine also influences *inter*-strand stacking of guanine with cytosine making 3'-G₄|C₃-3' interactions about 1.4 kcal mol⁻¹ stronger than for 5'-G₂|C₃-5' in S2. No differences in strength of stronger hydrogen bonds are observed for all canonical and oxidize Z-C pairs. It is also interesting to know if any compensation effects take place of previously discussed changes in base-base interactions for

higher ordered structures. Thus, dinucleotide steps were considered. The corresponding IIE values were estimated assuming pair-wise additivity supplemented with linear correlation corrections for many-body contributions (see Fig. 4). The most straightforward conclusion drawn from data collected in Table 1 is that the overall stabilization energy of all oxidized dinucleotide steps are significantly higher compared to canonical ones. The gain in IIE^{MFO} interaction energies is about 9 kcal mol^{-1} and significantly exceeds values of standard deviations. These changes were proved to be statistically significant. Thus, for dinucleotide steps oxidation has profound stabilization effect for all pairs in CTT originating from increase of *inter*- and *intra*-strand stacking of 8-oxoguanine. On the other hand interaction with TRF1 does not change significantly stabilities of dinucleotide steps constituting central telomeric triad.

<Figure 5 here>

<Table 1 here>

Structure to energy relationships

In previous sections there were discussed significant structural and energetic alterations as a consequence of oxidation and/or TRF1 protein binding to telomeric repeat unit. It is interesting to see if there are any direct structure to energy relationships not only on the level of base-base interactions but on bigger scale. Those parameters, which distributions were non-changed by two analyzed factors were excluded from the analysis, while for those distributions, which were proven to be statistically significant two dimensional contour maps were prepared. In Fig. 6 there are presented scattering plots characterizing actual correlation between the stabilization energy in the whole tetramer that forms central telomeric triad $\text{d}(\text{GpZpG})$ and values of *buckle*, *twist*, *shift* and *slide*. These distributions were over imposed on the contour map representing percentage of particular structure with given IIE value. For this reason the whole spectrum of values was divided into ten equal intervals and two-dimensional histograms were estimated and presented as smoothed contour maps. From such graphs it is clearly visible almost perfect energetical separation of $\text{d}(\text{G}_2\text{pG}_3\text{pG}_4)$ and $\text{d}(\text{G}_2\text{pZ}_3\text{pG}_4)$ trinucleotides. Besides, structural heterogeneities can also be seen. For example *buckle* can adopt values from interval $(-36^\circ, +36^\circ)$. The range of most frequently occurring buckle values is quite similar for canonical and oxidized trinucleotides, but there are structures characterized occasionally by values outside of this region. It is also visible that

buckle distributions of S4 system are significantly shifted toward higher values with respect of other oligomers. This is in good accord with conclusions drawn from Fig. 1. *Twist* distributions presented in Fig. 6 posses different properties. The most probable structures found in the range between 25° and 35° come from non-oxidized systems (S1 and S3). On the contrary much higher flexibilities with respect of *twist* values are observed for S2 and S4. Also the fluctuations of *shift* and *slide* can be momentous for all analyzed structures and cover range of more than 4Å for both of these step parameters. Interestingly, the discrepancies between *shift* distributions of S3 and S4 are well documented in Fig. 6. Much higher stiffness of this parameter is to be related to S1 and S3 structures. Similar conclusion can be drawn from slide distributions. In this case however, significantly lower values are observed for S3. Finally there is spectacular consequence of TRF1 binding to oxidized telomere repeat unit. Significant population of CTT structures found in S4 adopts positive values of *slide* and negative values of *shift*. This is observed only in the case of S4. The rest of populations are characterized by negative values of *slide* (<-1.0Å) and smaller than 1Å absolute values of *shift*. This is most direct consequence of TRF1 binding to oxidized telomere repeat unit, which distinguishes univocally such complex from canonical ones.

<Figure 6 here>

<Figure 7 here>

Conclusions

The B-DNA of d(A₁pG₂pG₃pG₄pT₅pT₆) sequence constituting telomeric repeat unit is significantly affected by both oxidation and binding of TRF1 protein. Presented data clearly identified structural and energetic consequences imposed on central telomeric triad by these two factor. The occurrence of 8-oxoguanine in CTT is the source of high stiffness if compared to non-modified oligomer. The origin of this property comes from significantly alteration of intermolecular interactions of 8-oxoguanine. Presence of this lesion in CTT leads to important increase of *intra*-strand stacking involving 8-oxoguanine in G₂/Z₃ pair by about 3.5 kcal mol⁻¹ if compared to G₂/G₃ pair. Also oxidized guanine influenced *inter*-strand stacking of guanine with cytosine making 3'-G₄|C₃-3' interactions about 1.4 kcal mol⁻¹ stronger than for 5'-G₂|C₃-5' in oxidizes oligonucleotides. The sequence dependent *inter*- and *intra* stacking involving 8-oxoguanine are observed. In the case of 8-oxoguanine present in d(ZpG) sequence significant

reduces of base-base interactions if compared with d(GpZ) dinucleotide steps. The increased stability observed for base-base interactions has cumulative character and is typical also for di- and tri-nucleotides. Although for some conformations there are observed non-negligible contributions coming from many-body terms still dinucleotides and the whole CTT are significantly more stable with respect of their canonical equivalents. The observed changes in the intermolecular interactions originate from structural alterations imposed by TRF1 binding to canonical and oxidized telomere B-DNA. First and most direct consequence of TRF1 binding to oxidized telomere repeat unit is alteration of *shift-slide* correlations. Only in case of TRF1 complex formed with oxidized CTT significant population adopts positive values of *slide* and negative values of *shift*. This in turn leads to large differences in purine-purine overlapping in oxidized structures. Thus, oxidized structures of telomere repeat unit are more sensitivity to interactions with protein ligands especially for Z_3/G_4 stacked pairs.

Supplementary material available

References

1. Fairall L, Chapman L, Moss H, de Lange T, Rhodes D (2001) *Mol Cell* 8:351–361
2. Lundblad V (2000) *Mutat Res* 451:227–240
3. Chong L, van Steensel B, Broccoli D, Erdjument-Bromage H, Hanish J, Tempst P, de Lange T (1995) *Science* 270:1663–1667
4. Broccoli D, Smogorzewska A, Chong L, de Lange T (1997) *Nat Genet* 17:231–235
5. Harley CB, Futcher AB, Greider CW (1990) *Nature* 345:458–460
6. von Zglinicki T, Serra V, Lorenz M, Saretzki G, Lenzen-Grossimlighaus R, Gressner R, Risch Steinhagen-Thiessen E (2000) *Lab Invest* 80:1739–1747
7. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, Harley CB, Shay JW, Lichtsteiner S, Wright WE (1998) *Science* 279:349–352
8. Court R, Chapman L, Fairall L, Rhodes D (2005) *EMBO Rep* 6:39–45
9. Hanaoka S, Nagadoi A, Nishimura Y (2005) *Protein Sci* 14:119–130
10. Tchirkov A, Lansdorp PM (2003) *Hum Mol Genet* 12:227–232
- 1.1 von Zglinicki T (2002) *Trends Biochem Sci* 27:339–344
12. Opresko PL, Fan J, Danzy1 S, David M, Wilson DM III, Bohr VA (2005) *Nucleic Acids Res* 33:1230–1239
13. Cysewski P, Czeleń P (2009) *J Mol Model* 15: 607–613
14. Cysewski P, Czeleń P (2007) *J Mol Model* 13:739–750
15. Naômé A, Schyman P, Laaksonen A and Vercauteren DP (2010) *J Phys Chem B* 114:4789–4801
16. Fujimoto H, Pinak M, Nemoto T O’neill P, Kume E, Saito K and Maekawa H, (2005) *J Comput Chem* 26:788–798
17. Dodson ML and Lloyd RS (2001) *Mutation Res* 487:93–108
18. Cheng X, Kelso C, Hornak V, de los Santos C., Grollman AP and Simmerling C (2005) *J Am Chem Soc* 127:13906-13918
19. Ishida H (2002) *J Biomol Struct Dyn* 19:839-851
20. Fujimoto H, Pinak M, Nemoto T, Sakamoto K, Yamada K, Hoshi Y, Kume E (2004) *J Mol Struct THEOCHEM* 681:1–8
21. Bruner SD, Norman DP, Verdine GL (2000) *Nature* 403:859-866
22. Hanaoka S, Nagadoi A, Nishimura Y (2005) *Protein Sci* 14:119-130
23. Pinak M (2003) *Comput Chem* 24:898–907
24. Pinak M (2003) *Comput Biol Chem* 27:431-441

- 2.5 Pinak M (2002) *J Mol Struct THEOCHEM* 583:189-197
26. Miller JH, Fan-Chiang CCP, Straatsma TP, Kennedy MA (2003) *J Am Chem Soc* 125:6331–6336
27. Zhang RB, Xu H, Qu ZW, Zhang XK, Ai CX, Zhang QY (2002) *J Mol Struct. THEOCHEM* 586:167-175
28. Tchirkov A, Lansdorp PM (2003) *Hum Mol Genet* 12:227–232
29. von Zglinicki T (2002) *Trends Biochem Sci* 27:339–344
30. Hanaoka S, Nagadoi A, Nishimura Y (2005) *Protein Sci* 14:119–130
31. Jansen HB, Ross P (1969) *Chem Phys Lett* 3:140–143
32. Boys SF, Bernardi F (1970) *Mol Phys* 19:553–566
33. Hobza P, Šponer J (1999) *Chem Rev* 99:3247-3276
34. Černý J, Hobza P (2005) *Phys Chem Chem Phys* 7:1624-1626
35. Hobza P, Sponer J, (2002) *J Am Chem Soc* 124:11802-11808
36. Cysewski P (2010) *Int J Quantum Chem.* doi:10.1002/qua.22435
37. Cysewski P (2008) *J Mol Struct THEOCHEM* 865:36–43
38. Cysewski P (2009) *J Mol Model* 15:597–606
39. Cysewski P (2009) *New J Chem* 33:1909-1917
40. van Gunsteren WF, Berendsen HJC (1999) *Angew Chem Int Ed Engl* 29:992-1023
41. Hansson T, Oostenbrink C, van Gunsteren WF (2002) *Curr Opin Struct Biol* 12:190-196
42. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham III TE (2000) *Acc Chem Res* 33:889-897
43. Karplus M, McCammon J (2002) *Acad Nat Struct Biol* 9:646-652
44. Cheatham TE III, Kollman PA (2000) *Annu Rev Phys Chem* 51:435-471
45. Shao J, Tanner SW, Thompson N, Cheatham TE III (2007) *J Chem Theory Comput* 3:2312-2334
46. Lyman E, Zuckerman DM (2006) *Biophys J* 91:164-172
47. Feher M, Schmidt JM (2003) *J Chem Inf Comput Sci* 43:810-818
48. Gabarro-Arpa J, Revilla R (2000) *Comput Chem* 24:696-698
49. Diekmann S (1989) *J Mol Biol* 205:787–791
50. Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B (1992) *Biophys J* 63:751–759
51. Lu XJ, Olson WK (2003) *Nucleic Acids Res* 31:5108–5121

52. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A Jr, Simmerling C, Wang B, Woods R (2005) *J Comput Chem* 26:1668–1688
53. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2004) *AMBER 10*. University of California, San Francisco
54. Werner HJ, Knowles PJ, Lindh R, Manby FR, Schutz M, Celani P, Korona T, Rauhut G, Amos RD, Bernhardsson A, Berning A, Cooper DL, Deegan MJO, Dobbyn AJ, Eckert F, Hampel C, Hetzer G, Lloyd AW, McNicholas SJ, Meyer W, Mura ME, Nicklaß A, Palmieri P, Pitzer R, Schumann U, Stoll H, Stone AJ, Tarroni R, Thorsteinsson T (2006) *MOLPRO*, a package of ab initio programs designed by Werner HJ and Knowles PJ, Version 2006.0 (Patch 2006.1). Cardiff UK
55. Nishikawa T, Okamura H, Nagadoi A, König P, Rhodes D, Nishimura Y (2001) *Structure* 9:1237-1251
56. Cysewski P (2010) *J Mol Model* (JMMO1145 - accepted for publication in this issue)
57. Azbel' MY (1995) *Phys Rev Lett* 75:168-171
58. Ponomarev SY, Thayer KM, Beveridge DL (2004) *PNAS* 101:14771–14775
59. Pérez A, Marchán I, Svozil D, Šponer J, Cheatham TE III, Laughton CA, Orozco M (2007) *Biophys J* 92:3817–3829

Tables

Table 1 Values of the most frequently occurring intermolecular interactions IIE^{MFO} [kcal mol⁻¹] in central d(G₂pZ₃pG₄) fragment of telomere repeat unit, where Z stands for G or 8oxoG. The values of standard deviation were estimated for populations comprising 100 snapshots. Symbol S1 represent canonical double strand, S2 designates oxidized one, S3 indicates canonical strand with bounded TRF1 and S4 denotes oxidized B-DNA in complex with TRF1

		S1	S2	S3	S4
d(GpZ)	G ₂ /Z ₃	-3.1±1.0	-7.4±1.1	-4.2±1.3	-7.0±1.3
	C ₂ /C ₃	-2.1±1.2	-1.8±1.3	-1.1±0.8	-0.9±1.2
	5'-G ₂ C ₃ -5'	-2.2±1.2	-2.0±1.3	-2.3±1.4	-2.3±1.0
	3'-Z ₃ C ₂ -3'	-3.0±1.6	-3.1±0.9	-2.8±0.7	-3.3±2.4
	5'-G ₂ -C ₂ -3'	-21.3±2.2	-20.4±2.2	-20.6±2.5	-21.5±2.4
	3'-Z ₃ -C ₃ -3'	-21.3±1.6	-20.7±2.1	-21.6±2.7	-21.2±3.2
	dinucleotide *	-50.5±2.8	-59.5±3.7	-53.0±3.5	-58.9±3.5
d(ZpG)	Z ₃ /G ₄	-3.8±0.6	-4.6±0.9	-3.6±0.9	-4.6±1.0
	C ₃ /C ₄	-2.6±1.0	-2.1±1.1	-1.3±0.9	-1.9±1.0
	5'-Z ₃ C ₄ -5'	-2.0±1.0	-2.2±0.8	-2.5±1.2	-2.3±0.9
	3'-G ₄ C ₃ -3'	-2.9±1.4	-2.3±2.0	-2.1±1.2	-2.1±1.0
	3'-G ₄ -C ₄ -3'	-21.0±2.1	-21.6±2.2	-21.0±2.5	-20.0±2.6
	dinucleotide	-52.7±2.8	-58.1±4.3	-50.7±3.1	-56.0±3.5

* corrected according to linear regressions presented in Fig. 4

Figure captions

- Fig. 1** Mean values of selected parameters defining conformations of nucleobases in telomere repeat unit
- Fig. 2** Schematic representation of mean values and corresponding standard deviations of parameters that are strongly affected by oxidation and/or complexation by TRT1 protein. In upper view normal style was used for representing values of *shift* (Å), in italics there are provided *roll* (°) values and numbers in gray represent *twist* (°) values. In the lower view *slide* (Å) and *tilt* (°) are presented in normal and bold style, respectively
- Fig. 3** The distributions of the overlapping area (in Å²) between polygons defined by base ring atoms on successive bases excluding areas of exocyclic atoms. Median values along with standard deviations are provided for both stacked pairs in central telomeric triad of analyzed oligomers. Symbol Z represent guanine (for S1 and S3) or 8-oxoguanine (in case of S2 and S4). Points correspond to every 10ps snapshots. Histogram plots for all distributions were supplied in edges of corresponding axis
- Fig. 4** Regression analysis between intermolecular interactions in dinucleotide steps estimated using pair-wise additivity (IIE(additive)) and full inclusion of all many body contributions (IIE(total)) [56]
- Fig. 5** Distributions of *intra*-strand stacking interactions in CTT of analyzed oligomers. In legends there are provided probabilities that the difference in the median values between the two groups (S3 with respect of S1 and S4 against S2) is statistically significant. In case of $P < 0.001$ it is excluded the possibility that the difference is merely due to random sampling variability
- Fig. 6** Distributions of the sum of intermolecular interactions energies in d(GpZpG) central telomeric triad as a function of values of selected parameters for all analyzed system. The color spectrum representing the total percentage of populations obtained from two-dimensional histograms has the same scale for all contour maps

Fig. 7 Correlations between *shift* and *slide* values characterizing Z_3/G_4 step for all analyzed oligonucleotide systems. Notation is the same as in Fig. 6

Figure 1
[Click here to download line figure: Fig1.doc](#)

Figure 1. Mean values of selected parameters defining conformations of nucleobases in telomere repeat unit.

Figure 2
Click here to download line figure: Fig2.doc

Figure 2. Schematic representation of mean values and corresponding standard deviations of parameters that are strongly affected by oxidation and/or complexation by TRT1 protein. In upper view normal style was used for representing values of *shift* (Å), in italics there are provided *roll* (°) values and numbers in gray represent *twist* (°) values. In the lower view *slide* (Å) and *tilt* (°) are presented in normal and bold style, respectively.

Figure 3
[Click here to download line figure: Fig3.doc](#)

Fig. 3. The distributions of the overlapping area (in $\text{\AA}^2$) between polygons defined by base ring atoms on successive bases excluding areas of exocyclic atoms. Median values along with standard deviations are provided for both stacked pairs in central telomeric triad of analyzed oligomers. Symbol Z represent guanine (for S1 and S3) or 8-oxoguanine (in case of S2 and S4). Points correspond to every 10ps snapshots. Histogram plots for all distributions were supplied in edges of corresponding axis.

Figure 4
[Click here to download line figure: Fig4.doc](#)

Figure 4. Regression analysis between intermolecular interactions in dinucleotide steps estimated using pair-wise additivity ($IIE(\text{additive})$) and full inclusion of all many body contributions ($IIE(\text{total})$) [56].

Figure 5
[Click here to download line figure: Fig5.doc](#)

Figure 5. Distributions of intra-strand stacking interactions in CTT of analyzed oligomers. In legends there are provided probabilities that the difference in the median values between the two groups (S3 with respect of S1 and S4 against S2) is statistically significant. In case of $P < 0.001$ it is excluded the possibility that the difference is merely due to random sampling variability.

Figure 6
[Click here to download colour figure: Fig6.doc](#)

Figure 6. Distributions of the sum of intermolecular interactions energies in d(GpZpG) central telomeric triad as a function of values of selected parameters for all analyzed system. The color spectrum representing the total percentage of populations obtained from two-dimensional histograms has the same scale for all contour maps.

Figure 7
[Click here to download colour figure: Fig7.doc](#)

Figure 7. Correlations between *shift* and *slide* values characterizing Z_3/G_4 step for all analyzed oligonucleotide systems. Notation is the same as in Fig.6.

Supporting materials

Figure S1. Energetic characteristics of probing distribution sets estimated for each of dinucleotide steps belonging to central telomeric triad of all four analyzed oligonucleotides. Algorithm was presented below in the footnote*. Symbol S1 stands for free non-modified B-DNA telomeric repeat unit d(A₁G₂G₃G₃T₅T₆) taken from central part of modelled 14-mer in water solution, S2 represents oxidized analogue d(A₁G₂Z₃G₄T₅T₆), S3 denotes non-oxidized telomere oligomer in complex with TRF1 and S4 refers to oxidized telomeric fragment bound to TRF1. Symbol Z stands for guanine (in case of S1 and S3) or 8-oxoguanine (for S2 and S4). Gray lines correspond to different probing sets of 36 structures, red lines represented averaged distributions (over 360 structures) and blue lines denotes selected the most representative set of structures used for further intermolecular interaction energies (IIE) estimation on *ab initio* level. All IIE values were estimated using Amber forcefield.

d(G₂pG₃) of S1

d(G₃pG₄) of S1

d(G₂pZ₃) of S2

d(Z₃pG₄) of S2

d(G₂pG₃) of S3

d(G₃pG₄) of S3

d(G2pZ3) of S4

d(Z3pG4) of S4

***Algorithm for generation o the most representative structures (MRS):**

1. the distributions of base step parameters (*shift*, *slide*, *rise*, *roll*, *tilt* and *twist*) were extracted from MD trajectories based on X3DNA program [53];
2. after normalisation of all values corresponding histograms were prepared for each base step parameter (after division of the whole span of values (from 0 to 1) into six equal intervals);
3. obtained population percentages were used as weighting coefficient for selection of six conformations for each parameter (thus in total 36 structures for each dinucleotide step);
4. the probing distribution sets were generated by random selection of structures in number corresponding to weighting coefficients;
5. ten probing distribution sets were obtained and IIE values were estimated for each conformation based on Amber forcefield [55];
6. obtained IIE distributions were presented in the form of histograms and most representative structures were defined as those probing distribution sets for which shapes of IIE histograms was closed to the averaged one (according to calculated RMSD values).

Figure S2. Median values of base pair parameters distributions estimated for trajectories collected during 20x2ns MD runs. Chains notation is the same as in Fig. S1. Solid lines with black symbols characterize CTT (central telomeric triad), while gray lines denote the rest of pairs within telomere repeat unit. For clarity, the magnitudes of standard derivations were indicated only for pairs involving Z_3 purine.

Figure S3. Median values of base step parameters distributions estimated for trajectories collected during 20x2ns MD runs. Notation is the same as in Fig.S1.

Table S1. Statistical characteristics of base pair parameters distributions for all pairs constituting telomere repeat unit in analyzed oligonucleotide chains. For notation refer to Fig.S1

		S1			S2			S3			S4		
		M±SD	K	S	M±SD	K	S	M±SD	K	S	M±SD	K	S
A1	Buckle	4.06 ± 11.25	0.04	-0.08	4.81 ± 11.10	0.09	-0.14	-1.57 ± 11.74	-0.05	-0.09	-3.72 ± 12.41	0.09	-0.10
G2		11.82 ± 9.59	0.06	-0.11	10.80 ± 9.08	0.10	-0.08	6.37 ± 9.51	0.12	-0.09	4.56 ± 9.50	0.02	-0.08
Z3		3.88 ± 10.27	0.00	-0.02	7.56 ± 10.28	-0.03	0.03	5.40 ± 10.79	0.23	-0.29	5.67 ± 10.96	0.02	-0.10
G4		-11.23 ± 9.88	0.04	0.20	-7.24 ± 10.83	-0.14	0.12	-1.40 ± 10.85	-0.15	0.09	-0.11 ± 11.16	0.01	-0.09
T5		-13.00 ± 9.86	-0.21	0.05	-10.95 ± 9.97	-0.04	0.06	-5.26 ± 10.01	0.07	0.06	-6.28 ± 9.43	0.13	0.08
T6		-8.25 ± 10.64	0.21	-0.03	-8.33 ± 10.82	0.23	0.08	0.54 ± 12.39	0.12	-0.04	-0.91 ± 11.52	0.11	0.08
A1	Opening	2.15 ± 6.14	2.28	0.52	2.62 ± 6.38	4.28	0.95	0.87 ± 6.14	1.09	0.15	2.97 ± 6.62	1.90	0.49
G2		1.97 ± 3.29	1.02	0.25	2.25 ± 3.33	0.89	0.30	1.19 ± 3.58	1.87	0.09	1.87 ± 3.68	0.86	0.26
Z3		2.28 ± 3.32	3.04	0.60	1.37 ± 3.43	5.67	0.78	0.96 ± 3.39	0.39	0.10	0.20 ± 3.40	2.01	-0.07
G4		2.24 ± 5.64	7.96	-2.07	2.73 ± 3.51	6.31	0.85	1.30 ± 3.33	0.44	0.09	1.50 ± 3.58	0.76	0.12
T5		0.51 ± 6.38	2.20	-0.79	1.02 ± 5.15	0.51	0.15	1.91 ± 5.15	0.47	0.05	1.77 ± 5.48	0.63	0.12
T6		1.71 ± 6.30	1.67	-0.11	2.33 ± 5.86	3.95	0.80	3.13 ± 5.14	1.34	0.31	3.37 ± 5.12	1.31	0.51
A1	Propeller	-5.95 ± 8.13	0.30	-0.08	-5.89 ± 8.49	0.60	-0.06	-9.81 ± 8.50	0.00	-0.10	-10.34 ± 8.73	0.11	-0.12
G2		-6.54 ± 7.81	0.13	0.06	-8.15 ± 7.51	0.08	0.02	-9.17 ± 8.29	-0.03	0.00	-11.38 ± 8.35	0.00	0.05
Z3		-11.76 ± 7.88	0.27	0.10	-11.63 ± 8.25	0.15	0.17	-7.82 ± 9.50	0.45	0.45	-7.32 ± 9.72	-0.01	0.13
G4		-15.91 ± 7.77	0.26	0.19	-15.42 ± 7.74	0.19	0.15	-11.80 ± 8.70	-0.05	0.21	-14.76 ± 7.31	0.06	0.00
T5		-14.28 ± 7.59	0.23	0.10	-15.37 ± 7.66	0.18	0.13	-13.29 ± 8.32	-0.08	-0.01	-15.20 ± 7.90	0.05	0.03
T6		-8.02 ± 8.57	0.19	0.08	-10.07 ± 8.47	0.10	-0.01	-7.37 ± 10.65	-0.13	0.09	-6.86 ± 9.20	0.17	-0.01
A1	Shear	-0.05 ± 0.29	0.04	0.08	-0.06 ± 0.29	0.08	0.05	-0.06 ± 0.30	0.03	0.13	-0.08 ± 0.30	0.05	0.05
A1		-0.19 ± 0.32	0.22	-0.03	-0.18 ± 0.32	0.19	-0.07	-0.16 ± 0.31	-0.01	0.01	-0.15 ± 0.32	0.10	-0.02
Z3		-0.29 ± 0.32	0.66	-0.07	-0.26 ± 0.33	0.95	-0.04	-0.28 ± 0.34	0.10	0.13	-0.27 ± 0.34	0.20	0.07
G4		-0.28 ± 0.32	0.42	-0.09	-0.24 ± 0.32	1.53	-0.22	-0.17 ± 0.33	0.90	0.18	-0.13 ± 0.32	1.16	0.28
T5		0.01 ± 0.27	0.06	-0.06	0.01 ± 0.27	-0.02	-0.06	-0.05 ± 0.28	0.20	-0.08	-0.06 ± 0.28	0.11	-0.05
T6		0.05 ± 0.28	0.01	-0.05	0.05 ± 0.28	0.01	-0.04	-0.07 ± 0.27	0.07	-0.06	-0.07 ± 0.27	-0.01	-0.06
A1	Stagger	0.08 ± 0.42	0.15	0.07	0.12 ± 0.42	0.09	0.01	0.09 ± 0.43	0.12	0.09	0.18 ± 0.45	0.01	0.03
G2		0.27 ± 0.40	0.08	0.00	0.28 ± 0.38	0.19	0.07	0.21 ± 0.38	0.09	0.00	0.18 ± 0.37	0.08	0.05
Z3		0.02 ± 0.41	0.23	0.18	0.09 ± 0.41	0.38	0.23	0.08 ± 0.45	0.22	0.11	0.00 ± 0.45	0.29	0.16
G4		-0.08 ± 0.37	0.36	0.10	-0.04 ± 0.38	0.22	0.12	0.12 ± 0.44	0.29	0.03	0.16 ± 0.43	0.49	0.27
T5		0.11 ± 0.39	0.21	0.08	0.09 ± 0.40	0.07	0.05	0.15 ± 0.44	0.04	0.01	0.12 ± 0.43	0.11	0.09
T6		-0.06 ± 0.40	0.18	-0.01	-0.10 ± 0.42	0.17	0.05	-0.15 ± 0.44	0.10	-0.08	-0.17 ± 0.43	0.11	0.04
A1	Stretch	0.05 ± 0.13	6.52	1.25	0.05 ± 0.14	6.75	1.29	0.02 ± 0.13	2.86	0.75	0.03 ± 0.14	4.21	0.98
G2		0.00 ± 0.12	0.97	0.37	0.00 ± 0.12	1.16	0.46	-0.03 ± 0.12	1.32	0.45	-0.01 ± 0.13	1.37	0.57
Z3		-0.05 ± 0.12	2.17	0.49	-0.05 ± 0.12	3.13	0.56	-0.06 ± 0.12	0.65	0.17	-0.08 ± 0.12	0.77	0.06
G4		-0.03 ± 0.16	6.35	-1.05	-0.02 ± 0.12	4.49	0.79	-0.04 ± 0.12	0.78	0.24	-0.03 ± 0.12	1.26	0.38
T5		0.06 ± 0.14	2.13	0.12	0.05 ± 0.12	1.07	0.49	0.04 ± 0.12	1.72	0.55	0.04 ± 0.13	2.48	0.76
T6		0.05 ± 0.14	3.13	0.31	0.05 ± 0.14	14.47	1.61	0.04 ± 0.13	3.82	0.79	0.04 ± 0.12	1.95	0.59

M – median value, SD – standard deviation, K – kurtosis, S – skewness

Table S2. Statistical characteristics of base step parameters distributions for all pairs constituting telomere repeat unit in analyzed oligonucleotide chains. For notation refer to Fig.S1

		S1			S2			S3			S4		
		M±SD	K	S	M±SD	K	S	M±SD	K	S	M±SD	K	S
A ₁ /G ₂	Rise	3.15 ± 0.30	0.53	0.33	3.17 ± 0.28	0.42	0.28	3.16 ± 0.30	0.31	0.31	3.15 ± 0.29	0.35	0.25
G ₂ /Z ₃		3.48 ± 0.31	0.20	0.24	3.39 ± 0.32	0.15	0.22	3.39 ± 0.32	0.38	0.36	3.36 ± 0.32	0.30	0.32
Z ₃ /G ₄		3.64 ± 0.32	0.10	0.12	3.59 ± 0.31	0.20	0.22	3.52 ± 0.36	0.13	0.21	3.50 ± 0.36	0.60	0.39
G ₄ /T ₅		3.38 ± 0.27	0.71	0.29	3.41 ± 0.28	0.36	0.26	3.47 ± 0.28	0.25	0.31	3.48 ± 0.29	0.80	0.36
T ₅ /T ₆		3.20 ± 0.27	0.42	0.34	3.23 ± 0.28	5.42	0.79	3.20 ± 0.29	2.41	0.48	3.20 ± 0.27	0.44	0.33
A ₁ /G ₂	Roll	2.10 ± 5.57	1.23	0.30	2.31 ± 5.60	0.40	0.12	2.04 ± 5.00	0.13	0.04	2.70 ± 5.34	0.23	0.00
G ₂ /Z ₃		4.37 ± 5.21	0.47	-0.12	4.79 ± 5.36	0.43	-0.11	0.83 ± 5.94	0.23	-0.16	0.64 ± 5.93	0.16	-0.08
Z ₃ /G ₄		6.10 ± 4.88	0.28	-0.07	5.57 ± 5.14	0.27	-0.07	0.78 ± 5.43	0.10	-0.13	0.22 ± 5.41	0.05	-0.07
G ₄ /T ₅		0.10 ± 4.49	0.29	-0.07	0.05 ± 4.52	0.32	-0.09	-3.11 ± 4.39	0.21	-0.21	-2.72 ± 4.61	0.47	-0.28
T ₅ /T ₆		0.21 ± 5.54	0.28	0.03	0.62 ± 5.60	0.48	0.08	-2.16 ± 5.59	0.80	-0.01	-1.58 ± 5.77	0.19	-0.21
A ₁ /G ₂	Shift	0.01 ± 0.84	-0.46	-0.16	-0.66 ± 0.72	0.05	0.30	-0.02 ± 0.73	-0.39	-0.01	-0.54 ± 0.72	-0.18	0.31
G ₂ /Z ₃		0.13 ± 0.68	-0.17	-0.05	0.70 ± 0.55	0.18	0.10	0.58 ± 0.73	-0.54	-0.15	1.11 ± 0.55	0.78	-0.41
Z ₃ /G ₄		-0.02 ± 0.58	0.32	-0.30	-0.30 ± 0.72	-0.26	0.06	-0.11 ± 0.69	0.34	-0.02	-0.46 ± 0.60	-0.21	0.01
G ₄ /T ₅		-0.46 ± 0.55	0.17	0.24	-0.35 ± 0.57	0.06	0.27	-0.33 ± 0.47	0.10	0.10	-0.26 ± 0.50	0.63	0.39
T ₅ /T ₆		0.56 ± 0.70	-0.26	-0.24	0.47 ± 0.69	-0.38	-0.13	0.94 ± 0.68	-0.14	-0.18	1.04 ± 0.62	1.80	-0.73
A ₁ /G ₂	Slide	-0.43 ± 0.69	-0.44	-0.18	-0.09 ± 0.61	0.48	-0.56	-0.14 ± 0.60	-0.27	-0.18	-0.01 ± 0.54	0.08	-0.13
G ₂ /Z ₃		-1.10 ± 0.72	-0.55	0.29	-1.05 ± 0.69	-0.40	0.30	-0.25 ± 0.72	-0.29	-0.21	-0.14 ± 0.69	-0.38	-0.27
Z ₃ /G ₄		-1.45 ± 0.59	1.24	0.92	-0.97 ± 0.74	-0.70	0.16	-1.07 ± 0.77	-0.11	0.59	-0.26 ± 0.86	-0.70	-0.26
G ₄ /T ₅		-0.87 ± 0.45	0.37	-0.29	-0.87 ± 0.47	0.31	-0.23	-0.91 ± 0.53	0.71	-0.58	-0.93 ± 0.47	0.85	0.14
T ₅ /T ₆		-0.07 ± 0.54	0.12	-0.44	-0.14 ± 0.54	0.08	-0.39	0.34 ± 0.59	0.38	-0.50	0.44 ± 0.54	1.97	-0.89
A ₁ /G ₂	Tilt	-2.63 ± 4.30	0.21	0.06	-3.12 ± 4.21	0.21	0.02	-0.97 ± 4.28	0.03	0.02	-0.40 ± 4.57	0.02	0.08
G ₂ /Z ₃		0.83 ± 4.44	0.07	0.07	1.63 ± 4.32	0.09	0.01	1.05 ± 4.56	0.00	0.03	3.44 ± 4.79	0.15	0.13
Z ₃ /G ₄		0.72 ± 4.39	0.10	-0.03	-0.67 ± 4.56	0.00	0.02	-1.04 ± 4.85	0.12	-0.08	-3.00 ± 5.05	0.10	0.09
G ₄ /T ₅		-0.16 ± 3.86	0.16	-0.03	0.18 ± 3.87	0.11	-0.05	0.90 ± 3.83	0.09	0.00	1.00 ± 4.07	0.15	0.00
T ₅ /T ₆		2.60 ± 3.99	0.19	0.00	2.57 ± 4.01	0.26	-0.04	3.32 ± 4.48	0.47	0.12	3.34 ± 4.27	0.58	-0.01
A ₁ /G ₂	Twist	30.66 ± 7.61	-0.11	-0.53	34.85 ± 5.17	0.00	-0.49	30.42 ± 5.71	-0.16	-0.12	32.73 ± 5.57	-0.27	-0.21
G ₂ /Z ₃		32.50 ± 5.51	-0.21	-0.02	27.36 ± 4.66	0.55	-0.30	35.25 ± 5.37	-0.14	0.02	30.77 ± 5.17	-0.10	-0.21
Z ₃ /G ₄		32.41 ± 4.69	0.00	0.27	36.30 ± 5.49	-0.34	-0.07	33.25 ± 5.75	0.48	-0.15	37.84 ± 5.77	-0.31	-0.05
G ₄ /T ₅		34.57 ± 4.03	0.49	-0.15	34.92 ± 3.98	0.23	-0.06	34.51 ± 4.45	-0.13	0.17	34.07 ± 3.67	0.44	0.19
T ₅ /T ₆		36.15 ± 5.20	0.45	-0.67	35.95 ± 5.51	0.13	-0.55	37.12 ± 4.20	0.81	-0.18	37.90 ± 4.06	0.84	-0.32