Thermodynamic characteristization of Di-, Tri-, and tetranucleotide loci in parthenogenetic lizards Darevskia unisexualis

Article *in* Russian Journal of Genetics · August 2009

DOI: 10.1134/S102279540908016X

CITATIONS
3

READS
9

2 authors:

Andrey Omelchenko
Severtsov Institute of Ecology and Evolution
15 PUBLICATIONS **54** CITATIONS

SEE PROFILE

Vitaly I Korchagin
Russian Academy of Sciences
43 PUBLICATIONS **168** CITATIONS

SEE PROFILE

ISSN 1022-7954, Russian Journal of Genetics, 2009, Vol. 45, No. 8, pp. 1004–1008. © Pleiades Publishing, Inc., 2009.
Original Russian Text © A.V. Omelchenko, V.I. Korchagin, 2009, published in Genetika, 2009, Vol. 45, No. 8, pp. 1143–1147.

SHORT COMMUNICATIONS

Thermodynamic Characteristization of Di-, Tri-, and Tetranucleotide Loci in Parthenogenetic Lizards *Darevskia unisexualis*

A. V. Omelchenko[a, b] and V. I. Korchagin[a]

[a]*Institute of Gene Biology, Russian Academy of Sciences, Moscow, 119334 Russia*
e-mail: dna@biogen.msk.su
[b]*Filippovich Teaching and Scientific Center for Biochemistry and Molecular Biology, Moscow Pedagogical State University, Moscow, 119882 Russia*
e-mail: biochem_mpgu@mail.ru

Received October 10, 2008

Abstract—Allelic polymorphism of three microsatellite loci from the genome of parthenogenetic lizard *Darevskia unisexualis* was characterized using analysis of free energy (Gibbs energy) of the DNA/DNA duplex formation within the stepwise mutational model. It was demonstrated that the number of microsatellite cluster monomericic units would change to decrease the mean free energy of the locus. In addition, based on the analysis of nucleotide composition, the GC content of each locus was evaluated, and belonging of the loci examined to certain isochore families was suggested.

DOI: 10.1134/S102279540908016X

Genome instability and microsatellite loci variation remain poorly studied in the organisms with clonal type of inheritance. It should be noted in this context that microsatellite loci are poorly studied with regard to elucidation of quantitative regularities at the level of primary and secondary DNA structure. In the present study, using numerical methods of applied informatics, we analyzed the nucleotide composition of individual microsatellite loci, containing microsatellite clusters differing in the repeat unit composition and the structure of the cluster. Furthermore, for allelic variants of these loci, calculation of such thermodynamic characteristic as free energy (Gibbs energy) was evaluated at the moment of the secondary structure (DNA/DNA duplexes) formation. Molecular genetic interpretation of microsatellite expansion dynamics at certain genome regions of clonally reproduced (parthenogenetic) lizards *Darevskia unisexualis* was performed.

Microsatellite sequences belonging to tandemly organized moderately repeated DNA of eukaryotic genomes attracts considerable attention as factors of genome instability. Microsatellite sequences vary in the cluster length and the size of the repetitive unit. They are characterized by high mutation rates ranging from 10^{-2} to 10^{-5} depending on the type of microsatellite [1]. One of the main mechanisms leading to the appearance, and providing the development of microsatellite, is the replication fork slippage. This process is based on the ability of microsatellite-containing sequences to form conformational DNA structures, like single-strand DNA hairpins, loops, i.e., double-strand DNA slippage structures (cause DNA polymerase lagging), and the so-called toroids (special double-stranded DNA structures with 81 bp per turn. The latter structures are actually the DNA/DNA duplexes and/or are stabilized at the expense of the formation of given duplexes in the loop-flanking DNA regions [2]. Calculation of thermodynamic characteristics (changes of entropy (ΔS), enthalpy (ΔH), Gibbs energy (ΔG), of DNA/DNA duplex formation is widely used for different nucleotide sequences [3]. However, there is a significant lack of data for microsatellites.

Investigation of the microsatellite loci nucleotide composition is also important because there is a distinct association between the primary structure of DNA molecule, structural genome organization, and functional compartmentalization of interphase nuclei [4]. This association most clearly manifests in thermodynamic classification of the genome regions for the so-called isochore families, i.e., discrete, extended, highly homogenous relative to the G + C content DNA regions. The isochors can be classified into a number of "light" and "heavy" (depending on the G + C content) families. These include the isochors of L1 and L2 families (< 40%), H1 family (~47%), H2 family (~52%), and H3 family (>52%) [5].

In this study, analysis of thermodynamic characteristics was performed using three genomic loci of parthenogenetic species *Darevskia unisexualis*, containing microsatellites of different types. Molecular genetic analysis of two of these loci, *Du214* (GenBank accession number EU252542) and *Du281* (accession number

AY442143), was carried out earlier [6, 7]. The *Du183* locus (accession number EU252539) was cloned and sequenced in the present study. The *Du183* polymorphism was typed with the help of PCR followed by electrophoresis of amplification products in native and denaturing gels (6% PAAG (40% PAA (acrylamide : bisacrylamide, 19 : 1) 50% urea, 37% formamide) for 4 h at 60°C using the lizard population sample ($N = 65$), as described earlier [7]. Sequencing of amplification products was performed according to the method of Sanger with the ABI PRISM®BigDye™ Terminator v.3.1 reagent kit and subsequent analysis of the reaction products on the automated sequencer DNA ABI PRISM 3100-Avant. Nucleotide sequences were aligned using the MegAlign 4.05 software program. Quantitative data on nucleotide composition and thermodynamic characteristics were obtained using self-elaborated OMIKO-MCS v1.1 software program. The program is based on the algorithm suggested in [8] and implying summation of free energy values for all pairs of nucleotides along the DNA chain, while free energy values for nucleotide pairs are taken from [9]. In this case, the free energy of DNA/DNA duplex formation is calculated considering the energy of hydrogen bounds formation upon Watson–Crick interaction of complementary nucleotide pairs, as well as considering the energy of stacking interaction of neighboring nucleotides. Since the algorithm is used for nucleotides, possible nucleotide pairs, discussed in [8] are generated in random order. In our program, free energy is calculated for each extended DNA linear chain. Because of this, we suggested the algorithm of sequential generation of all possible nucleotide pairs. It was also suggested that the mean free energy of DNA sequence should be calculated as the ratio between the summarized free energy of all possible nucleotide pairs and the total number of nucleotides. It is known that there is a standard deviation (error) between the calculated free energy values and those observed in biophysical experiments. Evaluation of this error showed that in many cases deviation from the regression line (on the comparative plot of predicted and observed ΔG) was not greater than 0.12 kcal/mol. In case of the analysis of different oligonucleotide DNAs, standard deviation between the predicted and experimental ΔG values constituted 0.14 kcal/mol [10]. As shown in [10], similar pattern is preserved in case of the analysis of extended DNA stretches. Calculations of absolute errors for the free energy of the base pairs formation based on literature data were also taken into consideration in our computing experiment carried out using the OMIKO-MCS v1.1 software program designed.

According to the data obtained, *Du183* has the size of 368 bp and incorporates a complex microsatellite consisting of different microsatellite clusters, which are tightly associated with each other, $(TG)_n$, $(AG)_n$, $(G)_n$, and $(AGG)_n$. The 558-bp *Du214* locus contains perfect microsatellite $(GT)_n$. The 566-bp *Du281* locus has tetranucleotide microsatellite $(GATA)_n$. In the genome of *D. unisexualis*, each of the loci mentioned is represented by several alleles, differing in the number of microsatellite units and in the presence stable single nucleotide substitutions within the cluster or in the immediately flanking regions. *Du183* is represented by three allelic variants, while *Du214* and *Du281* are represented by six alleles each. The G+C content for *Du183* 48.8%, for *Du214*, 40.9%, and for *Du281*, 40.7%. The change of the mean Gibbs energy (ΔG_m) for these loci showed that microsatellite expansion had higher effect on this thermodynamic characteristic for *Du183*, compared to *Du214* and *Du281*. For instance, for *Du183* the ΔG_m of the DNA/DNA duplex formation was –1672 cal/mol, while according to the calculation data without considering the effect of microsatellite cluster, $\Delta G_m = -1658$ cal/mol. For *Du214* the values of this index were –1629 and –1621 cal/mol, and for *Du281*, –1623 and –1635 cal/mol. Note that for *Du183* and *Du214* the increase of the number of microsatellite units brings nucleotide sequences of these loci to less energy intensive state, while microsatellite expansion in case of the *Du281* locus results in the increase of the ΔG_m value. Taking into consideration the data discussed, it should be noted that the effects displayed by different clusters of complex microsatellite of the *Du183* locus were different. Specifically, expansion of $(TG)_n$ dinucleotide cluster resulted in the decrease of the ΔG_m value ($\Delta G_m = -1652$ cal/mol). At the same time, the $(AG)_n$ dinucleotide expansion led to the increase of the ΔG_m value up to $\Delta G_m = -1647$ cal/mol. Enlargement of mono- and trinucleotide cluster led to substantial decrease of ΔG_m, to the value of $\Delta G_m = -1675$ cal/mol. Analogous calculation algorithm was applied to all allelic variants of microsatellite loci examined. Thermodynamic characteristics of the loci are demonstrated in the table, where allelic variants are organized in accordance to the increase of the sequence size for each of the loci. It follows from the table that the tendencies described were typical of all allelic variants of the loci examined.

Interpreting the data obtained in terms of molecular genetics, we consider the genome and the DNA molecule as a system, consisting of interrelated elements (nucleotides). Stability of the system is determined by free energy, which is responsible for the direction (forward or reverse) of chemical reaction. It also determines the conformation (bents and joints) of large DNA, RNA, and protein molecules [11]. Each system tends to the state with the lowest free energy. The free energy of two complementary nucleotide chains is minimum when they form a double helix. The highest the ability of microsatellite sequences to form the loops of one or several microsatellite units, the higher the negative value of ΔG upon formation of the DNA/DNA duplex in the loop-flanking regions. The formation of loops and slippage of DNA strands during replication are the key moments in the so-called stepwise mutation model, SMM), according to which changes in the length of microsatellite cluster occur sequentially due

Thermodynamic characteristic of the *Du183*, *Du214*, and *Du281* allelic variants

Locus*/allelic variant**	Microsatellite cluster	DNA size, nucleotides	G + C, %	Total ΔG of DNA/DNA duplex formation, cal/mol	Absolute calculation error for total ΔG of DNA/DNA duplex formation, cal/mol	Mean ΔG of DNA/DNA duplex formation, cal/mol	Absolute calculation error for mean ΔG of DNA/DNA duplex formation, cal/mol
Du183	$(TG)_{10}(AG)_7(G)_7(AGG)_{10}$	368	48.8	–615426	16555	–1672	44.98
2	$TGAG(TG)_4(AG)_8(G)_5(AGG)_7AAGAGG$	192	47.7	–318826	8576	–1660	44.65
3	$TGAG(TG)_4(AG)_7(G)_7(AGG)_{10}$	195	48.9	–324799	8737	–1665	44.79
1	$(TG)_{10}(AG)_7(G)_7(AGG)_{10}$	201	50	–337120	9069	–1677	45.11
Du214	$GTAA(GT)_{21}$	558	40.9	–909315	24461	–1629	43.82
2	$GTAA(GT)_{20}$	219	40.8	–355436	9561	–1622	43.63
4	$GTAA(GT)_{21}$	221	40.8	–358889	9654	–1623	43.66
6	$GTAA(GT)_{24}$	223	41	–362986	9764	–1627	43.77
1	$(GT)_{25}$	225	41.5	–366440	9857	–1628	43.79
3	$(GT)_{26}$	227	41.5	–369893	9950	–1629	43.82
5	$(GT)_{29}$	233	41.7	–380254	10229	–1631	43.87
Du281	$(GATA)_{10}TAGATA$	566	40.7	–902409	24275	–1623	43.66
4	$(GATA)_{10}TAGATA$	197	36.2	–311489	8379	–1581	42.53
1	$(GATA)_9GATGATATAGATA$	200	34.6	–313272	8427	–1566	42.13
6	$(GATA)_{11}TAGATA$	201	36	–317448	8539	–1579	42.48
3	$(GATA)_{10}GATGATATAGATA$	204	34.3	–319231	8587	–1564	42.07
5	$(GATA)_{11}GATGATATAGATA$	208	34.3	–325190	8748	–1563	42.04
2	$(GATA)_{12}GATGATATAGATA$	212	34	–331149	8908	–1562	42.02

Notes: * Full insert of recombinant clone;
** PCR amplificant produced on DNA of recombinant clone.

to the increase or reduction of the repeat length by one monomeric unit [12]. Based on these data, it can be suggested that the changes in microsatellite clusters investigated in the present study will tend to such number of microsatellite units, which will provide the lowest ΔG_m value to the whole cluster. Based on thermodynamic calculations, it can be expected that changes in the *Du183* and *Du214* loci will be directed to the enlargement of microsatellite clusters, while changes in *Du281* will tend to decrease of microsatellite cluster. In addition, dynamics of the ΔG_m variation among allelic variants of the loci can serve as an explanation of different polymorphism levels. Specifically, *Du214* and *Du281* have six allelic variants each, while *Du183* has three alleles. This finding can be explained in terms that microsatellite variation of *Du183* by one monomeric unit involves its several cluster, and in terms of $|\Delta G_m|$ and the module of total ΔG corresponds to changes of *Du214* and *Du281* microsatellites by several monomeric units (see the table). Hence, changes in the microsatellite monomer composition of the *Du183* locus, during equal time interval and by equal ΔG value, should be lower than in *Du214* and *Du281*. Thus, populations of *D. unisexualis* accumulate lower numbers of the *Du183* allelic variants, along with higher numbers of the *Du214* and *Du281* allelic variants.

Microsatellite sequences in general, and GT microsatellites in particular, are able to form numerous noncanonical DNA structures. These structures include G-quadruplexes, parallel duplexes, and triplexes. It is known that DNA polymerase can lag or delay at stable hairpins, triplexes, and quadruplexes, and the repeat number is critical for formation of different noncanonical structures [13]. However, looping of any microsatellite cluster region, which plays the key role in the stepwise mutation model, can take place only within the DNA region with linear structure. Hence, it is likely that the number of units in microsatellite clusters, where slippage of the replication fork happens more often, is such that they are able to form noncanonical structures only in very seldom cases.

Analysis of nucleotide compositions of the loci showed that these loci were stable in terms of the GC content irrespectively of the sizes of the locus fragments examined. Deviations in the GC content for any of allelic variant tested were insignificant (see the table). It can be thus suggested, that GC content of these loci can be preserved along the regions of 100 to 300 thousand nucleotides. The *Du183* locus can be thus attributed to the first family of heavy isochors, which is characterized by localization in the telomeric regions of metaphase chromosomes and early replication in the cell cycle [4]. In this case, *Du214* and *Du281* should belong to the family of "light" isochors, with different localization and characterized by late replication.

In conclusion, in this study, thermodynamic characteristization of three polymorphic microsatellite loci is presented. Based on calculation of the ΔG_m values, it is demonstrated that the numbers of monomeric units of microsatellite clusters will change to the decrease of the mean free energy of these sequences. Furthermore, the sizes of the *Du183* and *Du214* 2 microsatellites will increase, while the size of the *Du281* microsatellite will decrease. Based on thermodynamic calculations, molecular genetic interpretation of different polymorphism levels of the loci examined is given. The reason for the differences observed lies in different free energy (Gibbs energy) alteration rates upon single mutational event in each microsatellite in accordance with the stepwise model. Based on comparison and analysis of the GC content of each locus, belonging of the loci examined to certain isochore families is suggested.

ACKNOWLEDGEMENTS

We thank A. P. Ryskov for critical discussion and many helpful comments.

This work was supported by the Russian Foundation for Basic Research (grant nos. 08-04-0668-a, 07-04-00462-a, and ofi-08-04-12204), the Program of the Presidium of Russian Academy of Sciences "Molecular and Cellular Biology", Subprogram of the Presidium of the Russian Academy of Sciences (AE 142/200404-112) "Dynamics of Plant, Animal and Human Gene Pools", and the Programs of the President of the Russian Federation "Leading Scientific Schools" (grant no. NSh-2107.2008.4) and for support of young scientists (grant no. MK-4748.2007.4).

REFERENCES

1. Ellegren, H., Microsatellites: Simple Sequences with Complex Evolution, *Nat. Rev. Genet.*, 2004, vol. 5, pp. 435–445.
2. Wells, R.D., Molecular Basis of Genetic Instability of Triplet Repeats, *J. Biol. Chem.*, 1996, vol. 271, no. 6, pp. 2875–2878.
3. SantaLucia, J., Jr. and Hicks, D., The Thermodynamics of DNA Structural Motifs, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, vol. 33, pp. 415–440.
4. Patrushev, L.I. and Minkevich, I.G., The Problem of the Eukaryote Genome Size, *Usp. Biol. Khim.*, 2007, vol. 47, pp. 293–370.
5. Lercher, M.J., Smith, N.G.C., Eyre-Walker, A., and Hurst, L.D., The Evolution of Isochores: Evidence from SNP Frequency Distributions, *Genetics*, 2002, vol. 162, pp. 1805–1810.
6. Omel'chenko, A.V., Korchagin, V.I., Tokarskaya, O.N., and Sevast'yanova, G.A., Study of Microsatellite Dinucleotide Sequences in the Genome of Parthenogenetic Lizard *Darevskia unisexualis*, in *Perspektivnye napravleniya fiziko-khimicheskoi biologii* (Priority Directions of Physico-Chemical Biology), Proc. XVII Winter Youth Sci. School, Moscow, 2005, p. 60.
7. Korchagin, V.I., Badaeva, T.N., and Tokarskaya, O.N., Molecular Characterization of Allelic Variants of (GATA)*n* Microsatellite Loci in Parthenogenetic Lizards

Darevskia unisexualis (Lacertidae), *Gene*, 2007, vol. 392, pp. 126–133.

8. Owczarzy, R., Vallone, P.M., and Gallo, F.J., Predicting Sequence-Dependent Melting Stability of Short Duplex DNA Oligomers, *Biopolymers*, 1997, vol. 44, pp. 217–239.
9. SantaLucia, J., Jr., Allawi, H.T., and Seneviratne, P.A., Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability, *Biochemistry*, 1996, vol. 35, pp. 3555–3562.
10. SantaLucia, J., A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics, *Biochemistry*, 1998, vol. 95, pp. 1460–1465.
11. Siman, N., Nanotechnology and the Double Helix, *Sci. Am.*, 2007, no. 9, pp. 23–31.
12. Valdes, A.M., Slatkin, M., and Freimer, N.B., Allele Frequencies at Microsatellite Loci: The Stepwise Mutation Model Revisited, *Genetics*, 1993, vol. 133, pp. 737–749.
13. Siyanova, E.Yu. and Mirkin, S.M., Expansion of Trinucleotide Repeats, *Mol. Biol.* (Moscow), 2001, vol. 35, no. 2, pp. 208–223.