Universal Algorithm for Identification of Fractional Brownian Motion. A Case of Telomere Subdiffusion

Krzysztof Burnecki,[†] Eldad Kepten,[‡*] Joanna Janczura,[†] Irena Bronshtein,[‡] Yuval Garini,[‡] and Aleksander Weron[†]
[†]Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wroclaw University of Technology, Wroclaw, Poland; and [‡]Physics Department and Institute of Nanotechnology, Bar Ilan University, Ramat Gan, Israel

ABSTRACT We present a systematic statistical analysis of the recently measured individual trajectories of fluorescently labeled telomeres in the nucleus of living human cells. The experiments were performed in the U2OS cancer cell line. We propose an algorithm for identification of the telomere motion. By expanding the previously published data set, we are able to explore the dynamics in six time orders, a task not possible earlier. As a result, we establish a rigorous mathematical characterization of the stochastic process and identify the basic mathematical mechanisms behind the telomere motion. We find that the increments of the motion are stationary, Gaussian, ergodic, and even more chaotic—mixing. Moreover, the obtained memory parameter estimates, as well as the ensemble average mean square displacement reveal subdiffusive behavior at all time spans. All these findings statistically prove a fractional Brownian motion for the telomere trajectories, which is confirmed by a generalized *p*-variation test. Taking into account the biophysical nature of telomeres as monomers in the chromatin chain, we suggest polymer dynamics as a sufficient framework for their motion with no influence of other models. In addition, these results shed light on other studies of telomere motion and the alternative telomere lengthening mechanism. We hope that identification of these mechanisms will allow the development of a proper physical and biological model for telomere subdynamics. This array of tests can be easily implemented to other data sets to enable quick and accurate analysis of their statistical characteristics.

INTRODUCTION

Our work is motivated from one side by the growing interest in single molecule spectroscopy (1–16), in particular by single-particle tracking (SPT) in a context of anomalous diffusion (11,18–25) and from the other side by the discovery of how chromosomes are protected by telomeres and the enzyme telomerase (26). By performing a thorough statistical analysis on a very large data set of telomere motions, we set new perspectives, to our knowledge, on telomere dynamics and show how to systematically analyze future SPT results.

SPT measurements can provide new experimental knowledge even with basic analysis. For example, diffusion coefficients and qualitative behavior of the measured species can be found by analyzing the mean square displacement $\langle x^2(t)\rangle$ (MSD). However, there is much more to be obtained if one dives deeper into the statistical nature of the data, possibly even contradicting the results of simpler analysis. In the case of anomalous diffusion, i.e., when the MSD is not linear in t, the importance of thorough analysis is intensified as there are many mathematical and physical models that can give rise to similar forms of MSDs (27). A recent example of the strength of stochastic analysis is the study of the dynamics of Kv2.1 potassium channels in cellular membranes (28) that showed compartmentalization and binding.

A phenomenon observed in recent nanoscale single-molecule biophysics experiments is subdiffusion, which largely departs from the classical Brownian diffusion theory. Determining the origin anomalous diffusion in crowded fluids, e.g., in the cytoplasm of living cells (29), or in more controlled in vitro experiments (9), is a challenging problem. The most popular theoretical models that are commonly employed are continuous-time random walk (CTRW), obstructed diffusion (OD), fractional Brownian motion (FBM), fractional Levy stable motion (FLSM), and fractional Langevin equation (FLE) (30–40). These models can be divided into two categories: with short memory (CTRW, OD) and fractional with long (power law) memory (FBM, FLSM, FLE, and fractional autoregressive moving average (FARIMA)). The FARIMA model unifies the latter category (31).

To date, there is no standard analysis scheme that can identify the mathematical model behind a wide variety of measured anomalous trajectories. Each of these models has a unique definition, which implies unique characteristics and thus various tests have been proposed to differentiate between them (35–39). However, each of these tests covers only a limited set of characteristics, and hence does not capture the complete stochastic picture. It is therefore expected that a standardized testing scheme for different models that would encompass a variety of different experimental scenarios would be highly beneficial.

FBM statistically behaves identical to FLE in the overdamped limit. However, the noise in FBM is external and not coupled to fluctuations, in contrast to the FLE motion, for which a proper temperature is defined (40). FBM/FLE motion is the effective motion of a labeled, single particle

Submitted June 24, 2012, and accepted for publication September 10, 2012.

*Correspondence: Eldad.Kepten@biu.ac.il

Editor: Levi Gheber.

0006-3495/12/11/1839/9 $2.00

http://dx.doi.org/10.1016/j.bpj.2012.09.040

connected with a many-particle system, such as a monomer in a long polymer, or an individual particle in a single file (41).

One system that was suggested to obey such fractional dynamics is the motion of telomeres in the eukaryotic nucleus (42). Telomeres form the capping structures at the ends of each chromosome. They consist of a repeating DNA sequence and a specific protein complex surrounding it (named shelterin). The telomeres play an important role in maintaining the chromosomal integrity, preventing adhesion of chromosomes to one another, as well as degradation of the ends of the chromosomes through enzymes that are abundant in the nucleus (43–45). The telomeres lose a short piece of the DNA each time the chromosome is replicated. This shortening is believed to be related to aging, although the actual mechanisms are much more complex (46).

Because there are a significant number of telomeres (92 in humans) that are, generally speaking, uniformly distributed in the nucleus, they form excellent probes for investigating the dynamics of the DNA and its accompanying proteins called chromatin. The telomeres can be labeled in living cells through one of the native shelterin proteins using fluorescent proteins such as the green fluorescent protein. The first publication of telomeres diffusion appeared in (47).

During the last few years, the dynamics of telomeres was studied more carefully by analyzing their diffusion properties. This method can provide important information on the structure and function of the chromatin in general, and the telomeres themselves in particular. Previous study has found the telomeres motion to be subdiffusive up to 10^2 s with an anomalous exponent of $\alpha = 0.32 \pm 0.12 - 0.51 \pm 0.20$ and a higher anomalous exponent at longer times (48). Because anomalous diffusion can originate from various mechanisms (19,27), further analysis can shed light on the nuclear mechanisms behind telomere motion. Thus, a better picture of in vivo chromatin motion and nuclear structure can be obtained.

A follow-up analysis of short time motion of telomeres proved (42) that binding and CTRW models are not the origin of the subdiffusive regime. Displacements were found to be correlated in time and weak ergodicity convergence complied with the prediction for fractional processes. Thus, it was proposed that some correlative mechanism stands behind the local motion of chromatin and telomeres, specifically throughout the interphase. In this study, the possibility that the anomalous diffusion of telomeres stems from obstruction of motion in the nucleus by quasistatic obstacles was also looked into and found inadequate.

In parallel, a series of studies have connected telomere motion to their elongation and maintenance. This was proposed for telomerase-positive cancer cells (UMUC3) (49) and for U2OS cancer cells that use the alternative lengthening of telomeres (ALT) mechanism (50). In both cases, heterogeneity of the telomere population was claimed. It was proposed that this heterogeneity enables a motion-based control mechanism for telomere elongation. Recently, another study (51) showed that transcription activity also influences telomere motion.

In this work, we first present a universal, systematic, and intuitive algorithm for the identification and validation of FBM. The proposed algorithm tests all the fundamental characteristics of FBM. In addition, ergodicity is also tested for and its significance in the context of data analysis explained. Furthermore, the important property of the self-similarity of the process is validated with two independent tests. Compliance with the testing scheme given here proves that FBM is the only mathematical model suitable for the data tested.

To prove the ability of the proposed algorithms to test various data structures, we analyze an expanded version of the data presented in (48). The data are divided into three time domains, each with a different measurement rate and amount of trajectories. The telomere motion is found to follow the FBM process in over five time orders, a conclusion not possible with previous analysis. This, to our knowledge, sheds new light on the motion of telomeres measured in the previously mentioned works and on chromatin dynamics in general.

METHODS

Experimental methodology

A full description of the experimental procedure is given in the Supporting Material and in (48). In short, telomeres in U2OS osteosarcoma cells are labeled with a protein composed of a green fluorescent protein attached to a TRF-1 protein. Cells are then imaged from millisecond time frames and up to 30 min. Images are analyzed for telomere tracks with the use of Imaris and Image J software. Tracks are then analyzed for stochastic traits with in-house written scripts and programs.

The experimental noise was measured by measuring telomere motion after fixating the cells. It was found that the effective noise in all timescales is Gaussian with zero mean and a variance of 22 nm. For more details see the Supporting Material.

Identification and validation algorithm

It is natural to expect a universal identification algorithm for FBM that would encompass a variety of different experimental scenarios. Such an algorithm would give rise to confidence levels of results and enable benchmarking between different experiments and systems. Ultimately such an algorithm would direct the experimental process because it would indicate which tests should be applied to the analyzed data.

What are the requirements for such an algorithm? First of all, it should be applicable to any trajectory data, i.e., it should be universal. An SPT data structure is composed of independent trajectories for single particles, measured at a certain frame rate. Each trajectory can be in a single dimension or more. Both the number of trajectories and the amount of data points may vary between data sets. A good algorithm should work sufficiently well both in the case of many trajectories with a few data points or a few trajectories with many data points.

Second, such an algorithm should be systematic. It should identify all fundamental characteristics of a certain mathematical model. When looking at a single characteristic, it is possible to eliminate a certain model but not to validate one. However, if all fundamental characteristics of a model are verified, we can then confirm that the measured process is indeed of that