Physica A 392 (2013) 6375–6388

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# A novel family of space-filling curves in their relation to chromosome conformation in eukaryotes

Jan Smrek *, Alexander Y. Grosberg

*Center for Soft Matter Research and Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA*

## HIGHLIGHTS

- We show the construction of a novel family of space-filling curves with fractal domain boundaries of fractal dimension close to the dimension of the embedding space.
- These curves match the statistical properties of eukaryotic DNA in interphase nucleus.
- The curves are a very efficient tool in modeling diffusion and dynamics in the nucleus.

## ARTICLE INFO

*Article history:*
Received 3 May 2013
Received in revised form 7 August 2013
Available online 20 August 2013

*Keywords:*
Space-filling curve
Fractal
DNA
Polymer
Genome folding

## ABSTRACT

Spatial conformation of DNA chains during interphase in eukaryotic cell nucleus is relatively dense, yet unknotted and exhibits self-similar fractal properties. In this respect it resembles the space-filling curves of Hilbert, but differs in the experimentally accessible contact probability of distant loci. Here we construct space-filling curves with fractal domain boundaries of dimension close to that of the embedding space and show how these match the statistical properties and the contact probability of the DNA conformation. The present mathematical model should shed light on the statistical ensemble of unknotted dense polymers and ease the modeling of genome folding and related biological processes.



## 1. Introduction

Examples of curves visiting every point in a square or a cube were constructed over a century ago by Giuseppe Peano [1] and David Hilbert [2] in the context of set theory. Excitingly, the defining properties of these curves — space-filling, self-similar, and (in 3D) unknotted — are all implicated in the spatial organization of giant DNA chains in interphase chromosomes of eukaryotic cells [3–9]. Unfortunately, experiments on human [4], mouse [10], and drosophila [11] cells, as well as computer simulation of dense random unknotted polymers [4,12,13] revealed a major discrepancy between the known types of space-filling curves and chromatin, manifested, both, in contact probability between distant loci, and in surface roughness of subchains. Do self-similar curves exist with properties matching those of DNA? Here, we prove their existence by explicit construction, thus giving credence to the hypothesis that the statistical ensemble of conformations in folded chromatin is governed by simple physics and geometry principles.

To construct our novel one-parameter family of space-filling curves mimicking chromatin properties, we will consider a cubic (or square in 2D for the primer) lattice and an iterative process in which the lattice constant is decimated at every step by a given factor of $n$: $a_{k+1} = a_k/n$, with $n$ the parameter. In mathematical theory, the central issue is the nature of the $k \to \infty$ limit, along with associated questions of continuity. With our main interest in chromatin applications, we will focus on a

* Corresponding author. Tel.: +1 2129988966.
*E-mail address:* js5013@nyu.edu (J. Smrek).

0378-4371/$ – see front matter 
http://dx.doi.org/10.1016/j.physa.2013.08.014

finite level of decimation; accordingly, our curves will be self-similar over a finite range of scales, and we will call them fractal in a loose physical sense of the word. In this model, the chromatin fiber is represented by a self-avoiding path going through the centers of lattice plaquettes, always connecting nearest neighbors (i.e., never going along the diagonals). To account for the high and relatively uniform chromatin packing density, we assume the path Hamiltonian (i.e., visiting every lattice site within a given overall volume). That means the fractal dimension of our curves will be that of the embedding space, $d$.

Thus the formulated model immediately yields a useful insight into a quantity which is measured for the chromatin fiber and called the contact probability $P(s)$. Experimentally, it is the probability to find two loci separated by the genetic distance $s$ in spatial proximity of one another. In the model, we consider all possible curve segments of length $s$ (measured in lattice constants) and $P(s)$ is the fraction of these segments for which two ends occupy the centers of neighboring lattice plaquettes. It is observed that in a certain range of $s$ there is a power law $P(s) \sim s^{-\gamma}$, with the index $\gamma$ slightly above unity (1.08 for human cells [4] and 1.05 for mouse [10]). A useful intuition about $\gamma$ is afforded by the theorem [13] which relates this index to the fractal dimension $d_b$ of the surfaces separating the subchains: if we consider a subchain (or sub-path) of length $s$ (in lattice steps), we should look at the set of all plaquettes visited by this subchain, and measure total length (in 2D) or surface area (in 3D) where these plaquettes contact the ones not visited by the subchain. Treating this as "surface" motivates writing its length or area as $N_b(s) \sim r(s)^{d_b}$, with $r(s)$ the size (e.g., gyration radius or average end-to-end distance) of the $s$-long subchain; since the fractal dimension of the path is $d$, meaning $r(s) \sim s^{1/d}$, so $N_b(s) \sim s^{d_b/d}$, and

$$\gamma = 2 - \frac{d_b}{d}. \tag{1}$$

One quick way to prove the theorem is to note that $N_b(s)$ is the number of contacts between monomers $s'$ inside the subchain and $s''$ outside of it, which at large $N$ evaluates to $N_b(s) \sim \sum_{s''>s}^{N} \sum_{s'<s} P\left(s''-s'\right) \sim s^{2-\gamma}$; see also Refs. [13,7] for other proofs and a more nuanced discussion.

Thus, the observed value of the exponent $\gamma$ slightly above unity can be realized in the fractal space-filling curve only if the dimension of the boundary is just slightly below the embedding space dimension; it must be a very "thick", very wiggly boundary. Meanwhile, classical space-filling curves, such as Peano and Hilbert (see Appendix D), have smooth domain boundaries, which means $d_b = d-1$. One example of higher $d_b$, known to us, is in two dimensions the Gosper curve [14–16] with $d_b \approx 1.12915$ and its slight generalizations [17]. The corresponding $\gamma$ is nowhere near the observed value, quite apart from the fact that a 3D version of a curve with fractal boundary was not discovered.

## 2. Construction

Here we prove, by explicit algorithmic construction, the existence of space-filling curves with fractal boundaries and with $d_b$ arbitrarily close to $d$ from below, both in $d = 2$ and in $d = 3$. Our construction is almost clear from the figures, so only relatively few comments are needed.

Since the algorithm generates the curves iteratively, we only need to describe one iteration step. It starts with a lattice with square (or cubic) plaquettes $\mathcal{P}_k$, of side $a_k$, and replaces it with a tile $\mathcal{T}_{k+1}$ of the same area (volume) $a_k^d$; in other words, the new tile $\mathcal{T}_{k+1}$ must consist of $n^d$ plaquettes $\mathcal{P}_{k+1}$, each of the size $a_{k+1} = a_k/n$. The shape of the tiles $\mathcal{T}$ must satisfy several criteria and its choice (and existence) is by no means trivial (or obvious): (i) it must tessellate the space; (ii) it must allow for a Hamiltonian path traversing it via centers of all the $n^d$ plaquettes; (iii) the least trivial of all, it must allow for several different Hamiltonian paths, such that they connect smoothly between neighboring tiles.

For simplicity we begin with the curve building algorithm in 2D, postponing the straightforward 3D generalization towards the end of the note.

To solve the first of these problems, we draw inspiration from M. C. Escher, particularly drawings of his such as "Pegasus" (see, e.g., Ref. [18]). In the Escher case, each tile is based on a square, with opposite sides identically distorted. For our purposes it is sufficient to distort just one pair of sides and leave the other two sides straight. We illustrate this in Fig. 1 where we show one such (off-lattice) tile of Escher's for illustration, and one of our (on-lattice) tiles in 2D, with $n = 10$. This construction is easily generalized for larger values of $n$ (provided $n$ is even, but not divisible by 4, or $(n-2)/4 \in \mathbb{N}$; see Fig. 2 for $n = 10$, 14 and 18), eventually giving rise to a one-parameter family of space-filling curves with increasing $d_b$ approaching $d$ (i.e., two in this case):

$$d_b = \frac{\log\left(\frac{1}{2}n^2 - 3n + 4\right)}{\log n}. \tag{2}$$

It is calculated as the ratio of logarithms of the number of plaquettes that form the horizontal boundary and the scale factor of iteration $n$. For a more detailed calculation see Appendix A. The fractal dimension of the boundary is an increasing function of $n$, bounded by 2 from above; hence the curve with $\gamma$ arbitrarily close to one can be constructed using a tile of high enough $n$.

This choice of the tile shape does allow for at least four Hamiltonian paths, that are explicitly shown in Fig. 3 for $n = 14$ (with obvious generalization for other $n$) and labeled $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$; this way we obtain four "decorated tiles" $\mathcal{T}^{(\mathcal{A})}$, $\mathcal{T}^{(\mathcal{B})}$, $\mathcal{T}^{(\mathcal{C})}$, and $\mathcal{T}^{(\mathcal{D})}$. Many other Hamiltonian paths exist for this shape of the tile, but these four suffice to provide for all smooth connections, as we show next.

It is convenient to assign orientation to the entire curve and, accordingly, to its part within each plaquette. Then, every plaquette is traversed by the curve in one of twelve ways (or $2d$ possibilities to enter and $(2d-1)$ ways to exit, $2d(2d-1)$