Bill & Melinda Gates Foundation); the U.S. Agency for International Development (USAID); and the National Institute of Allergy and Infectious Diseases, NIH, AI33292 (D.R.B.). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the U.S. government. The authors declare competing financial interests. Protocol G Principal Investigators: G. Miiro, J. Serwanga, A. Pozniak, D. McPhee, O. Manigart, L. Mwananyanda, E. Karita, A. Inwoley, W. Jaoko, J. DeHovitz, L. G. Bekker, P. Pitisuttithum, R. Paris, and S. Allen.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1178746/DC1
Materials and Methods
SOM Text
Figs. S1 to S10
Tables S1 to S6
References

7 July 2009; accepted 26 August 2009
Published online 3 September 2009;
10.1126/science.1178746
Include this information when citing this paper.

Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden,[1,2,3,4]* Nynke L. van Berkum,[5]* Louise Williams,[1] Maxim Imakaev,[2] Tobias Ragoczy,[6,7] Agnes Telling,[6,7] Ido Amit,[1] Bryan R. Lajoie,[5] Peter J. Sabo,[8] Michael O. Dorschner,[8] Richard Sandstrom,[8] Bradley Bernstein,[1,9] M. A. Bender,[10] Mark Groudine,[6,7] Andreas Gnirke,[1] John Stamatoyannopoulos,[8] Leonid A. Mirny,[2,11] Eric S. Lander,[1,12,13]† Job Dekker[5]†

We describe Hi-C, a method that probes the three-dimensional architecture of whole genomes by coupling proximity-based ligation with massively parallel sequencing. We constructed spatial proximity maps of the human genome with Hi-C at a resolution of 1 megabase. These maps confirm the presence of chromosome territories and the spatial proximity of small, gene-rich chromosomes. We identified an additional level of genome organization that is characterized by the spatial segregation of open and closed chromatin to form two genome-wide compartments. At the megabase scale, the chromatin conformation is consistent with a fractal globule, a knot-free, polymer conformation that enables maximally dense packing while preserving the ability to easily fold and unfold any genomic locus. The fractal globule is distinct from the more commonly used globular equilibrium model. Our results demonstrate the power of Hi-C to map the dynamic conformations of whole genomes.

The three-dimensional (3D) conformation of chromosomes is involved in compartmentalizing the nucleus and bringing widely separated functional elements into close spatial proximity (*1*–*5*). Understanding how chromosomes fold can provide insight into the complex relationships between chromatin structure, gene activity, and the functional state of the cell. Yet beyond the scale of nucleosomes, little is known about chromatin organization.

[1]Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), MA 02139, USA. [2]Division of Health Sciences and Technology, MIT, Cambridge, MA 02139, USA. [3]Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, MA 02138, USA. [4]Department of Applied Mathematics, Harvard University, Cambridge, MA 02138, USA. [5]Program in Gene Function and Expression and Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA. [6]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [7]Department of Radiation Oncology, University of Washington School of Medicine, Seattle, WA 98195, USA. [8]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [9]Department of Pathology, Harvard Medical School, Boston, MA 02115, USA. [10]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA. [11]Department of Physics, MIT, Cambridge, MA 02139, USA. [12]Department of Biology, MIT, Cambridge, MA 02139, USA. [13]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: lander@broadinstitute.org (E.S.L.); job.dekker@umassmed.edu (J.D.)

Long-range interactions between specific pairs of loci can be evaluated with chromosome conformation capture (3C), using spatially constrained ligation followed by locus-specific polymerase chain reaction (PCR) (*6*). Adaptations of 3C have extended the process with the use of inverse PCR (4C) (*7*, *8*) or multiplexed ligation-mediated amplification (5C) (*9*). Still, these techniques require choosing a set of target loci and do not allow unbiased genomewide analysis.

Here, we report a method called Hi-C that adapts the above approach to enable purification of ligation products followed by massively parallel sequencing. Hi-C allows unbiased identification of chromatin interactions across an entire genome.We briefly summarize the process: cells are crosslinked with formaldehyde; DNA is digested with a restriction enzyme that leaves a 5′ overhang; the 5′ overhang is filled, including a biotinylated residue; and the resulting blunt-end fragments are ligated under dilute conditions that favor ligation events between the cross-linked DNA fragments. The resulting DNA sample contains ligation products consisting of fragments that were originally in close spatial proximity in the nucleus, marked with biotin at the junction. A Hi-C library is created by shearing the DNA and selecting the biotin-containing fragments with streptavidin beads. The library is then analyzed by using massively parallel DNA sequencing, producing a catalog of interacting fragments (Fig. 1A) (*10*).

We created a Hi-C library from a karyotypically normal human lymphoblastoid cell line (GM06990) and sequenced it on two lanes of an Illumina Genome Analyzer (Illumina, San Diego, CA), generating 8.4 million read pairs that could be uniquely aligned to the human genome reference sequence; of these, 6.7 million corresponded to long-range contacts between segments >20 kb apart.

We constructed a genome-wide contact matrix M by dividing the genome into 1-Mb regions ("loci") and defining the matrix entry m_{ij} to be the number of ligation products between locus i and locus j (*10*). This matrix reflects an ensemble average of the interactions present in the original sample of cells; it can be visually represented as a heatmap, with intensity indicating contact frequency (Fig. 1B).

We tested whether Hi-C results were reproducible by repeating the experiment with the same restriction enzyme (HindIII) and with a different one (NcoI). We observed that contact matrices for these new libraries (Fig. 1, C and D) were extremely similar to the original contact matrix [Pearson's $r = 0.990$ (HindIII) and $r = 0.814$ (NcoI); P was negligible ($<10^{-300}$) in both cases]. We therefore combined the three data sets in subsequent analyses.

We first tested whether our data are consistent with known features of genome organization (*1*): specifically, chromosome territories (the tendency of distant loci on the same chromosome to be near one another in space) and patterns in subnuclear positioning (the tendency of certain chromosome pairs to be near one another).

We calculated the average intrachromosomal contact probability, $I_n(s)$, for pairs of loci separated by a genomic distance s (distance in base pairs along the nucleotide sequence) on chromosome n. $I_n(s)$ decreases monotonically on every chromosome, suggesting polymer-like behavior in which the 3D distance between loci increases with increasing genomic distance; these findings are in agreement with 3C and fluorescence in situ hybridization (FISH) (*6*, *11*). Even at distances greater than 200 Mb, $I_n(s)$ is always much greater than the average contact probability between different chromosomes (Fig. 2A). This implies the existence of chromosome territories.

Interchromosomal contact probabilities between pairs of chromosomes (Fig. 2B) show that small, gene-rich chromosomes (chromosomes 16, 17, 19, 20, 21, and 22) preferentially interact with each other. This is consistent with FISH studies showing that these chromosomes frequently colocalize in the center of the nucleus

(*12*, *13*). Interestingly, chromosome 18, which is small but gene-poor, does not interact frequently with the other small chromosomes; this agrees with FISH studies showing that chromosome 18 tends to be located near the nuclear periphery (*14*).

We then zoomed in on individual chromosomes to explore whether there are chromosomal regions that preferentially associate with each other. Because sequence proximity strongly influences contact probability, we defined a normalized contact matrix M^* by dividing each entry in the contact matrix by the genome-wide average contact probability for loci at that genomic distance (*10*). The normalized matrix shows many large blocks of enriched and depleted interactions, generating a plaid pattern (Fig. 3B). If two loci (here 1-Mb regions) are nearby in space, we reasoned that they will share neighbors and have correlated interaction profiles. We therefore defined a correlation matrix C in which c_{ij} is the Pearson correlation between the ith row and jth column of M^*. This process dramatically sharpened the plaid pattern (Fig. 3C); 71% of the resulting matrix entries represent statistically significant correlations ($P \leq 0.05$).

The plaid pattern suggests that each chromosome can be decomposed into two sets of loci (arbitrarily labeled A and B) such that contacts within each set are enriched and contacts between sets are depleted. We partitioned each chromosome

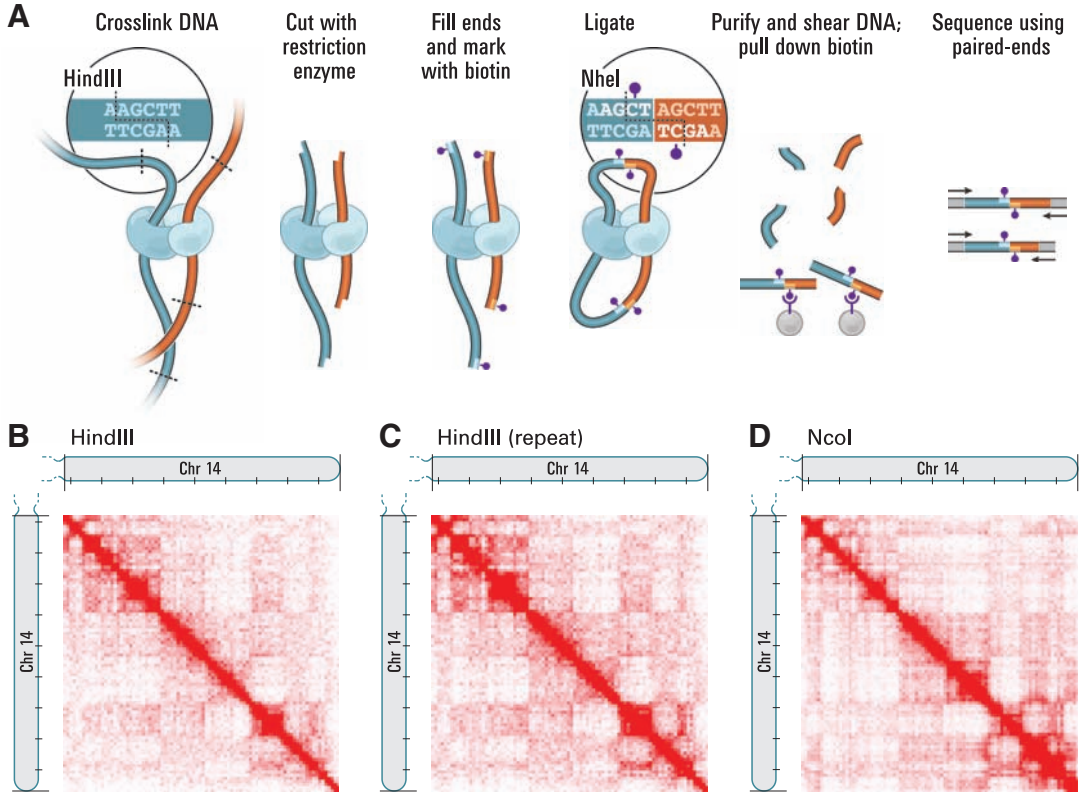


Fig. 1. Overview of Hi-C. (**A**) Cells are cross-linked with formaldehyde, resulting in covalent links between spatially adjacent chromatin segments (DNA fragments shown in dark blue, red; proteins, which can mediate such interactions, are shown in light blue and cyan). Chromatin is digested with a restriction enzyme (here, HindIII; restriction site marked by dashed line; see inset), and the resulting sticky ends are filled in with nucleotides, one of which is biotinylated (purple dot). Ligation is performed under extremely dilute conditions to create chimeric molecules; the HindIII site is lost and an NheI site is created (inset). DNA is purified and sheared. Biotinylated junctions are isolated with streptavidin beads and identified by paired-end sequencing. (**B**) Hi-C produces a genome-wide contact matrix. The submatrix shown here corresponds to intrachromosomal interactions on chromosome 14. (Chromosome 14 is acrocentric; the short arm is not shown.) Each pixel represents all interactions between a 1-Mb locus and another 1-Mb locus; intensity corresponds to the total number of reads (0 to 50). Tick marks appear every 10 Mb. (**C** and **D**) We compared the original experiment with results from a biological repeat using the same restriction enzyme [(C), range from 0 to 50 reads] and with results using a different restriction enzyme [(D), NcoI, range from 0 to 100 reads].

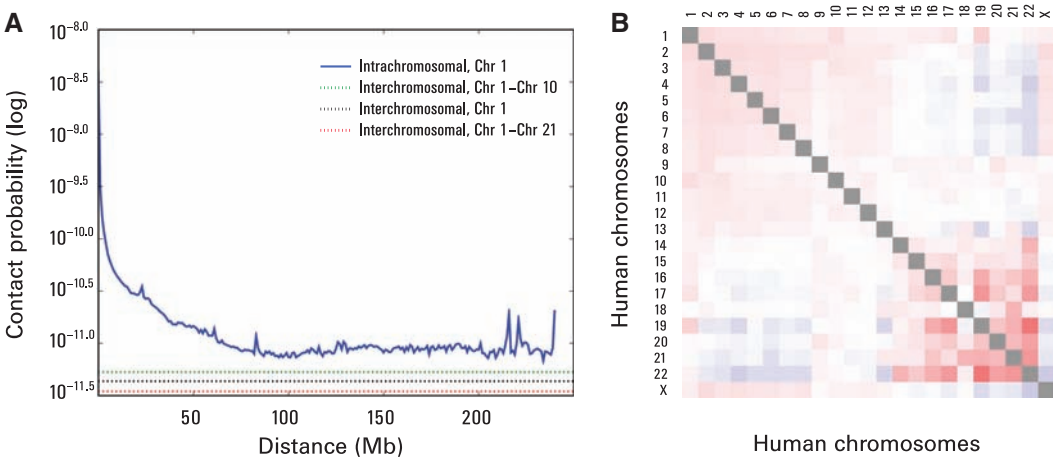


Fig. 2. The presence and organization of chromosome territories. (**A**) Probability of contact decreases as a function of genomic distance on chromosome 1, eventually reaching a plateau at ~90 Mb (blue). The level of interchromosomal contact (black dashes) differs for different pairs of chromosomes; loci on chromosome 1 are most likely to interact with loci on chromosome 10 (green dashes) and least likely to interact with loci on chromosome 21 (red dashes). Interchromosomal interactions are depleted relative to intrachromosomal interactions. (**B**) Observed/expected number of interchromosomal contacts between all pairs of chromosomes. Red indicates enrichment, and blue indicates depletion (range from 0.5 to 2). Small, gene-rich chromosomes tend to interact more with one another, suggesting that they cluster together in the nucleus.

in this way by using principal component analysis. For all but two chromosomes, the first principal component (PC) clearly corresponded to the plaid pattern (positive values defining one set, negative values the other) (fig. S1). For chromosomes 4 and 5, the first PC corresponded to the two chromosome arms, but the second PC corresponded to the plaid pattern. The entries of the PC vector reflected the sharp transitions from compartment to compartment observed within the plaid heatmaps. Moreover, the plaid patterns within each chromosome were consistent across chromosomes: the labels (A and B) could be assigned on each chromosome so that sets on different chromosomes carrying the same label had correlated contact profiles, and those carrying different labels had anticorrelated contact profiles (Fig. 3D). These results imply that the entire genome can be partitioned into two spatial compartments such that greater interaction occurs within each compartment rather than across compartments.

The Hi-C data imply that regions tend be closer in space if they belong to the same compartment (A versus B) than if they do not. We tested this by using 3D-FISH to probe four loci (L1, L2, L3, and L4) on chromosome 14 that alternate between the two compartments (L1 and L3 in compartment A; L2 and L4 in compartment B) (Fig. 3, E and F). 3D-FISH showed that L3 tends to be closer to L1 than to L2, despite the fact that L2 lies between L1 and L3 in the linear genome sequence (Fig. 3E). Similarly, we found that L2 is closer to L4 than to L3 (Fig. 3F). Comparable results were obtained for four consecutive loci on chromosome 22 (fig. S2, A and B). Taken together, these observations confirm the spatial compartmentalization

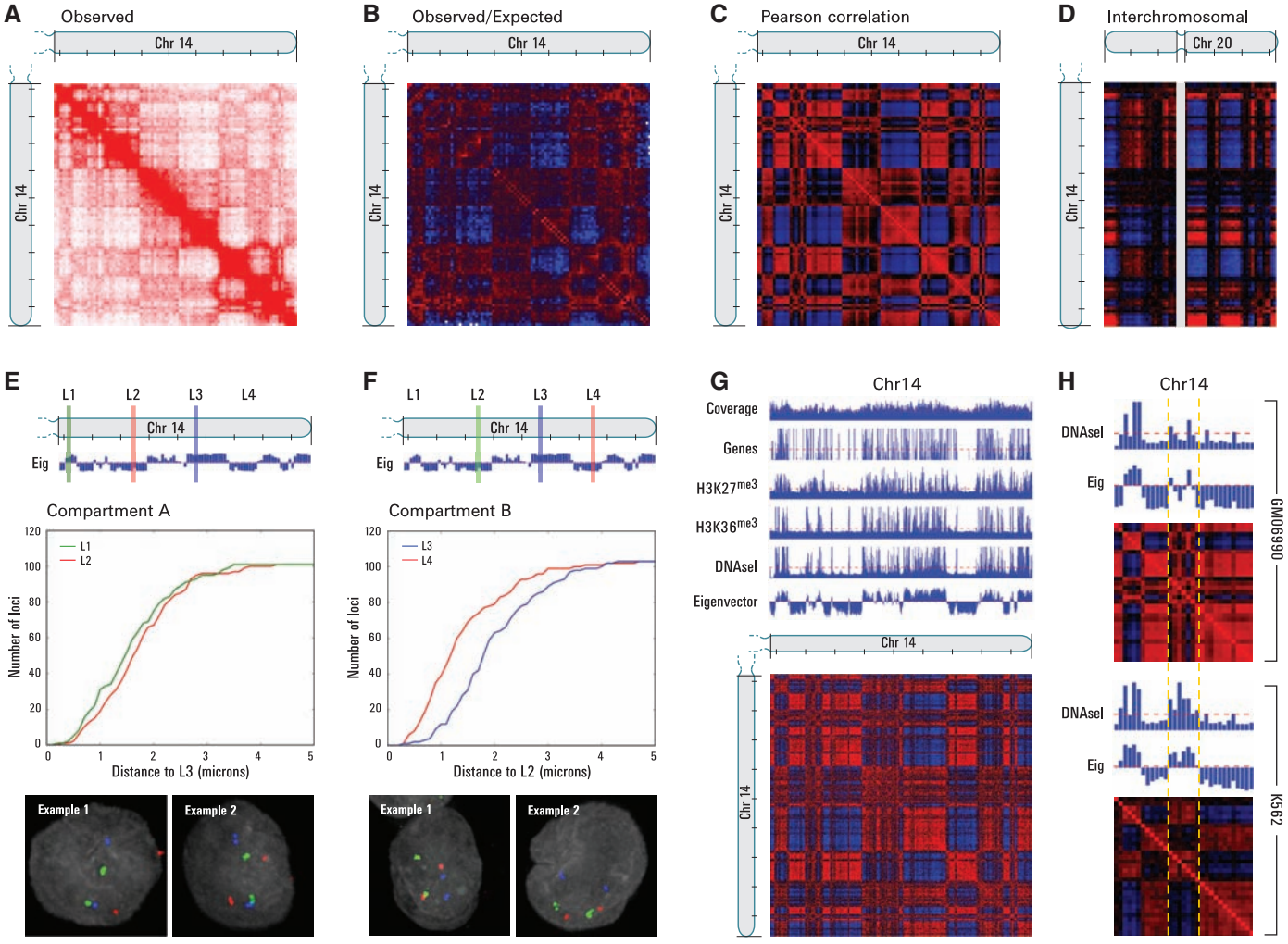


Fig. 3. The nucleus is segregated into two compartments corresponding to open and closed chromatin. (**A**) Map of chromosome 14 at a resolution of 1 Mb exhibits substructure in the form of an intense diagonal and a constellation of large blocks (three experiments combined; range from 0 to 200 reads). Tick marks appear every 10 Mb. (**B**) The observed/expected matrix shows loci with either more (red) or less (blue) interactions than would be expected, given their genomic distance (range from 0.2 to 5). (**C**) Correlation matrix illustrates the correlation [range from – (blue) to +1 (red)] between the intrachromosomal interaction profiles of every pair of 1-Mb loci along chromosome 14. The plaid pattern indicates the presence of two compartments within the chromosome. (**D**) Interchromosomal correlation map for chromosome 14 and chromosome 20 [range from –0.25 (blue) to 0.25 (red)]. The unalignable region around the centromere of chromosome 20 is indicated in gray. Each compartment on chromosome 14 has a counterpart on chromosome 20 with a very similar genome-wide interaction pattern. (**E** and **F**) We designed probes for four loci (L1, L2, L3, and L4) that lie consecutively along chromosome 14 but alternate between the two compartments [L1 and L3 in (compartment A); L2 and L4 in (compartment B)]. (E) L3 (blue) was consistently closer to L1 (green) than to L2 (red), despite the fact that L2 lies between L1 and L3 in the primary sequence of the genome. This was confirmed visually and by plotting the cumulative distribution. (F) L2 (green) was consistently closer to L4 (red) than to L3 (blue). (**G**) Correlation map of chromosome 14 at a resolution of 100 kb. The PC (eigenvector) correlates with the distribution of genes and with features of open chromatin. (**H**) A 31-Mb window from chromosome 14 is shown; the indicated region (yellow dashes) alternates between the open and the closed compartments in GM06990 (top, eigenvector and heatmap) but is predominantly open in K562 (bottom, eigenvector and heatmap). The change in compartmentalization corresponds to a shift in chromatin state (DNaseI).

of the genome inferred from Hi-C. More generally, a strong correlation was observed between the number of Hi-C reads m_{ij} and the 3D distance between locus i and locus j as measured by FISH [Spearman's $\rho = -0.916$, $P = 0.00003$ (fig. S3)], suggesting that Hi-C read count may serve as a proxy for distance.

Upon close examination of the Hi-C data, we noted that pairs of loci in compartment B showed a consistently higher interaction frequency at a given genomic distance than pairs of loci in compartment A (fig. S4). This suggests that compartment B is more densely packed (*15*). The FISH data are consistent with this observation; loci in compartment B exhibited a stronger tendency for close spatial localization.

To explore whether the two spatial compartments correspond to known features of the genome, we compared the compartments identified in our 1-Mb correlation maps with known genetic and epigenetic features. Compartment A correlates strongly with the presence of genes (Spearman's $\rho = 0.431$, $P < 10^{-137}$), higher expression [via genome-wide mRNA expression, Spearman's $\rho = 0.476$, $P < 10^{-145}$ (fig. S5)], and accessible chromatin [as measured by deoxyribonuclease I (DNAseI) sensitivity, Spearman's $\rho = 0.651$, P negligible] (*16*, *17*). Compartment A also shows enrichment for both activating (H3K36 trimethylation, Spearman's $\rho = 0.601$, $P < 10^{-296}$) and repressive (H3K27 trimethylation, Spearman's $\rho = 0.282$, $P < 10^{-56}$) chromatin marks (*18*). We repeated the above analysis at a resolution of 100 kb (Fig. 3G) and saw that, although the correlation of compartment A with all other genomic and epigenetic features remained strong (Spearman's $\rho > 0.4$, P negligible), the correlation with the sole repressive mark, H3K27 trimethylation, was dramatically attenuated (Spearman's $\rho = 0.046$, $P < 10^{-15}$). On the basis of these results we concluded that compartment A is more closely associated with open, accessible, actively transcribed chromatin.

We repeated our experiment with K562 cells, an erythroleukemia cell line with an aberrant karyotype (*19*). We again observed two compartments; these were similar in composition to those observed in GM06990 cells [Pearson's $r = 0.732$,

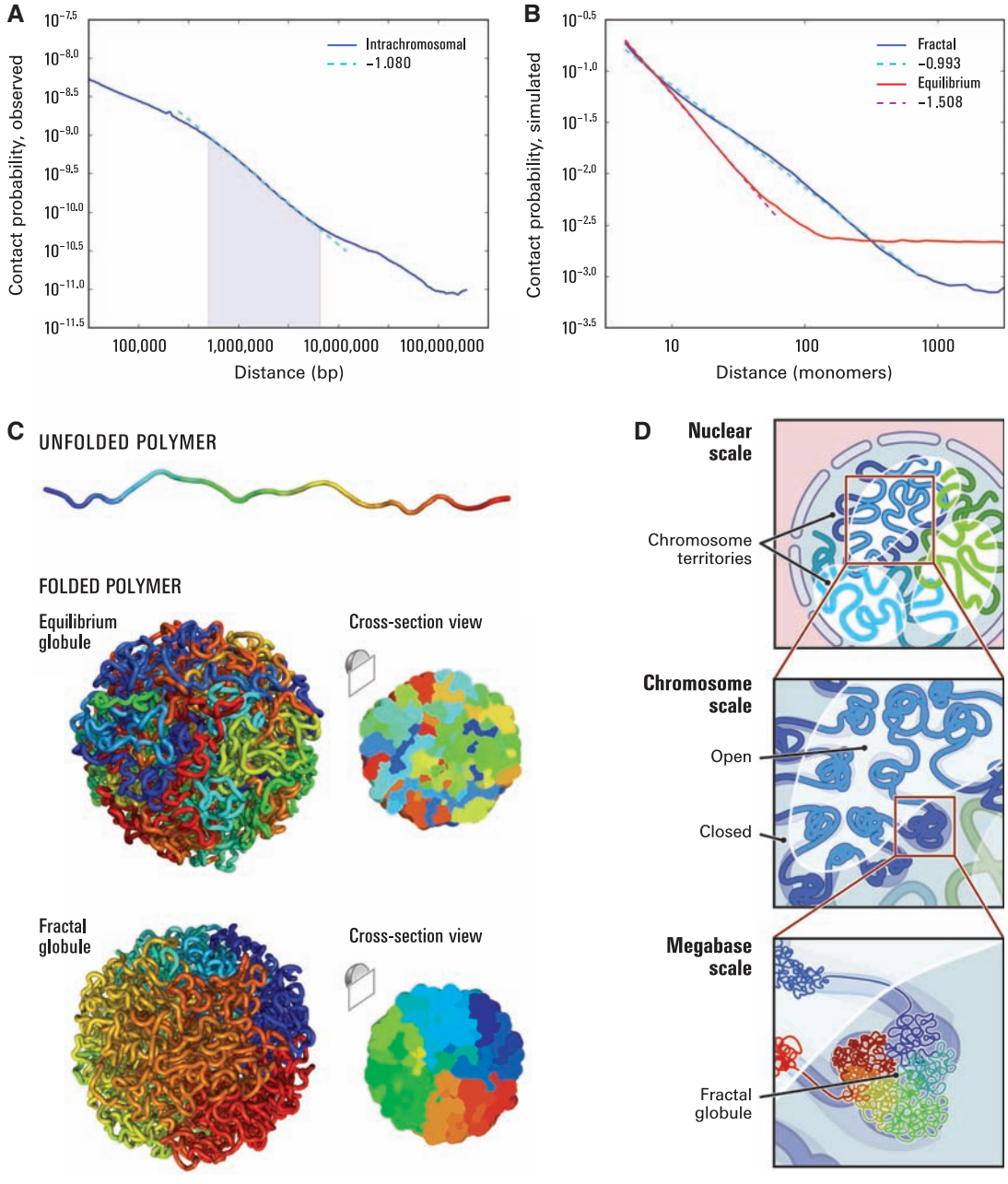


Fig. 4. The local packing of chromatin is consistent with the behavior of a fractal globule. (**A**) Contact probability as a function of genomic distance averaged across the genome (blue) shows a power law scaling between 500 kb and 7 Mb (shaded region) with a slope of −1.08 (fit shown in cyan). (**B**) Simulation results for contact probability as a function of distance (1 monomer ~ 6 nucleosomes ~ 1200 base pairs) (*10*) for equilibrium (red) and fractal (blue) globules. The slope for a fractal globule is very nearly −1 (cyan), confirming our prediction (*10*). The slope for an equilibrium globule is −3/2, matching prior theoretical expectations. The slope for the fractal globule closely resembles the slope we observed in the genome. (**C**) (Top) An unfolded polymer chain, 4000 monomers (4.8 Mb) long. Coloration corresponds to distance from one endpoint, ranging from blue to cyan, green, yellow, orange, and red. (Middle) An equilibrium globule. The structure is highly entangled; loci that are nearby along the contour (similar color) need not be nearby in 3D. (Bottom) A fractal globule. Nearby loci along the contour tend to be nearby in 3D, leading to monochromatic blocks both on the surface and in cross section. The structure lacks knots. (**D**) Genome architecture at three scales. (Top) Two compartments, corresponding to open and closed chromatin, spatially partition the genome. Chromosomes (blue, cyan, green) occupy distinct territories. (Middle) Individual chromosomes weave back and forth between the open and closed chromatin compartments. (Bottom) At the scale of single megabases, the chromosome consists of a series of fractal globules.

P negligible (fig. S6)] and showed strong correlation with open and closed chromatin states as indicated by DNAseI sensitivity (Spearman's $\rho = 0.455$, $P < 10^{-154}$).

The compartment patterns in K562 and GM06990 are similar, but there are many loci in the open compartment in one cell type and the closed compartment in the other (Fig. 3H). Examining these discordant loci on karyotypically normal chromosomes in K562 (*19*), we observed a strong correlation between the compartment pattern in a cell type and chromatin accessibility in that same cell type (GM06990, Spearman's $\rho = 0.384$, $P = 0.012$; K562, Spearman's $\rho = 0.366$, $P = 0.017$). Thus, even in a highly rearranged genome, spatial compartmentalization correlates strongly with chromatin state.

Our results demonstrate that open and closed chromatin domains throughout the genome occupy different spatial compartments in the nucleus. These findings expand on studies of individual loci that have observed particular instances of such interactions, both between distantly located active genes and between distantly located inactive genes (*8*, *20*–*24*).

Lastly, we sought to explore chromatin structure within compartments. We closely examined the average behavior of intrachromosomal contact probability as a function of genomic distance, calculating the genome-wide distribution $I(s)$. When plotted on log-log axes, $I(s)$ exhibits a prominent power law scaling between ~500 kb and ~7 Mb, where contact probability scales as s^{-1} (Fig. 4A). This range corresponds to the known size of open and closed chromatin domains.

Power-law dependencies can arise from polymer-like behavior (*25*). Various authors have proposed that chromosomal regions can be modeled as an "equilibrium globule": a compact, densely knotted configuration originally used to describe a polymer in a poor solvent at equilibrium (*26*, *27*). [Historically, this specific model has often been referred to simply as a "globule"; some authors have used the term "equilibrium globule" to distinguish it from other globular states (see below).] Grosberg *et al*. proposed an alternative model, theorizing that polymers, including interphase DNA, can self-organize into a long-lived, non-equilibrium conformation that they described as a "fractal globule" (*28*, *29*). This highly compact state is formed by an unentangled polymer when it crumples into a series of small globules in a "beads-on-a-string" configuration. These beads serve as monomers in subsequent rounds of spontaneous crumpling until only a single globule-of-globules-of-globules remains. The resulting structure resembles a Peano curve, a continuous fractal trajectory that densely fills 3D space without crossing itself (*30*). Fractal globules are an attractive structure for chromatin segments because they lack knots (*31*) and would facilitate unfolding and refolding, for example, during gene activation, gene repression, or the cell cycle. In a fractal globule, contiguous regions of the genome tend to form spatial sectors whose size corresponds to the length of the original region (Fig. 4C). In contrast, an equilibrium globule is highly knotted and lacks such sectors; instead, linear and spatial positions are largely decorrelated after, at most, a few megabases (Fig. 4C). The fractal globule has not previously been observed (*29*, *31*).

The equilibrium globule and fractal globule models make very different predictions concerning the scaling of contact probability with genomic distance s. The equilibrium globule model predicts that contact probability will scale as $s^{-3/2}$, which we do not observe in our data. We analytically derived the contact probability for a fractal globule and found that it decays as s^{-1} (*10*); this corresponds closely with the prominent scaling we observed ($s^{-1.08}$).

The equilibrium and fractal globule models also make differing predictions about the 3D distance between pairs of loci ($s^{1/2}$ for an equilibrium globule, $s^{1/3}$ for a fractal globule). Although 3D distance is not directly measured by Hi-C, we note that a recent paper using 3D-FISH reported an $s^{1/3}$ scaling for genomic distances between 500 kb and 2 Mb (*27*).

We used Monte Carlo simulations to construct ensembles of fractal globules and equilibrium globules (500 each). The properties of the ensembles matched the theoretically derived scalings for contact probability (for fractal globules, s^{-1}, and for equilibrium globules, $s^{-3/2}$) and 3D distance (for fractal globules $s^{1/3}$, for equilibrium globules $s^{1/2}$). These simulations also illustrated the lack of entanglements [measured by using the knot-theoretic Alexander polynomial (*10*, *32*)] and the formation of spatial sectors within a fractal globule (Fig. 4B).

We conclude that, at the scale of several megabases, the data are consistent with a fractal globule model for chromatin organization. Of course, we cannot rule out the possibility that other forms of regular organization might lead to similar findings.

We focused here on interactions at relatively large scales. Hi-C can also be used to construct comprehensive, genome-wide interaction maps at finer scales by increasing the number of reads. This should enable the mapping of specific long-range interactions between enhancers, silencers, and insulators (*33*–*35*). To increase the resolution by a factor of n, one must increase the number of reads by a factor of n^2. As the cost of sequencing falls, detecting finer interactions should become increasingly feasible. In addition, one can focus on subsets of the genome by using chromatin immunoprecipitation or hybrid capture (*36*, *37*).

References and Notes

1. T. Cremer, C. Cremer, *Nat. Rev. Genet.* **2**, 292 (2001).
2. T. Sexton, H. Schober, P. Fraser, S. M. Gasser, *Nat. Struct. Mol. Biol.* **14**, 1049 (2007).
3. J. Dekker, *Science* **319**, 1793 (2008).
4. T. Misteli, *Cell* **128**, 787 (2007).
5. S. T. Kosak, M. Groudine, *Genes Dev.* **18**, 1371 (2004).
6. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* **295**, 1306 (2002).
7. Z. Zhao *et al*., *Nat. Genet.* **38**, 1341 (2006).
8. M. Simonis *et al*., *Nat. Genet.* **38**, 1348 (2006).
9. J. Dostie *et al*., *Genome Res.* **16**, 1299 (2006).
10. Materials and methods are available as supporting material on *Science* Online.
11. H. Yokota, G. van den Engh, J. E. Hearst, R. K. Sachs, B. J. Trask, *J. Cell Biol.* **130**, 1239 (1995).
12. S. Boyle *et al*., *Hum. Mol. Genet.* **10**, 211 (2001).
13. H. Tanabe, F. A. Habermann, I. Solovei, M. Cremer, T. Cremer, *Mutat. Res.* **504**, 37 (2002).
14. J. A. Croft *et al*., *J. Cell Biol.* **145**, 1119 (1999).
15. J. Dekker, *J. Biol. Chem.* **283**, 34532 (2008).
16. P. J. Sabo *et al*., *Nat. Methods* **3**, 511 (2006).
17. J. R. Hesselberth *et al*., *Nat. Methods* **6**, 283 (2009).
18. T. S. Mikkelsen *et al*., *Nature* **448**, 553 (2007).
19. S. Naumann, D. Reutzel, M. Speicher, H. J. Decker, *Leuk. Res.* **25**, 313 (2001).
20. C. S. Osborne *et al*., *Nat. Genet.* **36**, 1065 (2004).
21. J. M. Brown *et al*., *J. Cell Biol.* **182**, 1083 (2008).
22. A. F. Dernburg *et al*., *Cell* **85**, 745 (1996).
23. L. S. Shopland *et al*., *J. Cell Biol.* **174**, 27 (2006).
24. P. Fraser, W. Bickmore, *Nature* **447**, 413 (2007).
25. P. G. d. Gennes, *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, NY, 1979).
26. C. Munkel, J. Langowski, *Phys. Rev. E* **57**, 5888 (1998).
27. J. Mateos-Langerak *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3812 (2009).
28. A. Y. Grosberg, S. K. Nechaev, E. I. Shakhnovich, *J. Phys. France* **49**, 2095 (1988).
29. A. Grosberg, Y. Rabin, S. Havlin, A. Neer, *Europhys. Lett.* **23**, 373 (1993).
30. B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, ed. 2, 1983).
31. O. A. Vasilyev, S. K. Nechaev, *Theor. Math. Phys.* **134**, 142 (2003).
32. G. Kolesov, P. Virnau, M. Kardar, L. A. Mirny, *Nucleic Acids Res.* **35**, W425 (2007).
33. E. M. Blackwood, J. T. Kadonaga, *Science* **281**, 60 (1998).
34. A. C. Bell, A. G. West, G. Felsenfeld, *Science* **291**, 447 (2001).
35. J. E. Phillips, V. G. Corces, *Cell* **137**, 1194 (2009).
36. M. H. Kuo, C. D. Allis, *Methods* **19**, 425 (1999).
37. A. Gnirke *et al*., *Nat. Biotechnol.* **27**, 182 (2009).

38. Supported by a Fannie and John Hertz Foundation graduate fellowship, a National Defense Science and Engineering graduate fellowship, an NSF graduate fellowship, the National Space Biomedical Research Institute, and grant no. T32 HG002295 from the National Human Genome Research Institute (NHGRI) (E.L.); a fellowship from the American Society of Hematology (T.R.); award no. R01HL06544 from the National Heart, Lung, and Blood Institute and R37DK44746 from the National Institute of Diabetes and Digestive and Kidney Diseases (M.G.); NIH grant U54HG004592 (J.S.); i2b2 (Informatics for Integrating Biology and the Bedside), the NIH-supported Center for Biomedical Computing at Brigham and Women's Hospital (L.A.M.), grant no. HG003143 from the NHGRI, and a Keck Foundation distinguished young scholar award (J.D.). We thank J. Goldy, K. Lee, S. Vong, and M. Weaver for assistance with DNaseI experiments; A. Kosmrlj for discussions and code; A. P. Aiden, X. R. Bao, M. Brenner, D. Galas, W. Gosper, A. Jaffer, A. Melnikov, A. Miele, G. Giannoukos, C. Nusbaum, A. J. M. Walhout, L. Wood, and K. Zeldovich for discussions; and L. Gaffney and B. Wong for help with visualization. We also acknowledge the ENCODE chromatin group at Broad Institute and Massachusetts General Hospital. Hi-C sequence data has been deposited at the GEO database (www.ncbi.nlm.nih.gov/geo/), accession no. GSE18199. Expression data are also available at GEO, accession no. GSE18350. Chromatin immunoprecipitation sequence (ChIP-Seq) data and DNAseI sensitivity data are available at the University of California Santa Cruz (UCSC) browser (http://genome.ucsc.edu/). Additional visualizations are available at http://hic.umassmed.edu. A provisional patent on the Hi-C method (no. 61/100,151) is under review.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5950/289/DC1
Materials and Methods
Figs. S1 to S32

1 September 2009; accepted 18 September 2009
10.1126/science.1181369

Science

Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander and Job Dekker

Science **326** (5950), 289-293.
DOI: 10.1126/science.1181369

Chromosomal Mapping

The conformation of the genome in the nucleus and contacts between both proximal and distal loci influence gene expression. In order to map genomic contacts, **Lieberman-Aiden *et al.*** (p. 289, see the cover) developed a technique to allow the detection of all interactions between genomic loci in the eukaryotic nucleus followed by deep sequencing. This technology was used to map the organization of the human genome and to examine the spatial proximity of chromosomal loci at one megabase resolution. The map suggests that the genome is partitioned into two spatial compartments that are related to local chromatin state and whose remodeling correlates with changes in the chromatin state.

www.sciencemag.org/cgi/content/full/326/5950/289/DC1

Supporting Online Material for

Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander,* Job Dekker*

*To whom correspondence should be addressed. E-mail: lander@broadinstitute.org (E.S.L.); job.dekker@umassmed.edu (J.D.)

Published 9 October 2009, *Science* **326**, 289 (2009)
DOI: 10.1126/science.1181369

This PDF file includes:

Comprehensive mapping of long-range interactions reveals folding principles of the human genome.

Lieberman-Aiden, Van Berkum, et al.

Supplemental Online Materials

Contents:

I. Materials and Methods

Hi-C method

Crosslinking of cells. Human cell line GM06990, an EBV-transformed lymphoblastoid cell line (Coriell, Camden, NJ), was cultured in RPMI1640, 15% fetal calf serum, 1% penicillin-streptomycin, and 2mM L-glutamine. Human erythroleukemia cell line K562 (ATCC, Manassas, VA) was cultured in DMEM, 10% fetal calf serum, 1% penicillinstreptomycin, and 2mM L-glutamine. One hundred million cells were spun down and resuspended in 45 ml fresh medium. Cells were fixed by adding 1.25 ml 37% formaldehyde and incubating for 10 minutes at room temperature (RT). The reaction was stopped by adding 2.5 ml 2.5 M glycine. The cell suspension was incubated for 5 minutes at RT, followed by 15 minutes on ice. The crosslinked cell suspension was split into 4 equal parts and centrifuged at 1500 rpm for 10 minutes. The supernatant was discarded and the cell pellets were stored at -80C.

Cell lysis and chromatin digestion. For cell lysis, 550 µl lysis buffer (500 µl 10 mM Tris-HCl pH8.0, 10 mM NaCl, 0.2% Ige cal CA630; 50 µl protease inhibitors (Sigma, St. Louis, MO) were added to one batch of cells (~ 25 million cells). Cells were incubated on ice for at least 15 minutes. Next, cells were lysed with a Dounce homogenizer by moving the pestle A up and down 10 times, incubating on ice for one minute followed by 10 more strokes with the pestle. The suspension was spun down for 5 minutes at 5000 rpm at RT. The supernatant was discarded and the pellet was washed

twice with 500 µl icecold 1x NEBuffer 2 (NEB, Ipswich, MA). The pellet was then resuspended in 1x NEBuffer 2 in a total volume of 250 µl and split into five 50 µl aliquots. Next, 312 µl 1x NEBuffer 2 was added per tube. To remove the proteins that were not directly crosslinked to the DNA, 38 µl 1% SDS was added per tube and the mixture was resuspended and incubated at 65°C for 10 minutes exactly. Tubes were put on ice and 44 µl 10% Triton X-100 was added and mixed carefully avoiding bubbles to quench the SDS. Chromatin was subsequently digested overnight at 37°C by adding 400 Units HindIII (NEB).

Marking of DNA ends and blunt-end ligation. Five tubes with digested chromatin were put on ice and tube 1 was kept separate and served as a 3C control. To fill in and mark the DNA ends, 1.5 µl 10 mM dATP, 1.5 µl 10 mM dGTP, 1.5 µl 10 mM dTTP, 37.5 µl 0.4 mM biotin-14-dCTP (Invitrogen, Carlsbad, CA) and 10 µl 5U/µl Klenow (NEB) were added to tubes 2-5. The mixtures were incubated at 37°C for 45 minutes and subsequently placed on ice. Enzymes were inactivated by adding 86 µl 10% SDS to tubes 1-5 and incubating all tubes at 65°C for 30 minutes. Tubes were placed on ice immediately. Five 15 ml tubes were prepared, each containing 7.61 ml ligation mix (745 µl 10% Triton X-100, 745 µl 10x ligation buffer (500 mM Tris-HCl pH7.5, 100 mM MgCl2, 100 mM DTT), 80 µl 10 mg/ml BSA, 80 µl 100 mM ATP and 5.96 ml water). Each digested chromatin mixture was transferred to a corresponding 15 ml tube. For normal 3C ligation 10 µl 1U/µl T4 DNA ligase (Invitrogen) was added to tube 1. For blunt-end ligation 50 µl 1U/µl T4 DNA ligase was added to tubes 2-5. All 5 tubes were incubated at 16°C for 4 hours.

DNA purification. To reverse crosslinks and to degrade protein, 50 µl 10 mg/ml proteinase K was added per tube and the tubes were incubated overnight at 65°C. The next day an additional 50 µl 10 mg/ml proteinase K was added per tube and the incubation was continued at 65°C for another 2 hours. Reaction mixtures were cooled to RT and transferred to five 50 ml conical tubes. The DNA was extracted by adding 10 ml phenol pH8.0, vortexing for 2 minutes and spinning for 10 minutes at 3,500 rpm. The supernatants were transferred to five new 50 ml conical tubes. Another DNA extraction was performed with 10 ml phenol pH8.0:chloroform (1:1). After vortexing and centrifugation for 10 minutes at 3,500 rpm, the supernatants were transferred to five 35 ml centrifugation tubes. The volume was brought to 10 ml per tube with 10 mM Tris pH8.0, 1 mM EDTA (1x TE). To precipitate the DNA, 1 ml 3M Na-acetate was added per tube and mixed well before adding 25 ml ice-cold 100% ethanol. Tubes were inverted several times to properly mix the contents and were incubated at -80°C for at least one hour. Next, the tubes were spun at 4°C for 20 minutes at 10,000xg. The supernatant was discarded and each DNA pellet was dissolved in 450 µl 1x TE and transferred to a 1.7 ml centrifuge tube. The DNA was extracted twice by adding 500 µl phenol pH8.0:chloroform (1:1), vortexing for 30 seconds and spinning at 14,000 rpm for 5 minutes at RT. After the second extraction, the supernatants (each ~400 µl) were transferred to five new 1.7 ml tubes and 40 µl 3M Na-acetate was added per tube and mixed. Next, 1 ml 100% ethanol per tube was added. After inverting the tubes several times, the tubes were incubated at - 80°C for at least 30 minutes. Tubes were spun at 18,000xg for 20 minutes at 4°C. The supernatant was discarded and the pellets were

washed once with 500 µl 70% ethanol. After centrifuging at 14,000 rpm for 5 minutes, the supernatant was discarded and the pellets were air-dried briefly prior to resuspending in 25 µl 1x TE. To degrade any purified RNA, 1 µl 1 mg/ml RNAse A was added per tube and incubated at 37°C for 15 minutes. The Hi-C contents of tubes 2-5 were pooled and tube 1 was kept separate as the 3C control.

Quality control HiC libraries. Both 3C and Hi-C libraries were checked for quality and quantified by running an aliquot on a 0.8% agarose gel. To confirm that the ligation process worked as intended, we used the fact that successful fill-in and ligation of HindIII sites (AAGCTT) should create sites for the restriction enzyme NheI (GCTAGC). We used PCR to amplify a ligation product formed from two nearby restriction fragments and determined that 70% of amplicons were cut only by NheI (Fig. S7). Sequences of the primers used for checking libraries are:

HindIII -1 GTTCATCTTGCTGCCAGAAATGCCGAGCCTG
HindIII-2 ATCCCAGCTGTCTGTAGCTTTAGAAAGTGGG
NcoI-1 ACCTGTTGTTTAATGAAGGGGCTCAGAAGC
NcoI-2 GTTTGCAGTGTGCTGTGCAGCATGTGTGTA

Removal of biotin from unligated ends. Biotin-14-dCTP at non-ligated DNA ends was removed with the exonuclease activity of T4 DNA polymerase. To this end 5 µg of Hi-C library was added to 1 µl 10 mg/ml BSA, 10 µl 10x NEBuffer 2, 1 µl 10 mM dATP, 1 µl 10 mM dGTP and 5 Units T4 DNA polymerase (NEB) in a total volume of 100 µl and incubated at 12°C for 2 hours. If possible, multiple 5 µg reactions were performed. Reactions were stopped by adding 2 µl 0.5 M EDTA pH8.0. DNA was subsequently purified with one phenol pH8.0:chloroform (1:1) extraction followed by ethanol precipitation. DNA pellets were resuspended and pooled in a total of 100 µl water.

Shearing and size selection. The DNA was sheared to a size of 300-500 basepairs with a Covaris S2 instrument (Covaris, Woburn, MA), Duty cycle 5, Intensity 5, Cycles/burst 200, time 60 secs for 4 cycles. The DNA ends were repaired by adding 14 µl 10x ligation buffer (NEB), 14 µl 2.5 mM dNTP mix, 5 µl T4 DNA polymerase (NEB), 5 µl T4 polynucleotide kinase (NEB), 1 µl Klenow DNA polymerase (NEB) and 1 µl water and was incubated at 20oC for 30 minutes followed by purification of the DNA with a Qiagen MinElute column (Qiagen, Valencia, CA). The DNA was eluted with 2x 15 µl 10 mM Tris pH8.0, 0.1 mM EDTA. Next, an 'A' was added to the 3' ends of the end repaired DNA by addition of 5 µl 10x NEBuffer2, 10 µl 1 mM dATP, 2 µl water and 3 µl Klenow (exo-) (NEB) . The reaction was incubated at 37°C for 30 minutes followed by 65°C for 20 minutes to inactivate Klenow (exo-). The reactions were cooled on ice and the volume was reduced to 20 µl with a speedvac. DNA was electrophoresed on a 1.5% agarose gel in 1X TAE for 3.5 hours at 80 V. The gel was stained with SYBR green (Lonza Walkersville, Basel, Switzerland), visualized on a DarkReader (Clare Chemical, Dolores, CO) and DNA between 300 and 500 base pairs was excised and purified with a gel extraction kit (Qiagen). The gel slices were solubilized with three volumes of Buffer QG (Qiagen) at RT and the DNA purified with QIAquick spin columns (Qiagen). The DNA was eluted twice with 50 µl 10mM Tris pH 8.0, 0.1 mM EDTA and the final volume was made up to 300 µl with 10 mM Tris pH 8.0, 0.1 mM EDTA. The DNA concentration was

measured with the Quant-iT assay (Invitrogen).

Biotin pull-down and Paired End sequencing. All subsequent steps were performed in DNA LoBind tubes (Eppendorf, Westbury, NY). The biotin tagged Hi-C DNA was bound to Dynabeads MyOne Streptavin C1 Beads (Invitrogen) as follows. Sixty µl of resuspended Streptavidin beads were washed twice with 400 µl Tween Wash Buffer (TWB) (5 mM Tris-HCl pH8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween) by incubating for 3 minutes at RT with rotation. After this and for all subsequent incubations or washes of Streptavidin beads, the beads were reclaimed by holding against a magnetic particle concentrator (Invitrogen) for 1 minute and the supernatant was removed. These reclaimed beads were then resuspended in 300 µl 2x Binding Buffer (BB) (10 mM Tris-HCl pH8.0, 1 mM EDTA, 2 M NaCl) and combined with 300 µl Hi-C DNA. The mixture was incubated at RT for 15 minutes with rotation. The supernatant was removed and the DNA bound Streptavidin beads were resuspended in 400 µl 1x BB and transferred to a new tube. The beads were then resuspended in 100 µl 1x ligation buffer, transferred to a new tube before a final resuspension in 50 µl 1x ligation buffer. Six picomoles of Illumina Paired End adapters (Illumina, San Diego, CA) per µg of Hi-C DNA (measured after Qiagen gel purification) were ligated to the Hi-C DNA for 2 hours at RT in the presence of, 1 mM ATP and 20U T4 DNA Ligase (Ambion, Austin, TX). The ligated Hi-C DNA was isolated by holding against the magnet and was washed with 400 µl of 1x TWB to remove non-ligated Paired End adapters. The beads were resuspended in a further 400 µl 1x TWB and the mixture was transferred to a new tube and the Streptavidin beads were recovered. This wash step was repeated with 200 µl 1x BB, then 200 µl 1x NEBuffer 2 and finally 50 µl 1x NEBuffer 2. The beads were resuspended in 50 µl 1x NEBuffer 2. Next, test PCR reactions were performed to determine the optimal PCR cycles needed to generate enough library for sequencing. Four trial PCR reactions, each containing 0.6 µl Streptavidin bead bound Hi-C library and Illumina PE1.0 and PE2.0 PCR primers (1.5 pmol each) in 10 µl 1x Phusion High Fidelity master mix with HF buffer (NEB), were set up to determine the number of cycles necessary to generate enough PCR product for sequencing. The temperature profile was 30 s at 98°C followed by 9, 12, 15 or 18 cycles of 10 s at 98°C, 30 s at 65°C, 30 s at 72°C and a final 7-minute extension at 72°C. The PCR reactions were run on a 5% polyacrylamide gel, stained with Sybr Green and the optimal cycle number was determined. A large-scale PCR was then set-up with the remainder of the Streptavidin bead bound Hi-C library with the number of PCR cycles determined by the trial PCR. 1% of the large scale PCR product was kept to run on a gel. The PCR product was purified by mixing with 1.8x volume Ampure beads (Beckman Coulter, Fullerton, CA). The mix was held against a magnet to separate the PCR product bound to the Ampure beads and the supernatant was discarded. The Hi- C library bound Ampure beads were washed twice with 1 ml 70% ethanol while the tube remained against the magnet. After air-drying the beads, the DNA was eluted by resuspending the beads in 50 µl of 10 mM Tris pH8.0, 0.1 mM EDTA. The tube was held against a magnet and the supernatant containing the purified PCR products was transferred to a new tube. Next, 1% of the Ampure bead purified PCR product was compared against the 1% aliquot of original PCR product on a 5% polyacrylamide gel. Finally, the Hi-C library was sequenced with Illumina paired end sequencing.

Chromatin immunoprecipitation

ChIP experiments were carried out as described previously (*S1, S2*). Briefly, chromatin from fixed cells was fragmented to a size range of 200–700 bases. Solubilized chromatin was immunoprecipitated with antibody against H3K27me3 (Upstate) or H3K36me3 (Abcam, Cambridge, MA). Antibody–chromatin complexes were pulled-down with protein A-sepharose, washed and then eluted. After cross-link reversal and proteinase K treatment, immunoprecipitated DNA was extracted with phenol-chloroform, ethanol precipitated, and treated with RNase. ChIP DNA was quantified with PicoGreen. Detailed information on sequencing and read alignment are described elsewhere (*S2*)

Mapping of DNAseI sensitivity

Cell lines (GM06990, Coriell and K562, ATCC) were cultured in humidified incubators at 37 °C in the presence of 5% CO2 according to the protocol provided by the source. Isolation of nuclei, DNaseI treatment, purification, and fractionation of small (<500bp) DNaseI double-cleaved fragments was performed, as described (*S3*). End-ligation of sequencing adapters (Illumina) and cycle sequencing (to 27bp) were performed, as described (*S4*). 27bp sequence reads were aligned to the human genome (NCBI build 37, UCSC HG18) with the Eland aligner (Illumina) (allowing 2 mismatches), and only reads mapping to unique genomic positions were utilized in downstream analyses. The density of DNaseI cleavages in a 150bp (i.e., ~nucleosome-size) sliding window (step 20bp, computed 5' to 3' across each chromosome individually) was computed for use in correlation analyses.

Expression analysis

Total RNA was extracted with QIAzol reagent following the miRNeasy kit's procedure (Qiagen), and sample quality was tested on a 2100 Bioanalyzer (Agilent, Palo Alto, CA). For oligonucleotide microarray hybridization, 1.5µg of RNA were labeled, fragmented and hybridized to an Affymetrix Human Genome U133 plus 2.0 Array. After scanning, the expression value for each gene was calculated with RMA (Robust Multi-Array) normalization. The average intensity difference values were normalized across the sample set. Probe sets that were absent in all samples according to Affymetrix flags were removed.

DNA FISH

3D DNA FISH was performed essentially as described (*S5*). BACs (Table S1) were obtained from the BACPAC Resource Center at Children's Hospital Oakland Research Institute in Oakland, CA. About 100 ng of nick translated probes (labeled with DIG, DNP or biotin) and 10ug of Cot-1 DNA were used in each hybridization. Image stacks (Z sections spaced 0.25 Km apart) were captured on an Olympus IX71 microscope (Olympus, Center Valley, PA) with a 100X/1.40 UPLS Apo objective and subsequently deconvolved with Deltavision SoftWorx software (Applied Precision, Issaquah, WA). 3D distance measurements were performed with the MeasurementPro module in Imaris (Bitplane, Saint Paul, MN). Specifically, measurements were taken from and to the perceived centers of each FISH spot. Example images of collapsed stacks were processed in Photoshop CS3 (Adobe, San Jose, CA).

II. Computational Analysis Overview

Read alignment and Heatmap generation.

Each end of the 76bp paired reads was aligned separately against the human hg18 reference sequence with Maq (http://maq.sourceforge.net/) using a mismatch threshold of 150. If both ends aligned successfully, the resulting pair was added to the interaction catalog. We confirmed that these reads tend to align near HindIII restriction sites with the expected orientation (Fig. S8, A and B).To produce heatmaps, the genome was divided into 1Mb loci (later, 100Kb loci) and each interaction was binned according to the location of both ends to produce the matrix M. Very few were identical at both ends, indicating that the effects of PCR bias are minimal. We compared alternative alignment strategies of multiple aligners and parameter settings and verified that no substantive differences were observed. When we randomly permuting one end of the reads, the resulting heatmap was essentially uniform; when we used reads derived from sheared genomic DNA instead of from a Hi-C library, the effects described in the paper all disappeared.

Presence of Chromosome Territories.

The total number of possible interactions at a given genomic distance was computed explicitly for each chromosome and compared to the actual number of interactions at that distance. (The possible number of pairs of genomic positions separated by d on a given chromosome is L_{c-d}, where L_c is the length of the chromosome.) To obtain the interchromosomal averages, the number of observed interactions between loci on a pair of chromosomes was divided by the number of possible interactions between the two chromosomes (the product of the number of loci on each chromosome). When multiple chromosome pairings were being averaged, such as in the computation of $I_n(s)$, the numerators and denominators were summed independently. The genome wide average, $I(s)$, is therefore the result of dividing the total number of interactions at a distance s by the number of possible interactions at distance s summed over all chromosomes. The bins in Fig 2A are linearly distributed.

Proximity of Chromosome Territories.

The expected number of interchromosomal interactions for each chromosome pair i,j was computed by multiplying the fraction of interchromosomal reads containing i with the fraction of interchromosomal reads containing j and multiplying by the total number of interchromosomal reads. The enrichment was computed by taking the actual number of interactions observed between i and j and dividing it by the expected value.

Correlation Analysis.

Intrachromosomal. The expected number of reads between two loci i,j was computed by calculating the distance between the midpoints of the two loci $s(i,j)$. This distance was then used as an argument to the function I(s) to compute the expected number of reads between the pair (Fig. S9). The entries of the observed/expected matrix M* was computed by taking each m_{ij} and dividing by $I(s(i,j))$. The corresponding entry of the correlation matrix was computed cij by taking the observed/expected value for every intrachromosomal locus pair including i (c_{ix})with every interchromosomal locus pair

including j (c_{xj}) and computing the Pearson correlation coefficient between the two resulting vectors. Superior results at low resolution may be obtained with the Spearman correlation coefficient, but the latter is not suitable for analyzing the sparse matrices which arise at higher resolution (100Kb).

Interchromosomal. We normalized for coverage, which does not exert a significant effect on intrachromosomal read counts but does exert a significant effect in the interchromosomal case. This is accomplished analogously to the proximity computation for chromosome territories. The expected number of interactions between each locus pair i,j is computed by multiplying the fraction of reads containing i with the fraction of reads containing j and multiplying by the total number of reads. (See coverage tracks in Fig 3G). The enrichment was computed by taking the actual number of interactions observed between locus i and locus j, m_{ij}, and dividing it by this expected value. The correlations are then computed as in the intrachromosomal case, comparing the enrichment values for all interchromosomal locus pairs involving either i or j but excluding any intrachromosomal locus pairs.

Principal Component Analysis.

Principal component analysis was performed as in (*S6*).

Genomic tracks.

We used UCSC gene annotations combined with DNAseI sensitivity and ChIP-Seq data. The total number of genic bases in a given locus was used in the gene density annotations shown. The DNAseI sensitivity data was generated from GM06990 and K562 cells. Raw DNAse data tracks were downloaded from the ENCODE UCSC browser; values within a given megabase or 100Kb locus was summed to produce the track shown. The ChIP-Seq data was generated using GM12878 cells, a cell line very closely related to GM06990. For ChIP-Seq data, the number of reads in each locus was plotted; see (*S7*).

Expression analysis.

Expression data for GM and K562 cells were collected with Affymetrix HGU133 2 Plus expression arrays (Affymetrix, Santa Clara, USA). Two experiments were performed for each cell type. We averaged expression data for all probes lying fully within each 1 Mb locus, including both experiments on the particular cell type. Probes overlapping the edges of the windows were not included in the analysis. We then grouped windows associated with either Compartment A or Compartment B, and computed the distribution of average expression for windows associated with each element type.

Polymer physics.

We plotted $I(s)$ on log-log axes. The bins in Fig 4A are logarithmically distributed, but the scaling exponents do not change significantly if linear binning is used. The theoretical derivation of the scaling for fractal globules is found below, as are details of the Monte Carlo simulations. Images were rendered with PyMol (http://pymol.sourceforge.net/).

III. Supplemental Derivation

Overview

Here we derive contact probability as a function of distance for fractal globules and, *en passant*, for finite iterations of Peano curves in *d* dimensions.

We illustrate the predictions of the theory with simulations exploring a variety of Peano curves and exhibiting a previously unexplored family of power-law scalings that emerge.

It would be ideal to illustrate the predictions of this derivation in the case of fractal globules as well. However, this is difficult because the fractal globule has not been observed before either in experiment (hence, we lack actual configurations to study) or *in silico* (hence, we lack simulated configurations to study). For this reason, we chose to simulate the condensation of a polymer using Monte Carlo methods in order to create instantiations of the fractal globule configuration for further analysis. Using these configurations, we are able to fully demonstrate the consequences of the forthcoming theoretical analysis. These simulations are described in the forthcoming section **(IV)**; this section focuses on the derivation itself and the straightforward application to Peano curves.

Derivation of Contact Probability Scaling for Fractal Globules and Peano Curves

We are interested in the contact probability $P(x)$ as a function of distance x along a fractal globule. We note that the same argument works well for finite iterations of Peano curves in an arbitrary number of dimensions, since their structure is analogous. As such, the argument here will be stated for d dimensions. Let us define $I_{actual}(x)$ as the number of actual interactions between loci separating by a distance x along the 1D polymer contour, and $I_{possible}(x)$ as the number of pairs of loci separated by a distance x along the polymer contour. Then by definition we have:

$$P_{contact}(x) = \frac{I_{actual}(x)}{I_{possible}(x)}$$

Let us compare the contact probability at two consecutive iterations of the space-filling fractal (see example in Fig S11). At the larger of the scales we have 2^d cubes, each of which contains $\frac{N}{2^d}$ monomers and at the smaller scale we have 2^{2d} cubes, each of which contains $\frac{N}{2^{2d}}$ monomers.

The total actual number of interactions satisfies

$$I_{actual} = (number\ of\ cubes) * (interactions/cube)$$

We get the following value at the large scale:

$$I_{actual}\left(\frac{N}{2^d}\right) \approx 2^d f\left(\frac{N}{2^d}\right)$$

where *f* is a function governing the number of local interactions per cube (blob) which we will discuss further below. At the small scale we obtain:

$$I_{actual}\left(\frac{N}{2^{2d}}\right) \approx 2^{2d} f\left(\frac{N}{2^{2d}}\right)$$

The number of possible interactions at the large scale is simply $\sim N^2$. At a smaller scale, the number of interactions is the product of the number of possible interactions within a cube (blob) times the number of such cubes 2^d. We get the following values at the large and small scale:

$$I_{possible}\left(\frac{N}{2^d}\right) \approx N^2$$

$$I_{possible}\left(\frac{N}{2^{2d}}\right) \approx 2^d \left(\frac{N}{2^d}\right)^2 = \frac{N^2}{2^d}$$

Combining $I_{actual}(x)$ and $I_{possible}(x)$, we obtain:

$$P_{contact}\left(\frac{N}{2^d}\right) = \frac{2^d}{N^2} f\left(\frac{N}{2^d}\right)$$

and

$$P_{contact}\left(\frac{N}{2^{2d}}\right) = \frac{2^{3d}}{N^2} f\left(\frac{N}{2^{2d}}\right)$$

Thus we have:

$$\frac{P_{contact}\left(\frac{N}{2^{2d}}\right)}{P_{contact}\left(\frac{N}{2^d}\right)} = 2^{2d} \frac{f\left(\frac{N}{2^{2d}}\right)}{f\left(\frac{N}{2^d}\right)}$$

Smooth and Interdigitating Cases

There are two cases to address: where the globules have smooth surfaces where interactions occur along the surface (such as in the Hilbert Curve), or where two globules interpenetrate as they meet, and interaction density is proportional to volume (such as DNA and chromatin (*S8*); see Fig S12). For smooth globules, *f(x)*, the local density of interactions, is proportional to the surface area in d - 1 dimensions, and thus scales with $x^{\frac{d-1}{d}}$. Thus we obtain:

$$\frac{P_{contact}\left(\frac{N}{2^{2d}}\right)}{P_{contact}\left(\frac{N}{2^{d}}\right)} = 2^{2d}\frac{c^{\frac{d-1}{d}}\frac{N^{\frac{d-1}{d}}}{2^{\frac{2d(d-1)}{d}}}}{c^{\frac{d-1}{d}}\frac{N^{\frac{d-1}{d}}}{2^{\frac{d(d-1)}{d}}}} = 2^{2d}\frac{2^{\frac{d(d-1)}{d}}}{2^{\frac{2d(d-1)}{d}}} = 2^{2d}2^{1-d} = 2^{d+1}$$

For interdigitating globules, *f(x)*, the local density of interactions, is proportional to the volume, and thus scales with *x*. Thus we obtain:

$$\frac{P_{contact}\left(\frac{N}{2^{2d}}\right)}{P_{contact}\left(\frac{N}{2^{d}}\right)} = 2^{2d}\frac{c\frac{N}{2^{2d}}}{c\frac{N}{2^{d}}} = 2^{d}$$

In both these cases we find that P exhibits scalefree behavior and is of the form kx^{α}. In general if we have $P(x) = kx^{\alpha}$ and $\frac{P\left(\frac{x}{2^d}\right)}{P(x)} = \beta$ then: $\frac{\left(\frac{x}{2^d}\right)^{\alpha}}{x^{\alpha}} = \beta$ yielding $\alpha = \frac{-\log_2 \beta}{d}$.

In summary, we find that $P_{contact}(x) = kx^{\alpha}$, where α is given by

$$\alpha_{smooth} = -(1+\tfrac{1}{d}) \text{ and } \alpha_{interdigitated} = -1.$$

The smooth case may be illustrated *in silico* using Peano curves (See Fig S13-S21).

The behavior of interphase DNA reflects the interdigitated case. Although this case is highly analogous - in terms of its hierarchical fractal structure - to the case of smooth-boundaried Peano curves, the resulting contact probability is, as we have seen, different. As such the interdigitated scaling cannot easily be illustrated using Peano curves. In the next section, we will explicitly simulate the fractal globule in order to illustrate the predicted scaling.

Appendix. Radius of Gyration of a Peano Curve

It is also worth noting that the radius of gyration of a finite iteration of a Peano curve of length *N* in *d* dimensions scales as $N^{1/d}$; in particular in three dimensions it is $N^{1/3}$, exactly like the fractal globule. The requisite sum is:

$$R_g = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\boldsymbol{r}_k - \boldsymbol{r}_{mean})^2}$$

where $\boldsymbol{r}_i$ is the position of the i^{th} position on the curve. This may be readily calculated by choosing coordinates such that $\boldsymbol{r}_{mean} = 0$, sending the curve to the limiting (Peano) iteration and integrating over the resulting continuous volume. For instance, in the three dimensional case, we obtain:

$$R_g = \sqrt{\frac{8}{N} \int_0^{\frac{\sqrt[3]{N}}{2}} \int_0^{\frac{\sqrt[3]{N}}{2}} \int_0^{\frac{\sqrt[3]{N}}{2}} x^2 + y^2 + z^2 \, dx \, dy \, dz} = N^{1/3}$$

To see the result in *d* dimensions, it is sufficient to approximate the region covered by the fractal globule as spherical, and perform the integration in polar coordinates.

IV. Supplemental Monte Carlo Analysis Description

We modeled polymer conformations that have statistical properties similar to those observed in chromatin at megabase length-scales. We considered two possible models of the polymer packing: the equilibrium globule and the fractal globule. Monte Carlo simulations were used to construct large ensembles of representative conformations for both models. Conformations in both ensembles have densities comparable to those of the interphase chromatin.

Among the many properties of these globules which we will explore are the scaling exponents for contact probability as a function of distance along the polymer contour. We will illustrate the -3/2 scaling known to be characteristic of an equilibrium globule, and the -1 scaling which we predicted in the previous section for the fractal globule.

Methods

The chromatin fiber was modeled by a polymer chain of N=4000 freely-joined spherical monomers connected by hard bonds. The distance between the centers of consecutive monomers is equal to their diameter, such that the chain is continuous (see Fig S23). These spheres thus define an excluded volume. The presence of excluded volume is important not only for taking into account steric interactions between the monomers, but also because it suppresses nonphysical, topology-violating moves where one fragment of the chain goes across the other one. We will occasionally turn off this excluded volume, in which case we refer to the chains as *phantom*. The excluded volume is the only form of interaction that we use in our simulations.

The dynamics of the polymer chain are simulated by a standard Metropolis Monte Carlo procedure that involves the following moves (Fig. S22) (*S9-S11*):

0. displacement of the terminal monomers;
1. random rotation of monomer i around axis connecting the i-1^{st} and i+1^{st} monomers;
2. generating a random conformation of 3 consecutive monomers, while keeping all the bond lengths constant.
3. rotating a fragment of the chain between a monomer i and one of the termini by a random angle.

The latter move has been used only for initial compaction of the chain, but was eliminated later to avoid knotting of the chain through topologically impossible moves. This is particularly important when simulating the crumpling of a fractal globule, which is governed primarily by topological factors. Moves that lead to collisions between the monomers are rejected, except in the phantom chain.

To obtain conformations of sufficiently high density, the polymer was confined into a spherical cage of radius R_0 and modeled by an exponentially increasing external potential:

$U(r) = e^{(r - R_0)\sigma}$

Equilibrium Globule

The *equilibrium globule* is the macroscopic state of a polymer reached after it has

collapsed in the presence of a poor solvent (*i.e.*, in the presence of attractive interactions between the monomers) or confined into a spherical cage (*S12*). We used the latter method to obtain our ensemble of equilibrium conformations. The protocol consists of two stages: (1) confinement and equilibration of the phantom chain in a small cage (*i.e.*, with excluded volume off); (2) equilibration of the non-phantom chain in a larger cage reflecting a realistic interphase volume. The initial phantom stage is essential for efficiency as it allows the chain to obtain entangled (knotted) conformations of the polymer that are part of the equilibrium ensemble but hard to achieve by equilibration of the confined non-phantom chain. Specifically, the first stage consists of 2500xN steps of gradual polymer confinement and 4000xN steps of equilibration in the cage of $R0 = 16$ (in the units of the bond length). The second stage involves 1500xN steps during which the excluded volume is gradually reintroduced. During this stage, the natural expansion of the chain in response to the presence of excluded volume is counteracted by compression into a cage of $R_0 = 11$, such that the radius of gyration of the chain stays approximately constant. Finally the chain is equilibrated for 1000xN further steps until the polymer density in the cage is uniform. We verify statistical properties of the conformations we obtained by comparing them with theory, and with the reported properties of the conformations obtained by full enumeration on a cubic lattice (*S13*).

Fractal Globule

The *fractal globule* (or *crumpled globule*) is a transient state of a collapsed or confined polymer. It has been suggested that this state should be very long-lived state due to the topological constraints which prevent rapid knotting (*S14*). Over a long period of time the fractal globule gradually transforms into the equilibrium globule through the reptation of the polymer ends. (It is possible that the genome suppresses this process via anchoring of telomeres or gelation.) To obtain conformations corresponding to the fractal globule we rapidly crumple the protein by adiabatically compressing spherical cage. The simulations begin with 3150xN steps in which the polymer is confined to a cage modeled by the external potential $U(r) = e^{(r - R0)\sigma}$. This cage "chases" the polymer since at every step we dynamically set $R_0 = 0.7R_{max}$ and $\sigma = R_0/6$, (R_{max} is the distance from the center of mass to the most remote monomer). Note that the "tail wagging" move (Move #3) is turned off after 175xN steps to avoid polymer knotting. After 3150xN steps the cage is set to a fixed radius $R_0 = 11$, $\sigma = 1.1$. In a second stage, we allow the confined polymer to settle for another 3850xN steps, enabling uniform polymer density in the cage to be obtained. Note that the time provided in this step is far too short to allow the chain to reach the equilibrium globular state. The scaling properties of the resulting conformations are very close to the properties of the fractal (crumpled) globule suggested by earlier theoretical work (*S14*) and by our own calculations (See above).

Results

The two globules have dramatically different conformations. Statistical properties of the equilibrium and fractal globules we obtained are shown in Fig S23; examples of the globules themselves are shown in Fig. S24

Consider the mean end-to-end distance *R(s)* for a fragment of contour length *s*. In the equilibrium globule, theory suggests that the chains traveling within the globule before touching the confining walls behave like Gaussian chains (*i.e.,* random walks)

with $R(s) \sim s^{1/2}$. This is the scaling we observed in our simulated equilibrium globules. At larger s ($s > R_0^2 \approx 100$) the end-to- end distance saturates due to perfect mixing of the monomers inside the globule (Fig S24).

The fractal globule, in contrast, shows a different scaling. According to the theory, the fractal globule consists of the hierarchy of blobs that do not interpenetrate. Each blobs constitutes a well-packed fractal globule itself. The volume of a fully packed globule (with a uniform polymer density) should be comparable to that of the total volume of the polymer monomers, *i.e.* $V \sim s$, suggesting that $R^3 \sim s$ and $R(s) \sim s^{1/3}$. This scaling in fact is an upper limit, since for any particular conformation a polymer does not fully fill the volume of a blob. In agreement with this argument, simulations show that $R(s) \sim s^{0.29\text{-}0.30}$ fits the data best, while $R(s) \sim s^{1/3}$ provides an upper limit that is closely approached (Fig S23).

Comparison of the scaling in the equilibrium and fractal globules clearly demonstrates the marked differences between the two configurations.

Similarly, the two models exhibit very different probability of a contact (loop) $P(s)$ between regions separated by distance s along the chain. The equilibrium globule demonstrates $P(s) \sim s^{-3/2}$ for small s, corresponding to the results for a Gaussian chain, and a uniform contact density for larger $s > R_0^2 \approx 100$. The fractal globule demonstrates a very different scaling of $P(s) \sim s^{-1}$, as predicted by the theory (see above), and in good agreement with the intra-chromosomal contact probability we obtained using Hi-C. This latter scaling makes the fractal globule a good statistical model for arrangement of interphase chromatin at the megabase scale.

Topological state of equilibrium and fractal globules

The two states are expected to have very different topologies. The fractal globule is the state of a collapsed polymer that lacks entanglements, i.e., it should have a largely unknotted conformation. The equilibrium globule, in contrast, has been shown to be highly knotted; only an exponentially small fraction of equilibrium globules are unknotted (*S15, S16*). These predictions were confirmed by our observations (Fig. S25,S26). Here we quantify the change in topological state using knot-theoretic analysis, and then illustrate its functional consequences on both global and local decondensation.

Knot Invariants.

We tested whether the ensembles of fractal and equilibrium globules we obtained by Monte Carlo simulations were consistent with theoretical expectations. Several groups have used the Alexander and Jones *(S15)* polynomials to detect knots in collapsed polymers (*S16*) and protein structures (*S17, S18*). We used a tool we previously developed (*S16*) to characterize the topological state of the conformations we obtained. (A version of this tool is available at http://knots.mit.edu.) Since knots are defined only on closed contours, the ends of the polymer need to be connected to test whether the polymer is knotted and to examine the complexity of the knots. To avoid additional crossings introduced by a procedure to connect polymer ends, we selected for this analysis only those conformations of the fractal and equilibrium globule that have both ends of the chain close to the surface of the globule ($|r|>11$ units).

We computed the values of the Alexander polynomial, a measure of knot complexity, for 29 fractal and 27 equilibrium globules (Fig S27). All of the equilibrium

globules exhibited extraordinarily high values of the knot complexity (from ~10^{20} to ~10^{30}) and are therefore highly knotted. In contrast, the fractal globules were either completely unknotted (> 20% of them) or showed only a few crossings (knot complexity ~1-100). The few crossings we observed may have been introduced by the large scale Move #3 that we used for initial polymer compression. As expected, our calculations show a dramatic difference between the two types of globules.

Global Expansion.

To illustrate the functional consequences of the differing degree of knottedness in the fractal vs. equilibrium globules, we simulated the effects of a change in solvent conditions by taking 50 fractal and 50 equilibrium globules and removing the outer wall constraining them. The fractal globules rapidly unraveled. In contrast, the equilibrium globules expanded briefly, but the expansion soon halted because of the large number of knots. (Fig S28,S29)

Local Expansion.

We further sought to demonstrate that the lack of knots in a fractal globule facilitated not only global, but local decondensation. We taking 36 fractal and 40 equilibrium globules and removed the outer wall constraining them, replacing this constraint with an attractive potential. We verified that the attractive potential did not destabilized the folded structure.

We then simulated the effects of changing the interaction term for a contiguous region on the polymer. Such a change might correspond to changes in solubility properties when a chromatin domain gains or loses an epigenetic mark. For fractal globules, this local change in potential led to complete unraveling of the local region. In equilibrium globules, some unraveling was observed, but it was largely suppressed by the presence of knots.

We quantified this effect by measuring the absolute distance from the perturbed monomers to the center of the globule over time. For fractal globules, this average distance was markedly larger than for equilibrium globules. (Fig S30,S31)

These results suggest that changes in solubility induced by such perturbations as the addition or removal of epigenetic marks may be sufficient to locally remodel chromatin and decondense the modified loci.

Appendix 1. Estimate of the Volume Fraction of Chromatin in Human Cells

In the simulations we sought to obtain an ensemble of structures that, *in their statistical properties,* resemble some of the features of chromatin arrangement in the cell. Below we demonstrate that chromatin occupies a significant fraction of cell volume, a property that we reproduced in simulations.

Taking the nuclear diameter of a tissue culture cell to be 5-10um, and assuming close to a spherical shape we obtain the volume in the range 50-500 um3, with a (geometric) mean of ~160 um3.

If we assume that the chromatin is built of DNA wrapped around nucleosomes, then we have 6x109bp/200bp=3x107 nucleosomes. Each may be approximated as a cylinder ~10nm in diameter and ~5nm in height, suggesting a volume of about 500nm3 each. The linker DNA after each nucleosome is about 50bps long, suggesting a volume of

about 50*.34nm*3.14*1nm2=50nm3. Thus the total volume of chromatin = 550x3x107 =16 um3, or ~10% (3-23%) of the nuclear volume. This strikingly large volume fraction is itself a significant underestimate, since we ignored, among other things, all other DNA-bound proteins. Note that any further packing or localization of chromatin inside the nucleus will increase local density.

In our simulations, the radius of the final crumpled globule was R≈12.5 and the volume V≈8000 cubic units. The total volume of the 4000 monomers, 1 unit in diameters each, is V≈2000. This implies a volume fraction of about 25%, which is consistent with the volume fraction estimated above.

Appendix 2. Monomer length in base pairs

Each monomer of the chain corresponds to a fragment of chromatin that equals the Kuhn length of the chromatin fiber, i.e. approximately twice the persistence length of the fiber. Although the persistence length of the chromatin fiber is unknown it can be estimated using the following arguments. DNA is packed into nucleosomes, where 150 bps are wrapped around the histone core and do not contribute to flexibility of the fiber. The linker DNA of about 50 bps that connects consecutive nucleosomes is bendable, and is the source of flexibility in the fiber. Since the persistence length of double-stranded DNA is 150 bps, an equally flexible region of the nucleosomal DNA should contain 3 linkers, i.e. 3 consecutive nucleosomes packing about 600 bps of DNA. The excluded volume of the nucleosomes, nucleosome interactions, and other DNA-bound proteins can make the fiber less flexible or prohibit certain conformation and may tend to increase the persistence length of the fiber. U

Using this estimated lower bound estimate for the persistence length, we obtain the Kuhn length of the equivalent freely-jointed chain to be 6 nucleosomes, or ~ 1200bp. A simulated chain of 4000 monomers corresponds to 4.8Mb of packed DNA. The size of each monomer was chosen such that its volume is equal to (or slightly above) that of 6 nucleosomes (V=6 x 600 nm^3); thus the radius of the spherical monomer is R=10nm. The diameter of each globule shown in Figure 4 is about 200 nm.

Appendix 3. Lattice analogues of the Fractal and Equilibrium Globule

As noted earlier, the fractal globule is hierarchical, self-similar, unentangled, and densely fills space; in these respects it analogous in structure to a finite iteration of a classic Peano curve. The equilibrium globule densely fills space, but shares none of the other structural properties of a fractal globule. It is therefore analogous to a Hamiltonian path on a lattice, a combinatorial object which satisfies no constraints beyond the requirement of visiting each point on the lattice exactly once (Fig S32).

V. Supplemental Figures

Fig. S1. The partitioning of chromatin into two spatial compartments is seen for all 23 chromosomes in GM06990. Correlation maps at a resolution of 1 Mb are shown for every chromosome (grey: unalignable, blue: centromeres). There is a strong correlation between the principal component (eigenvector), which reflects the compartmentalization inherent in the heatmaps, and the distribution of fixed features such as genes. The eigenvector also correlates with dynamic features such as open chromatin (DNAseI sensitivity), activating histone modifications (H3K36me3), repressive histone modifications (H3K27me3). At higher resolutions, the correlation to repressive marks is dramatically reduced.

Fig. S2. Confirmation of genome compartmentalization by 3D-FISH. (A, B) To confirm the compartmentalization of the genome, we selected FISH probes for four loci (L5, L6, L7, and L8) that lie consecutively along Chromosome 22 but alternate between the two compartments (L5, L7 in A; L6, L8 in B). (A) We observed that L5 (green) was in general closer to L7 (blue) than to L6 (red), despite the fact that L5 is closer to L6 than to L7 in the primary sequence of the genome. These results were observed both visually (right) and by plotting the cumulative distribution (middle). (G) In a second experiment, we observed that L6 (red) was consistently closer to L8 (green) than to L7 (blue).

Fig. S3. Hi-C interaction frequency correlates with 3D distance. Average inter-locus distance as determined by 3D FISH is compared to the number of reads with one end in each of the tested loci (blue dots). A strong correlation is observed (red).

Fig. S4. Compartment A is less compact then compartment B. Compartment B is more compact than compartment A. Read enrichment as a function of distance for interactions between loci in noncontiguous blocks belonging to the same compartment (A: blue; B: green). Compartment B is consistently more enriched at every inter-locus distance. Read enrichment is computed as number of reads divided by expected number of reads assuming random ligation.

Fig. S5. Cumulative distribution showing expression in compartment A (red) and compartment B (blue). The results demonstrate that genes in compartment B have markedly lower expression as compared to genes in compartment A.

Fig. S6. The partitioning of chromatin into two spatial compartments is seen for all 23 chromosomes in K562. Correlation maps at a resolution of 1 Mb are shown for every chromosome (grey: unalignable, blue: centromeres). There is a strong correlation between the principal component (eigenvector), which reflects the compartmentalization inherent in the heatmaps, and the distribution of fixed features such as genes. The eigenvector also correlates with dynamic features such as open chromatin (DNAseI sensitivity), activating histone modifications (H3K36me3), repressive histone modifications (H3K27me3). At higher resolutions, the correlation to repressive marks is dramatically reduced.

Fig. S7. PCR digest control. Hi-C ligation products can be distinguished from those produced in conventional 3C by PCR amplification identifying a ligation junction formed by two nearby fragments followed by digestion of the ligation site. Hi-C junctions are cut by NheI, not HindIII; the reverse is true for 3C junctions. 70% of Hi-C amplicons were cut by NheI confirming efficient marking of ligation junction. Two replicates were performed to ensure reliable quantification.

Fig. S8. Hi-C reads align near HindIII restriction sites with the correct orientation. (**A**) Reads from fragments corresponding to both intrachromosomal (blue) and interchromosomal (red) interactions align significantly closer to HindIII restriction sites as compared to randomly generated reads (green). Both the intrachromosomal reads and interchromosomal reads curves decrease rapidly as the distance from the HindIII site increases until a plateau is reached at a distance of ~500 bp. This corresponds to the maximum fragment size used for sequencing. (**B**) Hi-C sequences are expected to point (5'-3') in the direction of the ligation junction and therefore should align in the linear genome to the 3' end of HindIII restriction fragments. This tendency is reflected in ~80% of reads from both intrachromosomal (blue) and interchromosomal (red) interactions.

Fig. S9. The expected matrix. The average contact probability for a pair of loci at a given genomic distance produces an expectation matrix corresponding to what would be observed if there were no long-range structure.

Fig S10. Four distinct scaling regimes for contact probability at varying size scales.

Fig S11. Sketch of calculation for a 2D Hilbert Curve, comparing two consecutive iterations.

Fig S12. Illustration showing the smooth and interdigitated cases.

Fig S13-S21. Illustrations showing the accuracy of our predictions in the smooth case (the scaling exponent $\alpha_{smooth} = -(1+\frac{1}{d})$ using finite approximations of Peano curves. On the left, iterations of the curve are shown. On the right, we plot the scaling behavior. The periodicity observed in the scaling is a result of the size and dimensionality of the elementary motif.

Fig S13. Hilbert Curve in 2 Dimensions, $\alpha_{smooth} = -3/2$.

Fig S14. Peano Curve in 2 Dimensions, $\alpha_{smooth} = -3/2$.

Fig S15. Symmetrized Peano Curve in 2 Dimensions, $\alpha_{smooth} = -3/2$.

Fig S16. Quadratic Gosper Curve in 2 Dimensions, $\alpha_{smooth} = -3/2$.

Fig S17. Hilbert Curve in 3 Dimensions, $\alpha_{smooth} = -4/3$.

Fig S18. Peano Curve in 3 Dimensions, $\alpha_{smooth} = -4/3$.

Fig S19. Randomized Peano Curve in 3 Dimensions, α_{smooth} = -4/3. The elementary motif may be rotated in 3 different ways (or left intact) without affecting the starting position or ending position. To create the randomized curve, at each iteration, one of these four options is chosen for each subregion of the curve.

Fig S20. Hilbert Curve in many dimensions (2,3,4,6,9); $\alpha_{smooth} = -(1+\frac{1}{d})$.

Fig S21. Peano Curve in many dimensions (2,3,4,6); $\alpha_{smooth} = -(1+\frac{1}{d})$.

Fig. S22. Moves in the Monte Carlo procedure.

Fig. S23 Statistical properties of equilibrium and fractal globules. **Upper row**: The mean end-to- end distance vs. contour length, averaged over 100 conformations. For the crumpled globules we show the average within individual conformations (blue dots) and the average over the ensemble (black dots). The scaling $s^{0.29}$ provides the best fit to the data (solid red). The scaling $s^{1/3}$ shown by the dashed line constitutes an upper limit that is closely approached by individual conformations. **Lower row.** The probability of a contact between two points separated by contour length *s*. Simulated structures show very good agreement with the differing theoretical predictions for the two models.

Fig. S24 Examples of equilibrium and fractal globules. The polymer is colored in rainbow shades from red to blue along its contour length. Equilibrium globules demonstrate extensive mixing of the regions that are distant along the chain (have different colors). Fractal globules, in contrast, exhibit large monochromatic blocks, demonstrating little mixing of distant regions.

Fig. S25 Hierarchical structure of the fractal globule. Monochromatic domains at one level are isolated and repainted to reveal their domain organization at the next level. An important property of the fractal globule is its hierarchical organization, which is evident from the series of series of blow-ups (Fig MC.4) showing how individual domains isolated at one level, consist of well-separated (monochromatic) domains at the next level of folding.

Fig. S26 Subchains within fractal and equilibrium globules have vastly differing conformations. In a fractal globule, subchains correspond to compact spatial territories (left). In an equilibrium globule, subchains of a comparable size will wander randomly throughout the conformation; their spatial extent is equivalent to that of the globule as a whole.

Fig. S27 Different topological states of fractal and equilibrium globules. The distribution of the values of the Alexander polynomial, a knot invariant, computed for 29 fractal (green) and 27 equilibrium (red) globules. The Alexander polynomial characterizes the degree of complexity of the knot and equals 1 for unknotted chain, 9 for 3-1 knot, 25 for 4-1 knot, etc. For comparison, the most complex knot observed in proteins (6-1) has the value of 81. The polynomial has been computed for closed contours obtained by

connecting the ends as described in (*S18*). To avoid spurious knotting due to the end-joining procedure, only conformations with both ends on the surface of the globule have been used.

Fig. S28 Different expansion rates of fractal and equilibrium globules. When spatial constraints are removed, fractal globules unravel quickly (green); equilibrium globules expand briefly and then halt due to knotting.

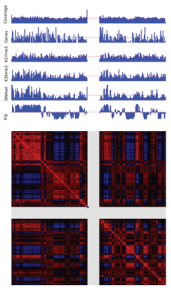
Fig. S29 Fractal globules expand readily when the compressive potential is removed; equilibrium globules remain tightly knotted. Here, examples are presented at the same size scale after an equal number of steps in the absence of a compressive potential. The fractal globule expands dramatically; the equilibrium globule is arrested early in its expansion due to knotting.

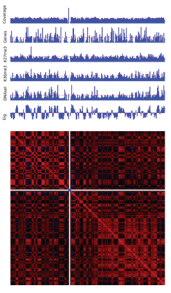
Fig. S30 Different expansion rates of local domains within fractal and equilibrium globules. Globules are stabilized by monomer attractions instead of hard boundaries. When the attraction is reversed and becomes repulsive for a subchain, the subchain bulges out of the fractal globule, but does not do so in an equilibrium globule. Here we plot the ratio of the mean absolute distance from the globule center of mass after repulsions are introduced vs. mean absolute distance from the globule center of mass before repulsions are introduced. Results are for subchains of length 700 in both fractal (green) or equilibrium (red) globules. These findings suggest that changes in solubility of a chromatin domain due to such factors as changes to epigenetic marks may be sufficient to induce local decondensation in a fractal globule.

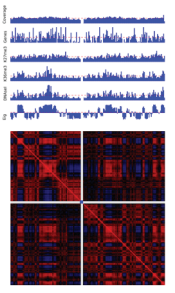
Fig. S31 A fractal globule subchain unravels when the stabilizing attractive potential is replaced by a repulsive potential.

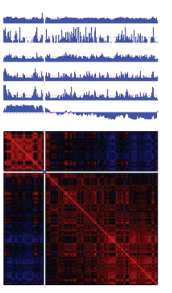
Fig S32. Comparison of a finite iteration of a Peano Curve (specifically, the Hilbert Curve) with an ordinary Hamiltonian Path in two dimensions. The former is analogous in structure to the fractal globule; the latter to an equilibrium globule. There is a stronger correspondence between one-dimensional position and d-dimensional position in the Hilbert Curve. In $d > 2$, Hamiltonian paths are highly knotted.

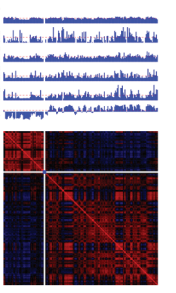
Lieberman-Aiden, Van Berkum et al.
Figure S1

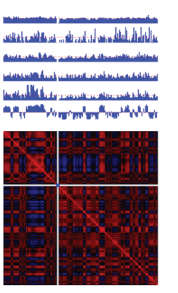

chromosome 1

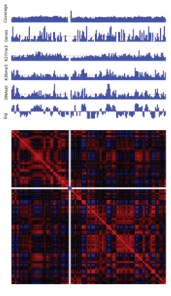

chromosome 2

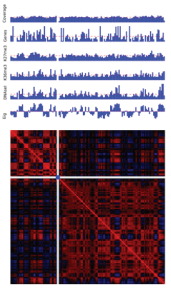

chromosome 3

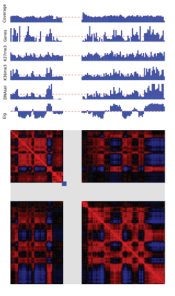

chromosome 4

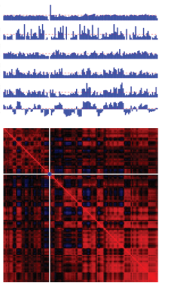

chromosome 5

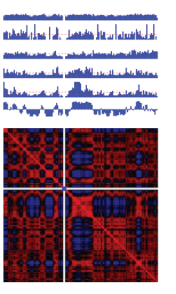

chromosome 6

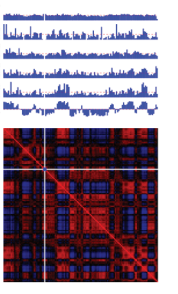

chromosome 7

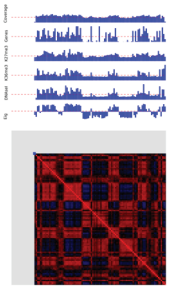

chromosome 8

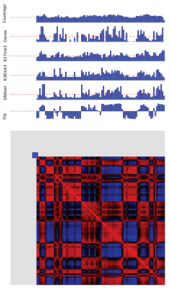

chromosome 9

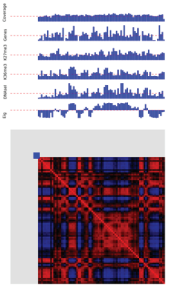

chromosome 10

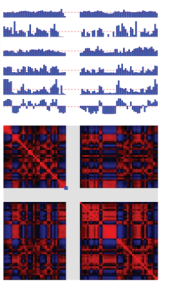

chromosome 11

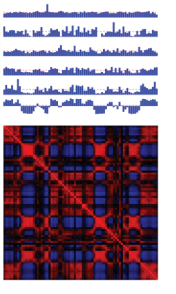

chromosome 12

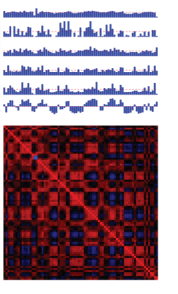

chromosome 13

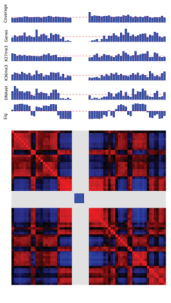

chromosome 14

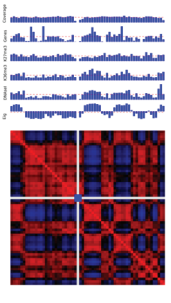

chromosome 15

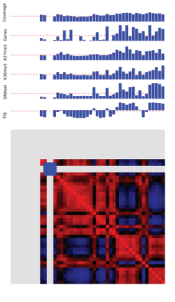

chromosome 16

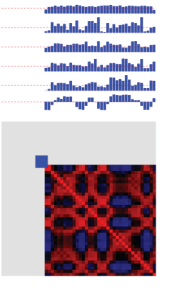

chromosome 17

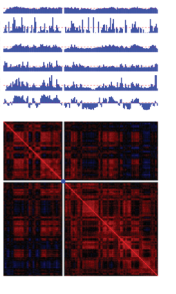

chromosome 18


chromosome 19

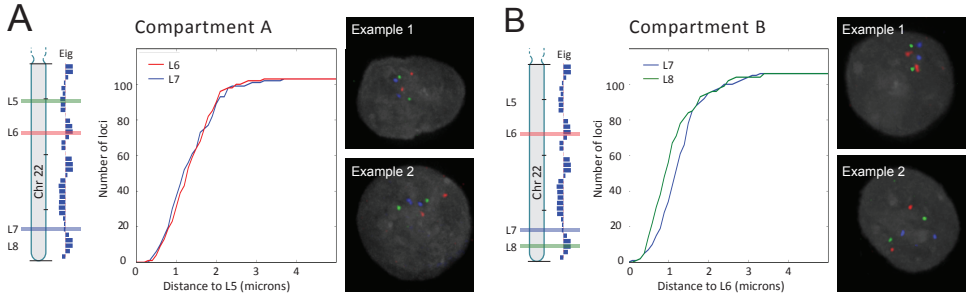

chromosome 20


chromosome 21

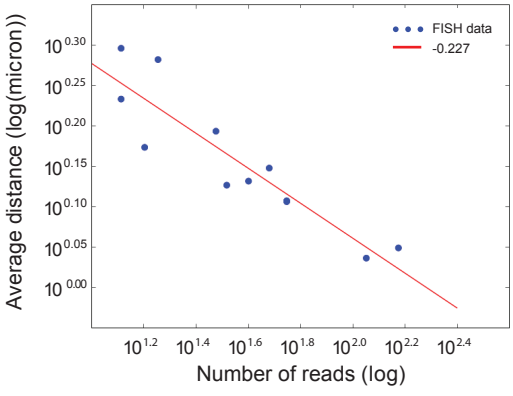

chromosome 22


chromosome 23

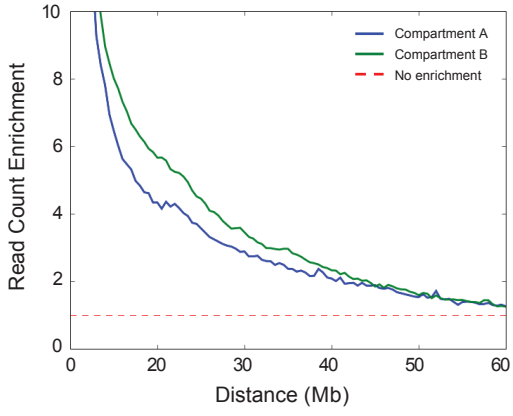
Lieberman-Aiden, Van Berkum et al.
Figure S2



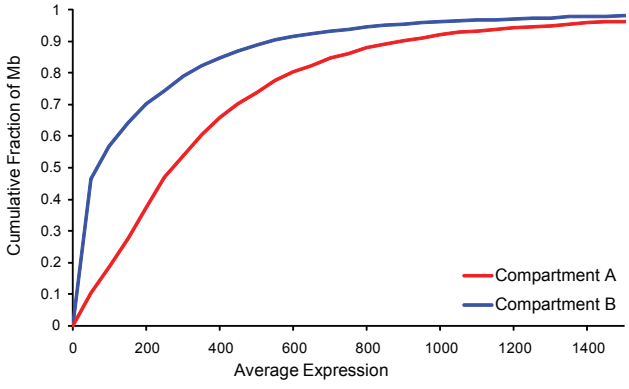
Lieberman-Aiden, Van Berkum et al.
Figure S3



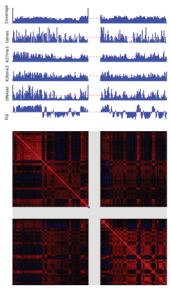
Lieberman-Aiden, Van Berkum et al.
Figure S4



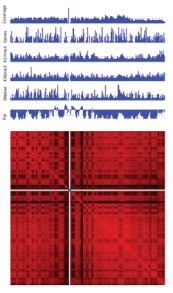
Lieberman-Aiden, Van Berkum et al.
Figure S5



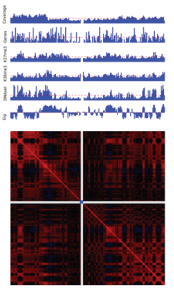
Lieberman-Aiden, Van Berkum et al.
Figure S6



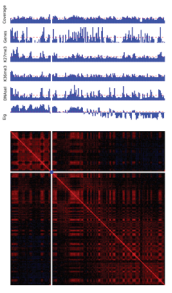
chromosome 1



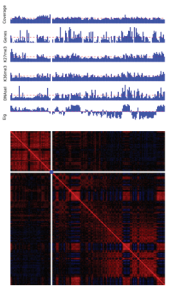
chromosome 2



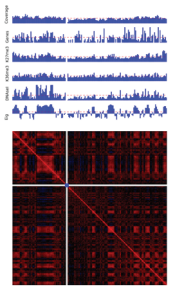
chromosome 3



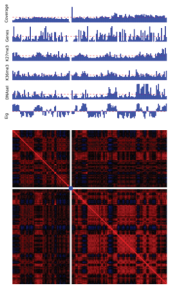
chromosome 4



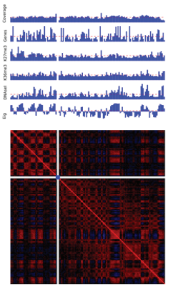
chromosome 5



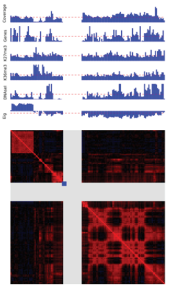
chromosome 6



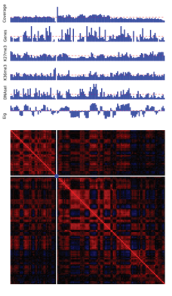
chromosome 7



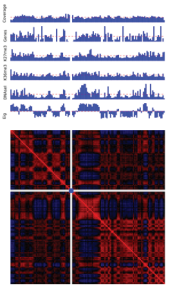
chromosome 8



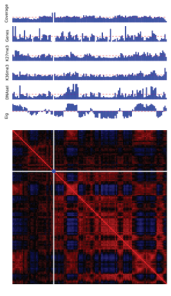
chromosome 9



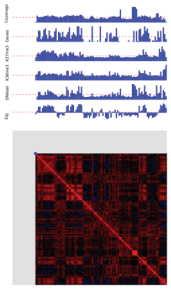
chromosome 10



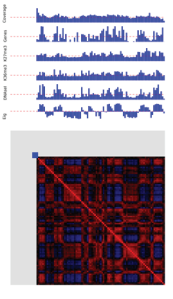
chromosome 11



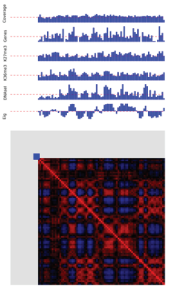
chromosome 12



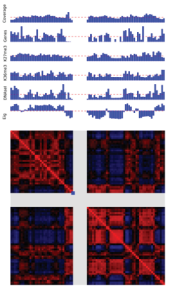
chromosome 13



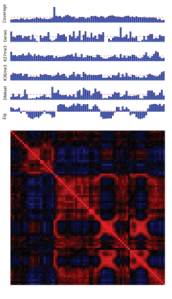
chromosome 14



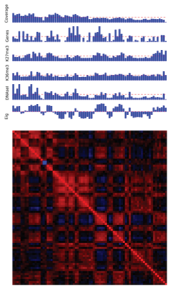
chromosome 15



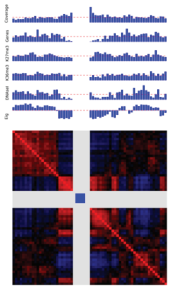
chromosome 16



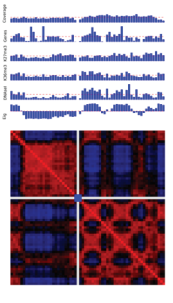
chromosome 17



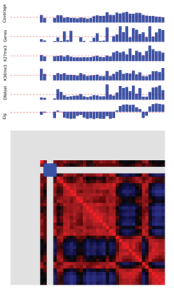
chromosome 18



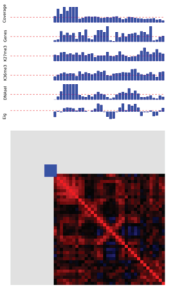
chromosome 19



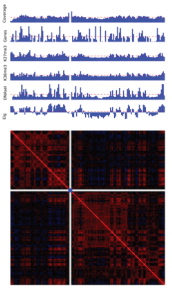
chromosome 20



chromosome 21

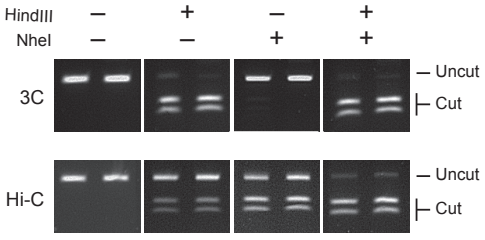


chromosome 22

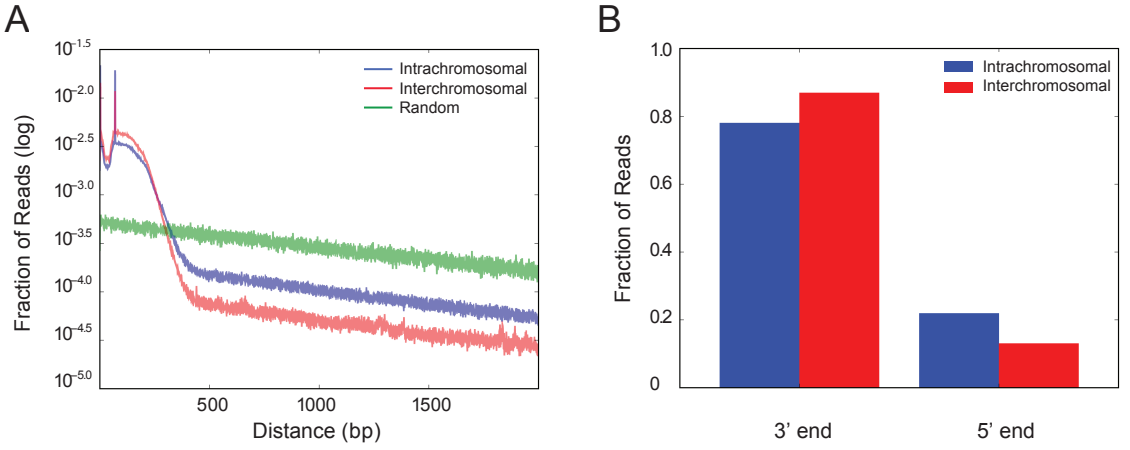


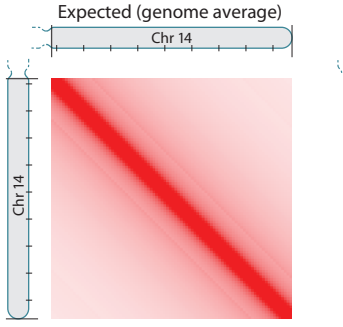
chromosome 23

Lieberman-Aiden, Van Berkum et al.
Figure S7



Lieberman-Aiden, Van Berkum et al.
Figure S8




Expected (genome average)
Chr 14
Chr 14

Lieberman-Aiden, Van Berkum et al.
Figure S10

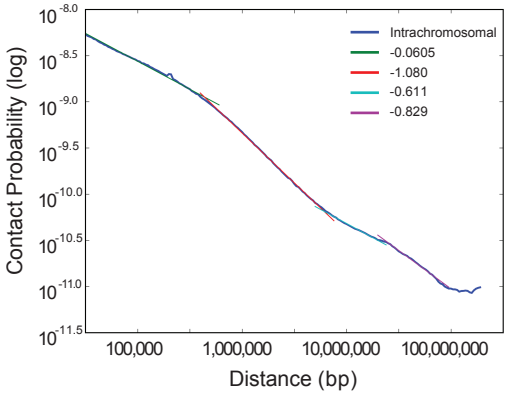
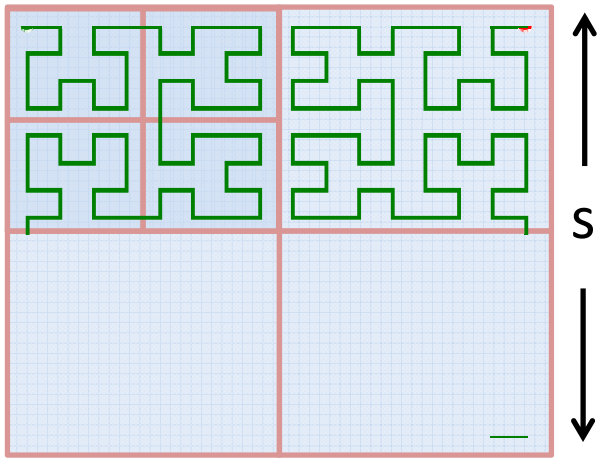


Fig S11. Calculation in 2D, Smooth Boundaries

Volume	Side	#Cubes	Int/Cube	#Interactions	#Possible Interactions	#I/#P	Change
N/4	s/2	4	4s/2 = 2s	8s	s^4	$8/s^3$	-
N/16	s/4	16	4s/4 = s	16s	$s^4/4$	$64/s^3$	8

Smooth

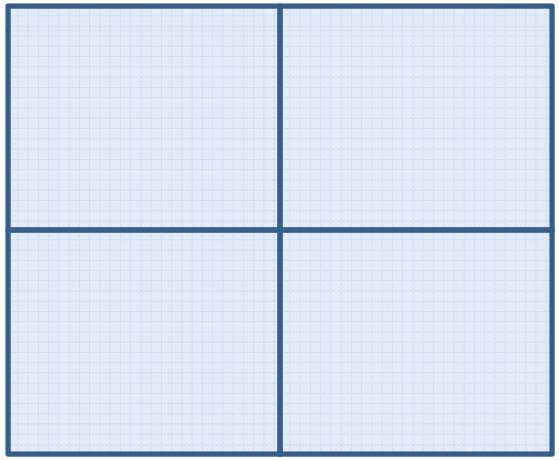


Total
Volume ~ N

$s^2=N$

Fig S12.
Two cases, Smooth and Interdigitating

Smooth



Interdigitating

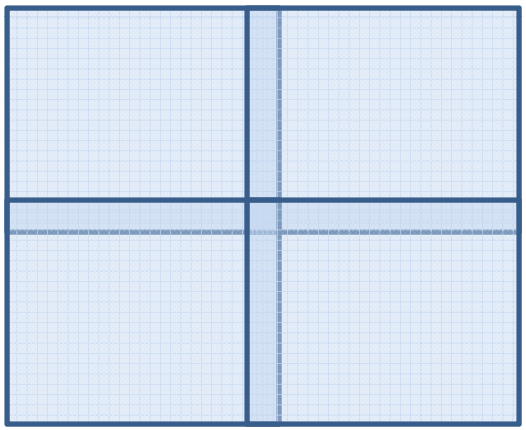
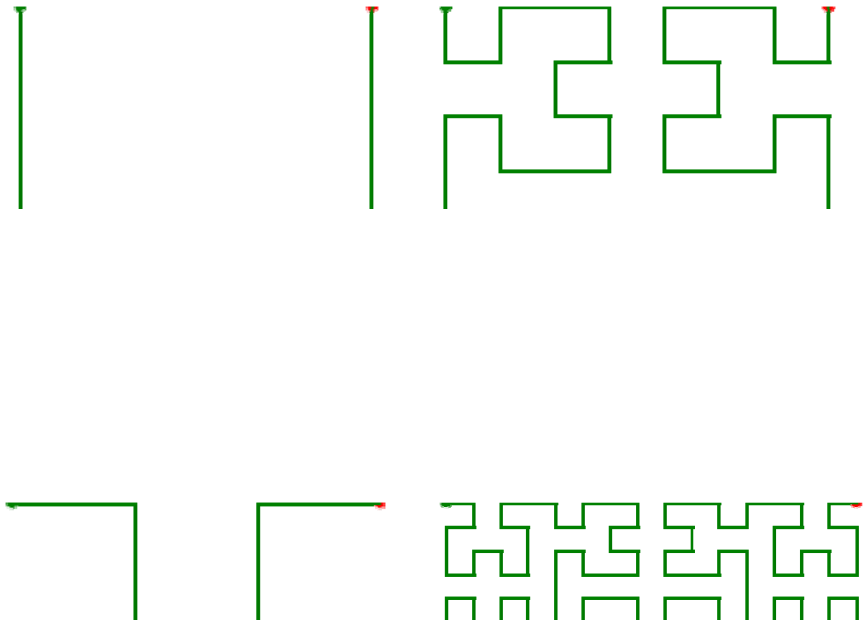


Fig S13. Hilbert Curve



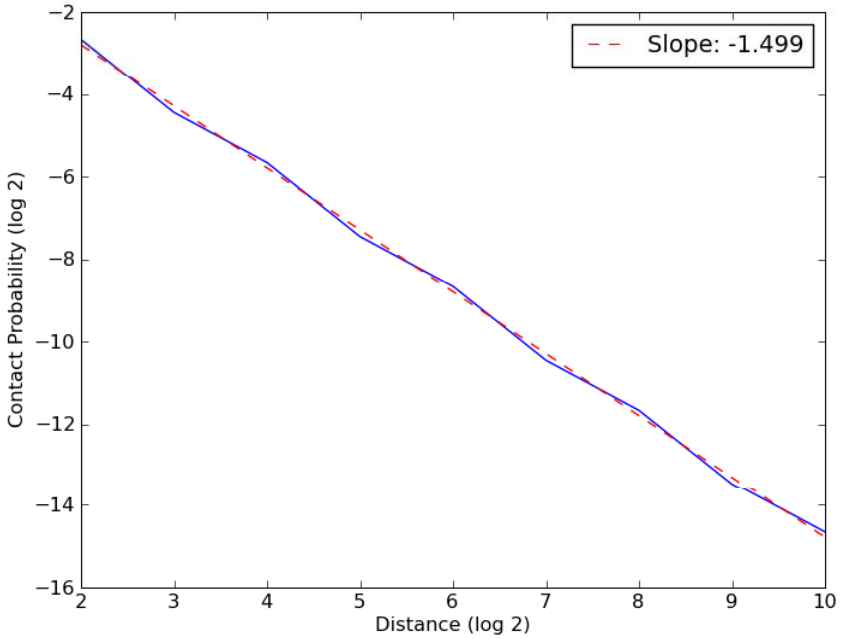
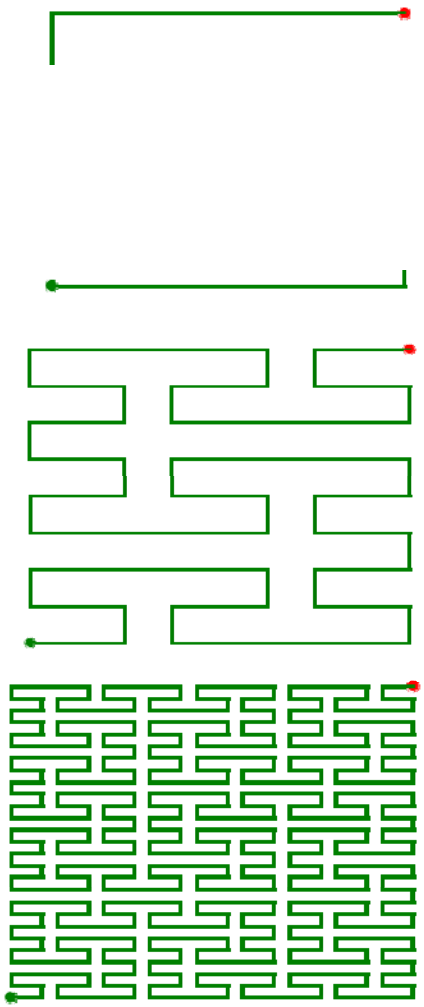


Fig S14. Peano Curve



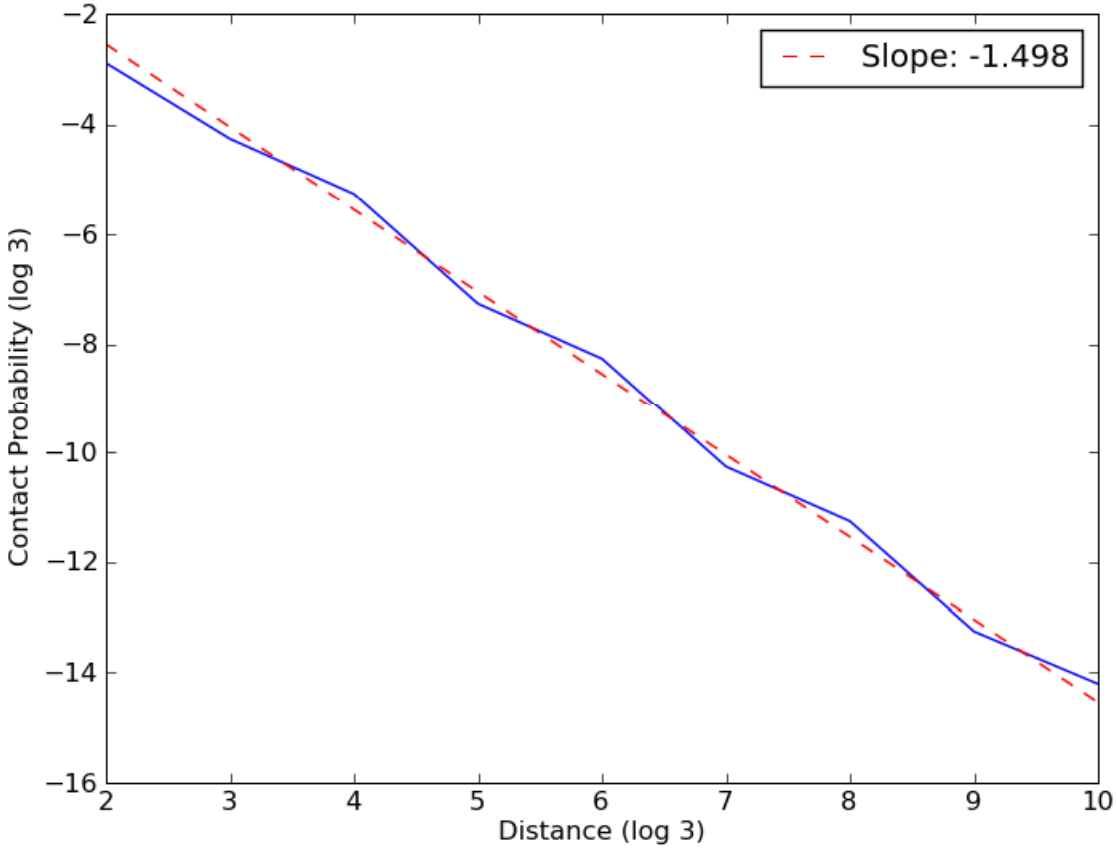
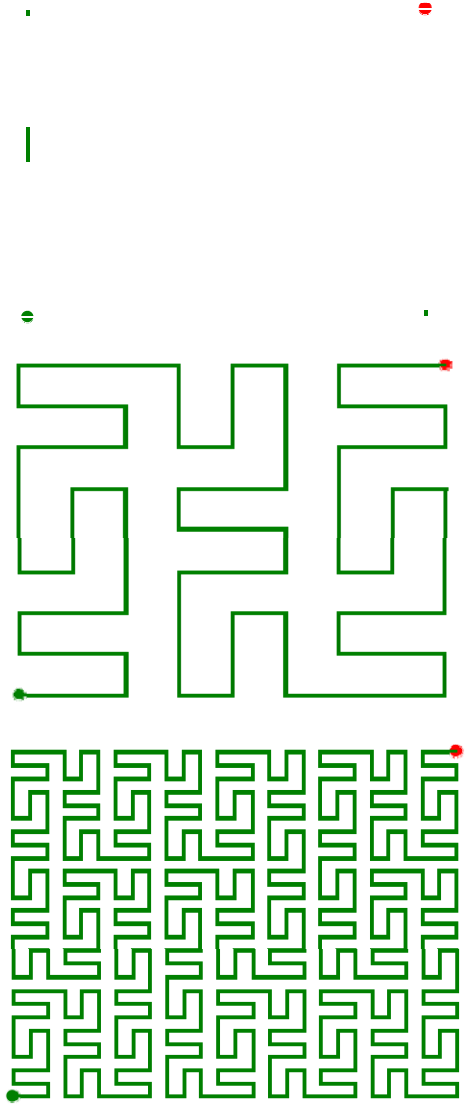


Fig S15. Symmetrized Peano Curve



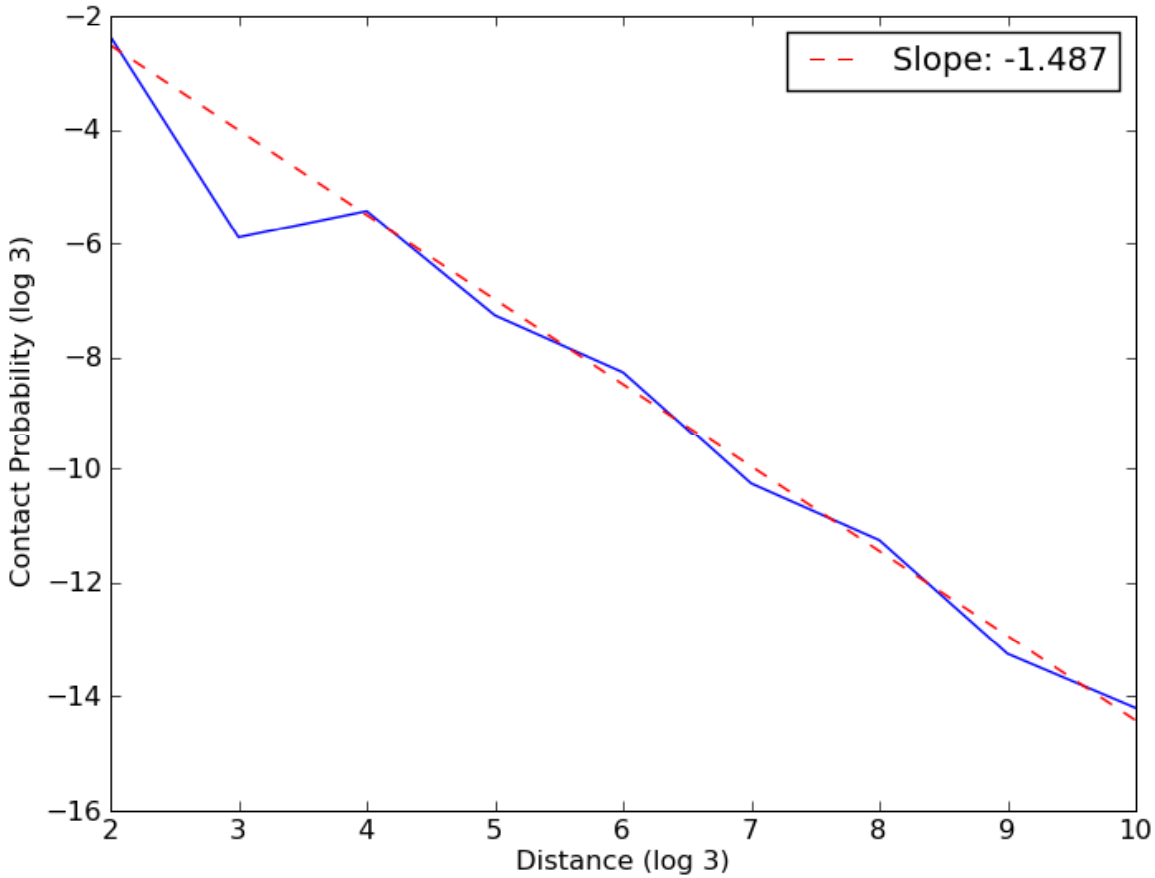
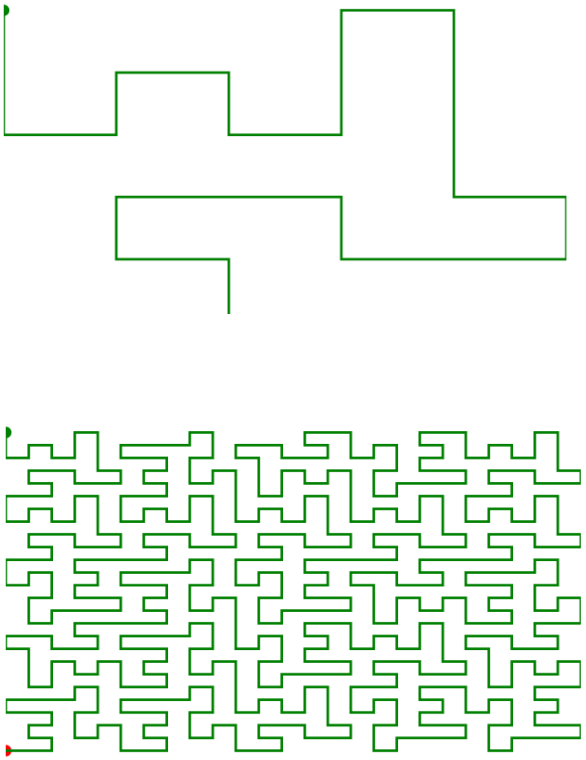


Fig S16. The Quadratic Gosper Curve



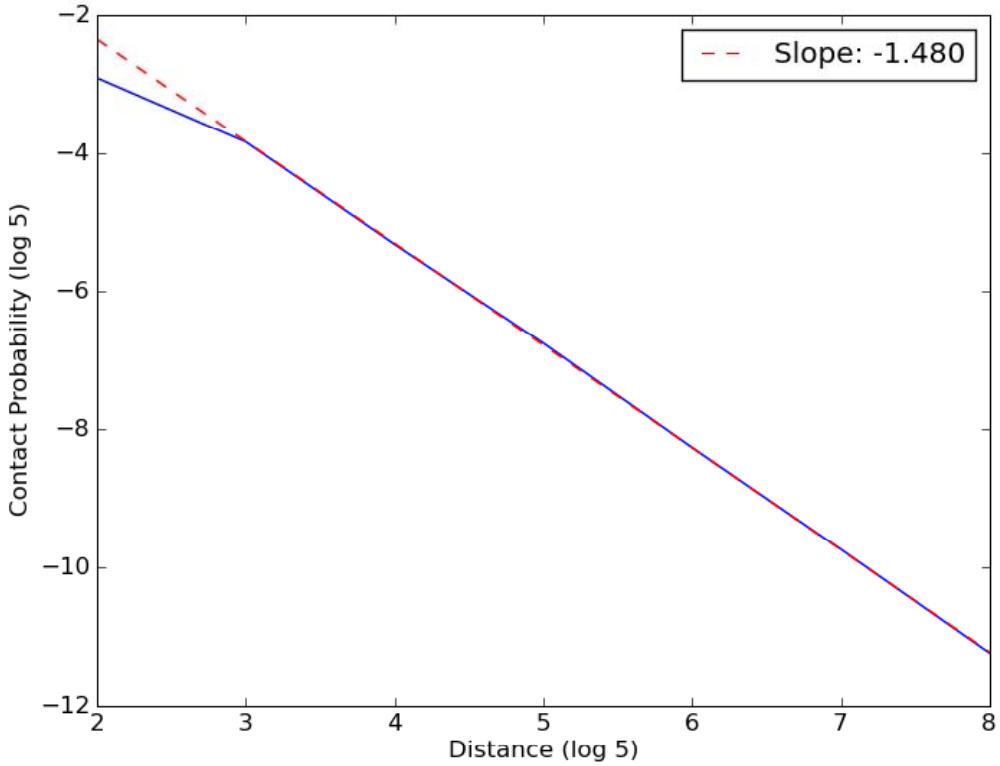
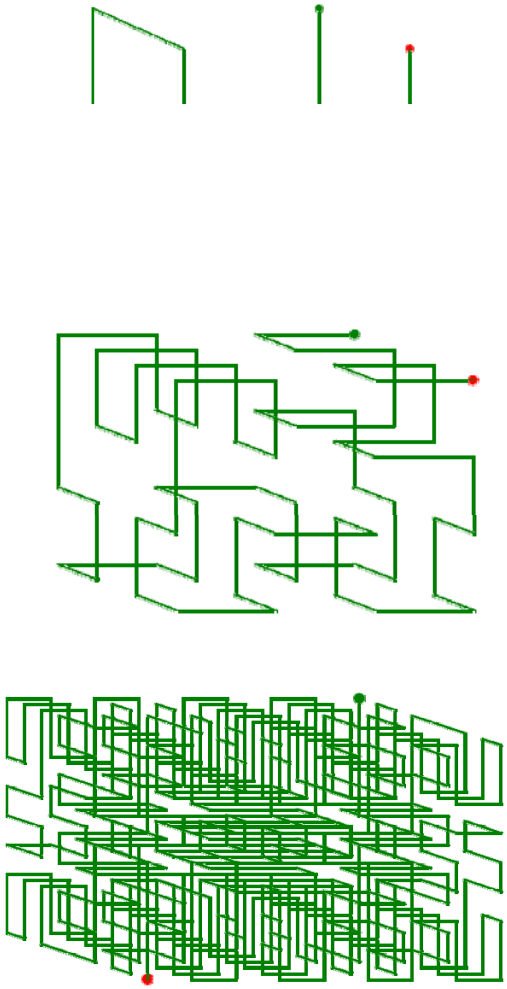


Fig S17. 3D Hilbert Curve



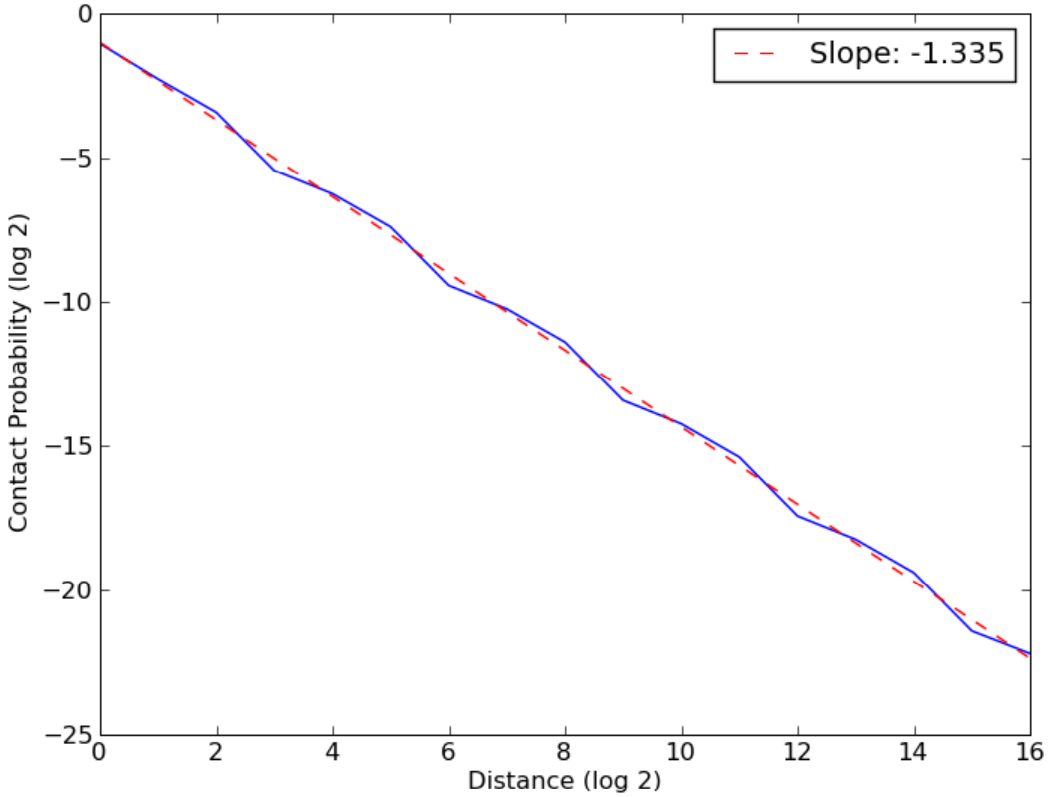
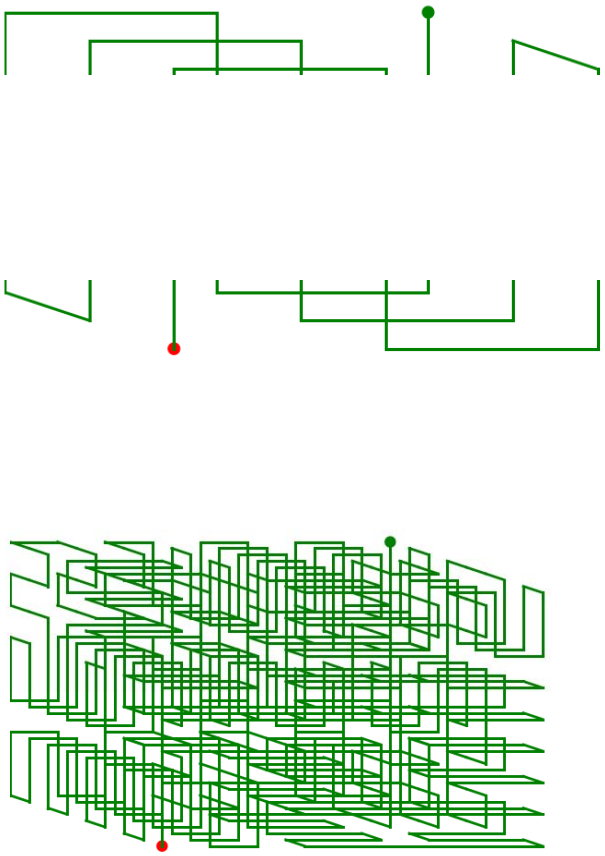


Fig S18. 3D Peano Curve



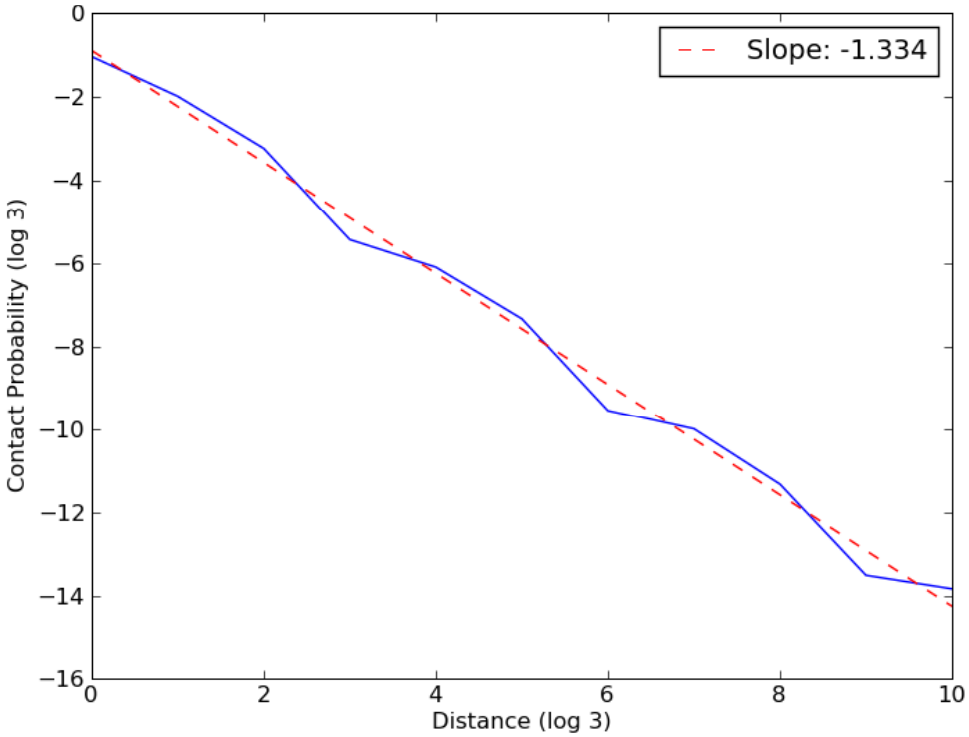
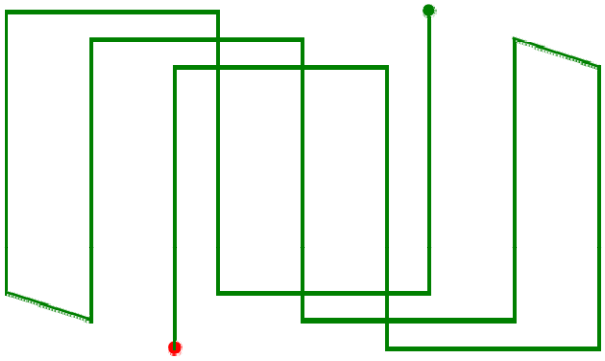
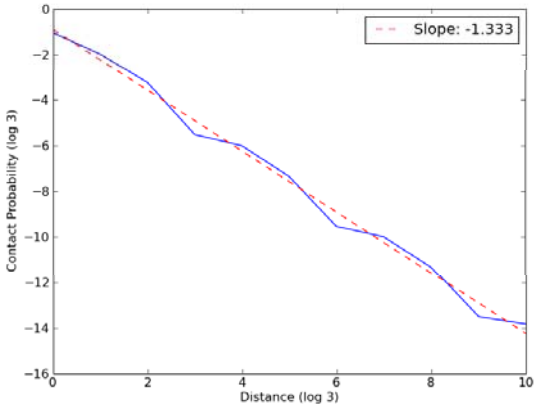
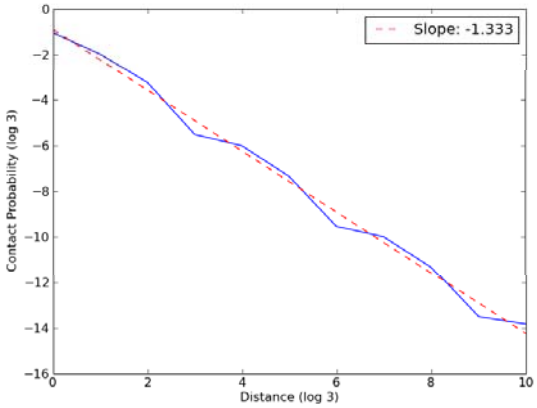
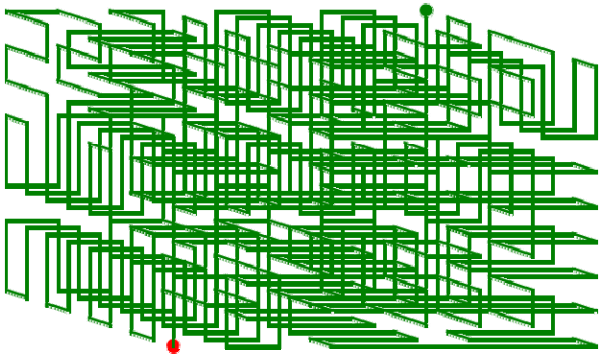


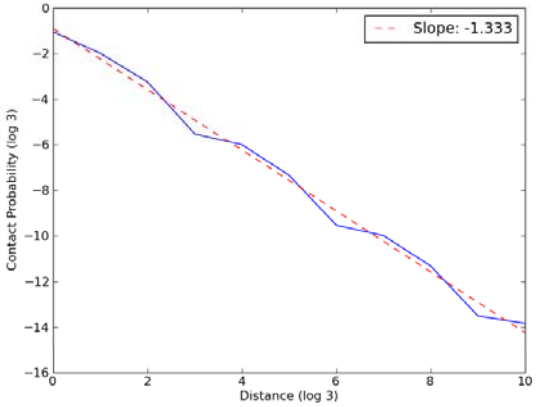
Fig S19. Randomized 3D Peano Curves











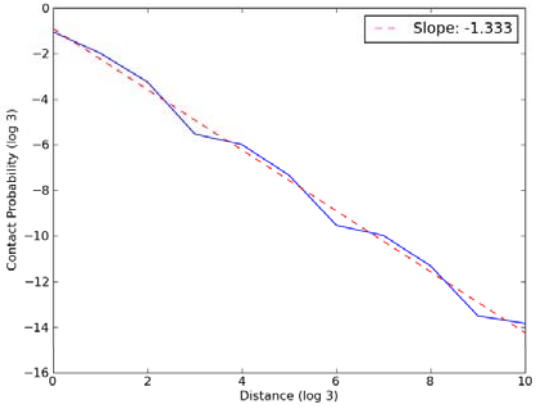
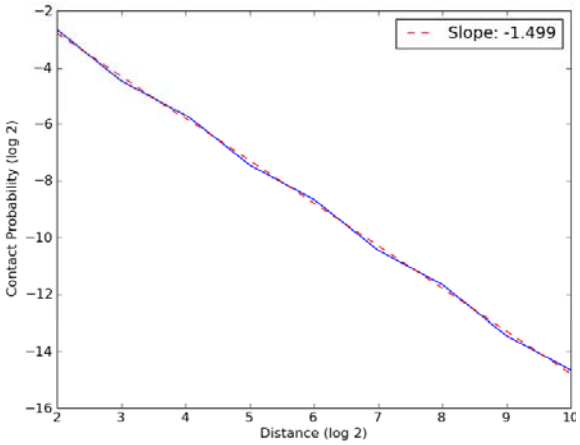
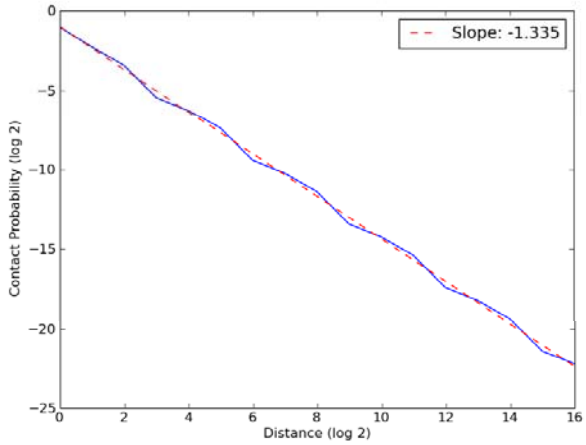


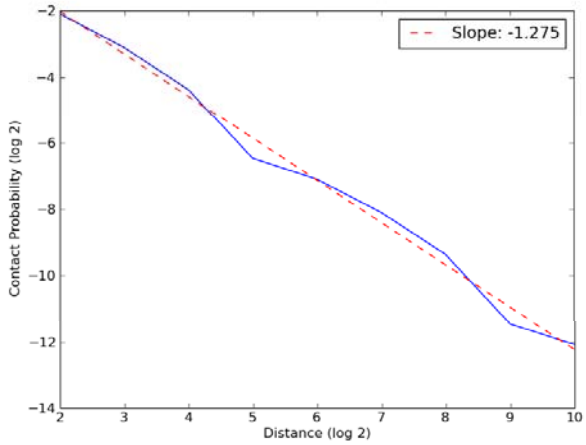
Fig S20. Hilbert Curve in d Dimensions



d=2

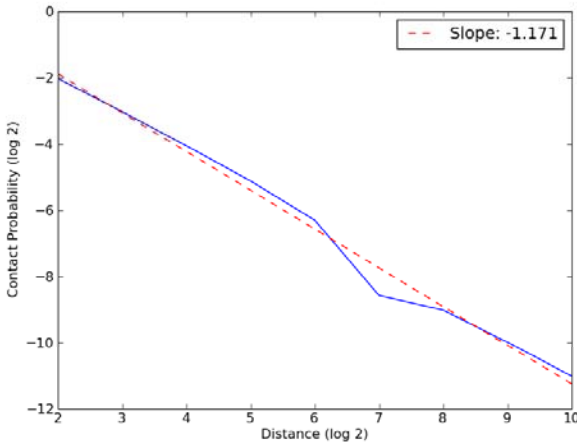


d=3

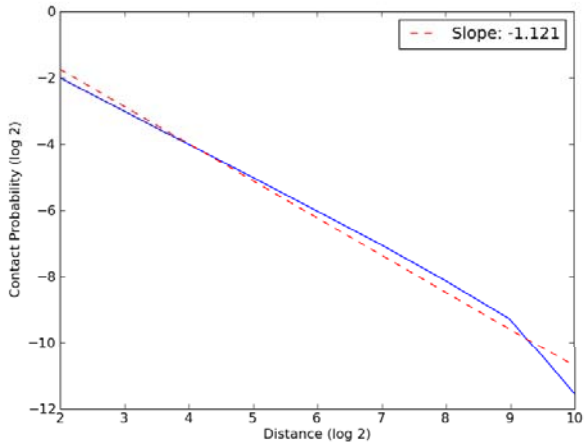


d=4

Dim	Experiment	Theory
2	-1.499	-1.5
3	-1.335	-1.333
4	-1.275	-1.25
6	-1.171	-1.166
9	-1.121	-1.111

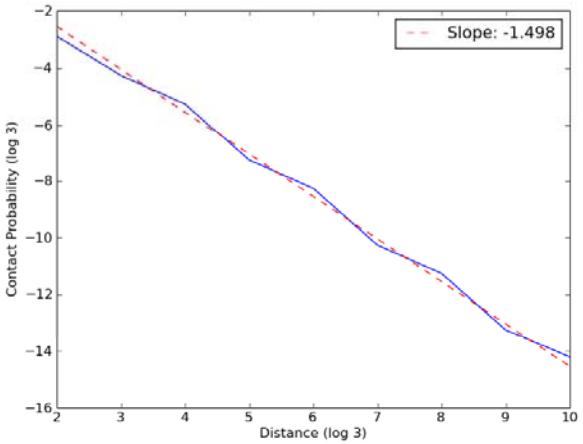


d=6

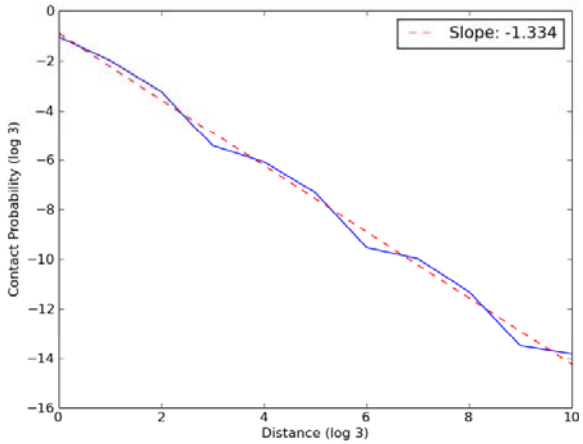


d=9

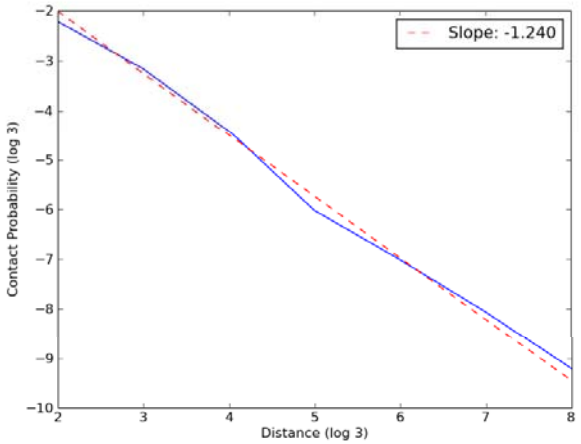
Fig S21. Peano Curve in d Dimensions



d=2

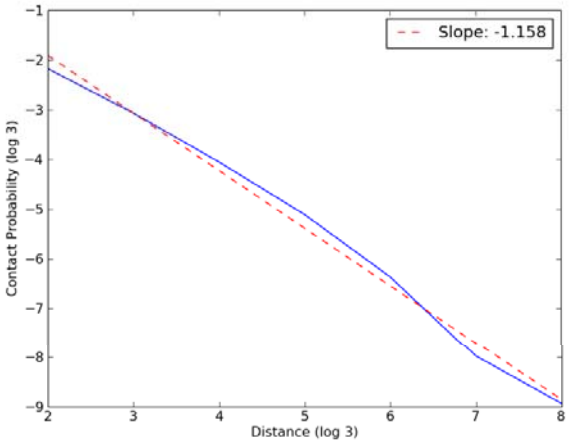


d=3



d=4

Dim	Experiment	Theory
2	-1.498	-1.5
3	-1.334	-1.333
4	-1.240	-1.25
6	-1.158	-1.166



d=6

Fig S22. Monte Carlo Moves

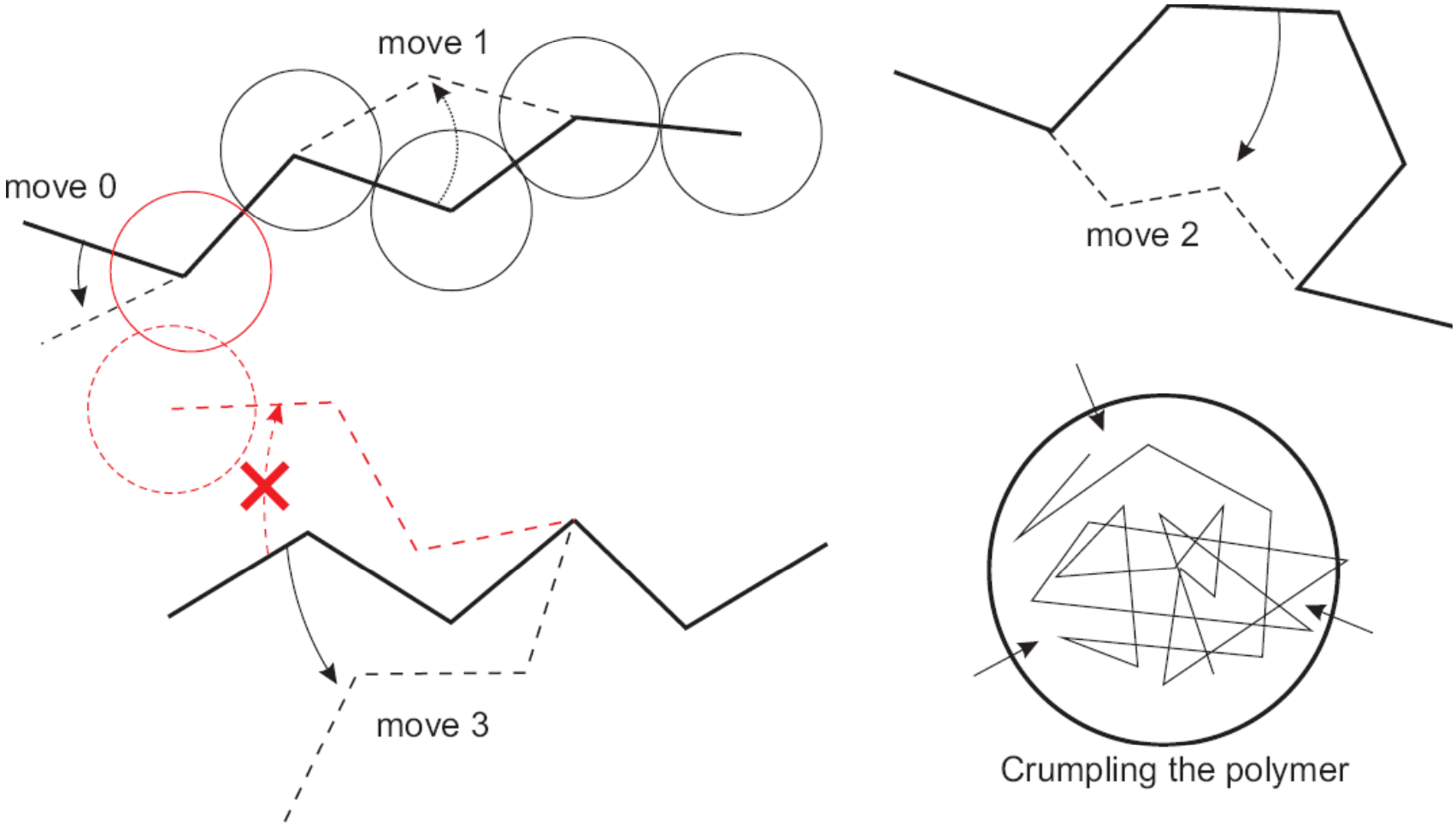


Fig S23. Statistics of Globules

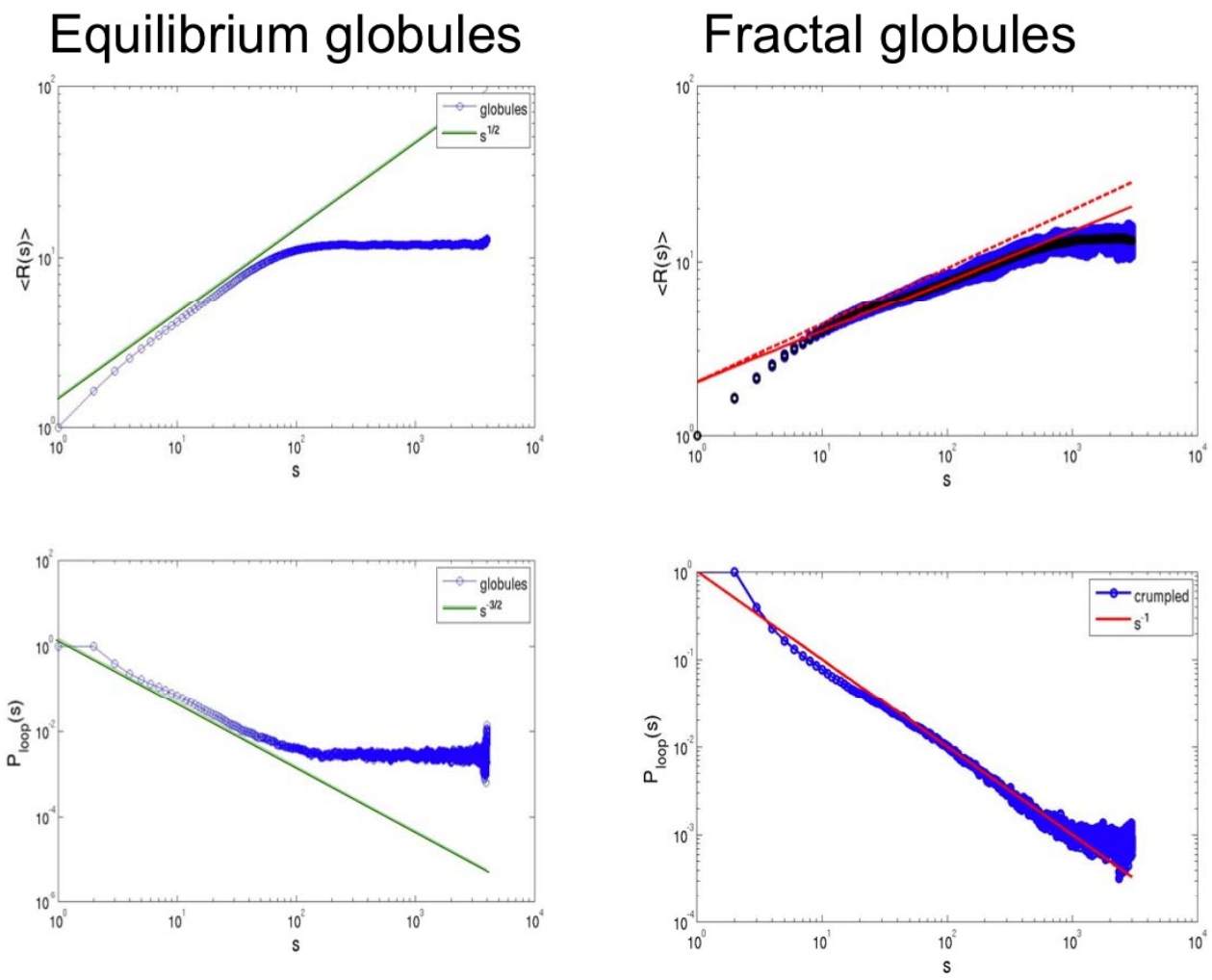


Fig S24. Examples of Globules

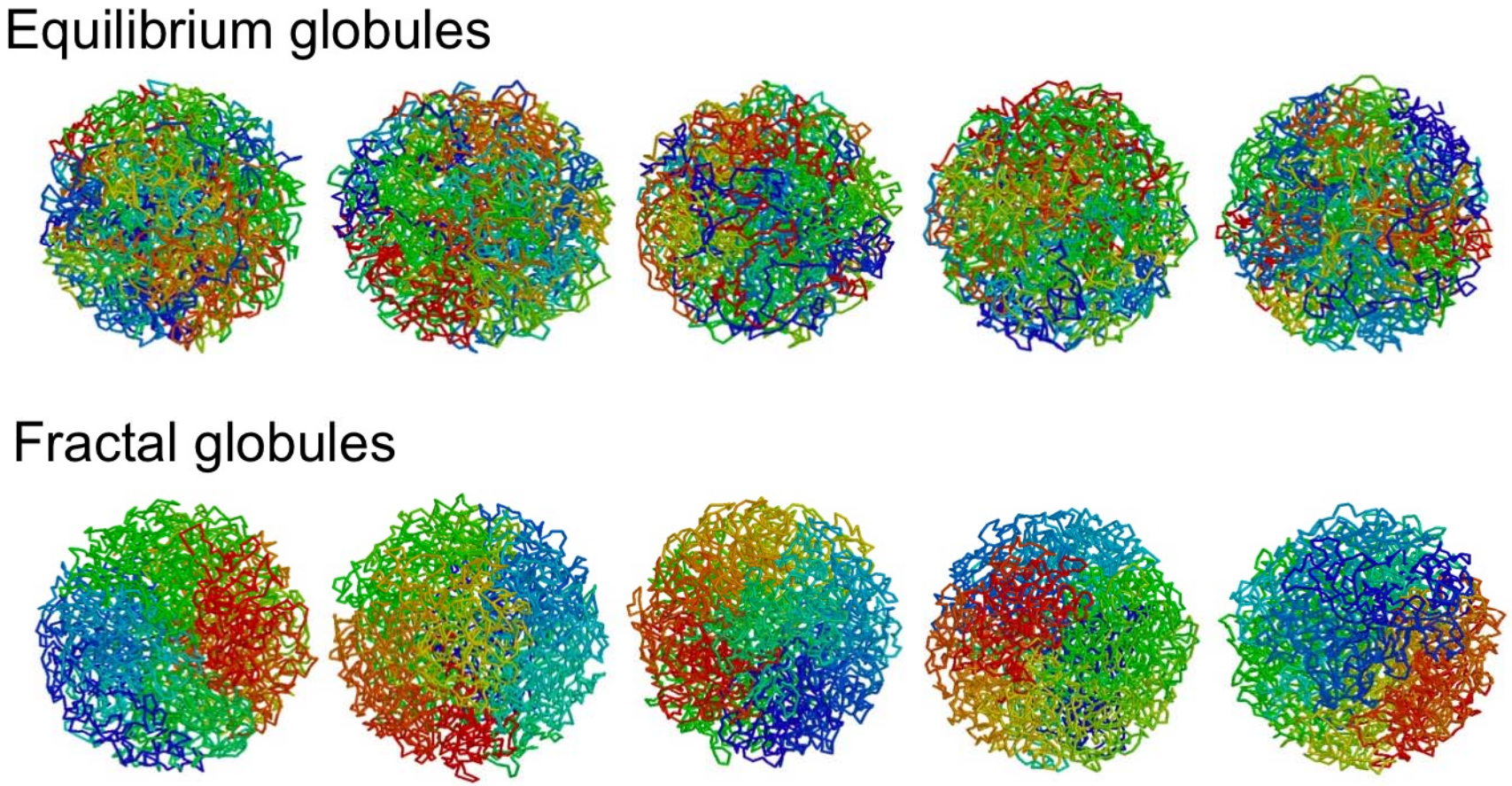


Fig S25. Self-similarity of Fractal Globule

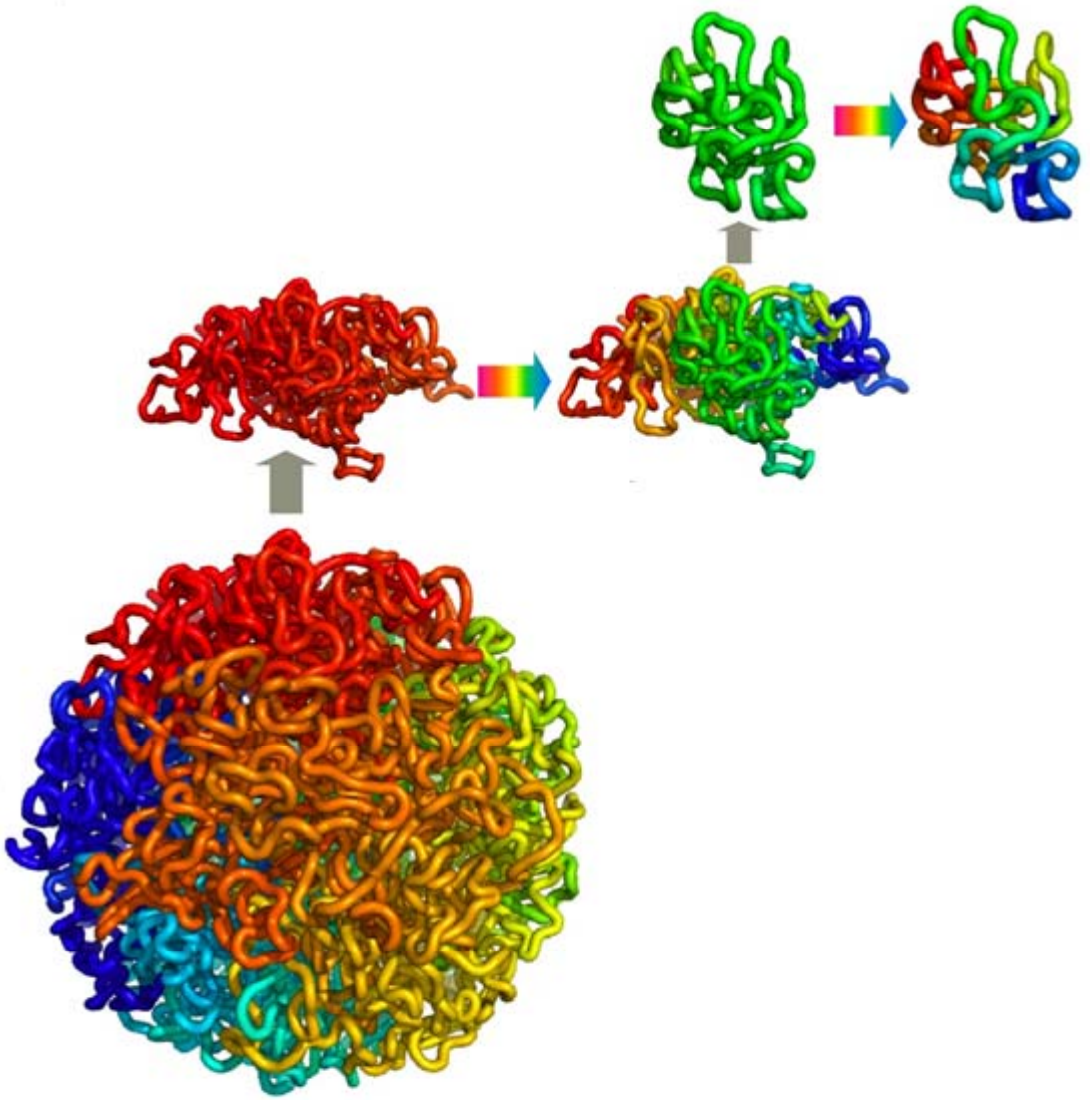
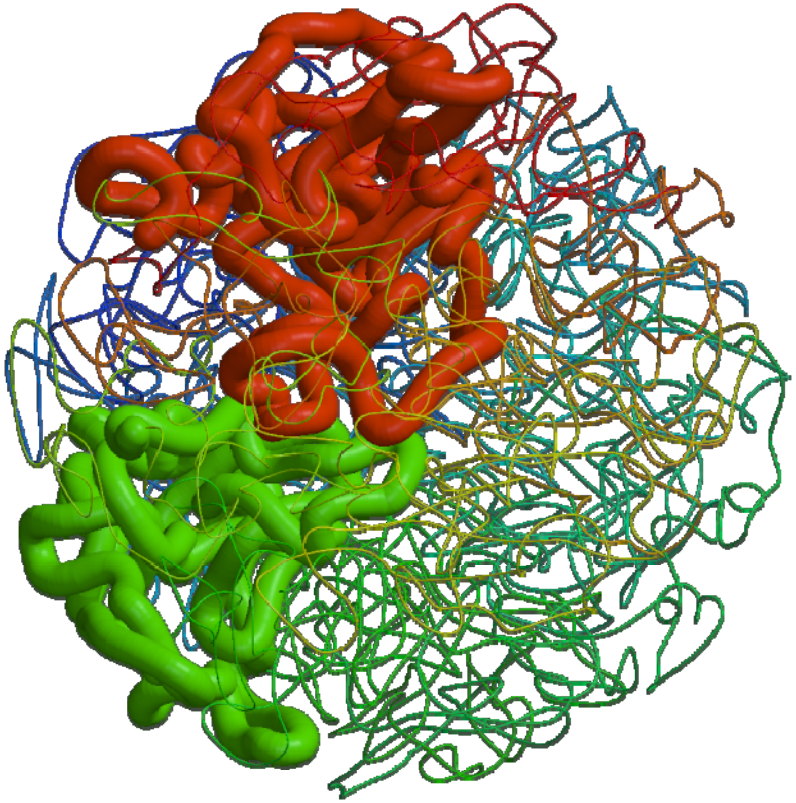
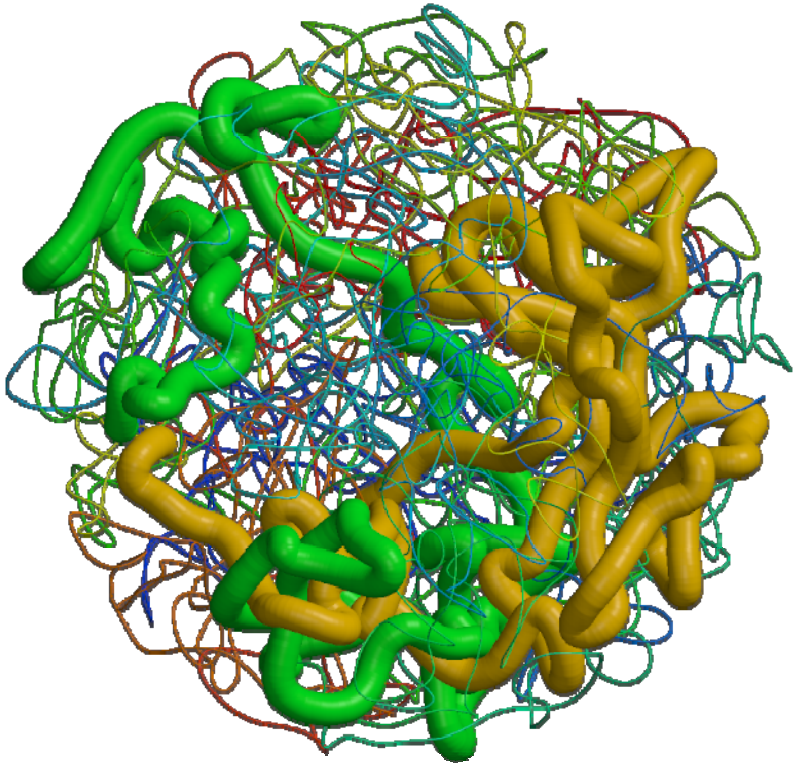


Fig S26. Structure of Subchains



Fractal Globule



Equilibrium Globule

Fig S27. Fractal globules lack knots

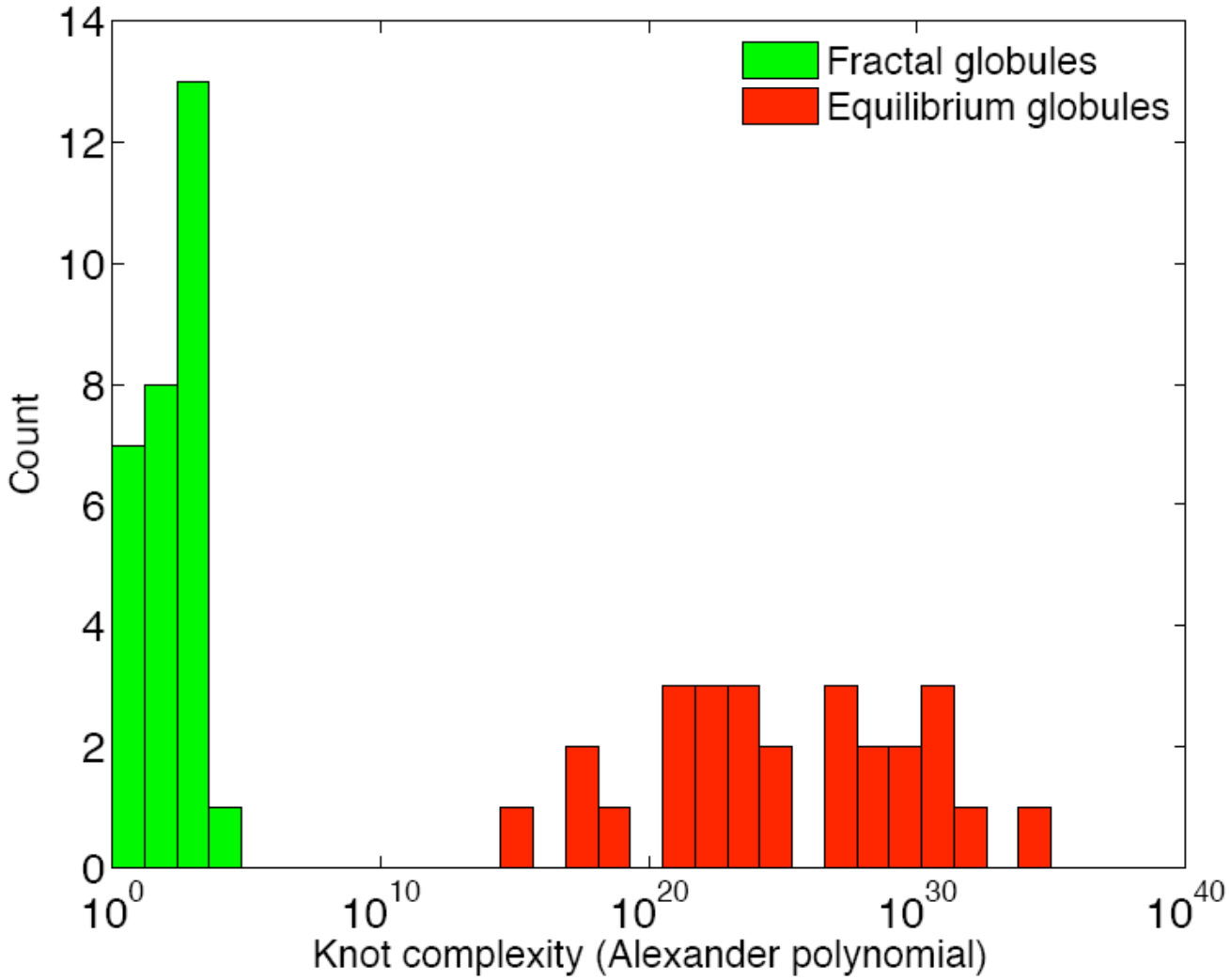


Fig S28. Global Expansion of Fractal Globules

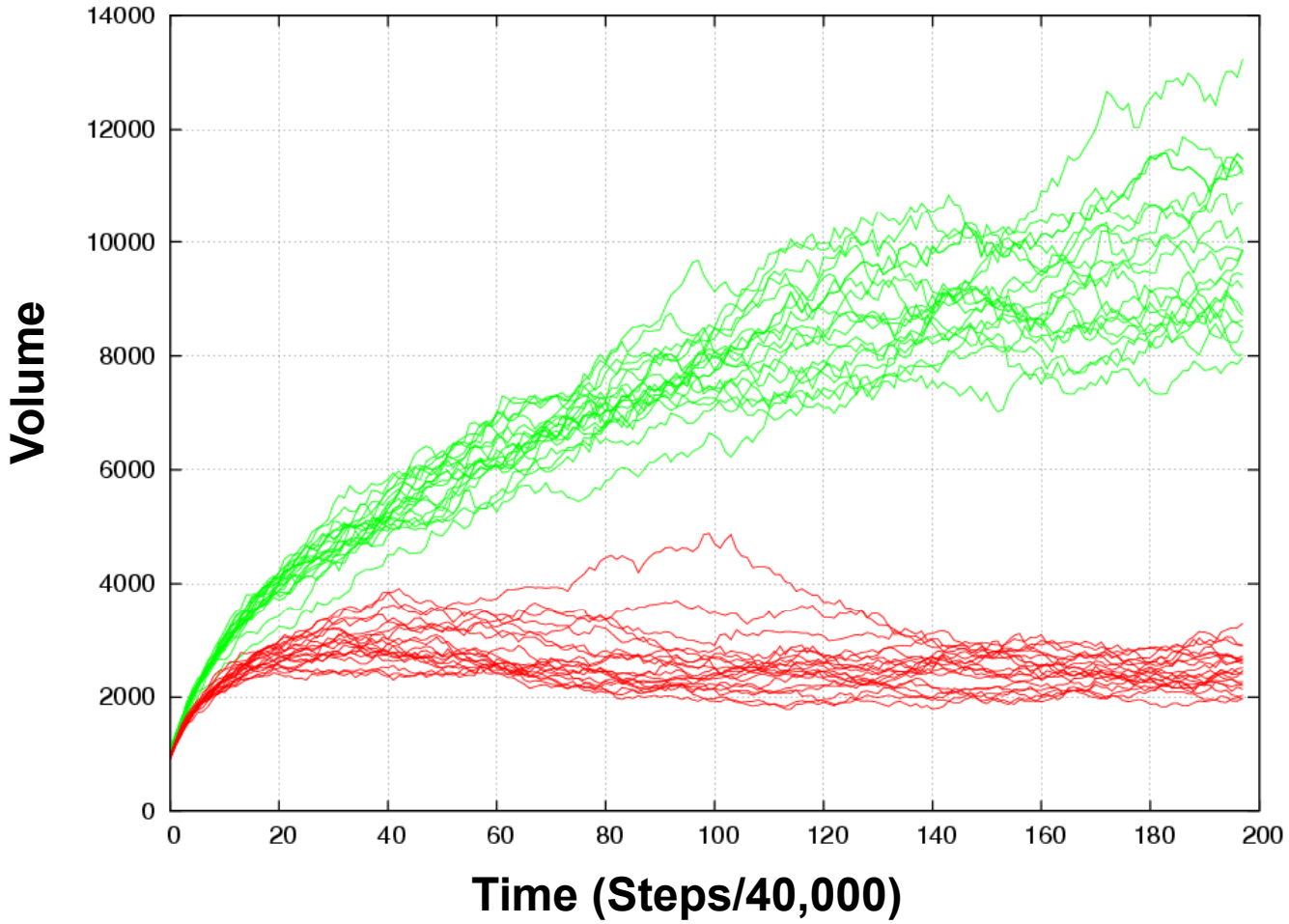
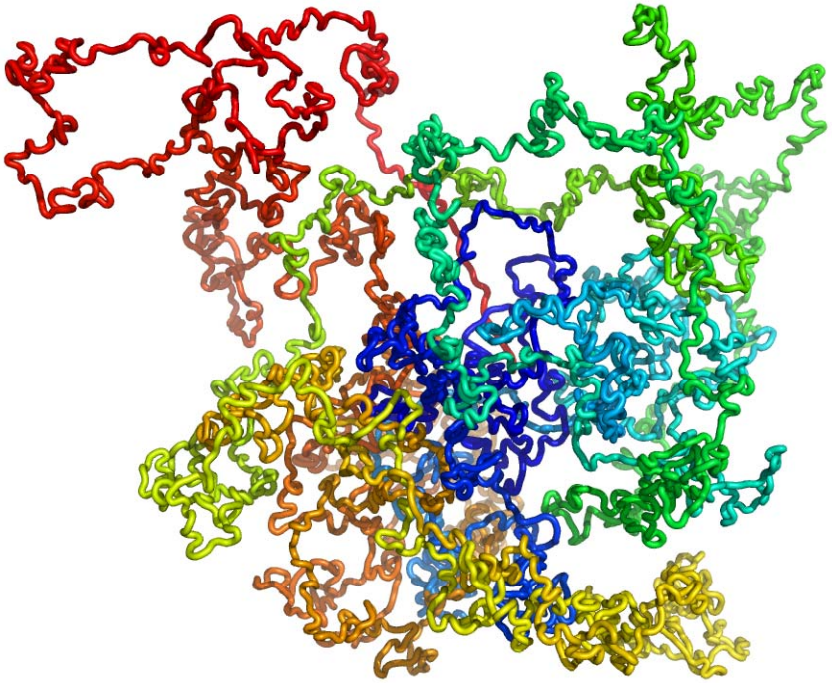
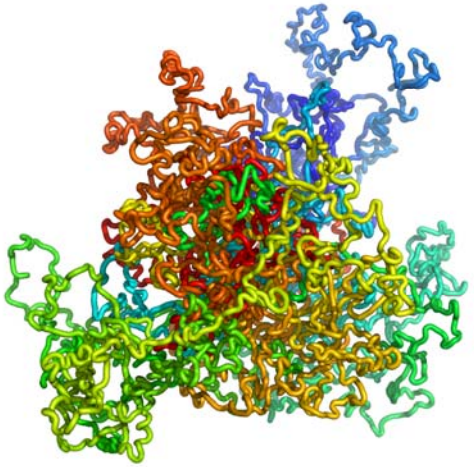


Fig S29. Global Expansion of Fractal Globules



Fractal Globule



Equilibrium Globule

Fig S30. Local Expansion of Fractal Globule Domains

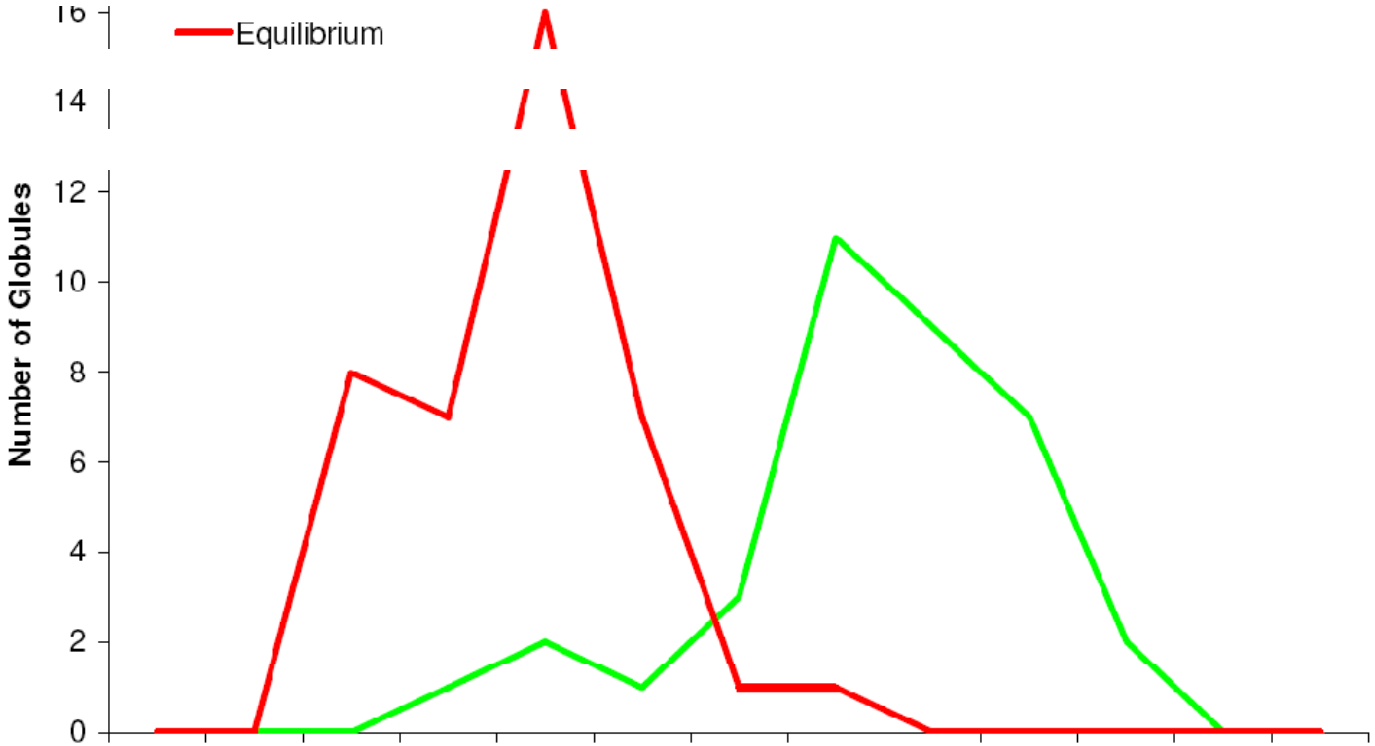


Fig S31. Local decondensation in a fractal globule

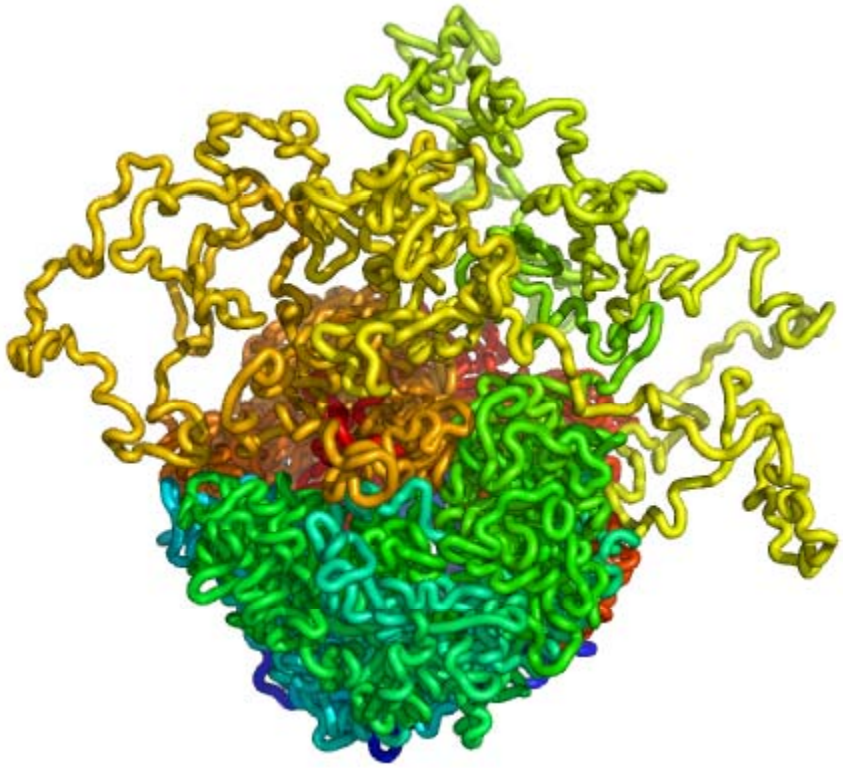
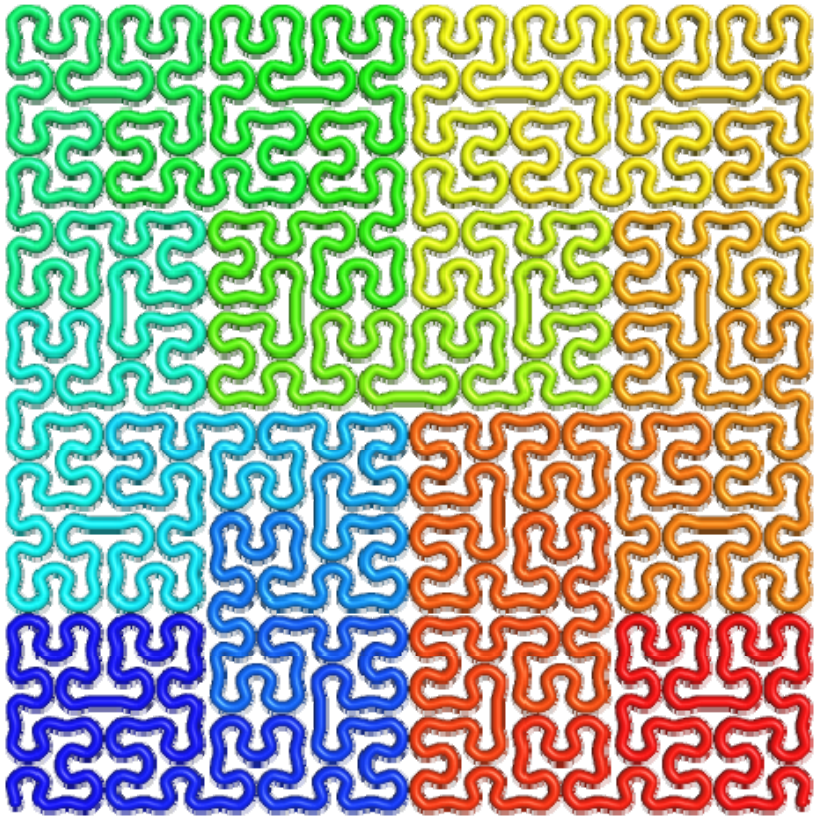
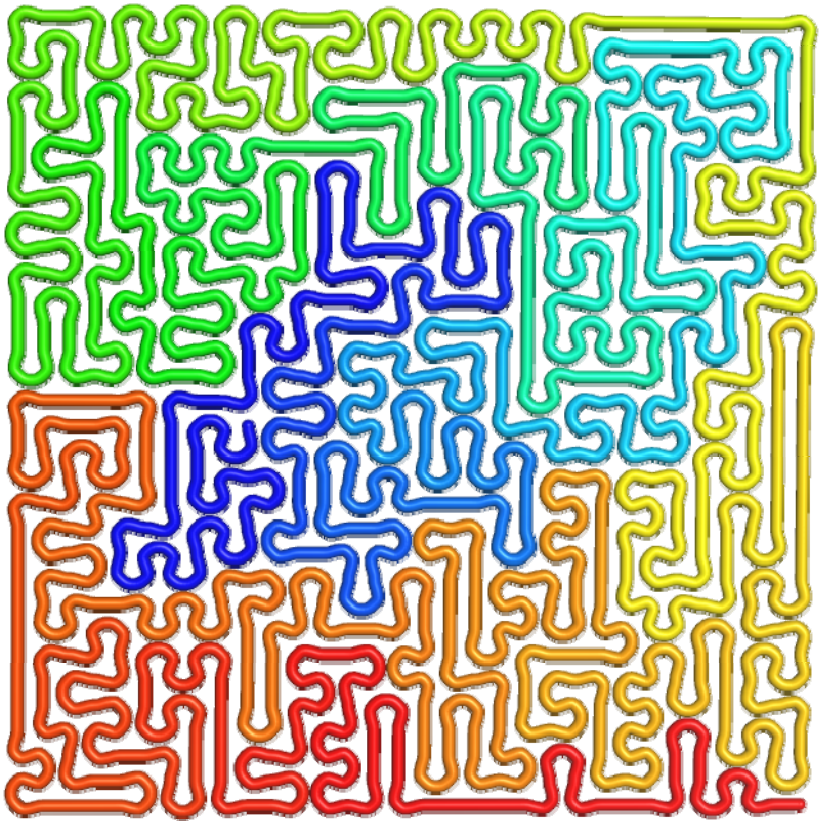


Fig S32. Peano Curve vs Hamiltonian Path



Peano Curve



Hamiltonian Path

VI. Supplemental Tables

Table S1: BAC clones used in 3D-FISH

BAC name	Alias	Chromosome	Start position	End position
RP11-68M15	L1	Chr14	22546692	22722266
RP11-91J1	L2	Chr14	45258185	45462464
RP11-79B13	L3	Chr14	67744258	67904880
RP11-88N20	L4	Chr14	86622674	86772926
RP11-22M5	L5	Chr22	20569761	20724994
RP11-79G21	L6	Chr22	26499393	26657386
RP11-49M22	L7	Chr22	43469285	43637241
RP11-66M5	L8	Chr22	46658148	46820598

VII. Supplemental References

S1. B. E. Bernstein *et al.*, *Cell* **120**, 169 (Jan 28, 2005).
S2. T. S. Mikkelsen *et al.*, *Nature* **448**, 553 (Aug 2, 2007).
S3. P. J. Sabo *et al.*, *Nat Methods* **3**, 511 (Jul, 2006).
S4. J. R. Hesselberth *et al.*, *Nat Methods* **6**, 283 (Apr, 2009).
S5. J. A. Croft *et al.*, *J Cell Biol* **145**, 1119 (Jun 14, 1999).
S6. A. L. Price *et al.*, *Nat Genet* **38**, 904 (2006).
S7. M. Ku *et al.*, *PLoS Genet* **4**, e1000242 (Oct, 2008).
S8. T. Vettorel, A. Y. Grosberg, K. Kremer, *Phys Biol* **6**, 25013 (2009).
S9. K. Binder, *Monte Carlo and molecular dynamics simulations in polymer science* (Oxford University Press, New York, 1995), pp. 587.
S10. J. Shimada, Kussell, E.L., Shakhnovich, E.I., *J Mol Biol* **308**, 79 (Apr 20, 2001, 2001).
S11. A. V. Vologodskii, Levene, S. D. , Klenin, K. V., Frank-Kamenetskii, M., , N. R. Cozzarelli, *J Mol Biol* **227**, 1224 (1992).
S12. A. I. U. Grosberg, Khokhlov, A.R., *Statistical physics of macromolecules*, AIP series in polymers and complex materials (AIP Press, New York, 1994), pp. 350.
S13. R. Lua, Borovinskiy, A.L., Grosberg, A. Y. , *Polymer* **45**, 717 (2004).
S14. A. Y. Grosberg, S. K. Nechaev, E. I. Shakhnovich, *J. Phys. France* **49**, 2095 (1988).
S15. O. A. Vasilyev, S. K. Nechaev, *Theoretical and Mathematical Physics* **134**, 142 (2003).
S16. G. Kolesov, P. Virnau, M. Kardar, L. A. Mirny, *Nucleic Acids Res* **35**, W425 (Jul, 2007).
S17. R. C. Lua, Grosberg, A. Y. , *PLoS Comput Biol* **2**, e45 (May 2006, 2006).
S18. P. Virnau, Mirny, L.A., Kardar, M. , *PLoS Comput Biol* **2**, e122 (Sep 15, 2006, 2006).
S19. Aubrey Jaffer provided the data for 4, 6, and 9 dimensional Hilbert curves shown in Fig S20, and for 4 and 6 dimensional Peano curves shown in Fig S21.