A predictive computational model of the dynamic 3D interphase yeast nucleus.

Hua Wong, Hervé Marie-Nelly, Sébastien Herbert, Pascal Carrivain, Hervé Blanc, Romain Koszul, Emmanuelle Fabre, Christophe Zimmer

A predictive computational model of the dynamic 3D interphase yeast nucleus

Hua Wong[1,4], Hervé Marie-Nelly[1,4,5], Sébastien Herbert[1,4,6], Pascal Carrivain[3,4], Hervé Blanc[2,4], Romain Koszul[2,4], Emmanuelle Fabre[2,4], Christophe Zimmer[1,4*]

[1]Institut Pasteur, Groupe Imagerie et Modélisation; URA CNRS 2582; Paris, France

[2]Institut Pasteur, Laboratoire Régulation Spatiale du Génome, Paris, France

[3]LPTMC, UMR 7600, Univ Pierre et Marie Curie, Paris, France

[4]CNRS GDR 3536, Univ Pierre et Marie Curie, Paris, France.

[5]Univ Pierre et Marie Curie, Cellule Pasteur, Paris, France

[6]Univ Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, France

*Corresponding author: czimmer@pasteur.fr

Summary

Background: Despite the absence of internal membranes, the nucleus of eukaryotic cells is spatially organized, with chromosomes and individual loci occupying dynamic, but non-random, spatial positions relative to nuclear landmarks and to each other. These positional preferences correlate with gene expression and DNA repair, recombination and replication. Yet the principles that govern nuclear organization remain poorly understood and detailed predictive models are lacking.

Results: We present a computational model of dynamic chromosome configurations in the interphase yeast nucleus that is based on first principles, and is able to statistically predict the positioning of any locus in nuclear space. Despite its simplicity, the model agrees with extensive previous and new measurements on locus positioning and with genome-wide DNA contact frequencies. Notably, our model recapitulates the position and morphology of the nucleolus, the observed variations in locus positions, and variations in contact frequencies within and across chromosomes, as well as subchromosomal contact features. The model is also able to correctly predict nuclear reorganization accompanying a reduction in rDNA transcription, and sites of chromosomal rearrangements tend to occur where the model predicted high contact frequencies.

Conclusions: Our results suggest that large-scale yeast nuclear architecture can be largely understood as a consequence of generic properties of crowded polymers rather than of specific DNA-binding factors, and that configurations of chromosomes and DNA contacts are dictated mainly by genomic location and chromosome lengths. Our model provides a quantitative framework to understand and predict large-scale spatial genome organization and its interplay with functional processes.

INTRODUCTION

Besides the one-dimensional information carried by the nucleotide sequence, the three-dimensional arrangement of the genome in the nucleus of eukaryotic cells emerges as an important determinant of gene expression, DNA repair, recombination and replication [1]. Although they lack any membrane apart from the nuclear envelope, nuclei from yeast to humans exhibit strong compartmentalization into nuclear bodies and other functionally distinct subdomains. In metazoans, chromosomes are confined to non-overlapping territories, whose relative positions in the nucleus are not random [2]. Individual loci occupy preferential, though dynamic, positions with respect to their chromatin territory or other nuclear landmarks such as the nuclear envelope or the nucleolus [3, 4]. These positioning patterns affect the transcriptional status of genes, as the nucleus appears compartmentalized in domains that either favor or silence gene expression [5]. Cancer-promoting chromosomal rearrangements arise from illegitimate fusions between broken parts of the same or distinct chromosomes. In mammals, some of these events occur preferably at pairs of genomic locations that are more frequently in contact, and are thus affected by spatial positioning of chromosomal regions [6-10]. Conversely, in yeast, many loci move to new subnuclear positions or change their dynamics upon changes in their expression or as a result of DNA breaks [5, 11-14].

Despite such functional relevance, the main factors and mechanisms that control dynamic nuclear organization are presently ill understood. These can be divided into two broad classes [15]. A first class includes protein complexes, such as CTCF, which bind to particular discrete DNA sequence motifs and promote their interaction with nuclear landmarks or distal loci, thus restricting subnuclear positions or forming loops or interchromosomal attachments [16, 17]. In a second class are generic effects arising from the properties of semi-flexible polymers (chromosomes) confined to the crowded nuclear volume. Generic properties have been proposed to explain, for example, the formation of chromosome territories, and the aggregation of large macromolecular complexes in the nucleus [18, 19]. While both specific factors and generic effects are present, their relative contributions remain unclear and a detailed predictive model of nuclear organization is not yet available.

The well-studied budding yeast *Saccharomyces cerevisiae* provides an attractive model to study nuclear organization and its functional relevance. Electron microscopy has revealed structural nuclear landmarks [20]; light microscopy has allowed to map the positions and dynamics of selected loci in individual nuclei [3, 21-25]; and a chromosome conformation capture (3C) assay [26] coupled with massive DNA sequencing (Hi-C) has provided a matrix of contact frequencies across the genome [27]. Despite its small diameter (~2 µm), the yeast nucleus is characterized by strong functional

compartmentalization [3, 5, 15]. The most prominent nuclear compartment is the nucleolus, the site of transcription of ribosomal RNA genes (rDNA), consisting of ~100-200 tandem repeats on the right arm of chromosome 12. In *S. cerevisiae,* the nucleolus is a single, crescent shaped structure abutting the nuclear envelope, encompassing roughly 1/3 of the nuclear volume, and excluding the bulk of the genome except for the rDNA [20, 22, 27]. What determines the position and shape of the nucleolus, as well as its segregation from the rest of the genome, remains unknown.

The budding yeast nucleus is further characterized by a distinct Rabl-like chromosome configuration, in which each chromosome's centromere is tethered by a single microtubule and the kinetochore complex to the spindle pole body (SPB), a multiprotein complex embedded in the nuclear envelope and located opposite of the nucleolus [3, 28]. Telomeres are tethered to the nuclear envelope via redundant pathways [29]. Their spatial position in the nucleus, as well as that of internal loci, correlates with genomic distance from the centromere [3, 22]. These data are qualitatively consistent with a relatively simple configuration of chromosomes governed by generic physical constraints [15]. By contrast, much more complex configurations, with the chromatin fiber criss-crossing the nucleus, have been proposed based on the reported colocalization of genes such as tRNA [30]. How exactly the chromatin fiber is organized in 3D yeast nuclear space thus remains unclear.

Contact frequencies measured by Hi-C have been used to construct a static, and more recently, a dynamic 3D model of yeast chromosomes in the nucleus [27, 31]. The dynamic model was found to be consistent with measurements of distances between telomeres [21]. However, both models [27, 31] assumed untested relationships between spatial distances and contact frequencies, and the dynamic model assumed an artificial nucleolar compartment [31]. As they rely on experimental data, such models cannot predict how nuclear organization changes in different experimental conditions.

Here, we present a predictive quantitative model of dynamic chromosome arrangements in the yeast nucleus. Our model is based on first principles rather than derived from imaging or Hi-C data. Nonetheless the model recapitulates observed patterns of intranuclear locus positioning and chromatin contacts across the genome. In addition, our model correctly predicted an alteration of nuclear architecture in response to a reduction of rDNA gene expression, and may be used to predict the propensity of different pairs of loci to undergo recombination.

RESULTS

Computational model of dynamic yeast chromosomes

We built a computational simulation of chromosome configurations and their dynamics in the yeast nucleus (**Fig. 1, Fig. S1, Movie S1**). Details are provided in **Materials and Methods.** We considered a nominal model and three control models (**Tables 1** and **2**). Briefly, we modeled the 16 chromosomes of haploid yeast as freely jointed chains of segments characterized by constant diameter, compaction and rigidity parameters (**Fig. 1A, Table 1**). Our nominal model reflects the specific nature of the rDNA chromatin (heteropolymer model). At the rDNA locus, ribosomal subunits are assembled cotranscriptionally, leading to strong accumulation of RNA and proteins [20]. To account for this, we increased the diameter of the rDNA segments, such that the effective volume occupied by rDNA was ~ 1/3 of the nucleus (**Fig. 1B, Table 1**). Each chromosome was linked to the SPB at its centromere by a single rigid microtubule (**Fig. 1C**). The telomeric ends were maintained near the nuclear envelope (represented by a spherical shell) by an outward force, but allowed to move freely along its surface. Assumed parameter values were based on the literature and are summarized in **Table 1**. Chromosomes were subjected to random thermal motions only (**Movie S1**). The model incorporated topological constraints, such that chain segments could not penetrate each other. From the simulated families of dynamic chromosome configurations (**Fig. S1I**), we then computed several features of nuclear organization, including intranuclear distributions for any locus (**Fig. S1J**), distances between any pair of loci (**Fig. S1K**), and contact frequencies between any pair of chromosomal regions (**Fig. S1L**). For comparisons, we also considered 3 control models: (i) a phantom model in which topological constraints were removed, (ii) a homopolymer model in which all chromosomes, including the rDNA, had the same properties (except their genomic length), and (iii) a microtubule-free model, in which centromeres were not linked to the SPB (**Table 2**).

Model recapitulates formation of the nucleolar compartment and quantitative locus positions

As a first test of our model, we compared the predicted subnuclear locations of selected loci to those obtained from imaging experiments [3, 22] (**Fig. 2**). These positions can be visualized as probability maps in a coordinate system ($R\cos\alpha$, $R\sin\alpha$), where R is the radial distance to the center of the nucleus, and α is the angle with respect to the axis joining the nuclear and nucleolar centers [3] (**Fig. S1J**). In our nominal model, centromeres occupied territories roughly halfway between the SPB and the nuclear center, while telomeres distributed themselves at the nuclear periphery, in accordance with observations (**Fig. 2A-E, G-K**). This is unsurprising since our model tethered centromeres to the SPB via microtubules and tethered telomeres to the nuclear envelope. By contrast, the subnuclear location of other loci was not built-in. Remarkably, our model predicted that the rDNA locus displayed a crescent-shape distribution abutting the nuclear envelope, and a position opposite the

SPB (**Fig. 2F**) [3, 20]. This morphology was strikingly similar to the rDNA territory determined by light microscopy (**Fig. 2L**) and to the dense nucleolus observed by electron microscopy [20]. The model predicted that all DNA except the rDNA is excluded from the nucleolus, such that the telomeres of even long arms cannot extend to the face opposite the SPB, in agreement with experimental data [3, 22]. By contrast, the control models failed to reproduce the territories of at least some loci: both the phantom model and the homopolymer model led to strikingly different patterns of localization (**Fig. 2A'-F',A''-F''**), while the microtubule-free model failed to position the rDNA opposite the SPB (**Fig. 2. A''',F'''**). Thus, our heteropolymer model qualitatively recapitulates experimentally observed features of nuclear organization, notably the morphology of the nucleolus, and the segregation of the rDNA from the rest of the genome.

We next turned to a more quantitative test of predicted locus positions. First, we analyzed the "absolute" intranuclear positions of selected loci. Because data from previous studies [3, 22] included mostly loci on different chromosomes, we performed new imaging experiments on 16 loci distributed along the right arm of chromosome 4, the second longest arm after the rDNA-carrying right arm of chromosome 12 (**Table S1**). Our data set encompassed 36 loci on 13 out of the 16 chromosomes, and included two loci on the right arm of chromosome 12. We first compared the predicted median angles α to measurements. Overall, predicted angles correlated remarkably well with the measurements (Pearson's r=0.87; $p<10^{-11}$), despite an underestimation of ≈18 deg (**Fig. 2M**). The model predicted that α correlated with genomic distance to the centromere (d_{CEN}) (**Fig. S2A**). This correlation had previously been observed for telomeres [22], but our new measurements showed that it also extends to internal loci, as predicted by the model (**Fig. S2B**). The model further predicted that loci on the rDNA carrying arm of chromosome 12 (GAL2 and rDNA) have larger α than other loci of similar d_{CEN} (**Fig. S2A**). This was again borne out by measurements (**Fig. S2B**).

We next considered the positioning of loci relative to each other. Since only the centromeres are tethered to a specific point, we reasoned that the most sensitive test of the model was for loci most distal from centromeres, i.e. the telomeres. We therefore compared our model's predictions against the 3D distances previously measured between 63 distinct pairs of subtelomeres [22], each of which contained one of three subtelomeres (6R, 10R, 4R), belonging to a short (122 Kb), middle (310 Kb) and long (1,050 Kb) chromosome arm, respectively. The correlation between predicted and measured distances was good (r=0.65, $p<10^{-7}$), despite an underestimation by the model of ~150 nm (**Fig. 2N**). The predicted distances strongly depended on the lengths of the two chromosome arms, in a manner approximately similar to that observed in experiments [22] (**Fig. S2C-H**). For short or medium arms, the two extremities of the same chromosome were predicted to be closer than for

pairs of arms with similar lengths on different chromosomes (**Fig. S2C,E**), also in agreement with measurements (**Fig. S2D,F**) [22].

In comparison to the nominal model, the homopolymer and phantom models both failed to explain the measured angles α (**Fig. S2J,K**), but predicted distances between telomeres also agreed with measurements (**Fig. S2N,O**). The microtubule-free model failed to explain both angles and distances (**Fig. S2L,P**) (**Table 2**). Thus, only the nominal model quantitatively accounts for absolute locus positions and the relative positions of telomeres.

Model recapitulates contact patterns of chromosomes and chromosome arms

While imaging provides detailed information on positions of a limited number of loci, Hi-C data allowed us to test the model on a genome-wide scale [27]. We first analyzed contact frequencies at the genomic scales of entire chromosomes (230-1500 Kb) (**Fig. 3**) and chromosome arms (80-1050 Kb) (**Fig. S3**). We considered the probability for a contact to occur between any pair of chromosomes (including within the same chromosome). In absence of any nuclear organization (i.e. if all pairs of loci in the genome randomly contact each other with uniform probability), this probability is the product of the chromosome's genomic lengths (**Fig. 3A**). In this case, only 14.0 % of all contacts should be intrachromosomal (*cis*) (**Fig. 3E**). In reality, 53.0 % of experimentally detected contacts are *cis* (**Fig. 3A,E**), indicating strong departure from a random collision scenario. Our nominal model, however, predicted 53.7 % of *cis* contacts, in remarkable agreement with experiments (**Fig. 3B,E**). The correlation between predicted and measured probabilities was very high ($r=0.99$, $\rho=0.96$, $p<10^{-50}$), indicating that the overall distribution of contacts among chromosomes was well recapitulated by our model (**Fig. 3D**). The model was also in good agreement with contact probabilities measured between pairs of chromosome arms (**Fig. S3A-D**). Here, 3 types of contacts exist: (i) between arms on different chromosomes, (ii) within each arm, (iii) between left and right arms of the same chromosomes. Unlike the random collision model, the nominal chromosome model predicted proportions of these 3 contact types in relatively good agreement with experimental data **(Fig. S3E)**.

However, we noted that, if taken separately, probabilities for interchromosomal contacts (*trans*) also agreed very well with the random collision scenario ($r=0.82$, $\rho=0.88$, $p<10^{-30}$) (**Fig. 3A,C**); a similar result held for chromosome arms ($r=0.87$, $\rho=0.88$, $p<10^{-50}$) (**Fig. S3A,C**). This suggested that, to first approximation, contacts between different chromosomes or arms occur indiscriminately. To remove this effect, we next computed contact probabilities averaged over the genomic length of the chromosome (or arms) (**Fig. 3F-I** and **Fig. S3F-I**). Because random collisions predicted uniform

average contact frequencies (**Fig. 3F, Fig. S3F**), any deviations from constancy reflect non-random nuclear organization. Overall, the average contact frequencies between chromosomes (*cis* + *trans*) predicted by the model still correlated very well with measurements (r=0.97, ρ=0.80, $p<10^{-50}$) (**Fig. 3G,H,I**). This was also true for contacts between arms (r=0.91, ρ=0.86, $p<10^{-50}$) (**Fig. S3G,H,I**). Taken separately, the 16 predicted *cis* contact frequencies were also in very good agreement with measurements (r=0.91, ρ=0.94, $p<10^{-5}$) (**Fig. 3I**). There was likewise good agreement for contacts within each of the 32 chromosome arms (r=0.69, ρ=0.73, $p<10^{-4}$) (**Fig. S3I**). Most significantly, the predicted *trans* contact frequencies of chromosomes, taken separately, also correlated very well with the measurements (r=0.84, ρ=0.72, $p<10^{-30}$) (**Fig. 3I**), as did predicted contacts between arms of different chromosomes (r=0.86, ρ=0.82, $p<10^{-50}$) (**Fig. S3I**). The *trans* frequencies agreed well, although less so, with those predicted by the homopolymer model (r=0.63, ρ=0.52 for chromosomes, r=0.79, ρ=0.73 for arms), but not with those predicted by the microtubule-free model (r=-0.004, p=0.97 ; ρ=-0.03, p=0.72 for chromosomes, r=0.04, p=0.42 ; ρ=0.12, p=0.01 for arms) (**Fig. S2R,U,S,V**) (**Table 2**). Note that the phantom model by definition did not predict any contacts.

Thus, our nominal model largely recapitulates genome-wide contact frequencies at the scale of entire chromosomes or chromosome arms.

Contact frequencies at subchromosomal scales

Next, we considered contact frequencies at the maximum resolution afforded by our model, ie. in *5* Kb bins. At this genomic resolution, the experimental contact matrix is very sparse, with on average only 1.4 contacts per bin, leading to strong counting noise (mean coefficient of variation (c.v.) $1.4^{-1/2}$=0.84, compared to 131 contacts per bin and c.v.<0.1 for the predicted matrix). The full predicted contact matrix correlated only weakly with its experimental counterpart (r=0.24) (**Fig. S4A,B**). However, trading off genomic resolution to reduce statistical noise strongly improved the correlation, which reached r=0.85 for bins of 75 Kb (**Fig. S4C**).

Both predicted and measured intrachromosomal matrices are dominated by a strong diagonal, i.e. by contacts between genomically proximal loci, as expected for polymers with persistence lengths larger than the genomic resolution. We first analyzed how contact frequencies *F* between pairs of loci on the same chromosome fall off with genomic separation *s* (in Kb). Our model predicted that *F(s)* decayed approximately as $s^{-1.5}$ between ~5 Kb and ~1000 Kb, and was roughly constant beyond (**Fig. 4A**). This is as expected for equilibrated confined polymers [32, 33] and differs from the s^{-1} decay observed in mammalian cells and attributed to an out-of-equilibrium fractal globule structure [9, 32].

The predicted $s^{-1.5}$ decay and the plateau above ~1000 Kb agreed well with the measurements (**Fig. 4B**).

We next considered intrachromosomal contacts. For each chromosome, the model predicted a contact pattern featuring a 'negative cross' centered on the centromere, indicating a segregation of centromeric regions from the rest of the chromosome (**Figs. 4C, S4A**). This feature was also apparent in the experimental data **(Fig. 4D, S4B).** For chromosome 12, which carries the rDNA locus, the model predicted a striking dearth of contacts between the genomic regions on either side of the rDNA **(Fig. 4E)**. This was also observed in the Hi-C data [27] (**Fig. 4F**).

Finally, we examined contacts between distinct chromosomes. The predicted contact patterns were mainly characterized by an enrichment between pericentromeric regions (**Fig. 4G, S4A**), in accordance with the Hi-C data [27] (**Fig. 4H, S4B**). Our model also predicted a weaker "butterfly" pattern characterized by a depletion of contacts between centromeres and non-centromeric regions of other chromosomes (**Fig. 4G, S4A**). This feature was barely discernable in the much noisier experimental map (**Fig. 4H, S4B**), thus a confirmation of this prediction may require new Hi-C data.

Thus, our model is able to recapitulate the main observed patterns of contact frequencies at sub-chromosomal scales in both *cis* and *trans*.

Predicting alterations of nuclear architecture:

To test our model's predictive power, we sought to simulate an alteration of nuclear organization amenable to experimental verification. Given the prominence of the nucleolar compartment and its origin in the high transcription of rDNA, we analyzed how the model responded to a reduction in the transcriptional activity of this locus. This was previously achieved experimentally by treating cells with rapamycin, resulting in an approximate halving of the nucleolar volume [22]. To model this effect, we diminished the diameter of rDNA segments to D_{rDNA}= 140 nm. The predicted rDNA territory had a reduced volume, but still resembled a crescent abutting the nuclear envelope opposite the SPB, in agreement with observations [22]. We analyzed the effect of the reduced nucleolus on the absolute and relative positions of telomeres (**Fig. 5**)**.** The model predicted that rapamycin treatment causes telomeres to extend to larger α than in absence of the drug and that the increase in α was larger for longer chromosome arms; this prediction was in good agreement with measurements for 6 telomeres whose arm lengths ranged from 85 Kb (Tel6R) to 1050 Kb (Tel4R) [22] (**Fig. 5A-N**). The predicted and measured changes in α agreed well (r=0.82, p=0.0011) (**Fig. 5N**). Similarly, the model predicted an increase in the distances of two telomere pairs (6R-3L and 6R-4R), in accordance with

previous measurements [22] (**Fig. 5O**). Thus, the model correctly predicted nuclear reorganization in response to a global change in rDNA transcription.

Discussion

We have presented a new computational model of global chromosome arrangements in the yeast nucleus. Unlike other recent models [27, 31, 34], ours did not rely on measured contact frequencies or light microscopy data, but instead used first principles and assumed few parameters all based on preexisting literature. The model represented chromosomes as confined (hetero)polymers undergoing passive Brownian dynamics, subject only to steric and topological constraints (**Fig. 1**). We did not assume specific DNA binding factors, except implicitly through the assumption of modified chromatin properties at the rDNA locus and via the tethering of telomeres and centromeres. In contrast to other models designed for human cells [33, 35, 36], ours made no provisions for loops. We also did not assume any active motions. Despite this simplicity, the model accounts qualitatively and quantitatively for key aspects of yeast nuclear organization: the morphology and position of the nucleolus, its exclusion of all DNA except the rDNA, the non-random positions of genes and telomeres relative to nuclear landmarks and to each other (**Fig. 2**), and patterns of contact frequencies across the genome, at the scale of chromosomes, chromosome arms, and at subchromosomal scales (**Figs. 3, 4, S3, S4**). Other models investigated here failed to explain all of these features simultaneously, although for some features the agreement was also good (**Table 2, Fig. S2**). Our nominal model successfully predicted a change in nuclear organization in conditions of reduced rDNA gene expression (**Fig. 5**). Nevertheless, it will be important to test its predictions in additional experiments that alter the mechanical constraints on chromosomes or chromosome structure, or in other yeast species.

Our results have implications for the mechanisms governing nuclear organization in yeast. Although we cannot rule out specific interactions binding chromatin loci to each other or to nuclear landmarks besides centromeres and telomeres, our data suggest that such interactions are not required to explain global large-scale organization of the nucleus. Similarly, our data suggest that no energy-dependent, e.g. molecular motor-driven, dynamics is required. Instead, our data support the notion that yeast nuclear organization can be understood by the sole properties of confined and topologically constrained polymers, combined with the tethering of centromeres and telomeres. More specifically, chromosome arrangements can be explained by entropic repulsion of topologically constrained chromosome arms [37], while the segregation of rDNA into a distinct nucleolar compartment is consistent with entropic phase separation of block heteropolymers [38] [39]. This

agrees with the notion that nuclear bodies arise by self-organization rather than by assembling onto preexisting scaffolds [40]. Our results are also in agreement with the conclusions of a different computational study (published during revision of this manuscript), which, however, did not self-consistently model the nucleolar compartmentalization [41].

Despite our model's success in recapitulating global features of nuclear organization, we acknowledge several limitations. First, the agreement with measurements is statistical and imperfect. Thus, other combinations of poorly known parameters (such as chromatin compaction and rigidity) or other equally simple models may explain observations even better. Furthermore, the experimental data considered here have only moderate spatial and genomic resolutions, and are based on large cell populations. Accordingly, specific rather than generic factors may still govern chromatin folding at smaller scales, or determine positions and dynamics of a subset of loci, or even exert global influence in a fraction of the cell population. Such specific factors may act during biological processes and underlie, e.g. the repositioning of inducible genes [42] or the clustering of replication origins [43]. To address this, more systematic explorations of the simulation parameter space will likely help improve modeling, but new imaging or Hi-C data with better spatial and genomic resolution are essential. Notwithstanding, our model may already facilitate the identification of specific DNA interactions from experimental data by providing a means to predict the unspecific effects.

In higher eukaryotes, substantial evidence points to the role of DNA-specific factors in mediating interactions between distal loci and organizing chromosomes into distinct large-scale domains [16, 17, 44]. To account for such observations, computational models may need to include specific interactions, but it is likely that entropic effects and polymer properties also need to be considered to understand nuclear organization in these organisms [18, 33, 45].

At the functional level, our model is likely to have direct implications in understanding where chromosomal rearrangements (which require physical interactions between distal loci) are most likely to occur. In support of this, we analyzed 96 known chromosomal breakpoints and found that predicted contact frequencies at the observed breakpoints were significantly higher than expected if breakpoint loci were randomly located in the genome (**Table S2**; **Fig. S5**). More experimental data are needed to refine these results in various classes of events. Nevertheless, this first analysis suggests that the model might be used to identify chromosomal regions that may spontaneously interact to generate chromosomal rearrangements. In the future, the model may also be used to examine constraints on gene repositioning during gene expression, the interaction of mating type loci, the formation of replication foci, and other functionally important aspects [42, 43, 46]. Overall, our

model constitutes a framework to quantitatively understand and predict global features of nuclear organization and some of its functional consequences in this important model organism.

Materials and methods

Computational model: The simulation used the open source physics engine (ODE), which allows to model the dynamics of rigid bodies subject to external forces and constraints arising from collisions and joints between distinct bodies (www.ode.org). We represented each of the 16 haploid yeast chromosomes as freely jointed chains of cylinders (**Fig. 1A**). The motion of each segment was governed by the discretized Langevin equation, which includes a randomly oriented force representing thermal agitation, and a viscous friction term, in addition to forces needed to respect the constraints arising from collisions and joints. The main parameters used in the model are listed in **Table 1**. In the homopolymer model, we assumed that the physical properties of chromatin are constant across the genome. Thus, the chains were uniquely characterized by only three parameters: cylinder length (Kuhn length L, in nm), which determines chain rigidity, cylinder diameter (D_0, in nm), and compaction, expressed as nucleotides per length (C, in bp/nm). The genomic length of each chromosome then uniquely determined the number of segments in each chain. Although the physical parameters (L, D_0, C) remain uncertain, we used values based on previous studies [47-49]. In the heteropolymer simulation, the N_{rDNA} segments corresponding to the rDNA locus were replaced by spheres of diameter D_{rDNA}. Except during the initial phase of the simulation (see below), we modeled the nuclear envelope as a sphere of radius R_0 = 1 µm, acting as an impenetrable boundary, and the SPB as a cylinder of diameter 120 nm embedded in the spherical shell (**Fig. 1C,D**). We modeled each of the 16 microtubule + kinetochore complexes as a single cylinder of length L_{MT} = 380 nm and diameter 25 nm (**Fig. 1D**) [50], one end of which remained in contact with the inner face of the SPB, while the other end was attached to the centromeric segment. To enforce telomere tethering to the nuclear envelope, we introduced a force that maintained the telomeric segments near the surface of the sphere without constraining their tangential motion. No other forces were introduced.

We initialized each simulation run with an artificial configuration in which chromosome chains were stretched out and parallel to each other, with their centromeres disposed at equal intervals along a circle of radius L_{MT} centered on the SPB, but in a random clockwise order (**Fig. S1A,B**). To accommodate this initial configuration, the initial nuclear envelope was modeled as a long cylinder of radius R_0 capped by two half spheres (**Fig. S1A**). During the initial phase of the simulation ($3.5\ 10^5$ time steps), which was not used to compute model predictions, the length of this cylinder was progressively reduced (**Fig. S1C**) to zero, until the nuclear envelope was a sphere of radius R_0

embedding the SPB (**Fig. S1D**). This initialization ensured that individual chains did not overlap and were not entangled with each other, as indeed the case for mitotic chromosomes. We plotted the time course of gyration radii and distances between chain extremities (telomeres) and verified that these quantities stabilized (i.e. fluctuated around a constant mean) before sampling the trajectories at large intervals (1 out of >1,000 time steps taken after 10^6 time steps) (**Fig. S1E,G**). We also checked that the autocorrelation of the sampled data was negligible for nonzero lags (**Fig. S1F,H**). Finally, to increase the sample size without excessive simulation time, we aggregated predicted loci and contact data from ~20 parallel simulation runs.

Strain constructions and microscopy: for fluorescent tagging of individual loci along chromosome 4 (**Table S1**), we constructed 16 new strains bearing Tet operator sequences near the locus of interest and expressing the repressor protein fused to GFP. Labeling of the nucleolus and nuclear pores, and spinning disc fluorescence microscopy were performed as previously described [3, 22].

Breakpoint analysis: We compiled a list of 96 breakpoints from the literature, which corresponded mainly to duplication and translocation events, and included 38 homologous and 56 non-homologous recombination events, 21 intrachromosomal and 75 interchromosomal events, 49 events from haploid and 47 from diploid cells (**Table S2**). These data were selected from experiments imposing as little constraints as possible on the position of at least one of the two breakpoint loci.

Acknowledgements

We acknowledge helpful discussions with O. Gadal, A. Grosberg, J-M. Victor, J. Mozziconacci, A. Rosa and A. Zidovska. We thank F. Mueller for comments on the manuscript. We thank the informatics services of Institut Pasteur for access to computing resources and P. Sorger for providing a yeast strain (CEN4). This work was funded by Institut Pasteur, Agence Nationale de la Recherche (grants ANR-09-PIRI-0024-1 and ANR-11-MONU-020-02), and Fondation pour la Recherche Médicale (Equipe FRM).

Table 1: Parameters of the nominal simulation

Parameter	Value
Number of chromosomes and microtubules	16
Kuhn length, L	60 nm
Nuclear radius, R_0	1 µm
Length of microtubules (+kinetochore), L_{MT}	380 nm
Compaction, C	83 bp/nm
Diameter of euchromatin segments, D_0	20 nm
Diameter of rDNA segments, D_{rDNA}	200 nm
Number of rDNA segments, N_{rDNA}	150
Segments can cross each other	No

Table 2: Simulations considered and comparison of their predictions. This table lists the nominal model and the three control models and indicates how their predictions agree with different measurements. Symbols + and - indicate good and poor agreement between model and measurements, respectively. NA, not applicable (no contact frequencies are scored for the phantom model).

Model	Difference with nominal simulation	Agreement with experimental data			
		Subnuclear probability maps	Absolute positions (α)	Relative telomere positions	Average inter-chromosomal contacts
Nominal	None	+	+	+	+
Phantom	Segments can cross	-	-	+	NA
Homopolymer	$D_{rDNA} = D_0$	-	-	+	+
Microtubule-free	No tethering to SPB	-	-	-	-

Figure Legends:

Figure 1: Computational model of the dynamic interphase yeast nucleus. (A) each chromosome is represented as a self-avoiding articulated chain of rigid segments. (B) the heterochromatic rDNA locus on chromosome 12 is represented by thicker segments (pink, displayed using surface smoothing) than the rest of the DNA (green). (C,D) A snapshot of the full model, showing each of the 16 chromosomes in a different color. The sphere represents the nuclear envelope. Two orthogonal views are shown: (C) view perpendicular to the axis joining the nuclear center to the SPB, (D): view along this axis and facing the SPB. (E) The SPB (white knob) and the 16 microtubules, represented each by one rigid segment. See also Fig. S1 and the animated simulation in Movie S1.

Figure 2: Model recapitulates nuclear compartmentalization, absolute and relative locus positions. (A-I**;** A'-F'; A''-F''; A'''-F''') Nuclear territories of selected genomic loci and the SPB, visualized as probability maps. (A-F): prediction by the nominal model; (G-L): measured from light microscopy data [3, 22, 51]; (A'-F'): predicted by the phantom model; (A''-F''): predicted by the homopolymer model. Probability maps were obtained and displayed as described in [3] and **Fig. S1J**. Dashed circles have a fixed radius of 1 mu. Territories shown in each column, from left to right, correspond to: the SPB (A,G,A',A'',A'''); the centromere of chromosome 4 (CEN4) (B,H,B',B'',B'''); the telomere on the 122 Kb long right arm of chromosome 6 (Tel6R) (C,I,C',C'',C'''); the telomere on the 310 Kb long right arm of chromosome 10 (Tel10R) (D,J,D',D'',D''') the telomere on the 1050 Kb long right arm of chromosome 4 (Tel4R) (E,K,E',E'',E'''), and an rDNA repeat (rDNA) (F,L,F',F'',F'''). Note the good agreement between observed territories and those predicted by the nominal model, but not the control models. (M,N): Quantitative comparison of predicted (nominal model) and measured absolute and relative locus positions. (M) Predicted vs. measured median angle α with respect to the nucleolar-nuclear axis. Each dot corresponds to a distinct locus as indicated by its number and **Table S1**. Red dots indicate loci along chromosome 4 that are new to this study, blue dots are data from previous work [3, 22]. (N) Predicted vs. measured median distances between 63 pairs of telomeres. Each dot corresponds to a different pair of telomeres. Pairs containing the reference telomere 6R, 10R or 4R have red, green and blue dots, respectively; the chromosome arm carrying the other telomere is indicated next to the dot (e.g. '3L' designates the left arm of chromosome 3). Squares indicate pairs of telomeres belonging to the same chromosome.See also Fig. S2.

Figure 3: Model recapitulates patterns of contacts among chromosomes. (A-C,F-H) contact frequency matrices for each pair of the 16 chromosomes. Chromosome numbers increase from top to bottom and from left to right. All matrices are displayed using the same logarithmic color scale, with dark colors indicating low probabilities, and bright colors high probabilities. (A,F) expected for random collisions; (B,G) predicted by the model; (C,H) measured [27]. (A-C) contact probabilities integrated over chromosomes (corresponding to the probability that a contact occurs between any pair of chromosomes). (F-H) averaged contact probabilities for each chromosome pair (probability per unit genomic length squared). (D,I) scatter plots of predicted vs. measured contact frequencies. Each of the 136 dots corresponds to a distinct pair of chromosomes. Blue dots denote *trans* interactions, red dots *cis* interactions. (E) proportions of *cis* and *trans* contacts expected for random collisions, predicted by the model, and measured. See Fig. S3 for corresponding analyses of contacts between chromosome arms.

Figure 4: Model recovers subchromosomal contact patterns. (A,C,E,G): model predictions. (B,D,F,H): measurements. (A,B): the solid curve shows the average intra-chromosomal contact frequency F as function of genomic separation s. The dotted and dashed lines indicate power laws s^{-1} and $s^{-1.5}$, respectively. Note the logarithmic scales. (C,D): contacts within chromosome 4; (E,F): contacts within chromosome 12; (G,H): contacts between chromosomes 4 and 15. All matrices are shown using a logarithmic color scale. The dotted lines indicate the position of the centromeres. See Fig. S4 for the entire genome-wide contact matrices.

Figure 5: Predicting alterations of nuclear organization under reduced rDNA transcription. (A-M) Intranuclear territories of 3 selected telomeres as predicted by the model (A-C,G-I) or measured [22] (D-F,K-M). (A-C) predicted by the nominal model with D_{rDNA}=200 nm. (D-F) measured experimentally, in absence of rapamycin [22]. (G-I) predicted by the model with D_{rDNA}=140 nm. (K-M) measured experimentally, in presence of rapamycin [22]. (N) measured vs. predicted median α of 6 telomeres in absence (blue) or presence (red) of rapamycine. Arrows indicate the change in α (predicted and measured) of each telomere upon treatment by rapamycine. The model correctly predicted an increase in α upon addition of rapamycin. (O) predicted vs. measured median distance between pairs of telomeres (6R, 3L) and (6R,4R) in each condition. Green dots: without rapamycin, red dots: with rapamycin. The model correctly predicted an increase in distances upon rapamycin treatment. See also Fig. S5.

478

479

480

References:

1. Misteli, T. (2005). Concepts in nuclear architecture. Bioessays *27*, 477-487.
2. Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb Perspect Biol *2*, a003889.
3. Berger, A.B., Cabal, G.G., Fabre, E., Duong, T., Buc, H., Nehrbass, U., Olivo-Marin, J.C., Gadal, O., and Zimmer, C. (2008). High-resolution statistical mapping reveals gene territories in live yeast. Nature Methods *5*, 1031-1037.
4. Heun, P., Laroche, T., Shimada, K., Furrer, P., and Gasser, S.M. (2001). Chromosome Dynamics in the Yeast Interphase Nucleus. Science *294*, 2181-2186.
5. Taddei, A., Schober, H., and Gasser, S.M. (2010). The budding yeast nucleus. Cold Spring Harb Perspect Biol *2*, a000612.
6. Misteli, T. (2010). Higher-order genome organization in human disease. Cold Spring Harb Perspect Biol *2*, a000794.
7. Wijchers, P.J., and de Laat, W. (2011). Genome organization influences partner selection for chromosomal rearrangements. Trends in genetics : TIG *27*, 63-71.
8. Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C., et al. (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature *484*, 69-74.
9. Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908-921.
10. Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107-119.
11. Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev *18*, 1119-1130.
12. Nagai, S., Dubrana, K., Tsai-Pflugfelder, M., Davidson, M.B., Roberts, T.M., Brown, G.W., Varela, E., Hediger, F., Gasser, S.M., and Krogan, N.J. (2008). Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. Science *322*, 597-602.
13. Mine-Hattab, J., and Rothstein, R. (2012). Increased chromosome mobility facilitates homology search during recombination. Nature cell biology *14*, 510-517.
14. Dion, V., Kalck, V., Horigome, C., Towbin, B.D., and Gasser, S.M. (2012). Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. Nature cell biology *14*, 502-509.
15. Zimmer, C., and Fabre, E. (2011). Principles of chromosomal organization: lessons from yeast. J Cell Biol *192*, 723-733.
16. Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.
17. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.
18. Rosa, A., and Everaers, R. (2008). Structure and Dynamics of Interphase Chromosomes. PLoS computational biology *4*.
19. Marenduzzo, D., Micheletti, C., and Cook, P.R. (2006). Entropy-driven genome organization. Biophys J *90*, 3712-3721.
20. Léger-Silvestre, I., Trumtel, S., Noaillac-Depeyre, J., and Gas, N. (1999). Functional compartmentalization of the nucleus in the budding yeast Saccharomyces cerevisiae. Chromosoma *108*, 103-113.

21. Bystricky, K., Laroche, T., van Houwe, G., Blaszczyk, M., and Gasser, S.M. (2005). Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. J Cell Biol *168*, 375-387.
22. Thérizols, P., Duong, T., Dujon, B., Zimmer, C., and Fabre, E. (2010). Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. Proceedings of the National Academy of Sciences *107*, 2025.
23. Marshall, W., Straight, A., Marko, J., Swedlow, J., Dernburg, A., Belmont, A., Murray, A., Agard, D., and Sedat, J. (1997). Interphase chromosomes undergo constrained diffusional motion in living cells. Curr. Biol *7*, 930–939.
24. Heun, P., Laroche, T., Raghuraman, M.K., and Gasser, S.M. (2001). The positioning and dynamics of origins of replication in the budding yeast nucleus. J Cell Biol *152*, 385-400.
25. Schober, H., Kalck, V., Vega-Palas, M.A., Van Houwe, G., Sage, D., Unser, M., Gartenberg, M.R., and Gasser, S.M. (2008). Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. Genome Res *18*, 261.
26. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.
27. Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363-367.
28. Rabl, C. (1885). über Zellteilung. Morphologisches Jahrbuch *10*, 214-330.
29. Hediger, F., Neumann, F.R., Van Houwe, G., Dubrana, K., and Gasser, S.M. (2002). Live Imaging of Telomeres. yKu and Sir Proteins Define Redundant Telomere-Anchoring Pathways in Yeast. Curr Biol *12*, 2076-2089.
30. Haeusler, R.A., and Engelke, D.R. (2004). Genome organization in three dimensions: thinking outside the line. Cell Cycle *3*, 273-275.
31. Tokuda, N., Terada, T.P., and Sasai, M. (2012). Dynamical Modeling of Three-Dimensional Genome Organization in Interphase Budding Yeast. Biophys J *102*, 296-304.
32. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.
33. Rosa, A., Becker, N.B., and Everaers, R. (2010). Looping probabilities in model interphase chromosomes. Biophys J *98*, 2410-2419.
34. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nature biotechnology *30*, 90-98.
35. Munkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T., and Langowski, J. (1999). Compartmentalization of interphase chromosomes observed in simulation and experiment. J Mol Biol *285*, 1053-1065.
36. Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E.M.M., Verschure, P.J., Indemans, M.H.G., Gierman, H.J., Heermann, D.W., and van Driel, R. (2009). Spatially confined folding of chromatin in the interphase nucleus. Proceedings of the National Academy of Sciences *106*, 3812.
37. De Gennes, P.G. (1980). Conformations of polymers attached to an interface. Macromolecules *13*, 1069-1075.
38. Leibler, L. (1980). Theory of microphase separation in block copolymers. Macromolecules *13*, 1602-1617.
39. Cook, P.R., and Marenduzzo, D. (2009). Entropic organization of interphase chromosomes. J Cell Biol *186*, 825-834.
40. Misteli, T. (2001). The concept of self-organization in cellular architecture. J Cell Biol *155*, 181-185.

41. Tjong, H., Gong, K., Chen, L., and Alber, F. (2012). Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. Genome Res *22*, 1295-1305.
42. Cabal, G.G., Genovesio, A., Rodriguez-Navarro, S., Zimmer, C., Gadal, O., Lesne, A., Buc, H., Feuerbach-Fournier, F., Olivo-Marin, J.C., Hurt, E.C., et al. (2006). SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. Nature *441*, 770-773.
43. Knott, S.R., Peace, J.M., Ostrow, A.Z., Gan, Y., Rex, A.E., Viggiani, C.J., Tavare, S., and Aparicio, O.M. (2012). Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. Cell *148*, 99-111.
44. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.
45. Finan, K., Cook, P.R., and Marenduzzo, D. (2011). Non-specific (entropic) forces as major determinants of the structure of mammalian chromosomes. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *19*, 53-61.
46. Bressan, D.A., Vazquez, J., and Haber, J.E. (2004). Mating type-dependent constraints on the mobility of the left arm of yeast chromosome III. J Cell Biol *164*, 361-371.
47. Dekker, J. (2008). Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. J Biol Chem *283*, 34532.
48. Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.
49. Cui, Y., and Denis, C.L. (2003). In vivo evidence that defects in the transcriptional elongation factors RPB2, TFIIS, and SPT5 enhance upstream poly(A) site utilization. Mol Cell Biol *23*, 7887-7901.
50. O'Toole, E.T., Winey, M., and McIntosh, J.R. (1999). High-voltage electron tomography of spindle pole bodies and early mitotic spindles in the yeast Saccharomyces cerevisiae. Mol Biol Cell *10*, 2017-2031.
51. Kunzler, M., Gerstberger, T., Stutz, F., Bischoff, F.R., and Hurt, E. (2000). Yeast Ran-binding protein 1 (Yrb1) shuttles between the nucleus and cytoplasm and is exported from the nucleus via a CRM1 (XPO1)-dependent pathway. Mol Cell Biol *20*, 4295-4308.

Figure 1

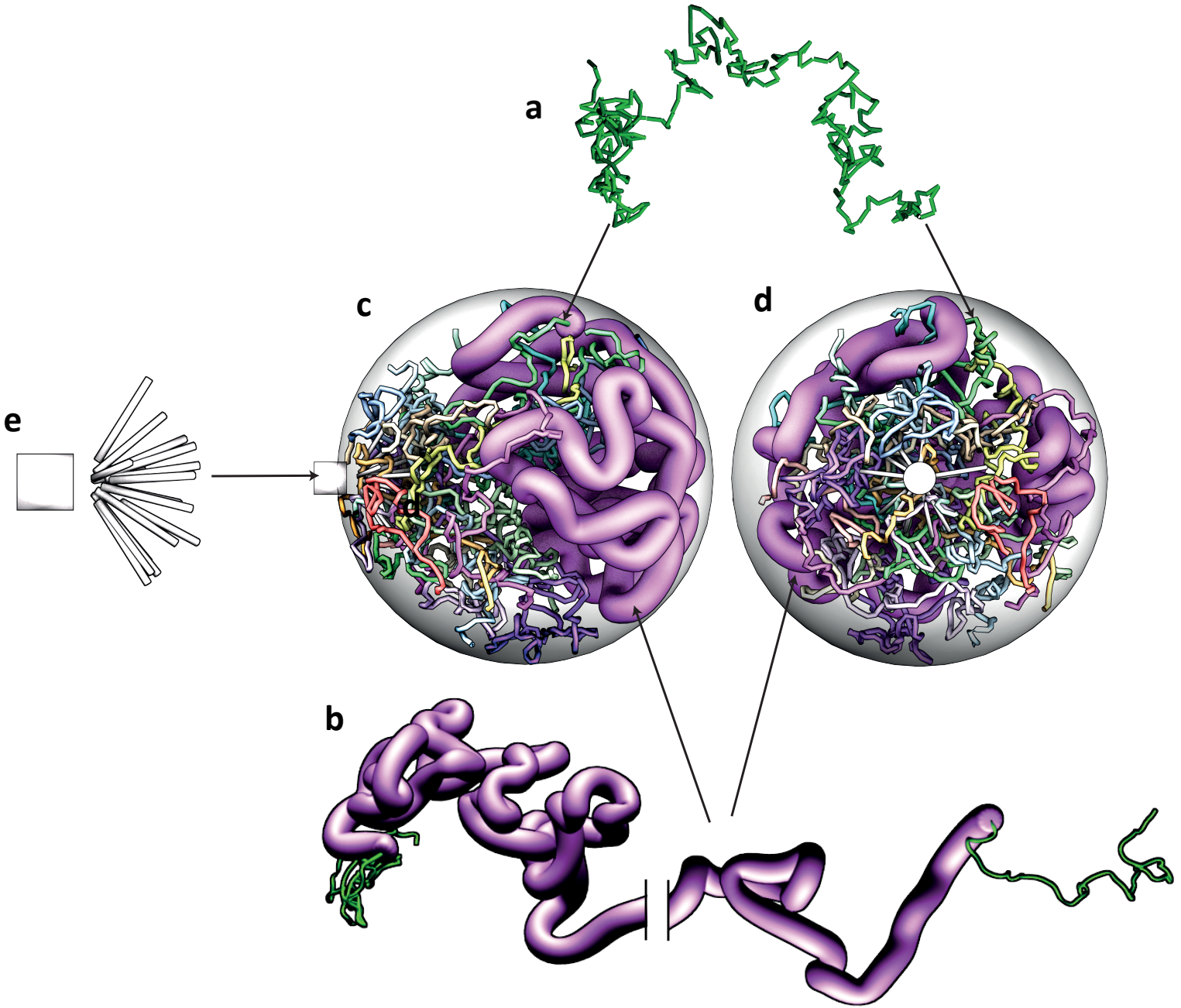

Figure 1

Figure 2

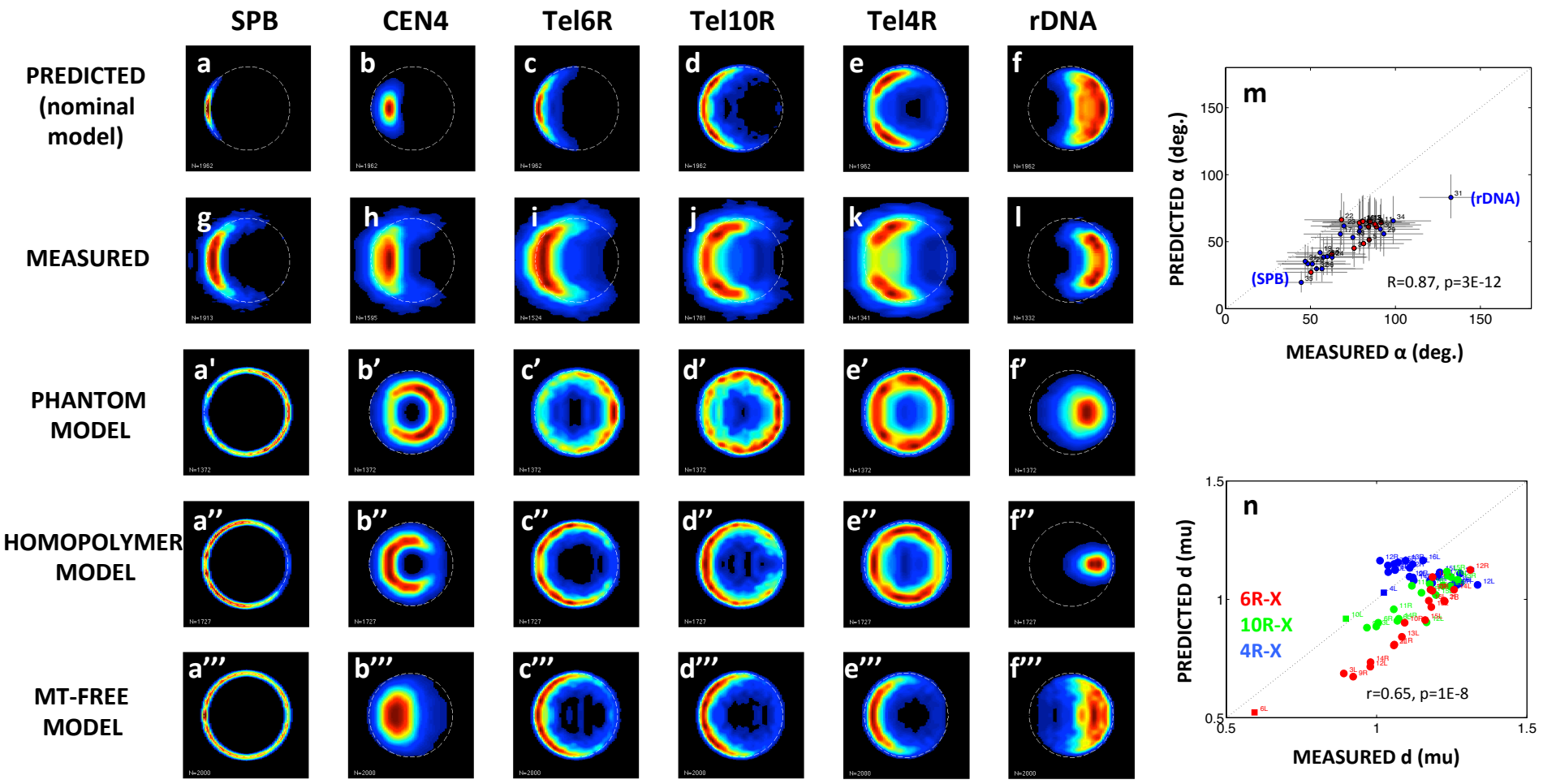

Figure 2

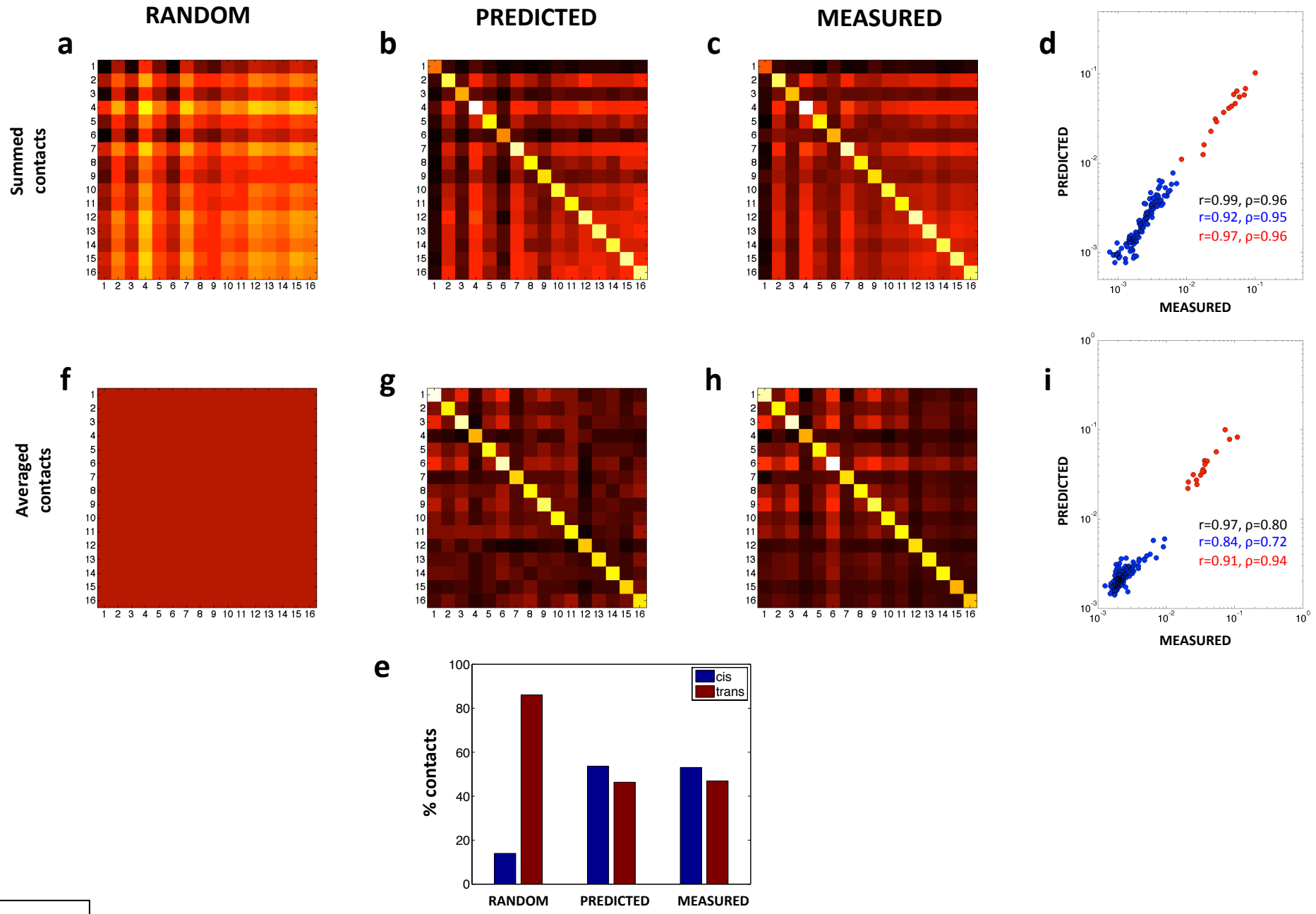
RANDOM
PREDICTED
MEASURED
a
b
c
d
Summed contacts
PREDICTED
MEASURED
r=0.99, ρ=0.96
r=0.92, ρ=0.95
r=0.97, ρ=0.96
f
g
h
i
Averaged contacts
PREDICTED
MEASURED
r=0.97, ρ=0.80
r=0.84, ρ=0.72
r=0.91, ρ=0.94
e
% contacts
cis
trans
RANDOM
PREDICTED
MEASURED

Figure 3

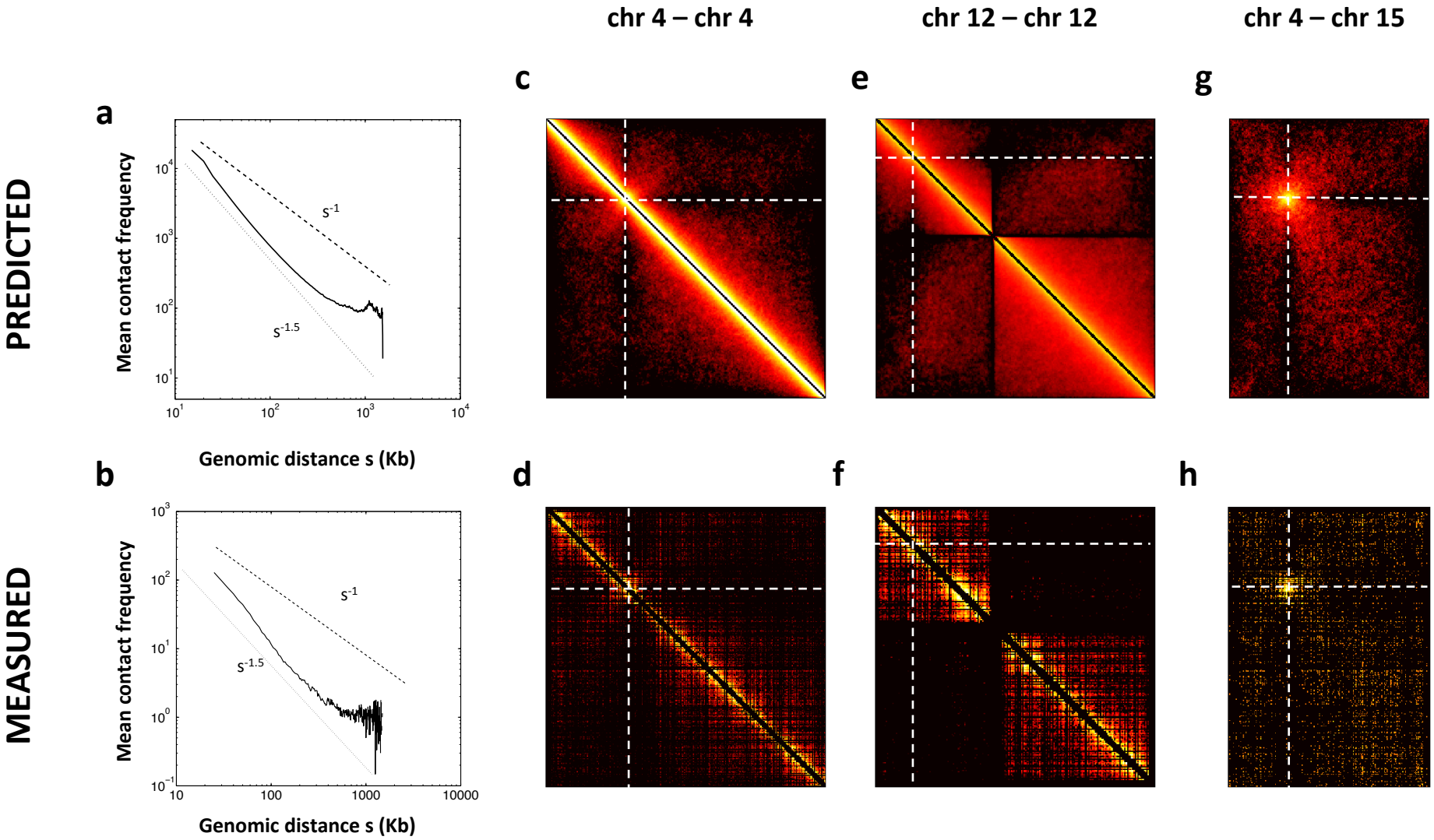
chr 4 – chr 4
chr 12 – chr 12
chr 4 – chr 15
PREDICTED
MEASURED
a
b
c
d
e
f
g
h
Mean contact frequency
Genomic distance s (Kb)
s^-1
s^-1.5

Figure 4

Figure 5

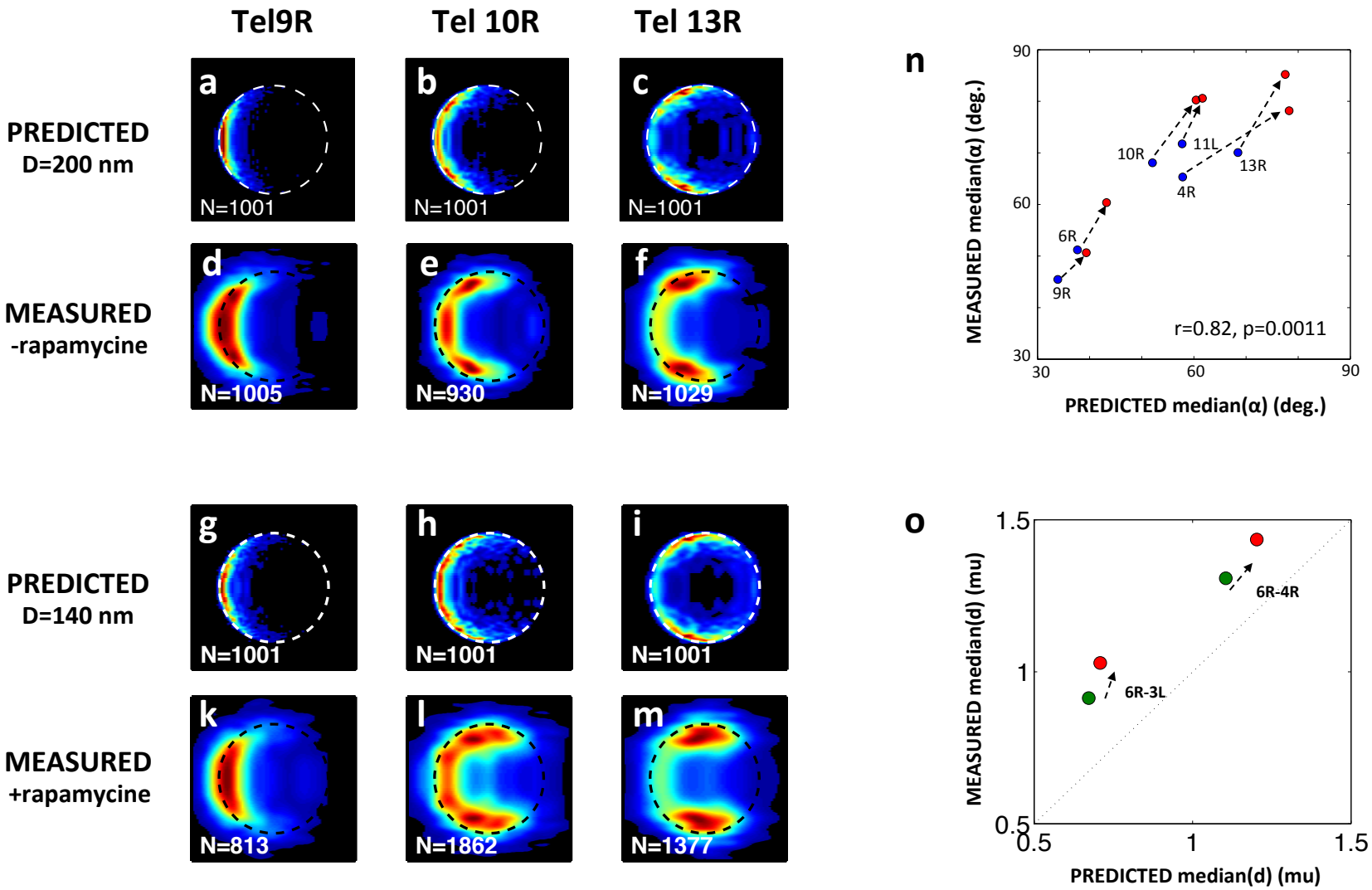

Figure 5

Supplemental Information for Wong et al.

Movie Legend:

Movie S1: Movie showing the dynamic simulation in progress. All 16 chromosomes are visible as moving chains tethered to the SPB on the left (not shown). The thick red structure represents the rDNA locus on the right arm of chromosome 12. The spherical nuclear envelope confines the chromosomes but is not shown.

Figure Legends

Figure S1: Generating testable predictions from the model. Data from many instances of the simulation (**a**) are used to produce several observables (**b**-**d**). (**b**) Intranuclear probability maps are obtained from N=1001 positions of an individual locus (here, URA3) sampled from the simulation trajectories, in a manner similar to that described for experimental data in Berger et al. Nat Meth (2008), but with an additional convolution with a Gaussian kernel of standard deviation 50 nm for better visualization. These positions can be defined either by the coordinates (R cos α, R sin α) (top panels) or by (R cos α', R sin α') (bottom panels) where R is the distance between the locus and the nuclear center, α (respectively α') is the elevation angles with respect to the axis joining the nuclear center to the mass center of the nucleus (respectively the SPB) (red dashed lines). The sampled positions in these two coordinate systems are shown as scatter plots in the middle. The heat maps on the right are probability densities obtained from these positions, with hot and cold colors indicating high and low probabilities, respectively. The dashed circle has a radius of 1 μm and represents the nuclear envelope. In experiments, the nucleolar mass center was used to define the central axis α = 0, thus the top coordinate system is relevant for comparisons with measurements (Berger et al. Nat Meth 2008). For the simulation used in this example, the mass center of the nucleolus was approximately aligned with the nuclear center and the SPB, therefore the positions in the two coordinate systems and the corresponding probability maps are similar. (**c**) Distribution of distances between a pair of loci obtained from 1001 snapshots of the simulation (here telomeres on the right arms of chromosomes 1 and 3). The curve shows the cumulative distribution function of distances. (**d**) Contact frequency matrix obtained by scoring contacts (collisions) between pairs of segments in the simulation. The map shows a histogram of collisions in bins corresponding to a predefined genomic resolution (here 5 Kb).

Figure S2: Locus territories predicted by the nominal model and control models, in a reference frame aligned with the SPB. Panels show intranuclear probability maps for the same loci and models as in Fig. 2, with the only difference that here the central axis (horizontal) is defined by the line joining the nuclear center to the SPB (instead of the nucleolar center) (see Fig. S1b). Note that the phantom and homopolymer models predict large rDNA territories at the center of the nucleus (**f',f''**), while the microtubule-free model predicts a widespread distribution in the nucleoplasm. Only the nominal model predicts an rDNA territory with a position (opposite the SPB) and morphology consistent with experimental data (**f**).

Figure S3: Absolute and relative locus positions as function of genomic location.
(**a,b**) Median angle with respect to central axis as function of the genomic distance to the centromere, predicted by the nominal model (**a**) and measured in experiments (**b**). Each dot corresponds to a distinct locus, indicated by the number (see **Table S1**). Red dots indicate loci along chromosome 4 for which measurements were obtained in this study. Blue dots correspond to loci for which positions were measured in previous studies (Berger et al. Nat Meth 2008, Thérizols et al. PNAS 2010). The rDNA and GAL2 loci are both located on the right arm of chromosome 12. (**c-h**) Median distance between telomeres as function of arm length, as predicted (**c,e,g**) or measured (Thérizols et al. PNAS 2010) (**d,f,h)**. Each panel shows the median distance between one of three reference telomeres and another telomere, as function of the arm length of the latter telomere. The reference telomeres are Tel6R (**c,d**), Tel 10R (**e,f**), and Tel4R (**g,h**). Panels **d,f** and **h** are identical to Fig. 1A,B,C in Thérizols et al. PNAS 2010.

Figure S4: Predicted vs measured position and contact data for nominal and control models. Predicted quantities are plotted against their measured counterparts for the nominal model (top row) and three control models (other rows, as labeled). Panels **a**, **e**, **i** and **l** are identical to **Fig. 2m**, **Fig. 2n**, **Fig. 3i**, and **Fig. S5i**, respectively and reproduced here for easier comparison. (**a-d**) Angles α for various genomic loci, plotted as in **Fig. 2m**. **(e-h)** Distances between telomeres, plotted as in **Fig. 2n**. (**i-k**) Average contact frequencies between pairs of chromosomes, plotted as in **Fig. 3i**. (**l-n**) Average contact frequencies between pairs of chromosome arms, plotted as in **Fig. S5i**. Note that in the phantom model, no contacts were scored because no collisions were detected.

Figure S5: Model recapitulates patterns of contact frequencies among the 32 chromosome arms. (a-c,f-h) contact frequency matrices for each pair of chromosome arms. The chromosome number and arm is indicated on the axes ('L' and 'R' indicates left and right arm, respectively). All matrices are displayed using the same logarithmic color scale, with dark color indicating low, and bright colors high probabilities. (**a,f**) contact frequencies expected for random collisions. (**b,g**) contact frequencies predicted by the model. (**c,h**) measured contact frequencies (Duan et al. Nature 2010). **(a-c)** contact probabilities integrated over each arm (corresponding to the probability that a contact occurs between distinct pairs of arms). **(f-h)** contact probabilities averaged over each arm (corresponding to a probability per unit genomic length). (**d,i**) comparison of predicted vs. measured contact frequencies. Each of the 528 dots corresponds to a distinct pair of chromosome arms. Blue dots denote contacts between arms on distinct chromosomes, red dots denote contacts within each of the 32 arms, green dots indicate contacts between the left and right arm of each of the 16 chromosomes. (**e**) proportions of these three types of contacts (as indicated by the corresponding color) in the experimental data, the model, and for random collisions.

Figure S6: Genome-wide contact matrices in experiment and simulation. (**a,b**) Genome-wide contact matrices, as predicted (**a**) and measured (Duan et al. 2010) (**b**). Each matrix shows the contact frequencies on a logarithmic color scale. Each matrix element (pixel) corresponds to a bin of 5 Kb. Boundaries between chromosomes are indicated by solid white lines. The positions of centromeres are indicated by dashed white lines. Chromosomes are ordered from 1 (top left) to 16 (bottom right). Each

chromosome is displayed from left (top left) to right (bottom right). The total number of contacts for each matrix is indicated. (**c**) Correlation between predicted and measured genome-wide contact matrices as function of genomic resolution (bin size). The correlation (Pearson's r) increases rapidly as the genomic resolution used to bin the contact data increases.

Figure S7: Predicting preferred breakpoint locations. Cumulative distribution functions of contact frequencies predicted by the model at 96 experimentally obtained breakpoint locations (solid black) and as expected for pairs of loci located randomly on the genome (dashed grey). The predicted contact frequencies at the observed breakpoints are significantly higher than compared to the random distribution (Kolmogorov-Smirnov test $p=1.4\ 10^{-4}$). This enrichment is also significant when separately considering homologous or non-homologous recombination events (p=0.013 and $p=8\ 10^{-4}$, respectively), or *cis* interactions ($p<10^{-10}$) (not shown). The enrichment is significant for haploid cells ($p=1.8\ 10^{-5}$), but not for *trans* interactions (p=0.6) or diploid cells (p=0.20) (not shown).

Figure S8: Initialization, equilibration and sampling.
(**a**) Initial configuration, showing the 16 chromosomes stretched out linearly, with each centromere attached by a radial microtubule to the SPB. To accommodate the longest chromosome, the nuclear envelope initially has a pill-like shape, with a diameter of 2 mu and a length of 83 mu (for better visualization, the figure is not to scale). (**b**) The chromosomes are arranged in random order around the SPB as shown here for two independent simulations. (**c**) As the simulation proceeds, the length of the nuclear envelope is progressively reduced. (**d**) After ~350,000 time steps, the nuclear envelope is a perfect sphere and no longer changes. (**e-h**) Time course and autocorrelation functions (ACF) of the gyration radius of chromosome 4 (**e,f**) and the distance between telomeres Tel4R and Tel6R (**g,h**). In panels (**e,g**), the dotted line indicates when the nuclear envelope is spherical, the dashed line indicates the time after which the simulation is sampled to generate predictions -earlier time steps are not sampled. Note that after this time point the gyration radius and inter-telomeric distance fluctuate around a constant mean value. (**f,h**): ACF of gyration radius and inter-telomeric distance as function of time lag, based on samples taken after time step 10^6. Note that the ACF becomes negligible for non-zero time lags, indicating that samples are not correlated and consistent with equilibration of the system.

Supplemental Table S1 : Experimental data on intranuclear locus positioning.

Each row in this table corresponds to a distinct genomic locus for which position measurements were obtained either in this study or in refs. [1, 2]. The position data are used in analyses of **Fig. 2** and **Fig. S4**.

Number	Locus name	Chromosome arm	Kb to centromere	Kb to telomere	Number of cells	Reference
1	CEN4	4R	0	1086	1595	This study; [3]
2	YDR042	4R	91	995	1562	This study
3	YDR068	4R	134	952	1249	This study
4	YDR095	4R	187	899	729	This study
5	YDR117	4R	235	850	1862	This study
6	YDR215	4R	445	641	850	This study
7	YDR234	4R	481	604	641	This study
8	YDR278	4R	567	518	341	This study
9	YDR297	4R	607	479	1470	This study
10	YDR336	4R	695	390	1500	This study
11	YDR336	4R	695	390	836	This study
12	YDR401	4R	822	263	510	This study
13	YDR422	4R	868	219	1471	This study
14	YDR467	4R	948	138	729	This study
15	YDR491	4R	985	101	666	This study
16	YDR514	4R	1023	63	1800	This study
17	Tel10R	10R	290	20	1781	[1]
18	Tel11L	11L	439.9	2.1	1995	[1]
19	Tel14R	14R	152.6	4.4	3328	[1]
20	Tel1L	1L	149.2	1.8	1837	[1]
21	Tel1R	1R	57.8	22.2	1272	[1]
22	Tel4R	4R	1030.2	19.8	2682	[1]
23	Tel5R	5R	421.2	8.8	2080	[1]
24	Tel6L	6L	133.2	14.8	2476	[1]
25	Tel6R	6R	120.9	1.1	1524	[1]
26	Tel7L	7L	485.7	6.3	1395	[1]
27	Tel9R	9R	68.2	16.8	4758	[1]
28	GAL1	2R	41	53	1857	[2]
29	GAL2	12R	139	787+rDNA	2083	[2]
30	HMO1	4R	362	724	3628	[2]
31	rDNA	12R	300-2300	600-2600	2663	[2]
32	RPS20	8L	30	75	5158	[2]
33	RPS5	10R	215	93	4211	[2]
34	SNR17A	15R	451	311	3547	[2]
35	SPB	NA	NA	NA	1913	[2]
36	URA3	5L	35	116	5472	[2]

1. Thérizols, P., Duong, T., Dujon, B., Zimmer, C., and Fabre, E. (2010). Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. Proceedings of the National Academy of Sciences *107*, 2025.
2. Berger, A.B., Cabal, G.G., Fabre, E., Duong, T., Buc, H., Nehrbass, U., Olivo-Marin, J.C., Gadal, O., and Zimmer, C. (2008). High-resolution statistical mapping reveals gene territories in live yeast. Nature Methods *5*, 1031-1037.
3. He, X., Asthana, S., and Sorger, P.K. (2000). Transient sister chromatid separation and elastic deformation of chromosomes during mitosis in budding yeast. Cell *101*, 763-775.

Supplemental Table S2 : List of breakpoints compiled from the literature. The Table is provided as a separate Excel spreadsheet.

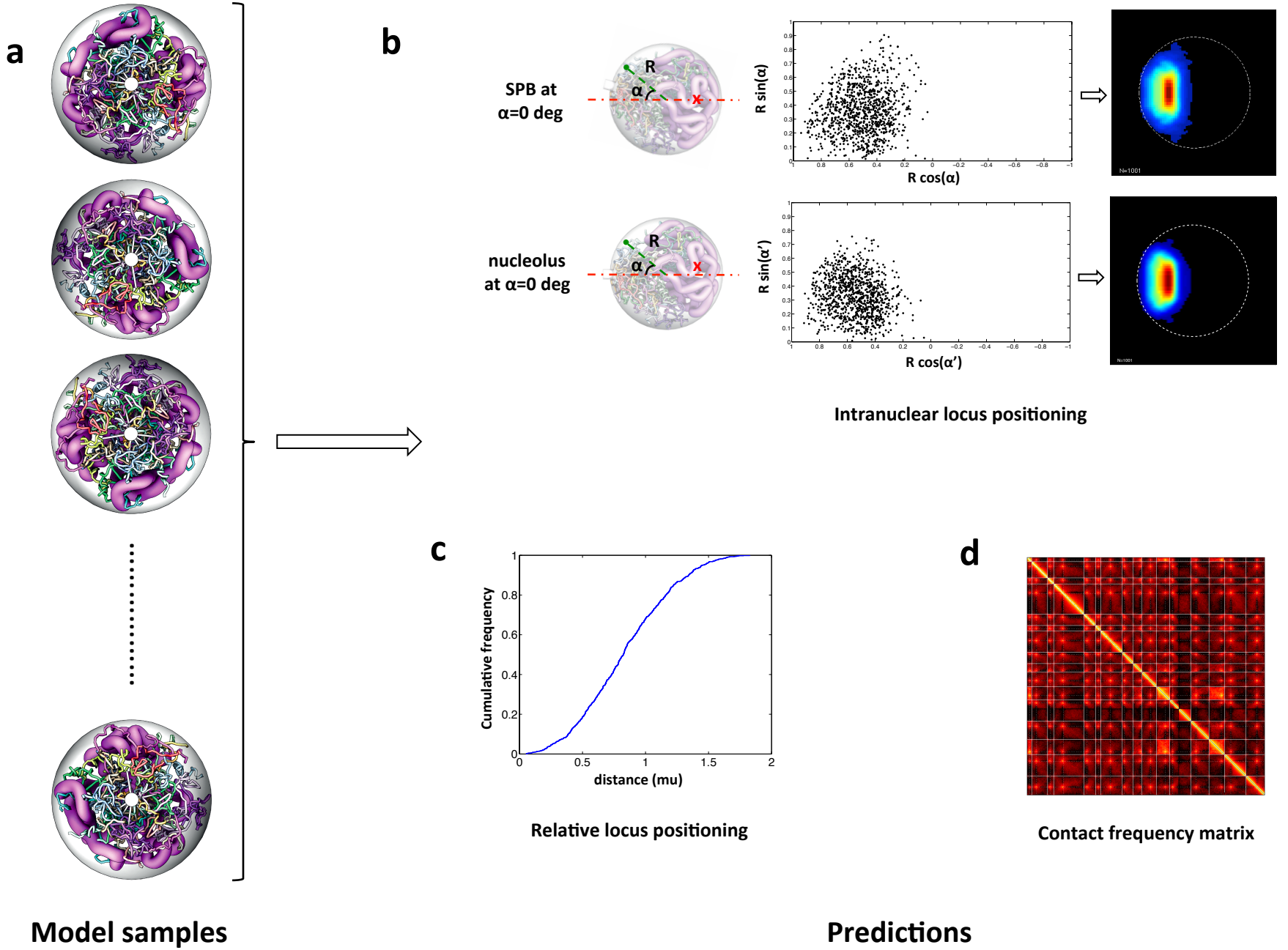
a
b
SPB at
α=0 deg
R
α
x
R sin(α)
R cos(α)
N=1001
nucleolus
at α=0 deg
R sin(α')
R cos(α')
Intranuclear locus positioning
c
Cumulative frequency
distance (mu)
Relative locus positioning
d
Contact frequency matrix
Model samples
Predictions

Figure S1

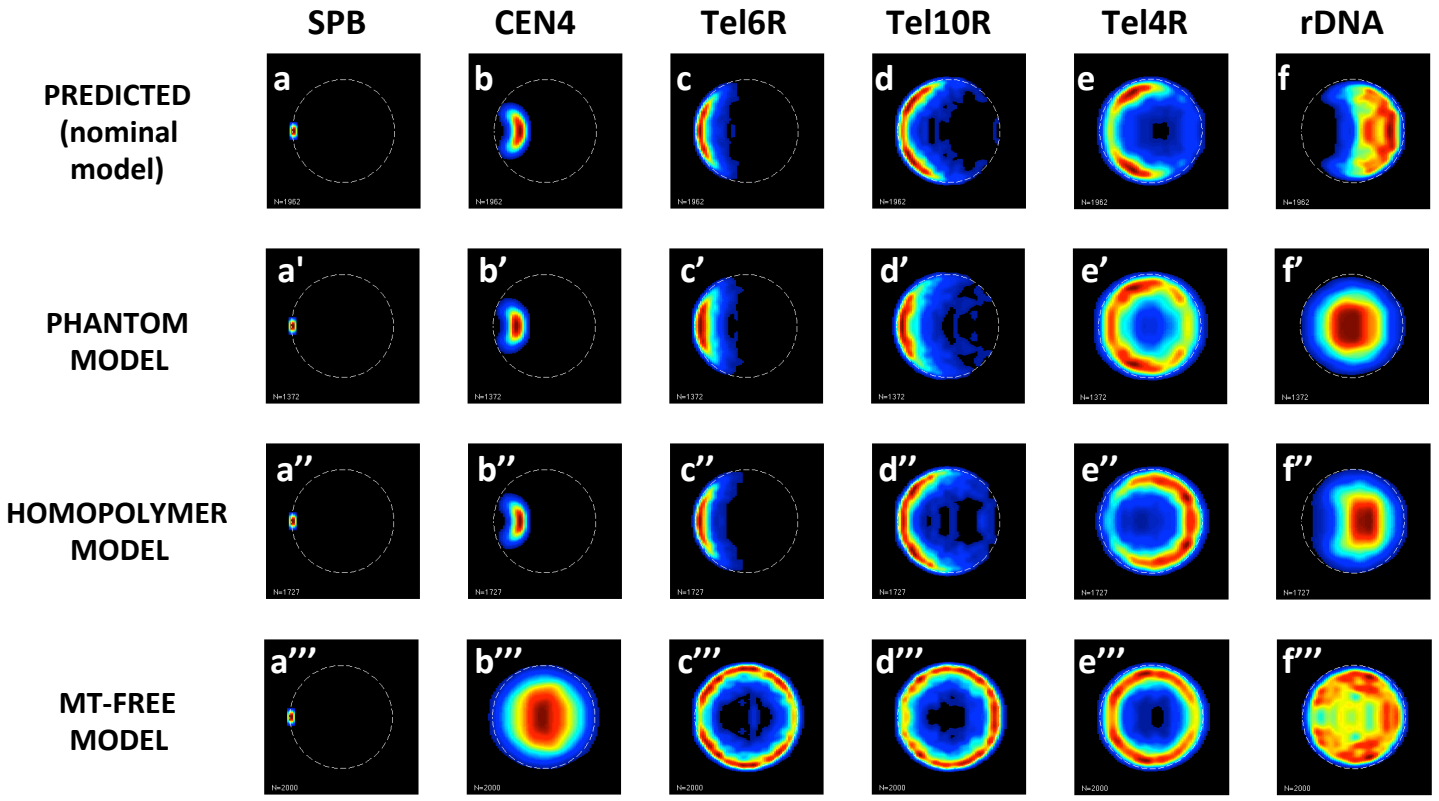
SPB
CEN4
Tel6R
Tel10R
Tel4R
rDNA
PREDICTED (nominal model)
PHANTOM MODEL
HOMOPOLYMER MODEL
MT-FREE MODEL
a
b
c
d
e
f
a'
b'
c'
d'
e'
f'
a''
b''
c''
d''
e''
f''
a'''
b'''
c'''
d'''
e'''
f'''

Figure S2

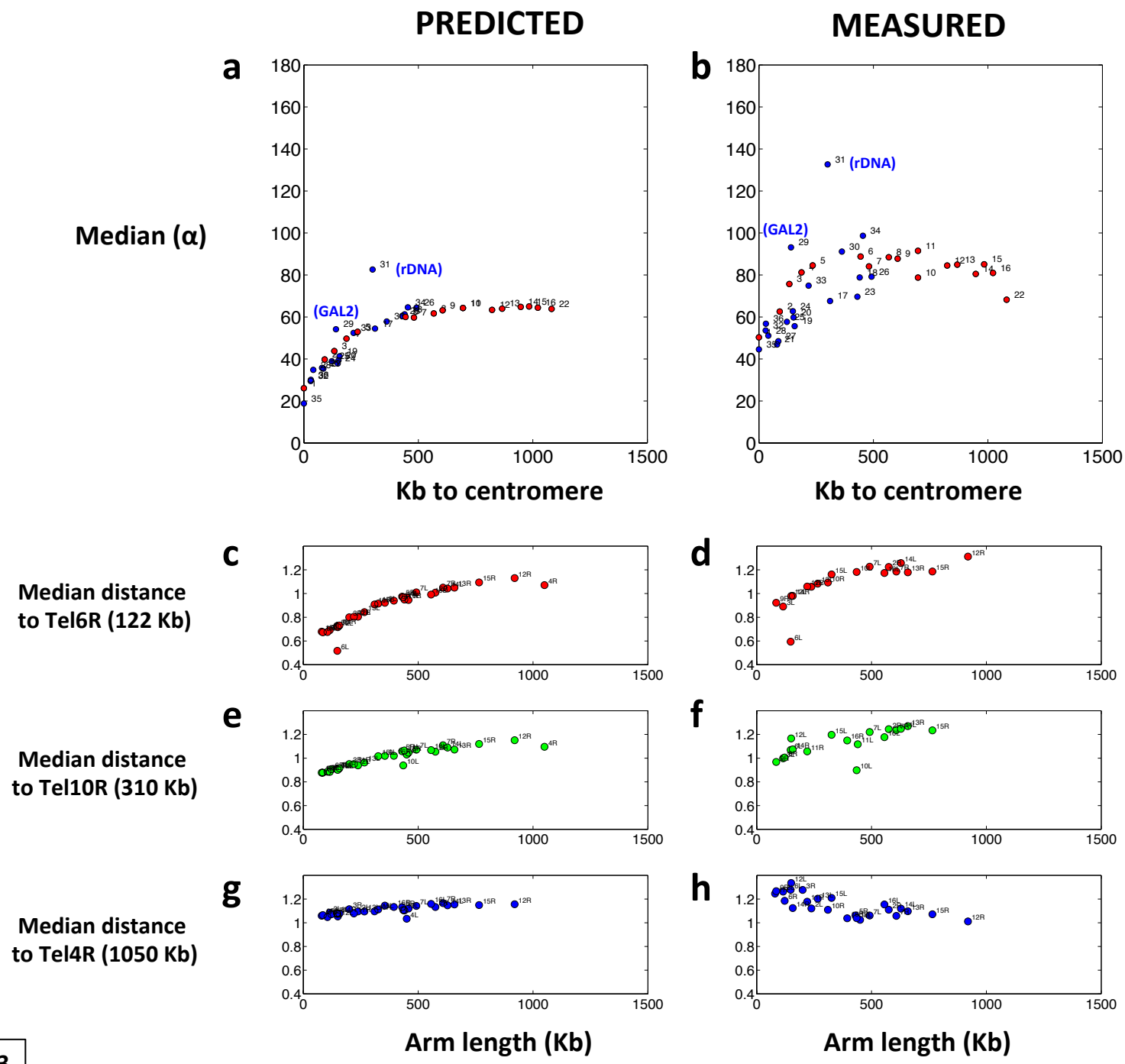
PREDICTED
MEASURED
a
b
Median (α)
(rDNA)
(GAL2)
Kb to centromere
c
d
Median distance to Tel6R (122 Kb)
e
f
Median distance to Tel10R (310 Kb)
g
h
Median distance to Tel4R (1050 Kb)
Arm length (Kb)

Figure S3

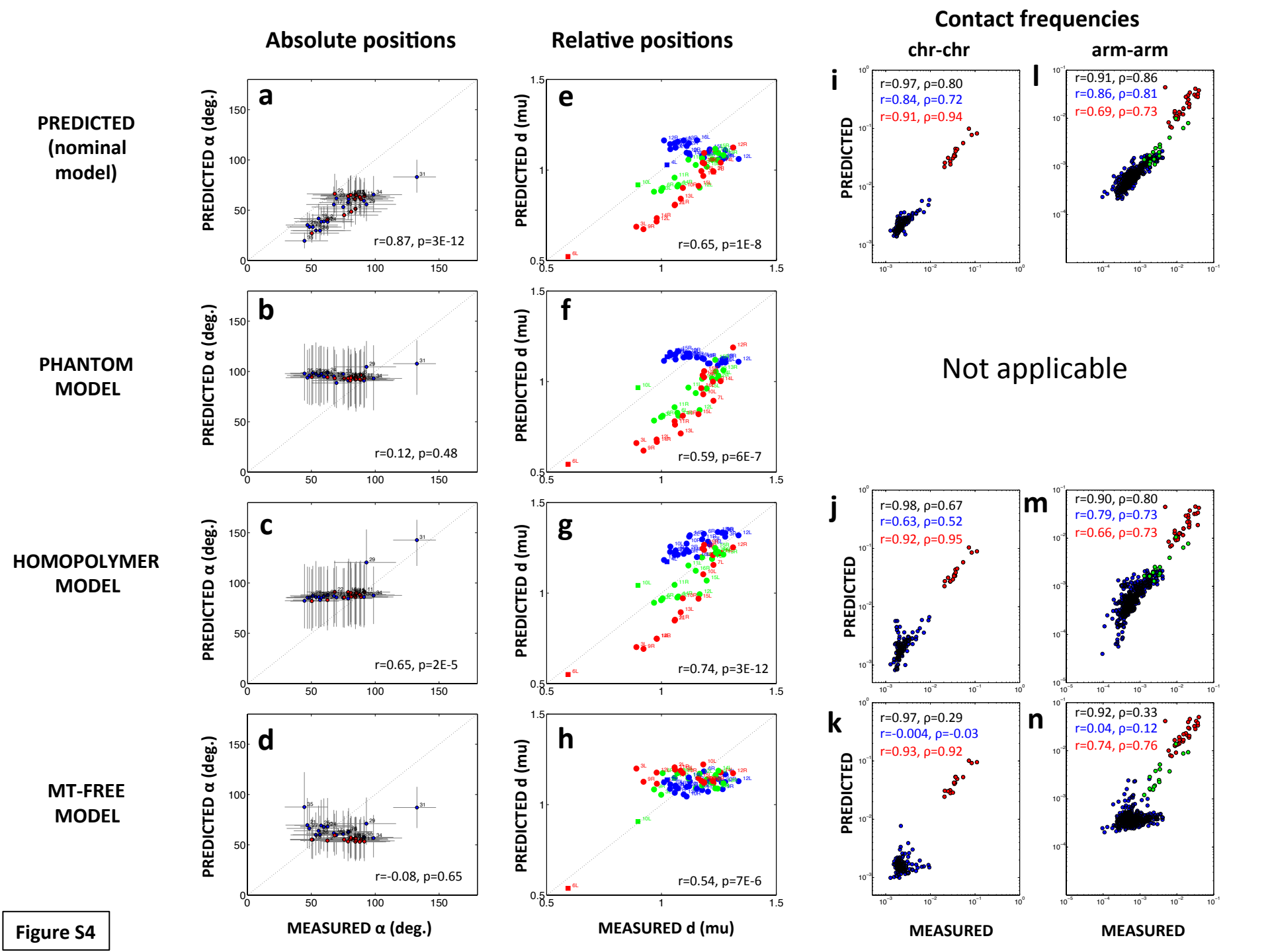
Absolute positions
Relative positions
Contact frequencies
chr-chr
arm-arm
PREDICTED (nominal model)
PHANTOM MODEL
HOMOPOLYMER MODEL
MT-FREE MODEL
a
b
c
d
e
f
g
h
i
j
k
l
m
n
PREDICTED α (deg.)
PREDICTED d (mu)
PREDICTED
MEASURED α (deg.)
MEASURED d (mu)
MEASURED
r=0.87, p=3E-12
r=0.12, p=0.48
r=0.65, p=2E-5
r=-0.08, p=0.65
r=0.65, p=1E-8
r=0.59, p=6E-7
r=0.74, p=3E-12
r=0.54, p=7E-6
r=0.97, ρ=0.80
r=0.84, ρ=0.72
r=0.91, ρ=0.94
r=0.91, ρ=0.86
r=0.86, ρ=0.81
r=0.69, ρ=0.73
Not applicable
r=0.98, ρ=0.67
r=0.63, ρ=0.52
r=0.92, ρ=0.95
r=0.90, ρ=0.80
r=0.79, ρ=0.73
r=0.66, ρ=0.73
r=0.97, ρ=0.29
r=-0.004, ρ=-0.03
r=0.93, ρ=0.92
r=0.92, ρ=0.33
r=0.04, ρ=0.12
r=0.74, ρ=0.76

Figure S4

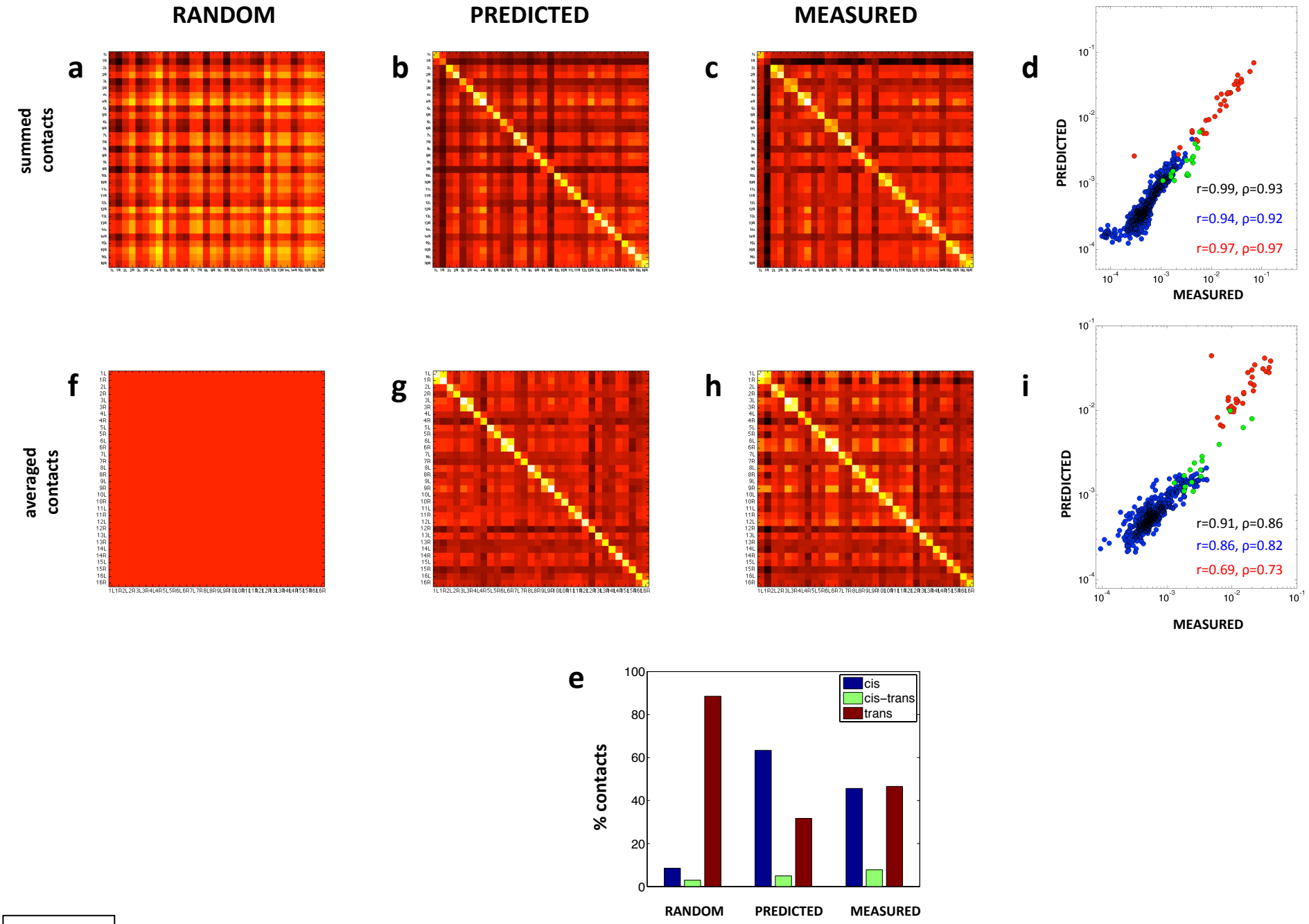
RANDOM
PREDICTED
MEASURED
summed contacts
averaged contacts
a
b
c
d
e
f
g
h
i
PREDICTED
MEASURED
r=0.99, ρ=0.93
r=0.94, ρ=0.92
r=0.97, ρ=0.97
r=0.91, ρ=0.86
r=0.86, ρ=0.82
r=0.69, ρ=0.73
% contacts
cis
cis–trans
trans
RANDOM
PREDICTED
MEASURED

Figure S5

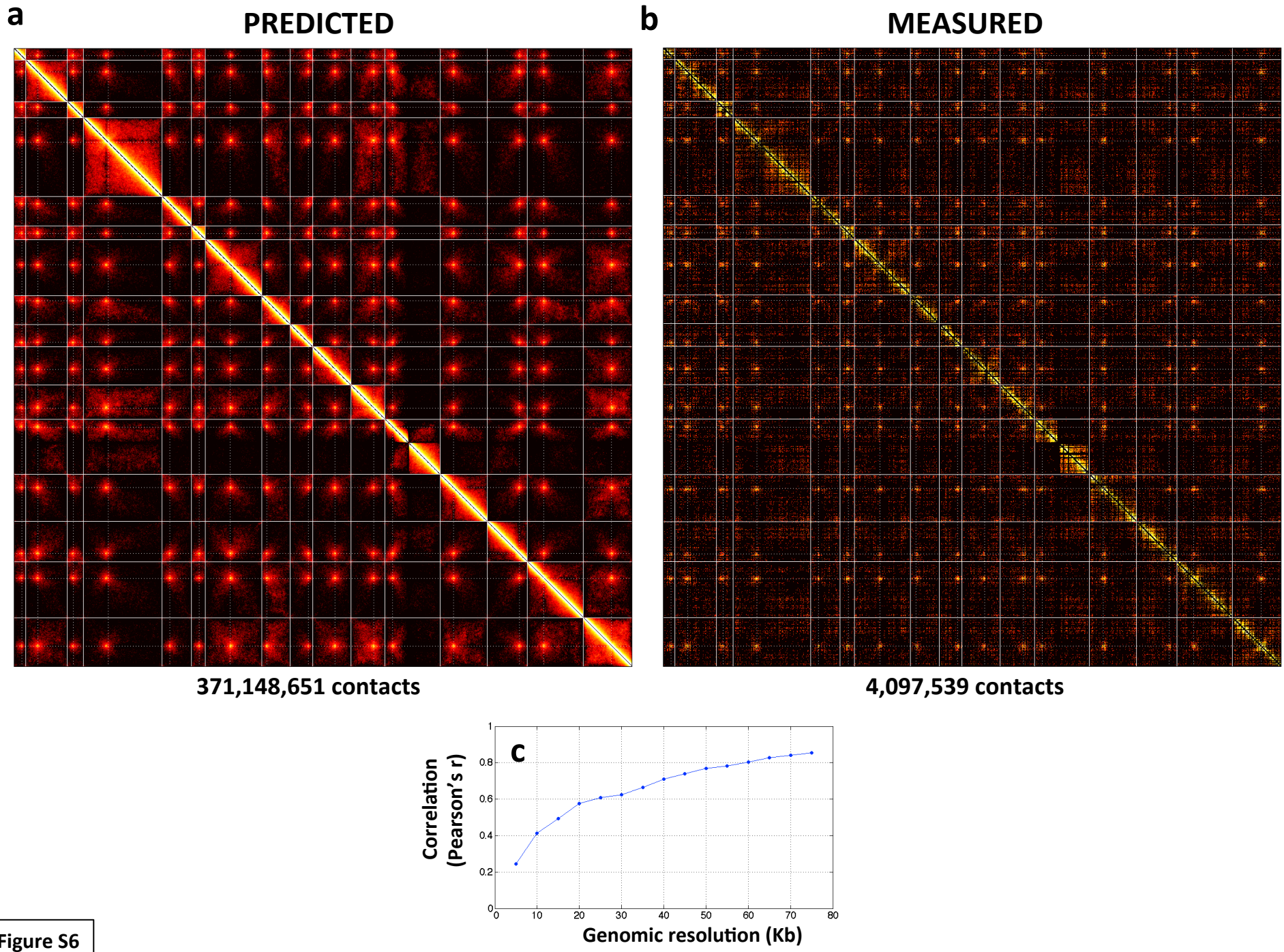
a
PREDICTED
b
MEASURED
371,148,651 contacts
4,097,539 contacts
c
Correlation (Pearson's r)
Genomic resolution (Kb)
0
0.2
0.4
0.6
0.8
1
0
10
20
30
40
50
60
70
80

Figure S6

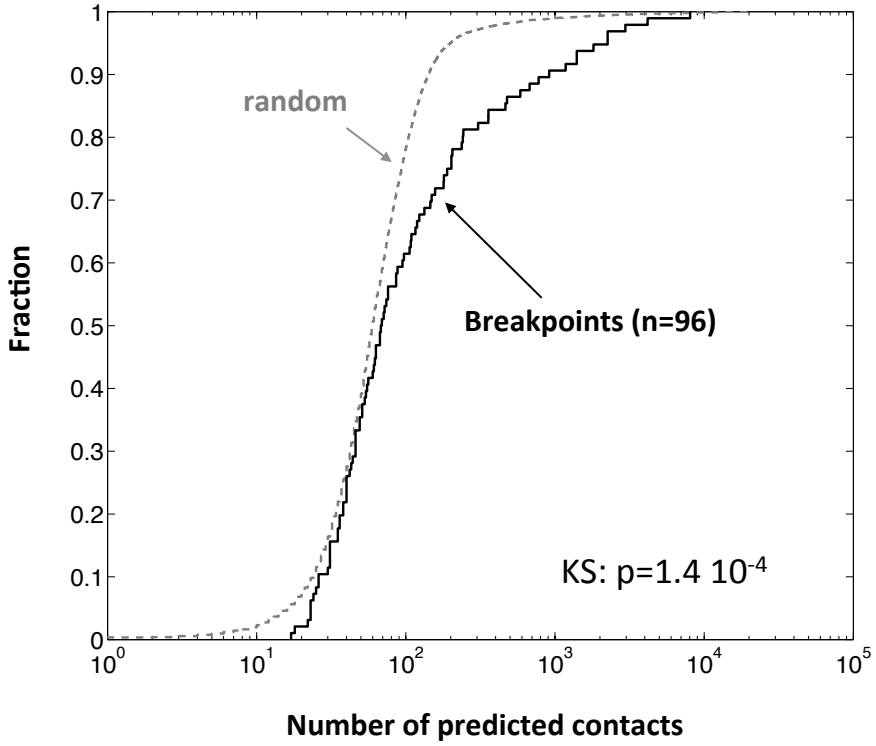
random
Breakpoints (n=96)
KS: p=1.4 10-4
Fraction
Number of predicted contacts
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
10^0
10^1
10^2
10^3
10^4
10^5

Figure S7

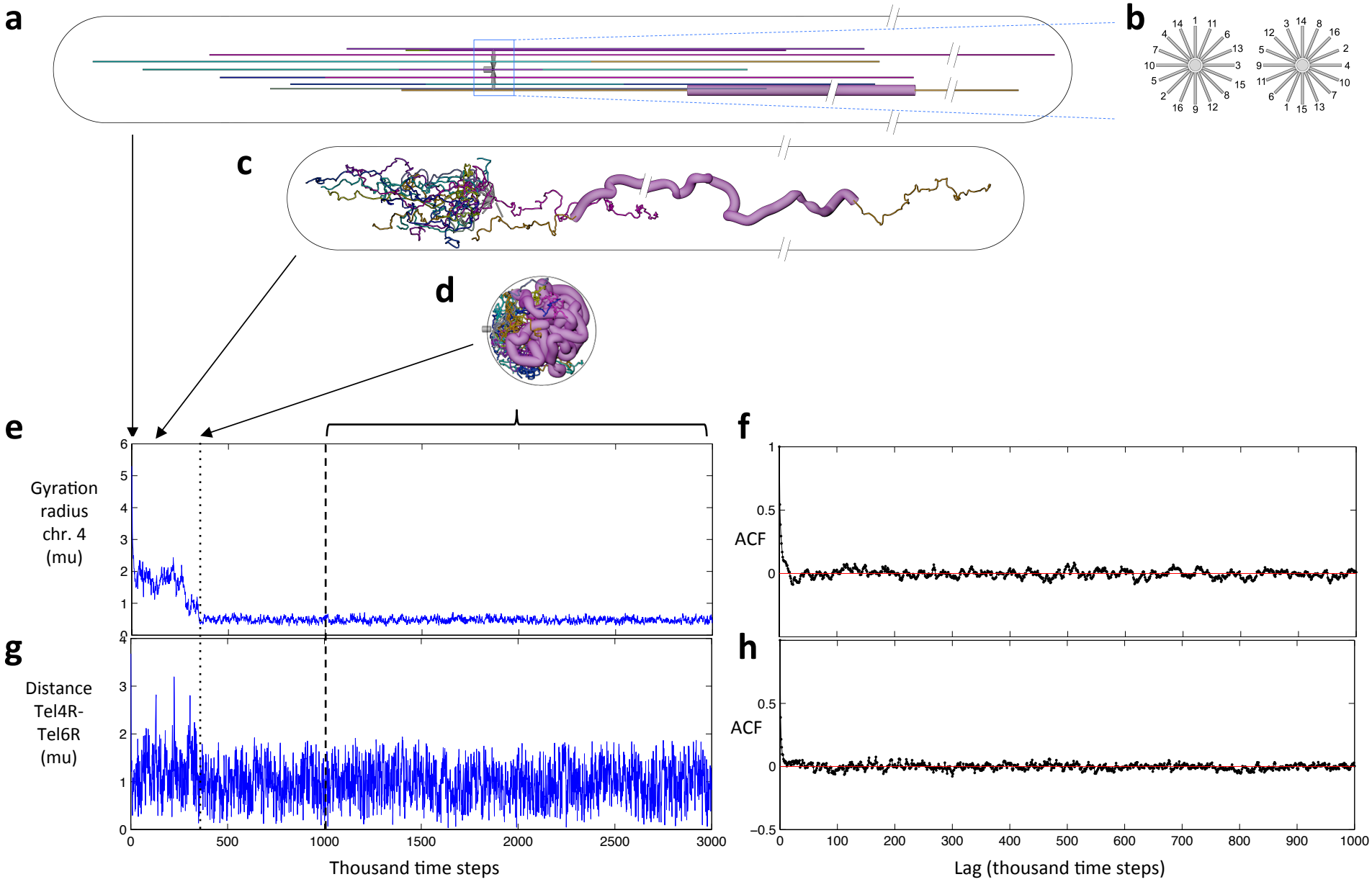
a
b
c
d
e
Gyration radius chr. 4 (mu)
f
ACF
g
Distance Tel4R-Tel6R (mu)
h
ACF
Thousand time steps
Lag (thousand time steps)

Figure S8