Simulating structurally variable Nuclear Pore Complexes for Microscopy

Maria Theiss[1,*], Jean-Karim Hériché[2], Craig Russell[1], David Helekal[3], Alisdair Soppitt[4], Jonas Ries[2], Jan Ellenberg[2], Alvis Brazma[1,*], Virginie Uhlmann[1,*]

1 European Molecular Biology Laboratory — European Bioinformatics Institute (EMBL-EBI), Hinxton, UK.
2 European Molecular Biology Laboratory (EMBL) — Cell Biology and Biophysics unit, Heidelberg, Germany.
3 EPSRC & MRC Centre for Doctoral Training in Mathematics for Real-World Systems, University of Warwick, Coventry, UK.
4 EPSRC Centre for Doctoral Training in Modelling of Heterogeneous Systems, University of Warwick, Coventry, UK.

Corresponding authors
*** theiss@ebi.ac.uk**
*** brazma@ebi.ac.uk**
*** uhlmann@ebi.ac.uk**

Abstract

The Nuclear Pore Complex (NPC) is a massive, 8-fold symmetrical protein assembly that enables macromolecular exchange between the cytoplasm and nucleus. The average architecture of proteins within the NPC has been resolved with pseudo-atomic precision, however observed NPC heterogeneities such as varying diameters, elongated shapes, 9-fold symmetry, and irregular shapes evidence a high degree of divergence from this average. Single Molecule Localization Microscopy (SMLM) images NPCs at protein-level resolution, whereupon image analysis software studies their heterogeneity. Antithetically, NPCs have been used as a reference standard for SMLM due to their abundance, size, and symmetry. Ground truth annotations would ultimately aid both lines of research, however SMLM noise impedes such annotations and complicates image analysis. Here, we introduce a pipeline that simulates structurally variable NPCs based on architectural models of the NPC. Users can specify one of several N- or C-terminally tagged NPC proteins, as well as geometric variability in radius, height, internal twist, and elongation. We represent the NPC as a spring-model such that arbitrary deforming forces, of user-defined magnitudes, simulate irregularly shaped variations. We provide an open-source simulation pipeline, as well as a reference dataset of simulated human NPCs, with multiple types of geometric variability and increasing magnitudes of deforming forces.

Introduction

Nuclear Pore Complexes (NPCs) enable all macromolecular exchange between cytoplasm and nucleus [1]. The ∼110 megadalton human NPC consists of ∼1000 individual proteins, called nucleoporins (Nups), with ∼30 symmetrically arranged types [2], [3]. Eight rotationally symmetric spokes enclose the transport axis, and scaffold Nups organise in three stacked rings along the nuclear envelope (NE) plane: the nucleoplasmic ring (NR) is almost point-symmetric to the cytoplasmic ring (CR). NR and CR constitute the outer rings (OR) that sandwich an inner ring (IR), which is point-symmetric to itself [4].

This structure has been solved with pseudo-atomic resolution. For instance human NPCs were partially resolved to 23 Å in isolated NEs [5], [6], and to 35 Å in their native environment [7]. A near-complete model of NPCs, in both isolated NEs and intact cells, was proposed based on the Alpha-fold structure prediction of Nups [8], [9]. However, these models achieve their high resolution by integrating information from multiple sources, such as Electron Microscopy (EM), Nup atomic structures, and protein interaction studies, effectively averaging out hundreds of NPCs and their variability [10].

Examples of this NPC variability are deviations from the 8-fold rotational symmetry [11]–[13], elliptical deformations [14]–[16], varying twist angle between CR and NR [17], and varying pore diameters [9], [13], [18], [19]. Furthermore, distance between CR and NR is inconsistently reported [17], [18], [20].

This natural NPC variability remains to be fully characterised. To this end, coarse-grained models and coarse grained elastic network models predicted possible deformations of the yeast NPC [21], [22]. Molecular dynamics have been used to simulate lateral tension on the NE, which dilated the model NPC [9]. Theoretical models aim to predict NPC heterogeneity and are oftentimes motivated by imaging studies [9], [19], but conversely, can form hypotheses that need to be tested *ex silico*.

For example Atomic Force Microscopy and EM have been used to study NPC shape variability (e.g. [11], [13]). They however cannot elucidate the interplay of different Nups of interest (NoIs) in structurally variable NPCs, as neither commonly tag specific NoI, and both lack the resolution to discern NoI at single pore level. Fluorescence light microscopy on the other hand can specifically tag one or more NoIs [23], however the diffraction limit of visible light is typically double the size of a human NPC [7], [24].

Single-Molecule Localisation Microscopy (SMLM) is a fluorescence light microscopy technique that circumvents this diffraction limit and can therefore be used to analyse NPC heterogeneity at single-pore level and for specific Nups [16], [18], [25], [26]. More precisely, SMLM illumination and buffer conditions cause fluorophores to blink stochastically such that only one fluorophore within a neighbourhood is active at a time; it is thus resolvable from its neighbours by fitting its centre [27]. The resulting movie is rendered into a coordinate map of fluorescent labels. SMLM has been shown to image 4 differently labelled Nups in 3D [23], [28]–[30] and to resolve Nups that are 12 nm apart [31], [32]. This results in a high-contrast, sparse, coordinate representation of each NPC, which can be fitted, classified, or averaged by downstream software [16], [18], [25].

SMLM however is prone to noise, aberrations, and artefacts from sample preparation, widefield fluorescence imaging, and image reconstruction. Notably, current labelling efficiency is below 80% of structures of interest, and repeated blinks cannot easily be assigned to one fluorophore [32], [33]. Consequently, it is near impossible to automatically or manually annotate structural variability of individual pores with high enough confidence. The ability of software to analyse NPC heterogeneity cannot be assessed without ground truth, whereas ground truth cannot be assigned without validated computational analysis. Simulated NPCs with known variability can solve this catch-22. Wu et al. [25], simulated heterogeneous NPCs by varying geometric parameters — NPC radius, distance between the ORs, and twist angle between the ORs; however, their approach assumes NPC variations are simple geometric shapes.

Multiple approaches have been developed to simulate sources of noise throughout the SMLM imaging pipeline, for instance sparse labelling and linkage error, and detector readout noise [34]–[39]. However, simulations of non-geometric shape irregularity of the underlying sample have been widely neglected. Such simulations are paramount to study the robustness of fitting, classification, or averaging software to hypothesised shape irregularity; they would allow to develop or benchmark software applications that aim to characterise the distribution and magnitude of shape irregularity. Furthermore, such simulations could help clarifying whether software can disentangle shape irregularity from imaging noise. Knowing the limits of image analysis software would allow to study the irregularity of real NPC shapes, and finally whether such shape irregularity has biological meaning.

We introduce a spring based model of the NPC to simulate both geometric variations and irregularly shaped variations for SMLM. We provide a wide range of Nup models for selection which are N- and C-terminally labelled, and from human NPCs in either NE isolates or intact cells [5]–[7]. Additionally, we provide a dataset of geometrically variable and irregularly shaped NPCs, which includes our ground truth coordinates and annotations, as well as simulated SMLM data based on the simulation engine of the Superresolution Microscopy Analysis Platform SMAP [39].

Methods

Spring models can simulate irregular but coupled deformations

We simulated the NPC as a partially connected spring model to which we apply deforming forces. This results in shape variations that are irregular due to randomness of deforming forces, yet coherent as springs ensure forces are distributed throughout the model. (Figure 1A and Figure 2 "Variability"). These simulations are performed on individual types of Nups, however we use "NPC" as an umbrella term for any NoI. Springs represent the elastic properties of proteins and membranes, whereby forces represent any reason the NPC is not observed in its perfectly geometrical shape. Upon application of a force, springs will be displaced from their resting lengths and oscillate to a new equilibrium. If oscillatory energy is lost to the environment, and if the applied force is static, the new equilibrium will also be static. Since SMLM has a low temporal resolution and usually images fixed cells, we indeed damp spring oscillation, and keep forces static. Consequently, we only use the static new equilibrium for downstream SMLM simulations.

NPC structural models function as input

We generated input models based on all human OR and IR models available on PDB to date (Table 1 and Table S2). These thoroughly validated state of the art models offer a resolution exceeding that of SMLM. Furthermore, they preserve relative positions of different Nups to each other allowing for future multi-colour simulations. SMLM usually captures NPCs in whole, but chemically fixed cells. However, models 5A9Q [5], 5IJN, and 5IJO [6] represent NPCs in isolated NEs, whereas 7PEQ and 7PER [7] represent NPCs in their native environment. It is therefore uncertain which models best represent NPCs under SMLM conditions. Furthermore, live-cell SMLM, in which NPCs are in their native environment, might become more common in the future [40]. We therefore made all mentioned human NPC models selectable by the user. Our pipeline allows for the addition of further models. Following, we detail how we integrate PDB models to our pipeline (Figure 2, 1st row; Figure 3).

Table 1. Input models

Model	Substructure	Origin	Resolution (Å)
5A9Q [5]	ORs	HeLa, isolated NE	23
7PEQ [7]	ORs	DLD-1, native environment	35
5IJN [6]	IR	HeLa, isolated NE	21.4
5IJO [6]	IR	HeLa, isolated NE	21.4
7PER [7]	IR	DLD-1, native environment	35

Implementing NPC structural models

Nomenclature

Nups within the NPCs in top-view are arranged in s near-identical "spokes" around a central channel, whereby s is the rotational symmetry of the NPC, 8 by default. "Rotational unit" (rot. unit) refers to all structurally resolved Nups in one of s spokes. The grouping of Nups into a rot. unit or spoke is not clearly defined (Compare Figure 3A, "rot. unit" and "PDB rot. unit"). The copy number g of one NoI in one rot. unit corresponds to its overall copy number m via $m = g \cdot s$. In side-view, it is accordingly arranged in g stacked "rings" that are parallel to the NE plane. The term "rings" is not to be confused with the terms "CR", "NR", or "IR", as each of the latter three might contain several "rings" of a specific Nup (Figure 3A). Important notation is summarised in Table S1.

Extracting Nup coordinates from PDB

PDB stores NPC models that contain all rot. units [5]–[7] (Figure 2). For each NoI we determined the indices of the most N- and C-terminal structurally resolved amino acids, as protein termini are common fluorophore attachment sites (Table S2). Furthermore, we note the unique identifier (ID) of each NoI within the rot. unit. If the same type of Nup occurs in g rings, it has a different ID for each ring. Finally, we load the pdb.gz file into ChimeraX [41], [42], where, for each NoI i with given IDs $j \in 1 \dots g$, we extract the Cα atom coordinates of both terminal amino acids within one PDB rot. unit (Table S3). Consequently, each NoI is represented by two (N- and C-terminal) groups of g coordinates. We use N-terminally tagged Nup107 from 5A9Q for a simulated dataset (Results). Example code to extract corresponding coordinates is shown in Table 2.

Table 2. ChimeraX command to extract Nup107 N-terminus from 5A9Q

"Get coordinates"	Rot. unit	IDs Nup107 ($g = 4$)	N-term. aa	Cα atom
`getcrd`	`#1.1`	`/M,D,V,4`	`:150`	`@CA`

Furthermore, we determine the x-y centre $\vec{x}_c = (x_c, y_c)$ of the NPC model (Suppl. section 2.1) and reformat ChimeraX output coordinates so that they are readable with our pipeline. Using said coordinates, $\vec{x}_i$, $i \in 1 \dots g$, of all g NoI within one rot. unit and $\vec{x}_c$, their relative angles α_i and radii r_i are determined (Figure 3C, Suppl. section 2.2). Furthermore, z positions z_i are directly read out from the atomic model, converted to nm, and offset so that $\min(z_i) = 0$. For each ring, Cartesian coordinates are generated by reflecting coordinates of the rot. unit s times around a new central axis $\vec{x}_0$ with $\vec{x}_0 = (0, 0)$ with an offset of $2\pi/s$. Coordinates of each NoI henceforth called "Nodes".

Layout of the spring model

The layout of the spring model is a simplification of real NPCs. For each ring, nodes are anchored to the centre via radial springs (Figure 3B). These springs approximate the NE, whereby interior pressure replaces exterior contraction-resisting tension [19]. Nodes are connected to their $2h$, $h \in 0 \dots \lfloor s/2 \rfloor$, neighbours via circumferential springs, which represent interactions within the NPC. h is 2 by default. Shorter circumferential springs are stiffer than longer circumferential springs, indicated by their spring constants k. Radial springs are less stiff than the shortest circumferential spring, as we assume protein-protein interactions are stronger than protein-membrane interactions. (Suppl. section 3).

Random forces deform the spring model

Random, static forces are applied to nodes to achieve irregular, static shapes. Forces are sampled from a multivariate normal distribution with covariance dependent on the distance between nodes. Consequently, forces on nearby nodes are more similar than forces on distant nodes (Figure 3D). This ensures continuity between separate NPC rings and considers that forces on real NPCs are likely not pointforces. More precisely, covariance between forces on nodes $\vec{x}_i$ and $\vec{x}_j$, with $i, j \in 1 \dots m$ are given by the RBF kernel (Eq. 1)

$$\mathrm{cov}(\vec{x}_i, \vec{x}_j) = e^{-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2\sigma^2}} \tag{1}$$

$\|\vec{x}_i - \vec{x}_j\|^2$ is the squared Euclidean distance between node $\vec{x}_i$ and $\vec{x}_j$, and σ a free parameter that we couple to the size of the model by setting it to a fraction of the mean radius $\bar{r}$. In comparison to a linear covariance between forces and distance, the RBF kernel ensures more similarity between forces on nearby nodes, preserving local structures. This is particularly important for nearby nodes on different rings. Simultaneously, forces on distant nodes are

less similar, allowing for more irregular deformations. Empirically, RBF free parameter $\sigma = 0.5\bar{r}$ produced NPC deformations that best fit this description (Figure 1B).
The CR and NR in real NPCs are ∼60 nm apart, attach to different sides of the NE and do not directly interact with each other. However, the cytoplasmic and nucleoplasmic side of the NE merges at the nuclear pore, and the CR and NR are connected via the IR [7]. To capture this incomplete coupling, we transiently set their z-distance to RBF free parameter σ. We neglect z otherwise, as we do not compute deformations in z.
In addition, radii r_i of all nodes are transiently set to the mean radius $\bar{r}$ of all rings (Eq. 2). Radial distances are therefore equalised, which preserves concentricity of rings during deformation.

$$\bar{r} = \sum_{i=1}^{g} \frac{r_i}{g} \tag{2}$$

Lastly, the magnitude (mag) of deformation D_{mag} is user-definable, such that the effects of different levels of deformation can be studied (Figure 1B). D_{mag} [nm] covers ∼99.7 % of the sampling range (= 3 standard deviations = 3σ). Following, it is converted to variance σ'^2 via $\sigma'^2 = (D_{\text{mag}}/3)^2$. Next, cov is updated: $\text{cov}' = \sigma'^2 \cdot \text{cov}$. Subsequently, scalar forces f_i with $i \in 1 \ldots m$ are sampled $f_i \sim \mathcal{N}(0, \text{cov}')$. Optionally, f_i may be modified to include no-noise deformations (see Elongation, Suppl. section 5.3). Lastly, f_i is transformed into a vector $\vec{f_i} = (f_{x_i}, f_{y_i})$ that points in radial direction to avoid rotations (Eq. 3). Any set of forces with no net-torque would avoid rotations, however radial forces are simple to compute and intuitively understand. Predominantly radial forces stemming from the NE are the most parsimonious explanation for observed dilations of the NPC, particularly as no motor activity of scaffold Nups has been observed [9], [19].

$$\vec{f_i} = \frac{f_i \vec{x}_i}{\|\vec{x}_i - \vec{x}_0\|} \tag{3}$$

The final deformation of the model is computed for each ring separately (Suppl. section 4).

Simulating geometric variability

Geometric variability can be set as a fixed parameter or randomised

Geometric variability of the NPC is commonly observed in imaging data [9], [11]–[20]. We simulate differences in radius, ring distance and twist angle between CR and NR, rotational symmetry, and elongations (See Figure 4A for an overview, excluding ring distance, and Table 3). Either of these values X can be kept as standard input, or redefined by the user. Additionally, a standard deviation σ can be user-input for any parameter, except for the discrete rotational symmetry. If a standard deviation is given, a new value X' for each model NPC is drawn from a normal distribution $X' \sim \mathcal{N}(X, \sigma^2)$.

Radius

A new mean radius $\bar{r}'$ is set manually and calculated similarly as in Wu et al., 2021 [25]. New radii r'_i with $i \in 1 \ldots g$ are calculated according to Eq. 4

$$r'_i = \frac{\bar{r}'}{\bar{r}} r_i \tag{4}$$

The angle between Nups is thereby kept constant. Therefore, Nups are further apart in dilated NPCs (Figure 4A “constricted”).

Ring distance

Distance d between CR and NR can be changed to a manually preset value and calculated similarly as in Wu et al., 2021 [25]. d is the difference between the average z positions of the CR ($\bar{z}_{\mathrm{CR}}$) and NR ($\bar{z}_{\mathrm{NR}}$) (Eq. 5).

$$d = \bar{z}_{\mathrm{CR}} - \bar{z}_{\mathrm{NR}} \tag{5}$$

A new distance d' is set manually. z coordinates of all CR rings z_{CR} are updated to z'_{CR} (Eq. 6):

$$z'_{\mathrm{CR}} = z_{\mathrm{CR}} - (d - d') \tag{6}$$

Symmetry

For NPCs of varying symmetries, we keep the geodesic distance between Nups constant and preserve the absolute difference of radii between rings, as we assume the space taken up by each Nup does not change with its copy number. Consistently, NPCs with 9- or 10-fold rotational symmetry exhibit a greater diameter [11]–[13]. Unlike for varying radii, the angle between Nups changes. Computations are performed separately for each NPC ring (Figure 4A "9-fold", Suppl. section 5.1).

Twist angle

We define twist angle θ as the difference between the mean CR angles and the mean NR angles within one spoke, assuming a spoke contains the most acute possible θ. (Figure 3A, Figure 4A "Twisted", Suppl. section 5.2).

Elongation

Elongated NPCs are generated by approximating them with an ellipse, whose semiminor/semimajor axis ratio q is defined by the user. The NPC is randomly rotated relative to this ellipse, and a force $\vec{f}_{e_i}$ on a node i is computed from the distance between i and the closest point on the ellipse that lies on a line between node i and the NPC centre. $\vec{f}_{e_i}$ are then added to other deforming forces $\vec{f}_i$ (Figure 4A "Elongated", Suppl. section 5.3).

Features of simulated NPCs

After the application of deforming forces, the NPC model might deviate from its input geometric parameters, such as input radius $\bar{r}$ and semiminor/semimajor axis ratio q of an elongated NPC (Figure 4C). Any software aimed at analysing deformed NPCs can be benchmarked based on two questions. Firstly, and more ambitiously, whether this software is able to retrieve parameters of the non-deformed NPCs from their deformed counterparts. When applied to real images, such hypothetical software would aim to reverse and quantify the effects of both imaging noise and biological distortion. This software would be benchmarked against the original input-parameters that define the model-NPC before distortion. The second question is whether software can reverse, circumvent, or quantify the effects of solely imaging noise. To enable a fair benchmark of the latter, we extract features of deformed model NPCs. To this end, we fit a circle and an ellipse to the deformed model. This is computed for the xy-coordinates of each whole NPC model, as well as separately for the CR and NR, and lastly for each individual ring g of the model NPC.

Extracting features

We fit a circle based on least-squares [43] and an ellipse (Figure 4B, Suppl. section 6) to the model, and for both we extract centre coordinates (x_c, y_c) and the residual sum of squares (RSS) (for RSS of an ellipse see Suppl. section 7). For the fitted circle, we extract its radius. For the fitted ellipse we extract semiminor- and semimajor axes b and a, ratio $q = a/b$, and a rotational angle ρ. Assuming optimal fit, higher values of the RSS indicate deviations from the fitted shape. However, if the RSS is computed for a fit to the whole NPC model, or to CR, IR, or NR, these may contain several rings of varying radii. The RSS would therefore exceed 0 even for a geometrically shaped NPC. We therefore recommend to use the RSS of an equivalent NPC with mag 0 as baseline. The relationship between input parameters and features after deformation is demonstrated in Figure 4C.

Output files and image data simulation

For each simulation experiment, NPC coordinates are saved in a CSV file that can be loaded into photophysics simulation software (e.g. [39]). A metadata file with properties of simulated NPCs is generated, as well as a CSV files containing features of circles and ellipses fitted to whole simulated NPCs, and a CSV file with such features for CR, NR, and individual rings. (Suppl. section 8).

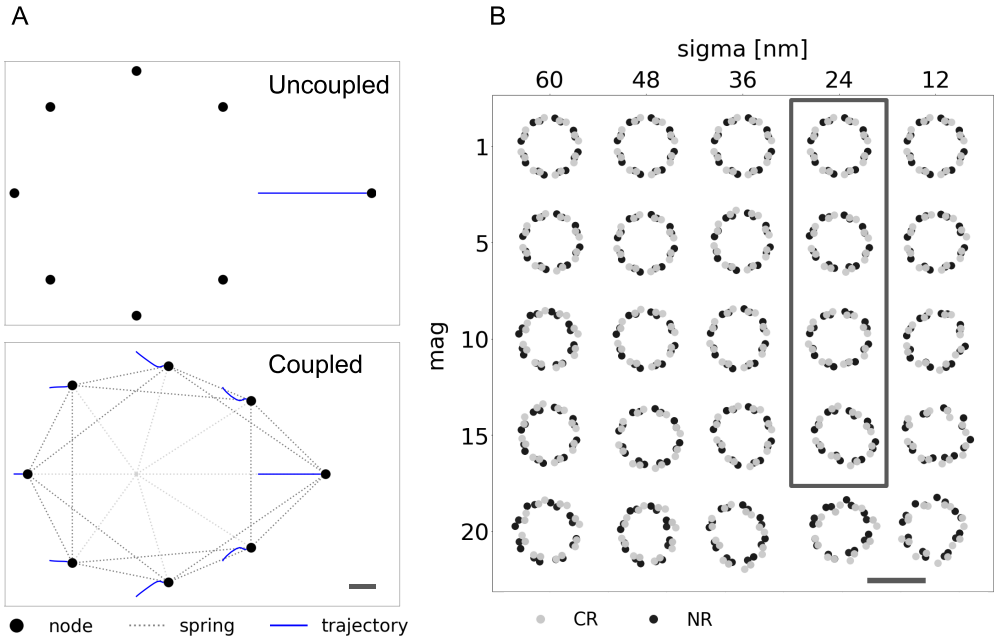

Figure 1. A: Behaviour of a spring model upon application of a point force. Purely geometric models are incapable of propagating a point force, whereas a spring model allows for irregular shape variations. Scale bar 10 nm **B:** RBF σ vs deformation mag. Large values of σ prohibit non-convex deformation, while deformations are too irregular for small values of σ. We set σ to $r/2$ by default, which is 24 nm in this case. Boxed are reasonable values for σ and deformation mag. Scale bar 100 nm, Nup107 N-terminus of 5A9Q [5].

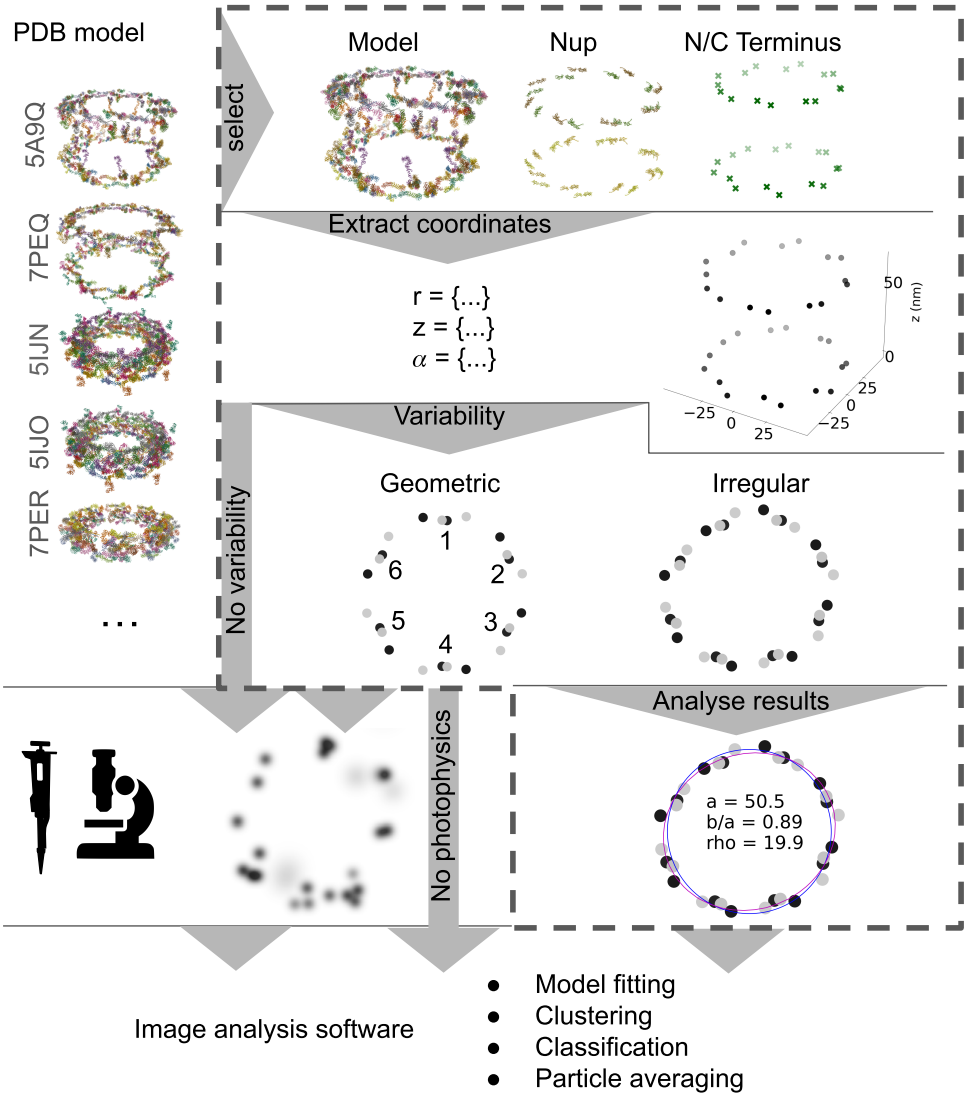

Figure 2. Overall workflow from selection of the underlying model, to image simulation, to benchmarking of image analysis software. Steps included in our pipeline are boxed with a dotted line. PDB models that are already integrated to our pipeline are listed as examples: 5A9Q [5]; 7PEQ, 7PER [7]; 5IJN, 5IJO[6]. From top to bottom: A PDB model, Nup, and N- or C terminus can be selected by the user. Its coordinates are extracted. Geometric or irregularly shaped variability can optionally be added. Coordinates are passed to microscopy simulation software (e.g. [39]). Features of irregularly shaped NPCs are extracted as ground truth to benchmark Image analysis software.

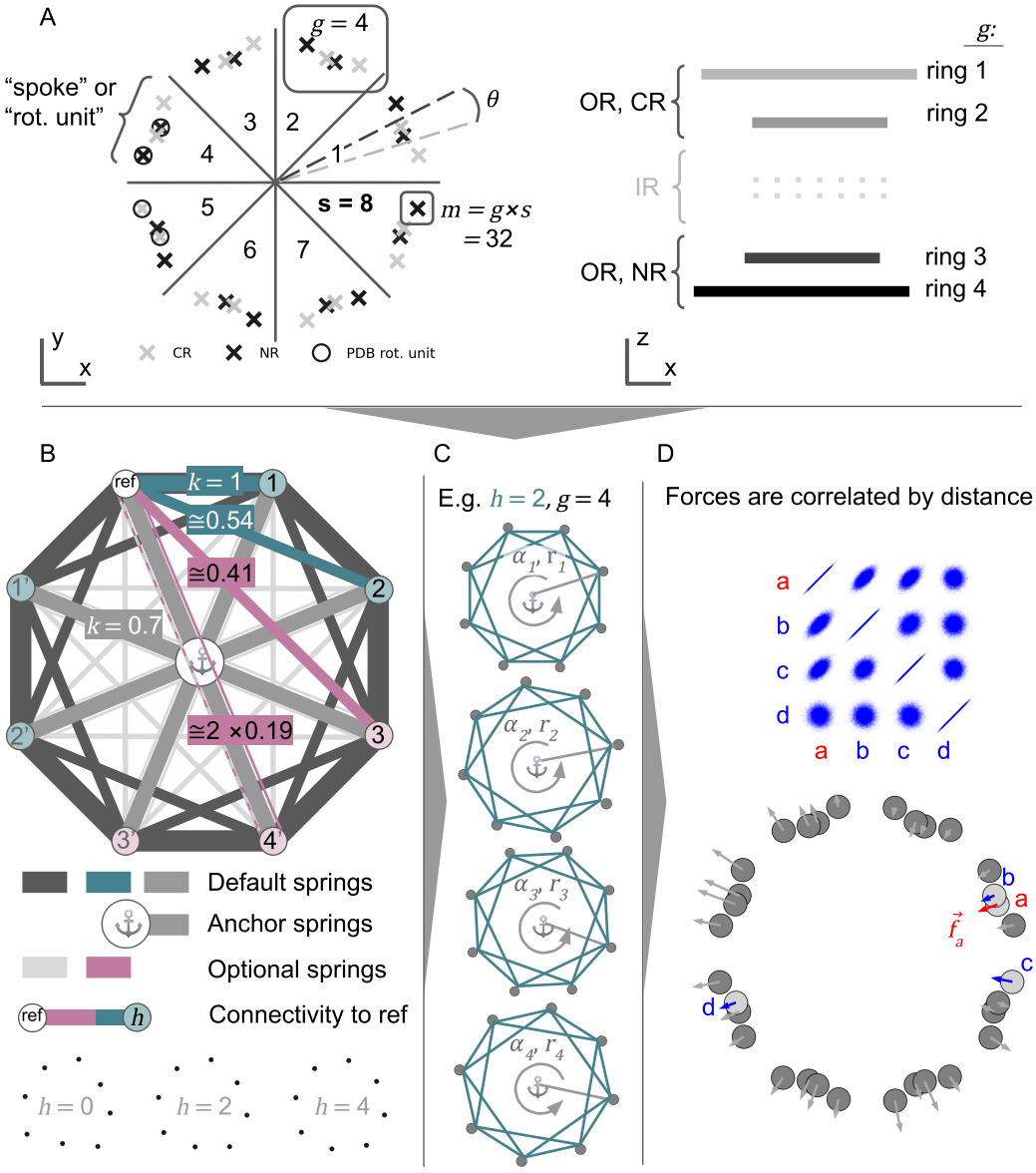

Figure 3. Layout of the spring model. **A:** Coordinates of an example Nup (N terminus of Nup107 in PDB 5A9Q [5]) in top view (left) and side view (right). Nup107 occurs in $g = 4$ rings and has symmetry $s = 8$, corresponding to a total copy number of $m = g \cdot s = 32$. CR and NR are offset by a twist angle θ. A rot. unit according to the PDB assembly is indicated. The rot. unit can be modified by offsetting the CR or NR by $\pm 2\pi/s$ such that θ is the most acute (Suppl. section 5.2). Right: Difference between OR, NR, CR, IR, and rings in side view. The IR is shown as faint dotted lines, as Nup107 does not occur in the IR. **B:** Spring connectivity of one of g rings. By default, circumferential springs connect each node to their nearest $h = 2$ neighbours in clockwise (cw) and anticlockwise (acw) direction. cw default springs are shown in dark gray, and in teal for a reference node ref. Optional cw circumferential springs for ref, with $h = 3$ and $h = 4$, are shown in magenta. Spring constants k are shown on top of example springs and are decreasing with circumferential spring length. $h = 4$ connects opposing nodes with two springs whose spring-constant is halved to be equivalent to a single spring. All nodes are connected to a central anchor via springs with $k = 0.7$. Bottom: Effect of deformation mag 15 on nodes with $h = 0, 2, 4$. **C:** Each ring g is simulated separately and is defined by its radius r, rotation α, and z-position (z not shown). **D:** Continuity between independent rings is ensured with deforming forces that are covariant by distance. Top: 10000 forces (scalar values) sampled on exemplary nodes a, b, c, and d.

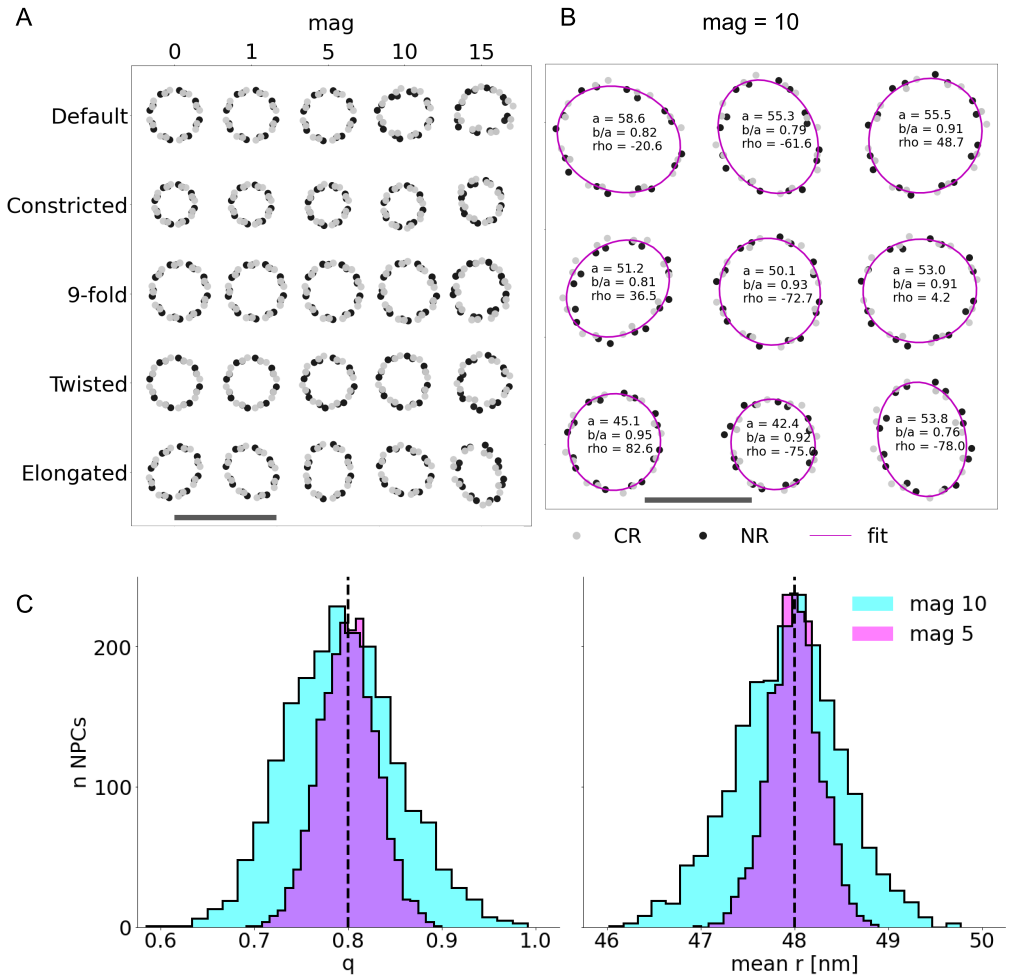

Figure 4. **A:** Different types of geometric deformation at increasing mag. Scale bar 200 nm. **B:** NPCs with randomly varying degrees of elongation ($q \sim \mathcal{N}(1, 0.1)$) and varying radii ($\bar{r'} \sim \mathcal{N}(\bar{r}, 5)$ [nm]). Shown is a fitted ellipse with semimajor axis value a, axes ratio q, and rotation ρ°. Scale bar 100 nm. **C:** NPCs with input elongation $q = 0.8$ (left, dotted line) or mean radius $\bar{r} = 48nm$ (right, dotted line) were deformed by mag 10 or 5, 2000 NPCs and 25 bins each. q of the deformed model was determined by fitting an ellipse, the new mean radius $\bar{r}$ was determined by fitting a circle.

Results

Simulated Dataset

We used PDB model 5A9Q [5] for simulations due to its high pseudo-atomic resolution, and selected N-terminally labelled Nup107 (Tables 2, S3, S2), as it is frequently studied (e.g. [18]). We simulated six classes: reference NPCs, smaller radius, smaller CR-NR distance, elongated NPCs, 14° twist angle, and 9-fold symmetry; details are in Table 3 and examples in Figure 4A. For each class, we simulated NPCs with deformation mag 0, 1, 5, and 10; 500 NPCs each, and generated output files according to Methods "Output files and Image simulation".

Table 3. Default and modified geometric parameters of simulated Nup107 (N terminus)

Parameter	Symbol	Default	Modified value	Justification
Mean radius	$\bar{r}$	48 nm	38.4 (80% of 48 nm)	Observed [18]
Mean dist NR to CR	$\bar{d}$	58.6 nm	30 nm	To resemble Nup133, to test z-resolution [18]
Elongation (short/long axis)	q	1	0.835 (0.8 at NPC)	Commonly observed[14]–[16]
twist angle	θ	9.6°	14°	Observed[17]
Symmetry	s	8-fold	9-fold	Commonly observed [11]–[13]

Simulated Image Acquisition

Photophysics were simulated in SMAP [39] with image parameters representing a good-quality experiment adopted from Wu et al., 2021 [25]. Simulation parameters are shown in Table 4. Simulated imaging data were exported as CSV files.

Table 4. Photophysics simulation parameters from Wu et al., 2021 [25]

SMLM property	value	units/notes
Photon count	12000	photons/pixel/localization
Background	260	
Labelling efficiency	67	%
Re-activation	4.1	
Lifetime	2	frames
Linkage error (free)	3	nm
EM on	True	simulates EMCCD sensor
Number of frames	100000	

Code and Data availability

All described code is available at `https://github.com/uhlmanngroup/SimulateNPCs`. Simulated data is available at BioStudies (Accession number S-BSST842).

Discussion

SMLM simulation software encompasses the entire imaging pipeline [34]–[39] including basic variations of NPC geometry [25]. However, we are aware of no approach that simulates irregularly shaped NPCs. Here, we introduce a toolbox that simulates irregularly shaped NPCs as well as a wide range of geometric NPC variability reported in literature and anecdotally. This geometric variability encompasses elongated NPCs [14]–[16], varying radii [9], [13], [18], [19] and rotational symmetry [11]–[13], ring distance [17], [18], [20] and twist angle [17]. Finally, we used our software to generate a reference dataset that includes deformed ground truth NPCs, annotated features, as well as corresponding simulated imaging data using the SMLM simulation engine of SMAP [39]. Among other applications, our software and dataset can be used to benchmark fluorescence microscopy clustering, classification, and averaging software.

Pseudo-atomic NPC models serve as input to our pipeline, of which we have integrated human NPCs in membrane isolates [5], [44] and in their native environment [7]. This can be extended to NPCs of other species such as yeast [19], [45], [46], especially since recent advances in integrative modelling [9], [10] promise a wider gamut of resolved NPCs. Our pipeline simplifies image simulation of variable NPCs based on these models, as it only requires Nup coordinates of one rot. unit.

Nup coordinates are transformed into a spring model to which static deforming forces are applied to achieve irregular, yet coherent shapes. These irregular shape simulations resemble coarse grained elastic models [21], [22] in that a model of the NPC is subjected to deforming forces. Crucially however we do not intend to predict NPC variability, but instead replicate observed and hypothesised variability. This allows for a computationally light-weight model specifically tailored for microscopy simulations. For example, we only simulate fluorescently labelled Nups as nodes and employ heuristics such as forces correlated by distance on separately simulated NPC rings. Therefore, absolute position of Nups in the NE or to neighbouring non-labelled Nups do not currently influence deforming forces. This simplifies generalisation to other radially symmetrical structures. We currently apply deforming forces only in lateral direction due to the lower axial resolution of standard 3D SMLM systems.

Anecdotal evidence and modelling however point towards irregularly-shaped NPC variations along its axis, such as saddle-like shapes [21], [22]. Furthermore, NPCs could be imaged in side-view, or SMLM techniques with isotropic resolution could be employed [30]. We currently simulate varying distances between rings in z. This could easily be extended to further geometric variability, such as decoupling between NR and CR geometry and non-parallel arrangements of NR and CR. Long term, the modular architecture of our pipeline permits to integrate a full-3D deformable model.

Furthermore, our software can offer insights into how image analysis software performs on structures that are not the NPC, but share its properties, such as protein copy number or distance between fluorophores. Analogously, NPCs are used as reference standard in light microscopy [32]. Our simulations are particularly suited for this, as the exact geometry and protein copy number of the reference NPC is adaptable. Additionally, our method is generalisable to other circular structures, such as the cross section of cilia [47], or to filamentous structures by defining several anchor points.

Simulations can be used to benchmark the capabilities of image analysis software to fit, cluster, classify, or average structures of interest. We offer two types of ground truth: NPC class labels generated before the addition of shape irregularity, and features computed after. Benchmarking software for their ability to retrieve original input classes tests their capability to reverse the effects of both biological distortion and imaging noise. Few such software applications exist [47]. This task is particularly ambitious if several geometrically variable classes underlie irregular deformations. Our software enables the prototype development of such biological denoising software. The second type of ground truth, namely features

computed from irregularly shaped NPCs, is aimed at software that promises to reverse, circumvent, or quantify the effects of only imaging noise. This benchmark is applicable to most software that aims to classify or analyse NPC heterogeneity [16], [18], [25].

Both benchmarking tasks will be influenced by the magnitude of shape irregularity added to the model NPCs. However, the amount and behaviour of shape irregularity in real NPCs is not well known, as for quantification EM lacks the Nup-specific contrast, SMLM the resolution, and both lack a sufficient signal to noise ratio. We therefore provide simulated data of varying magnitudes of shape irregularity and allow for the smoothness of shape irregularity to be adjusted. This allows to simulate any hypothetical shape irregularity that may be found in wild type and mutant NPCs of any species, and under any environmental condition. Therefore, we can study the robustness of image analysis software to different degrees of shape irregularity, particularly in response to extreme values. Eventually, this may allow to quantify shape irregularity in real data in an informed manner.

In more general terms, our approach combines microscopy simulations with molecular physics simulations. Notably, it is not limited to SMLM simulations, as several other microscopy techniques work on the basis of attaching fluorescent or electron-dense labels to a structure, locations of which can be expressed by coordinates. Such coordinates generated by our node and spring model can therefore be passed to simulation pipelines of other microscopy techniques. More such combinations of simulation or modelling techniques could emerge in the future, for instance combining Brownian dynamics or molecular dynamics with any kind of imaging simulations. Such interfaces would provide image simulations with more realistic ground truths, and dynamics simulations with a direct comparison to real data.

Contributions

Conceptualisation & design of work: M.T., J.-K. H., C.R., D.H., A.S., J.R., J.E., A.B., and V.U.; Data curation: M.T. and J.-K. H.; Data analysis & interpretation: M.T., J.-K. H., C.R., A.B., and V.U.; Writing – original draft: M.T., J.-K. H., C.R., A.B., and V.U.; Writing - review & editing: M.T., J.-K. H., C.R., D.H., J.R., A.B., and V.U.

Acknowledgements

We thank Jeremy Tempkin, Wanlu Zhang, Vilma Jimenez Sabinina, Zhichao Miao, Gerard Kleywegt, and Jan Kosinski for discussions. Special thanks to Yu-Le Wu and Philipp Hoess for discussions and help with SMAP. This work was supported by Wellcome Trust grant 212962/Z/18/Z.

References

[1] G. Paci, J. Caria, and E. A. Lemke, "Cargo transport through the nuclear pore complex at a glance," *Journal of Cell Science*, vol. 134, no. 2, jcs247874, Jan. 2021, ISSN: 0021-9533. DOI: 10.1242/jcs.247874. [Online]. Available: http://jcs.biologists.org/lookup/doi/10.1242/jcs.247874.

[2] R. Reichelt, "Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components," *The Journal of Cell Biology*, vol. 110, no. 4, pp. 883–894, Apr. 1990, ISSN: 0021-9525. DOI: 10.1083/jcb.110.4.883. [Online]. Available: http://www.jcb.org/cgi/doi/10.1083/jcb.110.4.883.

[3] T. U. Schwartz, "The Structure Inventory of the Nuclear Pore Complex," *Journal of Molecular Biology*, vol. 428, no. 10, pp. 1986–2000, May 2016, ISSN: 00222836. DOI: 10.1016/j.jmb.2016.03.015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0022283616002011.

[4] D. H. Lin and A. Hoelz, "The Structure of the Nuclear Pore Complex (An Update)," *Annual Review of Biochemistry*, vol. 88, no. 1, pp. 725–783, Jun. 2019, ISSN: 0066-4154. DOI: 10.1146/annurev-biochem-062917-011901. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev-biochem-062917-011901.

[5] A. von Appen, J. Kosinski, L. Sparks, A. Ori, A. L. DiGuilio, B. Vollmer, M.-T. Mackmull, N. Banterle, L. Parca, P. Kastritis, K. Buczak, S. Mosalaganti, W. Hagen, A. Andres-Pons, E. A. Lemke, P. Bork, W. Antonin, J. S. Glavy, K. H. Bui, and M. Beck, "In situ structural analysis of the human nuclear pore complex," *Nature*, vol. 526, no. 7571, pp. 140–143, Oct. 2015, ISSN: 0028-0836. DOI: 10.1038/nature15381. [Online]. Available: http://www.nature.com/articles/nature15381.

[6] J. Kosinski, S. Mosalaganti, A. von Appen, R. Teimer, A. L. DiGuilio, W. Wan, K. H. Bui, W. J. H. Hagen, J. A. G. Briggs, J. S. Glavy, E. Hurt, and M. Beck, "Molecular architecture of the inner ring scaffold of the human nuclear pore complex," *Science*, vol. 352, no. 6283, pp. 363–365, Apr. 2016, ISSN: 0036-8075. DOI: 10.1126/science.aaf0643. [Online]. Available: http://www.sciencemag.org/cgi/doi/10.1126/science.aaf0643.

[7] A. P. Schuller, M. Wojtynek, D. Mankus, M. Tatli, R. Kronenberg-Tenga, S. G. Regmi, P. V. Dip, A. K. R. Lytton-Jean, E. J. Brignole, M. Dasso, K. Weis, O. Medalia, and T. U. Schwartz, "The cellular environment shapes the nuclear pore complex architecture," *Nature*, vol. 598, no. October, 2021, ISSN: 0028-0836. DOI: 10.1038/s41586-021-03985-3.

[8] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021, ISSN: 14764687. DOI: 10.1038/s41586-021-03819-2. [Online]. Available: http://dx.doi.org/10.1038/s41586-021-03819-2.

[9] S. Mosalaganti, A. Obarska-kosinska, M. Siggel, C. E. Zimmerli, K. Buczak, F. H. Schmidt, M.-t. Mackmull, W. Hagen, G. Hummer, M. Beck, J. Kosinski, C. B. Unit, E. Molecular, A. Arbor, and E. Molecular, "Title : Artificial intelligence reveals nuclear pore complexity," 2021.

[10] V. Rantos, K. Karius, and J. Kosinski, "Integrative structural modeling of macromolecular complexes using Assembline," *Nature Protocols*, 2021, ISSN: 17502799. DOI: 10.1038/s41596-021-00640-z.

[11] J. E. Hinshaw and R. A. Milligan, "Nuclear pore complexes exceeding eightfold rotational symmetry," *Journal of Structural Biology*, vol. 141, no. 3, pp. 259–268, Mar. 2003, ISSN: 10478477. DOI: 10.1016/S1047-8477(02)00626-3. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1047847702006263.

[12] A. Löschberger, C. Franke, G. Krohne, S. van de Linde, and M. Sauer, "Correlative super-resolution fluorescence and electron microscopy of the nuclear pore complex with molecular resolution," *Journal of Cell Science*, vol. 127, no. 20, pp. 4351–4355, Jan. 2014, ISSN: 1477-9137. DOI: 10.1242/jcs.156620. [Online]. Available: http://jcs.biologists.org/lookup/doi/10.1242/jcs.156620%20https://journals.biologists.com/jcs/article/doi/10.1242/jcs.156620/259211/Correlative-super-resolution-fluorescence-and.

[13] G. J. Stanley, A. Fassati, and B. W. Hoogenboom, "Atomic force microscopy reveals structural variability amongst nuclear pore complexes," *Life Science Alliance*, vol. 1, no. 4, pp. 1–13, 2018, ISSN: 25751077. DOI: 10.26508/lsa.201800142.

[14] M. Beck, V. Lučić, F. Förster, W. Baumeister, and O. Medalia, "Snapshots of nuclear pore complexes in action captured by cryo-electron tomography," *Nature*, vol. 449, no. 7162, pp. 611–615, Oct. 2007, ISSN: 0028-0836. DOI: 10.1038/nature06170. [Online]. Available: http://www.nature.com/articles/nature06170.

[15] T. Maimon, N. Elad, I. Dahan, and O. Medalia, "The human nuclear pore complex as revealed by cryo-electron tomography," *Structure*, vol. 20, no. 6, pp. 998–1006, 2012, ISSN: 09692126. DOI: 10.1016/j.str.2012.03.025. [Online]. Available: http://dx.doi.org/10.1016/j.str.2012.03.025.

[16] T. A. Huijben, H. Heydarian, A. Auer, F. Schueder, R. Jungmann, S. Stallinga, and B. Rieger, "Detecting structural heterogeneity in single-molecule localization microscopy data," *Nature Communications*, vol. 12, no. 1, p. 3791, Dec. 2021, ISSN: 2041-1723. DOI: 10.1038/s41467-021-24106-8. [Online]. Available: http://dx.doi.org/10.1038/s41467-021-24106-8%20http://www.nature.com/articles/s41467-021-24106-8.

[17] H. Heydarian, M. Joosten, A. Przybylski, F. Schueder, R. Jungmann, B. V. Werkhoven, J. Keller-Findeisen, J. Ries, S. Stallinga, M. Bates, and B. Rieger, "3D particle averaging and detection of macromolecular symmetry in localization microscopy," *Nature Communications*, vol. 12, no. 1, p. 2847, Dec. 2021, ISSN: 2041-1723. DOI: 10.1038/s41467-021-22006-5. [Online]. Available: http://dx.doi.org/10.1038/s41467-021-22006-5%20http://www.nature.com/articles/s41467-021-22006-5.

[18] V. J. Sabinina, M. J. Hossain, J.-k. Hériché, P. Hoess, B. Nijmeijer, S. Mosalaganti, M. Kueblbeck, A. Callegari, A. Szymborska, M. Beck, J. Ries, and J. Ellenberg, "Three-dimensional superresolution fluorescence microscopy maps the variable molecular architecture of the nuclear pore complex," *Molecular Biology of the Cell*, vol. 32, no. 17, J. Lippincott-Schwartz, Ed., pp. 1523–1533, Aug. 2021, ISSN: 1059-1524. DOI: 10.1091/mbc.E20-11-0728. [Online]. Available: https://www.molbiolcell.org/doi/10.1091/mbc.E20-11-0728.

[19] C. E. Zimmerli, M. Allegretti, V. Rantos, S. K. Goetz, A. Obarska-Kosinska, I. Zagoriy, A. Halavatyi, G. Hummer, J. Mahamid, J. Kosinski, and M. Beck, "Nuclear pores dilate and constrict in cellulo.," *Science (New York, N.Y.)*, vol. 9776, eabd9776, 2021, ISSN: 1095-9203. DOI: 10.1126/science.abd9776. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/34762489.

[20] A. Curd, J. Leng, R. Hughes, A. Cleasby, B. Rogers, C. Trinh, M. Baird, Y. Takagi, C. Tiede, C. Sieben, S. Manley, T. Schlichthaerle, R. Jungmann, J. Ries, H. Shroff, and M. Peckham, "Nanoscale pattern extraction from relative positions of sparse 3D localisations," 2020. DOI: 10.1101/2020.02.13.947135.

[21] C. Wolf and M. R. Mofrad, "On the octagonal structure of the nuclear pore complex: Insights from coarse-grained models," *Biophysical Journal*, vol. 95, no. 4, pp. 2073–2085, 2008, ISSN: 15420086. DOI: 10.1529/biophysj.108.130336. [Online]. Available: http://dx.doi.org/10.1529/biophysj.108.130336.

[22] T. R. Lezon, A. Sali, and I. Bahar, "Global motions of the nuclear pore complex: Insights from elastic network models," *PLoS Computational Biology*, vol. 5, no. 9, 2009, ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1000496.

[23] Y. Li, W. Shi, S. Liu, U. Matti, D. Wu, and J. Ries, "Global fitting for high-accuracy multi-channel single-molecule localization," *bioRxiv*, vol. 7, 2021. DOI: 10.1101/2021.09.22.461230.

[24] E. Abbe, "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung.," *Archiv für mikroskopische Anatomie*, 1873.

[25] Y.-l. Wu, P. Hoess, A. Tschanz, U. Matti, M. Mund, and J. Ries, "Maximum-likelihood model fitting for quantitative analysis of SMLM data," pp. 1–39, 2021.

[26] Y.-l. Wu, A. Tschanz, L. Krupnik, and J. Ries, "Quantitative Data Analysis in Single-Molecule Localization Microscopy," *Trends in Cell Biology*, vol. 30, no. 11, pp. 837–851, Nov. 2020, ISSN: 09628924. DOI: 10.1016/j.tcb.2020.07.005. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S096289242030146X.

[27] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–796, Oct. 2006, ISSN: 1548-7091. DOI: 10.1038/nmeth929. [Online]. Available: http://www.nature.com/articles/nmeth929.

[28] I. Testa, C. A. Wurm, R. Medda, E. Rothermel, C. Von Middendorf, J. Fölling, S. Jakobs, A. Schönle, S. W. Hell, and C. Eggeling, "Multicolor fluorescence nanoscopy in fixed and living cells by exciting conventional fluorophores with a single wavelength," *Biophysical Journal*, vol. 99, no. 8, pp. 2686–2694, 2010, ISSN: 15420086. DOI: 10.1016/j.bpj.2010.08.012.

[29] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy," *Science*, vol. 319, no. 5864, pp. 810–813, Feb. 2008, ISSN: 0036-8075. DOI: 10.1126/science.1153529. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18174397%20https://www.sciencemag.org/lookup/doi/10.1126/science.1153529.

[30] C. v. Middendorff, A. Egner, C. Geisler, S. W. Hell, and A. Schönle, "Isotropic 3D Nanoscopy based on single emitter switching," *Optics Express*, vol. 16, no. 25, p. 20 774, Dec. 2008, ISSN: 1094-4087. DOI: 10.1364/OE.16.020774. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?uri=oe-16-25-20774.

[31] R. Diekmann, M. Kahnwald, A. Schoenit, J. Deschamps, U. Matti, and J. Ries, "Optimizing imaging speed and excitation intensity for single-molecule localization microscopy," *Nature Methods*, vol. 17, no. 9, pp. 909–912, Sep. 2020, ISSN: 1548-7091. DOI: 10.1038/s41592-020-0918-5. [Online]. Available: http://dx.doi.org/10.1038/s41592-020-0918-5%20http://www.nature.com/articles/s41592-020-0918-5.

[32] J. V. Thevathasan, M. Kahnwald, K. Cieśliński, P. Hoess, S. K. Peneti, M. Reitberger, D. Heid, K. C. Kasuba, S. J. Hoerner, Y. Li, Y.-L. Wu, M. Mund, U. Matti, P. M. Pereira, R. Henriques, B. Nijmeijer, M. Kueblbeck, V. J. Sabinina, J. Ellenberg, and J. Ries, "Nuclear pores as versatile reference standards for quantitative superresolution microscopy," *Nature Methods*, vol. 16, no. 10, pp. 1045–1053, Oct. 2019, ISSN: 1548-7091. DOI: 10.1038/s41592-019-0574-9. [Online]. Available: http://www.nature.com/articles/s41592-019-0574-9.

[33] G. Jacquemet, A. F. Carisey, H. Hamidi, R. Henriques, and C. Leterrier, "The cell biologist's guide to super-resolution microscopy," *Journal of Cell Science*, vol. 133, no. 11, jcs240713, Jun. 2020, ISSN: 0021-9533. DOI: 10.1242/jcs.240713. [Online]. Available: http://jcs.biologists.org/lookup/doi/10.1242/jcs.240713.

[34] J. Sinkó, R. Kákonyi, E. Rees, D. Metcalf, A. E. Knight, C. F. Kaminski, G. Szabó, and M. Erdélyi, "TestSTORM: Simulator for optimizing sample labeling and image acquisition in localization based super-resolution microscopy," *Biomedical Optics Express*, vol. 5, no. 3, p. 778, Mar. 2014, ISSN: 2156-7085. DOI: 10.1364/BOE.5.000778. [Online]. Available: https://www.osapublishing.org/boe/abstract.cfm?uri=boe-5-3-778.

[35] V. Venkataramani, F. Herrmannsdörfer, M. Heilemann, and T. Kuner, "SuReSim: simulating localization microscopy experiments from ground truth models," *Nature Methods*, vol. 13, no. 4, pp. 319–321, Apr. 2016, ISSN: 1548-7091. DOI: 10.1038/nmeth.3775. [Online]. Available: http://www.nature.com/articles/nmeth.3775.

[36] T. Novák, T. Gajdos, J. Sinkó, G. Szabó, and M. Erdélyi, "TestSTORM: Versatile simulator software for multimodal super-resolution localization fluorescence microscopy," *Scientific Reports*, vol. 7, no. 1, p. 951, Dec. 2017, ISSN: 2045-2322. DOI: 10.1038/s41598-017-01122-7. [Online]. Available: http://www.nature.com/articles/s41598-017-01122-7.

[37] J. Griffié, T. A. Pham, C. Sieben, R. Lang, V. Cevher, S. J. Holden, M. Unser, S. Manley, and D. Sage, "Virtual - SMLM , a virtual environment for real - time interactive SMLM acquisition," *bioRxiv*, 2020. DOI: https://doi.org/10.1101/2020.03.05.967893.

[38] D. Sage, T. A. Pham, H. Babcock, T. Lukes, T. Pengo, J. Chao, R. Velmurugan, A. Herbert, A. Agrawal, S. Colabrese, A. Wheeler, A. Archetti, B. Rieger, R. Ober, G. M. Hagen, J. B. Sibarita, J. Ries, R. Henriques, M. Unser, and S. Holden, "Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software," *Nature Methods*, vol. 16, no. 5, pp. 387–395, 2019, ISSN: 15487105. DOI: 10.1038/s41592-019-0364-4. [Online]. Available: http://dx.doi.org/10.1038/s41592-019-0364-4.

[39] J. Ries, "SMAP: a modular super-resolution microscopy analysis platform for SMLM data," *Nature Methods*, vol. 17, no. 9, pp. 870–872, Sep. 2020, ISSN: 1548-7091. DOI: 10.1038/s41592-020-0938-1. [Online]. Available: http://dx.doi.org/10.1038/s41592-020-0938-1%20http://www.nature.com/articles/s41592-020-0938-1.

[40] R. Lincoln, M. L. Bossi, M. Remmel, E. D'Este, A. N. Butkevich, and S. W. Hell, "A general design of caging-group free photoactivatable fluorophores for live-cell nanoscopy," *bioRxiv*, p. 2021.11.15.468659, 2021. [Online]. Available: https://www.biorxiv.org/content/10.1101/2021.11.15.468659v1%7B%5C%%7D0Ahttps://www.biorxiv.org/content/10.1101/2021.11.15.468659v1.abstract.

[41] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, and T. E. Ferrin, "UCSF ChimeraX: Meeting modern challenges in visualization and analysis," *Protein Science*, vol. 27, no. 1, pp. 14–25, 2018, ISSN: 1469896X. DOI: 10.1002/pro.3235.

[42] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin, "UCSF ChimeraX: Structure visualization for researchers, educators, and developers," *Protein Science*, vol. 30, no. 1, pp. 70–82, 2021, ISSN: 1469896X. DOI: 10.1002/pro.3943.

[43] *circle-fit library GitHub Repository*. [Online]. Available: https://github.com/AlliedToasters/circle-fit.

[44] A. von Appen and M. Beck, "Structure Determination of the Nuclear Pore Complex with Three-Dimensional Cryo electron Microscopy," *Journal of Molecular Biology*, vol. 428, no. 10, pp. 2001–2010, May 2016, ISSN: 00222836. DOI: 10.1016/j.jmb.2016.01.004. [Online]. Available: http://dx.doi.org/10.1016/j.jmb.2016.01.004%20https://linkinghub.elsevier.com/retrieve/pii/S0022283616000231.

[45] M. Allegretti, C. E. Zimmerli, V. Rantos, F. Wilfling, P. Ronchi, H. K. Fung, C. W. Lee, W. Hagen, B. Turoňová, K. Karius, M. Börmel, X. Zhang, C. W. Müller, Y. Schwab, J. Mahamid, B. Pfander, J. Kosinski, and M. Beck, "In-cell architecture of the nuclear pore and snapshots of its turnover," *Nature*, no. August 2019, 2020, ISSN: 14764687. DOI: 10.1038/s41586-020-2670-5.

[46] S. J. Kim, J. Fernandez-Martinez, I. Nudelman, Y. Shi, W. Zhang, B. Raveh, T. Herricks, B. D. Slaughter, J. A. Hogan, P. Upla, I. E. Chemmama, R. Pellarin, I. Echeverria, M. Shivaraju, A. S. Chaudhury, J. Wang, R. Williams, J. R. Unruh, C. H. Greenberg, E. Y. Jacobs, Z. Yu, M. J. de la Cruz, R. Mironska, D. L. Stokes, J. D. Aitchison, M. F. Jarrold, J. L. Gerton, S. J. Ludtke, C. W. Akey, B. T. Chait, A. Sali, and M. P. Rout, "Integrative structure and functional anatomy of a nuclear pore complex," *Nature*, vol. 555, no. 7697, pp. 475–482, Mar. 2018, ISSN: 0028-0836. DOI: 10.1038/nature26003. [Online]. Available: http://dx.doi.org/10.1038/nature26003%20http://www.nature.com/articles/nature26003.

[47] X. Shi, G. Garcia, Y. Wang, J. F. Reiter, and B. Huang, "Deformed alignment of super-resolution images for semi-flexible structures," *PLoS ONE*, vol. 14, no. 3, pp. 1–12, 2019, ISSN: 19326203. DOI: 10.1371/journal.pone.0212735.

[48] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M.

Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. Vázquez-Baeza, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, ISSN: 1548-7091. DOI: 10.1038/s41592-019-0686-2. arXiv: 1907.10121. [Online]. Available: http://www.nature.com/articles/s41592-019-0686-2.

[49] C. Chatfield, *Simple Method for Distance to Ellipse*, 2017. [Online]. Available: https://blog.chatfield.io/simple-method-for-distance-to-ellipse/.

Supplementary Material

1 Notation

Information on notation can be found in Table S1.

Table S1. Important notation

Name	Variable	Note
Symmetry	s	Default $s = 8$
Copy number of Nup, mass of node	m	
Number of rings	g	$m = gs$. Not the same as CR, NR, IR
Coordinates of node i	$\vec{x}_i = (x_i, y_i)$	
Central anchor coordinates	$\vec{x}_0$	At (0, 0)
Centre coordinates	$\vec{x}_c$	Not necessarily at (0,0)
Indices	i, j, l	
Radius	r	
Nup of interest i	nup_i	
Number of connected neighbours of a node	$2h$	
Spring constant	k	
Magnitude (mag) of deformation	D_{mag}	
Standard deviation, RBF kernel free parameter	σ	
Arithmetic mean of x	$\bar{x}$	
Ring distance, damping constant	d	
Angle of Nup i within rot. unit	α_i	
Azimuthal angle between CR and NR	θ	Also called "twist angle"
Argument of parametric form of of ellipse and evolute, time	t	Eq. 36, 37
Rotational angle of ellipse	ρ	
Rotation whole NPC	δ	
Generic max count	n	
Semiminor axis ellipse	b	
Semimajor axis ellipse, acceleration	a	
Semiminor/semimajor axis ratio ellipse	q	$q = b/a$
Perimeter	p	
Scale factor, Polynomial coefficient in implicit form of ellipse	C	
Sign	sgn	$+, -, 0$
Euclidean distance between $\vec{x}_i$ and $\vec{x}_j$	$\lVert\vec{x}_i - \vec{x}_j\rVert$	
Geodesic distance of a circle segment	d_{segment}	segment described as angle

2 Extracting Nup Coordinates from PDB

2.1 Finding the model centre

We find the model centre $\vec{x}_c$ by calculating the midpoint of a Nup i with coordinates $\vec{x}_i = (x_i, y_i)$, $i \in 1 \ldots s$, and the opposite Nup of the same type and on the same ring with index $j = \text{mod}(i + 0.5s, s)$ (Eq. 7). This assumes s of the underlying model is even. We verify $\vec{x}_c$ for multiple opposing Nup pairs.

$$\vec{x}_c = 0.5(\vec{x}_i + \vec{x}_j) \tag{7}$$

2.2 Finding radii and angles of Nups within one rot. unit

One arbitrary $n\vec{u}p_i$ out of $i \in 1 \ldots g$, is set as reference: $n\vec{u}p_{ref} = (x_{ref}, y_{ref})$. α_i of $n\vec{u}p_i = (x_i, y_i)$ relative to $\vec{x}_c$ and $n\vec{u}p_{ref}$ is calculated via Eq. 8. Anticlockwise angles are positive.

$$\alpha_i = atan2(y_i - y_c, x_i - x_i) - atan2(y_{ref} - y_c, x_{ref} - x_c) \tag{8}$$

If $\alpha_i > \pi$ or if $\alpha_i < -\pi$ it is represented as the explementary angle $\alpha + 2\pi$ or $\alpha - 2\pi$ respectively. Finally, mean angle $\bar{\alpha}$ is computed (Eq. 9) and used to zero-centre α_i (Eq. 10).

$$\bar{\alpha} = \sum_{i=1}^{g} \frac{\alpha_i}{g} \tag{9}$$

$$\alpha_i' = \alpha_i - \bar{\alpha} \text{ for } i \in 1 \dots g \tag{10}$$

r_i is the Euclidean distance between $n\vec{u}p_i$ and $\vec{x}_c$ (Eq. 11) and is converted to nm.

$$r_i = \|n\vec{u}p_i - \vec{x}_c\| \tag{11}$$

3 Layout of the spring model

Nodes are connected to their $2h$, $h \in 0 \dots \lfloor s/2 \rfloor$, neighbours via circumferential springs. Higher values of h require higher deforming forces to achieve sufficiently irregular deformations. Computation time would therefore increase due to these higher forces, and due to the additional number of springs. By contrast, small values for h lead to overly jagged deformations (Figure 3C). h can be set by the user and is 2 by default. Furthermore, spring constant k of the shortest circumferential spring is 1 and decreases inversely proportional with spring length. If h is set to $\lfloor s/2 \rfloor$, the nodes are fully connected. Consequently, if s is even, opposite nodes are connected by two springs. k for these springs are therefore halved to make them equivalent to a single spring. Furthermore, spring constant k of radial springs is 0.7, a value lower than that of the shortest circumferential spring, as we assume protein-protein interactions to be stronger than protein-membrane interactions. A lower value would lead to more drift upon deformation and concomitantly to longer computation times.

4 Ordinary differential equations compute deformation

Deformations are computed for each ring separately. Nodes are therefore $i \in 1 \dots s$. To calculate the position $\vec{x}_i = (x_i, y_i)$ of a node i in response to a force $\vec{f_i} = (f_{x_i}, f_{y_i})$, Newton's second law of motion (Eq. 12) can be consulted:

$$\vec{f_i} = m_i \cdot \vec{a}_i \tag{12}$$

Henceforth, we assume the mass of the node $m_i = 1$. Acceleration of the node $\vec{a}_i$ is the second derivative of $\vec{x}_i$ over time t, which changes Eq. 12 to Eq. 13:

$$\frac{d^2\vec{x}_i}{dt^2} = \vec{f_i} \tag{13}$$

This only holds true for a disconnected node. Hooke's spring law states that a force stored in a spring between nodes i and j depends on spring constant k_{ij}, and the displacement of the spring from its resting length l_{ij}. In a spring system with s nodes, the force $\vec{f_{is}}$ from the system on a node i is the sum of the forces stored in each spring (Eq. 14). Here, $j = 0$ for an anchor node at $\vec{x}_0$ and $j \in 1 \dots s$ for the NPC nodes.

$$\vec{f_{is}} = \sum_{j=0, j \neq i}^{s} k_{ij} \frac{\vec{x}_i - \vec{x}_j}{\|\vec{x}_i - \vec{x}_j\|} \left(l_{ij} - \|\vec{x}_i - \vec{x}_j\|\right) \tag{14}$$

The acceleration, and therefore position $\vec{x}_i$ of a node i over time thus depends on both an external force $\vec{f}_i$, and the force $\vec{f}_{is}$ stored in the system (Eq. 15):

$$\frac{d^2\vec{x}_i}{dt^2} = \vec{f}_i + \vec{f}_{is} \tag{15}$$

Due to the conservation of energy, displacing a node from its equilibrium would result in continuous movements of the node. Damping $-d\frac{d\vec{x}_i}{dt}$ needs to be applied to the node, so that it reaches a new equilibrium (Eq. 16). d is a damping constant and $\frac{d\vec{x}_i}{dt}$ is the velocity of the node.

$$\frac{d^2\vec{x}_i}{dt^2} = \vec{f}_i + \sum_{j=0,j\neq i}^{s} k_{ij}\left(\frac{\vec{x}_i - \vec{x}_j}{\|\vec{x}_i - \vec{x}_j\|}\left(l_{ij} - \|\vec{x}_i - \vec{x}_j\|\right)\right) - d\frac{d\vec{x}_i}{dt} \tag{16}$$

We are interested in static variability. To obtain solutions at equilibrium, we numerically solve the system using scipy.integrate.solve_ivp [48] until time t, where t is large.

5 Geometric variability

5.1 Symmetry

To determine the new radius r' of an NPC with deviant symmetry s, an isosceles triangle between two neighbouring Nups and the NPC centre is constructed. The angle at the centre is $\beta = 2\pi/s$. The other two corners both have angle $\gamma = 0.5(\pi - \beta)$. Half the distance between neighbouring Nups $b_{\frac{1}{2}}$ is calculated via $b_{\frac{1}{2}} = r \cdot sin(\pi/8)$, where r is the radius of the 8-fold symmetric NPC ring. The updated radius r' now corresponds to the hypotenuse of the right triangle between a Nup, the half-point to its neighbouring Nup, and the centre (Eq. 17).

$$r' = \frac{b_{\frac{1}{2}}}{cos(\gamma)} \tag{17}$$

If a Nup occurs in multiple rings $i \in 1 \ldots g$, each with radii r_i, their initial ($\bar{r}$) and updated ($\bar{r'}$) arithmetic means are computed. Next, $r'_i = \triangle r_i + \bar{r'}$ with $\triangle r_i = r_i - \bar{r}$. Consequently, radial distances between Nups in different rings are preserved.
For all rings, angles α_i of Nups within one rot. unit are updated to preserve geodesic distance of adjacent Nups. First, geodesic distance of Nups in an NPC with $s = 8$ is calculated via $d_{\alpha_i} = \alpha_i \cdot \bar{r}$. Updated angles α'_i are calculated via $\alpha'_i = d_{\alpha_i}/\bar{r'}$ (Figure 4, "Symmetry").

5.2 Twist Angle

To assign Nups to CR or NR, we compute their average z position $\bar{z}$. rings i with $i \in 1 \ldots g$ are assigned to the CR, if $z_i > \bar{z}$ and to the NR if $z_i < \bar{z}$. If $n > 1$ rings of the NoI occur in the CR, the arithmetic mean $\bar{\alpha}_{\mathrm{CR}}$ of the ring angles is used for further calculations (Eq. 18). $\bar{\alpha}_{\mathrm{NR}}$ is computed analogously.

$$\bar{\alpha}_{\mathrm{CR}} = \sum_{i=1}^{n} \frac{\alpha_{i,\mathrm{CR}}}{n} \tag{18}$$

Furthermore, a rot. unit does not always contain Nups with the most acute (closest to 0) angular distance $\theta = \bar{\alpha}_{\mathrm{NR}} - \bar{\alpha}_{\mathrm{CR}}$. Therefore, we probe the next clockwise- (cw) $\bar{\alpha}_{\mathrm{CR,cw}} = \bar{\alpha} - 2\pi/8$ and anticlockwise (acw) angle $\bar{\alpha}_{\mathrm{CR,acw}} = \bar{\alpha} + 2\pi/8$, and set the final twist angle θ to whichever $\bar{\alpha}_{\mathrm{NR}} - \bar{\alpha}_{\mathrm{CR,cw}}$, $\bar{\alpha}_{\mathrm{NR}} - \bar{\alpha}_{\mathrm{CR}}$, or $\bar{\alpha}_{\mathrm{NR}} - \bar{\alpha}_{\mathrm{CR,acw}}$ is the most acute.

This angle θ can then be changed to user-defined θ' (Eq. 19).

$$\begin{aligned} \alpha'_{i,\mathrm{CR}} &= \alpha_{i,\mathrm{CR}} + \frac{\theta - \theta'}{2} \\ \alpha'_{i,\mathrm{NR}} &= \alpha_{i,\mathrm{NR}} - \frac{\theta - \theta'}{2} \end{aligned} \tag{19}$$

5.3 Elongation

Elongated NPCs are simulated by computing deforming forces from an ellipse with user-defined semiminor/semimajor axis ratio q. To this end, the perimeter p_j of each NPC ring $j \in 1 \dots g$ is determined by $p_j = r_j 2\pi$. Next, semiminor- and semimajor axes b_j and a_j of an ellipse with perimeter $\sim p_j$ are determined (Eq. 20).

$$\begin{aligned} p_j &\approx 2\pi \sqrt{\frac{a_j^2 + b_j^2}{2}} \\ a_j &\approx \frac{\frac{p_j \sqrt{2}}{2\pi}}{\sqrt{1+q^2}} \\ b_j &\approx q \cdot a_j \end{aligned} \tag{20}$$

The resulting ellipse can be described with Eq. 21, where x and y are Cartesian coordinates.

$$\frac{x^2}{a_j^2} + \frac{y^2}{b_j^2} = 1 \tag{21}$$

NPC Cartesian coordinates are converted to polar coordinates and rotated by angle $\delta \sim U[-\pi, \pi]$ to vary the alignment of nodes with the later superimposed ellipse. Next, Polar coordinates are converted back to Cartesian coordinates.
To project NPC coordinates onto the resulting ellipse, a scale factor C_i needs to be found for each node $i \in 1 \dots s$ such that $C_i \cdot (x_i, y_i) = (x, y)$. C_i can be computed (Eq. 22) by substitution in Eq. 21.

$$\begin{aligned} \frac{(C_i \cdot x_i)^2}{a_j^2} + \frac{(C_i \cdot y_i)^2}{b_j^2} &= 1 \\ C_i &= \sqrt{\frac{a_j^2 b_j^2}{x_i^2 b_j^2 + y_i^2 a_j^2}} \end{aligned} \tag{22}$$

Difference $\triangle \vec{x}_i = (\triangle x_i, \triangle y_i)$ between the ellipse and NPC coordinates is computed (Eq. 23).

$$\triangle \vec{x}_i = (\triangle x_i, \triangle y_i) = C_i(x_i, y_i) - (x_i, y_i) \tag{23}$$

Next, the compressive ($\mathrm{sgn}(x) < 0$) or expanding ($\mathrm{sgn}(x) > 0$) forces $\vec{f}_{e_i}$ on a node i are determined (Eq. 24).

$$\vec{f}_{e_i} = \mathrm{sgn}(\triangle \vec{x}_i \cdot \vec{x}_i) \, \|\triangle \vec{x}_i\| \tag{24}$$

$\vec{f}_{e_i}$ therefore deform the NPC into an ellipse-like shape and are added to $\vec{f}_i$ from Eq. 3.

6 Fitting an Ellipse to analyse deformed NPCs

A standard ellipse has a centre $(0, 0)$, and axes b and a that align with the x and y axes of a Cartesian coordinate system (Eq. 25).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{25}$$

This ellipse can be rotated anticlockwise by ρ and translated to a new centre (x_c, y_c), expanding Eq. 25 to Eq. 26.

$$\frac{((x - x_c)cos(\rho) + (y - y_c)sin(\rho))^2}{a^2} + \frac{((x - x_c)sin(\rho) - (y - y_c)cos(\rho))^2}{b^2} = 1 \tag{26}$$

Eq. 26 can be written in polynomial form and its coefficients can substituted with A to F (Eq. 27).

$$\begin{aligned}
&AX^2 + BXY + CY^2 + DX + EY + F = 0 \\
&\text{with} \\
&A = a^2sin^2(\rho) + b^2cos^2(\rho) \\
&B = 2(b^2 - a^2)sin(\rho)cos(\rho) \\
&C = a^2cos^2(\rho) + b^2sin^2(\rho) \\
&D = -2Ax_c - By_c \\
&E = -Bx_c - 2Cy_c \\
&F = Ax_c^2 + Bx_cy_c + Cy_c^2 - a^2b^2
\end{aligned} \tag{27}$$

Assuming $F = -1$ this yields Eq. 28.

$$AX^2 + BXY + CY^2 + DX + EY = 1 \tag{28}$$

A matrix M is constructed from the x, y coordinates of model NPC nodes, whereby columns represent the indeterminates of Eq. 28 and rows the nodes $1 \ldots n$. Depending on which part of the NPC the ellipse is fitted to, n is either all nodes of model NPC ($n \in 1 \ldots m$), or all nodes within the CR or NR, or all nodes within a ring ($n \in 1 \ldots s$).

$$M = \begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_ny_n & y_n^2 & x_n & y_n \end{bmatrix} \tag{29}$$

Coefficients are written in column vector form $\vec{b}$ (Eq. 30) and optimised using least squares (Eq. 31).

$$\vec{b} = \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} \tag{30}$$

$$\arg\min_b \left\| M\vec{b} - \vec{1} \right\|^2 \tag{31}$$

a and b are then determined according to 32.

$$a, b = \frac{-\sqrt{2(AE^2 + CD^2 - BDE + (B^2 - 4AC)F)((A + C) \pm \sqrt{(A - C)^2 + B^2})}}{B^2 - 4AC} \tag{32}$$

In Eq. 32, either of a or b can corresponds to the semimajor axis. We therefore redefine a' as semimajor, and b' as semiminor (Eq. 33).

$$\begin{aligned} a' &= \max(a, b) \\ b' &= \min(a, b) \end{aligned} \tag{33}$$

Furthermore, we determine rotation ρ (Eq. 34).

$$\rho = \begin{cases} arctan(\frac{1}{B}(C - A - \sqrt{(A-C)^2 + B^2}) & \text{for } B \neq 0, \\ 0 & \text{for } B = 0, A < C, \\ 0.5\pi & \text{for } B = 0, A > C. \end{cases} \tag{34}$$

Next, ρ is expressed as a value between $[-0.5\pi, 0.5\pi]$(Eq. 35).

$$\begin{aligned} \rho' &= \begin{cases} \rho - 0.5\pi & \text{if } A < 0 \text{ and if } \rho \geq 0, \\ \rho + 0.5\pi & \text{if } A < 0 \text{ and if } \rho < 0. \end{cases} \\ &\text{and} \\ \rho &= \begin{cases} \rho' & \text{if } A < 0, \\ \rho & \text{otherwise} \end{cases} \end{aligned} \tag{35}$$

Consequently, ρ is the angle the right-hand semimajor axis forms with respect to a horizontal line. Fitted ellipses are shown in Figure 4B.

7 Approximating the RSS of the fitted ellipse

As for the fitted circle, the RSS of nodes to the fitted ellipse is a measure of deviation from it. However, the fitting procedure described in Suppl. section 6 does not straight-forwardly return the sum of squared Euclidean distances of nodes to the ellipse. We therefore estimate this measure by approximating the closest point on the fitted ellipse for each node [49], and then summing up the squared Euclidean distances of all pairs of nodes and closest points.
Model NPC and fitted ellipse are first transformed into standard ellipse form. For this, the model NPC is rotated by $-\rho$ around the centre of the fitted ellipse (x_c, y_c) and then translated so that its centre of mass is at $(0, 0)$. It consequently aligns with a standard ellipse with semimajor axis a' and semiminor axis b' of the fitted ellipse, henceforth called a and b. Therefore, the RSS between this rotated and shifted model NPC to the standard ellipse with a, b is equivalent to the RSS between the model NPC and fitted ellipse. Furthermore, projecting all m nodes with coordinates $\vec{x} = (x_i, y_i)$, $i \in 1 \ldots m$ onto the 1st quadrant $(x'_i = |x_i|, y'_i = |y_i|$, $\vec{x}' = (x'_i, y'_i))$ does not change the RSS. Any point on the ellipse can be approximated with a circle of the same curvature, whereby the centre of all such circles form a curve called evolute. The closest point on the ellipse $\vec{p}_i = (p_{x_i}, p_{y_i})$ of any point $\vec{x}' = (x'_i, y'_i)$ is therefore the tangential point between a circle around $\vec{x}'$ and the ellipse, or alternatively the circle approximating the ellipse at this point with corresponding centre $\vec{e}_i = (e_{xi}, e_{yi})$ on the evolute.
Both the ellipse and evolute can be parameterized based on an angle t (Eq. 36, Eq. 37).

$$\begin{aligned} p_x &= a \cdot cos(t) \\ p_y &= b \cdot sin(t) \end{aligned} \tag{36}$$

$$
\begin{aligned}
e_x &= (a^2 - b^2)\frac{cos(t)^3}{a} \\
e_y &= (b^2 - a^2)\frac{sin(t)^3}{b}
\end{aligned}
\tag{37}
$$

Consequently, for a point $\vec{x}_i'$, t for the closest point $\vec{p}$ needs to be approximated, which is done iteratively:
As all nodes have been projected to the 1st quadrant, t is initialised as the bisector of the 1st quadrant, $t_0 = \pi/4$ rad. The corresponding point on the ellipse is $\vec{p}(t_0) = (p_x(t_0), p_y(t_0))$, whereby $p_x(t_0) = a \cdot cos(t_0)$ and $p_y(t_0) = b \cdot sin(t_0)$. The centre of curvature of $\vec{p}(t_0)$ is at $\vec{e}(t_0)$, which is computed analogously using Eq. 37. The radius of curvature at t_0 is therefore $|\vec{r}(t_0)|$ with $\vec{r}(t_0) = (r_x(t_0), r_y(t_0)) = (p_x(t_0) - e_x(t_0), p_y(t_0) - e_y(t_0))$.
Furthermore, vector $\vec{q}(t_0)$ connects $\vec{e}(t_0)$ with $\vec{x_i'}$. $\vec{q}(t_0) = (q_x(t_0), q_y(t_0)) = (x_i' - e_x(t_0), y_i' - e_y(t_0))$. The arc length $\triangle c(t_0)$ between $\vec{r}(t_0)$ and $\vec{q}(t_0)$ on a circle around $\vec{e}(t_0)$ with radius $|\vec{r}(t_0)|$ is Eq. 38.

$$
\begin{aligned}
&\triangle c(t_0) = |\vec{r}|(t_0) \cdot arcsin(\frac{|\vec{r}(t_0) \times \vec{q}(t_0)|}{|\vec{r}|(t_0) \cdot |\vec{q}|(t_0)}) \\
&\text{with} \\
&\vec{r}(t_0) \times \vec{q}(t_0) = r_x(t_0) \cdot q_y(t_0) - r_y \cdot q_x(t_0)
\end{aligned}
\tag{38}
$$

Following, a new t_1 is found that approximately minimises $\triangle c(t_0)$ (Eq. 39, 40).

$$
\triangle t = \frac{\triangle c(t_0)}{\sqrt{a^2 + b^2 - (p_x(t_0))^2 - (p_y(t_0))^2}} \tag{39}
$$

$$
t_1 = t_0 + \triangle t \tag{40}
$$

Next, it is ensured t_1 remains in the 1st quadrant (Eq. 41).

$$
\begin{aligned}
t_1' &= \min(0.5\pi, \max(0, t_1)) \\
t_1 &= t_1'
\end{aligned}
\tag{41}
$$

t_1 updates $\vec{e}(t_1)$ and $\vec{p}(t_1)$. $\triangle_c(t_0)$ is the arc-length on the circle that approximates curvature at $\vec{p}(t_0)$, not on the ellipse. Therefore $\vec{p}(t_1)$ might not correspond to the closest point to $\vec{x_i'}$. However, when the above stated procedure is repeated for $j \in 0 \ldots n$ iterations, the circle approximating curvature at $\vec{p}(t_j)$ increasingly resembles the ellipse at the closest point to $\vec{x}'$. We set n to 2 and define $\vec{p}(t_2)$ as the closest point to $\vec{x_i'}$. $\vec{p}(t_2)$ is next projected onto the original quadrant of node $\vec{x}_i$ $p_x = p_x(t_2) \cdot \text{sgn}(x_i), p_y = p_y(t_2) \cdot \text{sgn}(y_i)$. Henceforth, we call the closest point $\vec{p}$ to $\vec{x_i}$ "$\vec{p_i}$".
Finally, the RSS is computed with Eq. 42, where n is the number of nodes.

$$
\text{RSS} = \sum_{i=1}^{n} (\vec{x}_i - \vec{p}_i)^2 \tag{42}
$$

For a circular NPC, this RSS is almost the same as for the fitted circle. For an elongated NPC, this RSS is smaller than that of a fitted circle. When all nodes are close to the ellipse, the RSS is close to 0.

8 Output files

Simulated NPCs are saved as CSV files that can be loaded into photophysics simulation software. Each node is described by coordinates x, y, z and an index per NPC. We do not introduce further rotation, translation or tilt to the NPC, as most downstream photophysics simulation software includes this functionality. CSV files can be loaded into photophysics simulation software (e.g. [39]).

9 Simulated Nups

N-terminus and C-terminus respectively denote the first and last structurally resolved amino acid, whose coordinates are used for simulations. n in CR, NR, IR, and in bridge denote the number of rings in each. Bridge is a linker between OR and IR Table S2.

Table S3. ChimeraX commands to extract rot. unit coordinates

Model	Nup	Terminus	Command
5A9Q	Nup43	N	getcr #1.1/0,R,9,I:4@CA
		C	getcr #1.1/0,R,9,I:380@CA
	Nup160	N	getcr #1.1/1,S,a,J:41@CA
		C	getcr #1.1/1,S,a,J:1201@CA
	Nup37	N	getcr #1.1/2,T,b,K:9@CA
		C	getcr #1.1/2,T,b,K:326@CA
	Sec13	N	getcr #1.1/6,X,F,O:14@CA
		C	getcr #1.1/6,X,F,O:304@CA
	Seh1	N	getcr #1.1/7,Y,G,P:1@CA
		C	getcr #1.1/7,Y,G,P:322@CA
	Nup85	N	getcr #1.1/Q,8,Z,H:9@CA
		C	getcr #1.1/Q,8,Z,H:475@CA
	Nup155	N	getcr #1.1/A,B:20@CA
		C	getcr #1.1/A,B:730@CA
	Nup133	N	getcr #1.1/3,U,C,L:518@CA
		C	getcr #1.1/3,U,C,L:1156@CA
	Nup96	N	getcr #1.1/5,W,E,N:277@CA
		C	getcr #1.1/5,W,E,N:751@CA
	Nup107	N	getcr #1.1/M,D,V,4:150@CA
		C	getcr #1.1/M,D,V,4:924@CA
7PEQ[1]	Nup133	N	getcrd /?C:518@CA
		C	getcrd /?C:1156@CA
	Nup107	N	getcrd /?D:150@CA
		C	getcrd /?D:924@CA
	Nup96	N	getcrd /?E:333@CA
		C	getcrd /?E:922@CA
	Sec13	N	getcrd /?F:14@CA
		C	getcrd /?F:304@CA
	Seh1	N	getcrd /?G:5@CA
		C	getcrd /?G:320@CA
	Nup85	N	getcrd /?H:20@CA
		C	getcrd /?H:651@CA
	Nup43	N	getcrd /?I:4@CA
		C	getcrd /?I:380@CA
	Nup160	N	getcrd /?J:78@CA
		C	getcrd /?J:1195@CA

	Nup37	N	getcrd /?K:18@CA
		C	getcrd /?K:324@CA
7PER[1]	Nup205	N	getcrd /D,J,V,P:9@CA
		C	getcrd /D,J,V,P:1692@CA
	Nup155	N	getcrd /E,K,Q,W:20@CA
		C	getcrd /E,K,Q,W:1375@CA
	Nup93	N	getcrd /C,I,O,U:173@CA
		C	getcrd /C,I,O,U:815@CA
	Nup54	N	getcrd /F,L,R,X:128@CA
		C	getcrd /F,L,R,X:493@CA
	Nup58 "P58P45"	N	getcrd /G,M,S,Y:248@CA
		C	getcrd /G,M,S,Y:418@CA
	Nup62	N	getcrd /H,N,T,Z:334@CA
		C	getcrd /H,N,T,Z:502@CA
5IJN	Nup205	N	getcrd #1.1/D,J,P,V:9@CA
		C	getcrd #1.1/D,J,P,V:1692@CA
	Nup54	N	getcrd #1.1/F,L,R,X:128@CA
		C	getcrd #1.1/F,L,R,X:493@CA
	Nup58	N	getcrd #1.1/G,M,S,Y:248@CA
		C	getcrd #1.1/G,M,S,Y:418@CA
	Nup155	N	getcrd #1.1/A,B,E,K,Q,W:20@CA
		C	getcrd #1.1/A,B:863@CA
			getcrd #1.1/E,K,Q,W:1375@CA
	Nup93	N	getcrd #1.1/C,I,O,U:1@CA
		C	getcrd #1.1/C,I,O,U:815@CA
	Nup62	N	getcrd #1.1/H,N,T,Z:334@CA
		C	getcrd #1.1/H,N,T,Z:502@CA
5IJO	Nup54	N	getcrd /F,L,R,X:128@CA
		C	getcrd /F,L,R,X:493@CA
	Nup188	N	getcrd /J,V:1@CA
		C	getcrd /J,V:1564@CA
	Nup205	N	getcrd /D,P:9@CA
		C	getcrd /D,P:1692@CA
	Nup155	N	getcrd /A,B,E,K,Q,W:20@CA
		C	getcrd /A,B:863@CA
			getcrd /E,K,Q,W:1375@CA
	Nup93	N	getcrd /C,I,O,U:1@CA
		C	getcrd /C,I,O,U:815@CA
	Nup58 "P58P45"	N	getcrd /G,M,S,Y:248@CA
		C	getcrd /G,M,S,Y:418@CA
	Nup62	N	getcrd /H,N,T,Z:334@CA
		C	getcrd /H,N,T,Z:502@CA

[1] Command for PDB file with single rot. unit. Output verified on Biological Assembly with all rot. units (not shown).

Table S2. Input models in detail

Model	Nup	N-term.	C-term.	n in CR	n in NR	n in IR	n in bridge	n total
5A9Q	Nup43	4	380	2	2			4
	Nup160	41	1201	2	2			4
	Nup37	9	326	2	2			4
	Sec13	14	304	2	2			4
	Seh1	1	322	2	2			4
	Nup85	9	475	2	2			4
	Nup155	20	730				2	2
	Nup133	518	1156	2	2			4
	Nup96	277	751	2	2			4
	Nup107	150	924	2	2			4
7PEQ	Nup133	518	1156	2	2			4
	Nup107	150	924	2	2			4
	Nup96	333	922	2	2			4
	Sec13	14	304	2	2			4
	Seh1	5	320	2	2			4
	Nup85	20	651	2	2			4
	Nup43	4	380	2	2			4
	Nup160	78	1195	2	2			4
	Nup37	18	324	2	2			4
7PER	Nup205	9	1692			4		4
	Nup155	20	1375			4		4
	Nup93	173	815			4		4
	Nup54	128	493			4		4
	Nup58 “P58P45”	248	418			4		4
	Nup62	334	502			4		4
5IJN	Nup205	9	1692			4		4
	Nup54	128	493			4		4
	Nup58	248	418			4		4
	Nup155	20	2 x 863 4 x 1375			4	2	6
	Nup93	1	815			4		4
	Nup62	334	502			4		4
5IJO	Nup54	128	493			4		4
	Nup188	1	1564			2		2
	Nup205	9	1692			2		2
	Nup155	20	2 x 863 4 x 1375			4	2	6
	Nup93	1	815			4		4
	Nup58 “P58P45”	248	418			4		4
	Nup62	334	502			4		4