PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Integrative spatiotemporal modeling of biomolecular processes: Application to the assembly of the nuclear pore complex

Andrew P. Latham[a,b], Wanlu Zhang[c], Jeremy O. B. Tempkin[a,b], Shotaro Otsuka[c,1,2], Jan Ellenberg[c], and Andrej Sali[a,b,3]

Affiliations are included on p. 9.

Edited by Gerhard Hummer, Max-Planck-Institut fur Biophysik, Frankfurt am Main, Germany; received August 8, 2024; accepted February 6, 2025 by Editorial Board Member J. A. McCammon

Dynamic processes involving biomolecules are essential for the function of the cell. Here, we introduce an integrative method for computing models of these processes based on multiple heterogeneous sources of information, including time-resolved experimental data and physical models of dynamic processes. First, for each time point, a set of coarse models of compositional and structural heterogeneity is computed (heterogeneity models). Second, for each heterogeneity model, a set of static integrative structure models is computed (a snapshot model). Finally, these snapshot models are selected and connected into a series of trajectories that optimize the likelihood of both the snapshot models and transitions between them (a trajectory model). The method is demonstrated by application to the assembly process of the human nuclear pore complex in the context of the reforming nuclear envelope during mitotic cell division, based on live-cell correlated electron tomography, bulk fluorescence correlation spectroscopy–calibrated quantitative live imaging, and a structural model of the fully assembled nuclear pore complex. Modeling of the assembly process improves the model precision over static integrative structure modeling alone. The method is applicable to a wide range of time-dependent systems in cell biology and is available to the broader scientific community through an implementation in the open source *Integrative Modeling Platform (IMP)* software.

integrative modeling | nuclear pore complex | molecular dynamics

Describing the dynamic processes involving biomacromolecules is essential for understanding the function of the cell. Thus, individual techniques, such as Förster resonance energy transfer, (1, 2) NMR, (3, 4) and physics-based molecular simulation, (5, 6) have been pioneered to characterize biomolecular processes. These methods have been successful in investigating complex phenomena, such as protein folding (7–10). However, describing the rearrangements of multicomponent protein systems, including the process of complex assembly, remains challenging because of the scale in space and time as well as the relative lack of information.

One approach to overcoming this challenge is integrative modeling. Integrative modeling benefits from simultaneously using varied information, including from different experimental data, statistical preferences, physical theories, and other prior models (11–13). By maximizing the input information, integrative modeling maximizes the accuracy, precision, and completeness of the output models. Static structures of protein complexes are already routinely determined by integrative modeling (14–22). For example, the nuclear pore complex (NPC) structure was computed by considering previously determined structures of its components, chemical cross-links between them, a cryoelectron microscopy map of the entire complex, and other information. (23–32) The human NPC consists of ∼1,000 copies of ∼30 different proteins, called nucleoporins (Nups). It spans the nuclear envelope and facilitates controlled macromolecular transport between the nucleus and the cytoplasm (33–35). However, despite the success of integrative modeling of static structures, no integrative method has yet been developed for modeling macromolecular processes, including the process of complex assembly (36).

Here, we present and illustrate an integrative spatiotemporal modeling method. First, a set of "heterogeneity models" are computed at each time point. Each heterogeneity model describes the biomolecular composition of the system and, optionally, an assignment of each component to its coarse location in the fully assembled complex. Next, for each heterogeneity model, a "snapshot model" is computed. A snapshot model is a set of static structural models computed to maximize agreement with both its corresponding heterogeneity model and other information about the system for the corresponding time point. Trajectories are then composed by selecting and connecting snapshot models

Significance

Biological systems are constantly in motion; therefore, describing the large-scale rearrangements of biological macromolecules is crucial to understanding their function. Yet, there is a dearth of experimental and computational approaches to study the dynamics of large macromolecular systems. In this paper, we introduce integrative spatiotemporal modeling, a generalization of integrative structural modeling of static protein structures. Our approach combines static and dynamic information from multiple experimental and computational sources. Thus, integrative spatiotemporal modeling produces a model of a macromolecular process that is more accurate, precise, and complete than possible based on any one source of information alone. The approach is demonstrated by its application to the assembly process of the nuclear pore complex in the context of the nuclear envelope.

This article is a PNAS Direct Submission. G.H. is a guest editor invited by the Editorial Board.

[1] Present address: Max Perutz Labs, Vienna Biocenter Campus, Vienna 1030, Austria.

[2] Present address: Medical University of Vienna, Center for Medical Biochemistry, Vienna 1090, Austria.

[3] To whom correspondence may be addressed. Email: sali@salilab.org.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2415674122/-/DCSupplemental.

between neighboring time points; these trajectories are scored by their fit to the input information. This final set of weighted trajectories is called the "trajectory model." This depiction of a dynamic process facilitates informing its model by the data created at discrete time points during the process. Such data are available for the assembly of the human NPC during mitosis (37–44).

Next, we outline the integrative spatiotemporal modeling workflow, followed by a description of our trajectory model of the NPC assembly pathway, a description of its agreement with the input information, an experimentally validated prediction made by our trajectory model of the NPC assembly pathway, and a comparison between our method and an alternative approach of modeling each snapshot independently, without considering connections between the snapshot models.

1. Integrative Spatiotemporal Modeling Workflow

Integrative spatiotemporal modeling of biomolecular processes is described by way of applying it to the assembly of the human NPC in the context of the nuclear envelope during cell division (*SI Appendix*, Table S1). The approach generalizes our method for integrative modeling of static structures. Our integrative modeling is an instance of statical modeling (45, 46), consisting of three steps: i) gathering information, ii) defining the model representation, scoring the models, as well as searching for good scoring models, and iii) validating the models (11–13).

While this integrative modeling method could in principle generate any type of model from any type of input information, the search process may be too large for sufficiently efficient computation. One approach to reduce the computational cost is to apply a composite workflow, which breaks one modeling problem into multiple instances of statistical modeling, where the previous model serves as a prior model in the next modeling stage. For example, in modeling of static structures, it is common to compute models of subunits before assembling them to model the full complex.

Our integrative spatiotemporal modeling follows a composite workflow to combine varied input information (Fig. 1*A*) into a model of a biomolecular process that represents a trajectory as a series of snapshot models (Fig. 1*B*). Three integrative modeling stages are applied in succession to compute i) heterogeneity

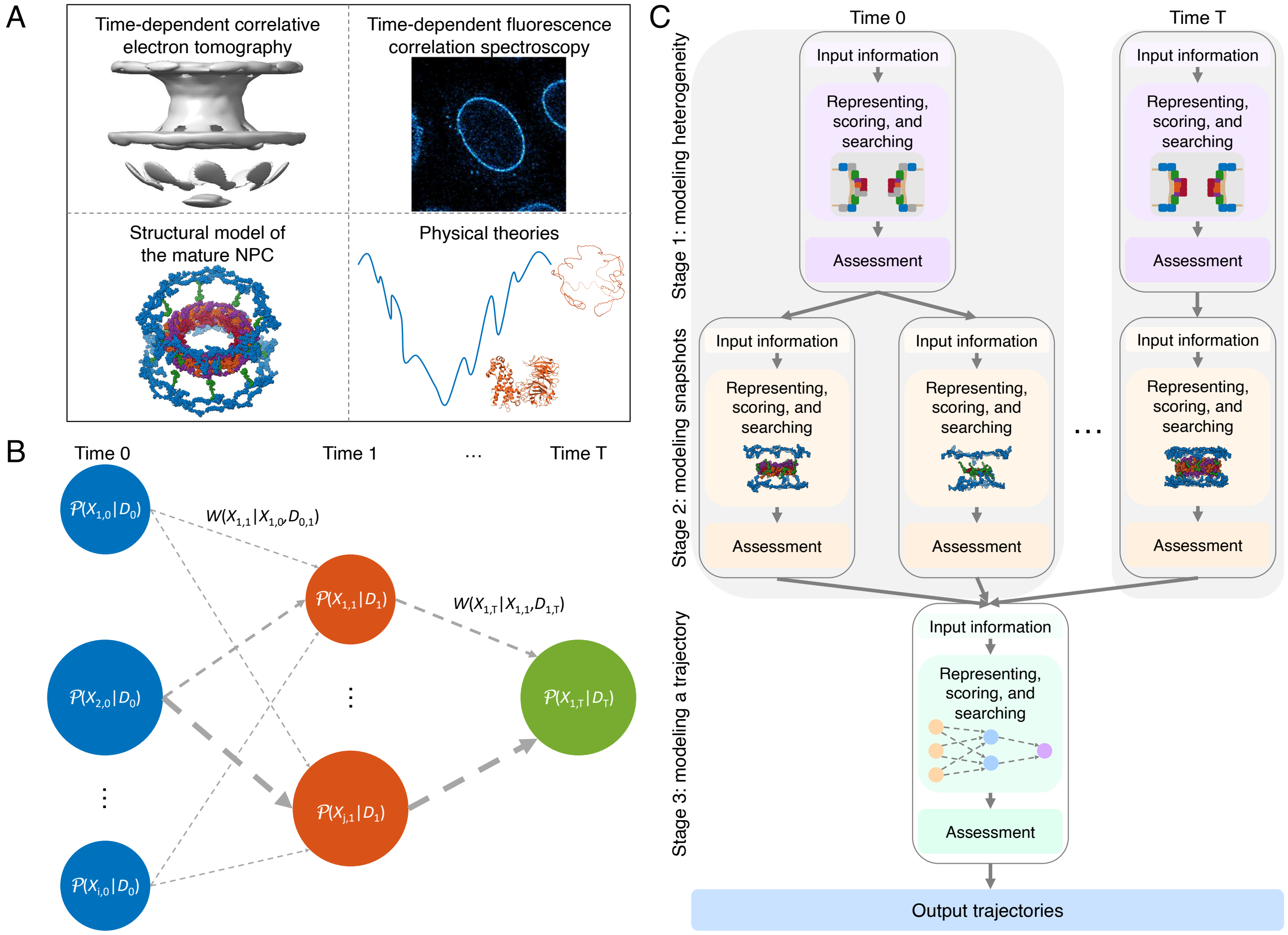

Fig. 1. Illustration of integrative spatiotemporal modeling. (*A*) The trajectory model of the NPC is computed based on structural models of the mature NPC, correlative ET, FCS, and physical theories. (*B*) The integrative spatiotemporal modeling workflow produces a weighted set of these trajectories. The weights of both individual snapshot models and transitions between these snapshot models contribute to the trajectory score (Eq. **3**). The weights of individual snapshot models are represented by the size of the nodes, while the weights of the transitions are represented by the width of the arrows. (*C*) The composite workflow for integrative spatiotemporal modeling proceeds through three stages to produce heterogeneity models, snapshot models, and a trajectory model. Each modeling stage proceeds through three steps: i) gathering information, ii) representing, scoring, and searching for models, and iii) assessing alternative models. For the NPC assembly process, the final time point (T) is the fully assembled structure, which has only one heterogeneity model. The output of our composite workflow is a set of weighted trajectories describing the biological process.

models at each time point, ii) a snapshot model for each heterogeneity model, and iii) a trajectory model (Fig. 1*C*). Ultimately, the workflow produces a set of weighted trajectories scored by their agreement with input information.

1.1. Stage 1: Modeling Heterogeneity. For modeling the NPC assembly process, heterogeneity models include the copy number of each protein as well as the subcomplex location assignment, which couples each protein in a time point along the assembly process to its coarse location in the mature NPC.

In the first step of heterogeneity modeling, information about the process of interest is gathered. Heterogeneity modeling utilizes the structural model of the mature NPC (PDB: 5a9q, 5ijn) (47, 48) and fluorescence correlation spectroscopy (FCS)-calibrated quantitative fluorescence microscopy data (37) (*SI Appendix*, Table S2). The structure of the mature NPC constrains the maximum copy number for each Nup and informs the division of NPC into subcomplexes, which are assumed to assemble together. Time-dependent information about the Nup copy number is provided by FCS data.

In the second step of heterogeneity modeling, the model representation is chosen, a scoring function is generated, and alternative models are searched for. Each heterogeneity model specifies the biomolecular composition of the assembling NPC and an assignment of each subcomplex to its coarse location in the mature NPC. Time points were selected based on the availability of ET maps: 5 min, 6 min, 8 min, 10 min, and 15 min after anaphase onset as well as the mature pore structure. The NPC subcomplexes were defined based on a structural model of the mature pore (*SI Appendix*, Table S3). Experimental FCS data were available for all subcomplexes except the connecting complex with Nup155, which was therefore coupled to the first appearing Y-complex on the same side of the nuclear envelope.

Each heterogeneity model was scored based on the data likelihood of the Nup copy numbers, $P(D_t|N_t)$:

$$P(D_t|N_t) \propto \prod_l \frac{1}{\sigma_l\sqrt{2\pi}} e^{-\frac{(n_{t,l}-\mu_l)^2}{2\sigma_l^2}}, \qquad [1]$$

where D_t is the FCS data at time t, N_t is the heterogeneity model at time t, l indexes the list of tagged Nups, $n_{t,l}$ is the copy number of Nup l in the heterogeneity model at time t, μ_l, and σ_l are respectively the mean and variance of the average copy number of fluorescently tagged Nups estimated in a previous study (37). D_t depends on 4 Nups (Nup107, Nup93, Nup205, and Nup62).

To create a set of alternative heterogeneity models at each time point, we applied the following four-step procedure. First, we enumerated eightfold symmetric subsets of the mature NPC subcomplexes. Second, we ranked each enumeration of the Nup copy numbers according to its likelihood (Eq. **1**). All enumerations where the copy number of the Nup205 complex in the inner ring was greater than or equal to that of the Nup188 complex in the inner ring were considered, to reflect the FCS data. Third, we selected the top 4 scoring Nup copy enumerations. Fourth, for each one of these enumerations, we also enumerated the subcomplex location assignments for the Y-complex. Other subcomplexes with multiple copy numbers were randomly assigned an order of subcomplex location assignment before modeling. In total, this procedure resulted in 15, 20, 22, 18, 15, and 1 heterogeneity models for, respectively, 5 min, 6 min, 8 min, 10 min, and 15 min after anaphase onset and the mature pore.

Heterogeneity models were assessed by comparing them to experimental copy number data. The 4 Nups used for FCS scoring (Nup107, Nup93, Nup205, and Nup62) provide an assessment of how well the models agree with input information, while FCS data on 2 Nups (Nup188 and Seh1) were intentionally excluded from computing heterogeneity models and served as an independent validation of the model.

1.2. Stage 2: Modeling Snapshots. Each snapshot model represents a set of integrative structural models with compositions defined by its corresponding heterogeneity model. It is computed by our previously established procedures for integrative modeling of static protein structures (11–13).

In the first step of snapshot modeling, information about the process of interest is gathered. Each snapshot model is computed based on its corresponding heterogeneity model, live-cell correlated electron tomography (ET) maps, (38) a model of the mature NPC structure, (47, 48) and physical principles. The heterogeneity model informed the composition of the NPC at the corresponding time point. The live-cell correlated ET maps taken at successive time points along the assembly pathway informed the overall size and shape of the assembling NPC as well as the size and shape of the nuclear envelope as a function of time. The model of the mature NPC structure informed Nup positions and the rigid representation of the NPC subcomplexes. Physical principles informed excluded volume and Gō-like (49) terms.

In the second step of snapshot modeling, the model representation is chosen, a scoring function is generated, and alternative models are searched for. Briefly, each snapshot model includes as a set of structural models, consisting of previously defined rigid subcomplexes (*SI Appendix*, *Representation of Snapshot Models* and Table S3). The rigid subcomplexes are a subset of those in the mature NPC. The included subcomplexes are identified by the corresponding heterogeneity model. The structural models are coarse grained with 10 residues per bead.

Snapshot models were scored using standard scoring functions for integrative modeling of static structures (*SI Appendix*, *Scoring Snapshot Models* and Table S2):

$$P(X_{N_t,t}|D_t) \propto P(D_t|X_{N_t,t})P(X_{N_t,t}), \qquad [2]$$

where $P(D_t|X_{N_t,t})$ is the data likelihood for a structural model, and $P(X_{N_t,t})$ is the prior probability for the structure and its composition. These likelihoods and priors reflected the ET maps (*SI Appendix*, Fig. S1), FCS data, Gō-like terms, (49) and excluded volume.

Snapshot models were sampled using Markov Chain Monte Carlo simulations with temperature replica exchange (*SI Appendix*, *Sampling Snapshot Models*).

Each snapshot model was assessed by statistical checks of sampling exhaustiveness and by comparing the model to input information used in modeling (*SI Appendix*, *Assessment of Snapshot Models*).

1.3. Stage 3: Modeling a Trajectory. In the final stage, the snapshot models are selected and connected into a trajectory model.

In the first step of trajectory modeling, information about the process of interest is gathered. Trajectory modeling uses snapshot models and physical principles of biomolecular dynamics. Snapshot models inform transitions between sampled time points. Their scores and physical principles of biomolecular dynamics are to score the trajectories.

In the second step of trajectory modeling, the model representation is chosen, a scoring function is generated, and alternative models are searched for. In the trajectory model, a trajectory is represented by an ordered sequence of snapshot models, with one snapshot model at each time point. In other words, the trajectory is $\mathcal{X} = \{(X_{0,N_0}), \ldots, (X_{15,N_{15}}), (X_{\mathrm{MP},N_{\mathrm{MP}}})\}$, where MP is the mature pore.

Next, we created a scoring function for trajectory models. Scoring only individual snapshot models along a trajectory generally cannot utilize all information about the process; in particular, it cannot reflect any information about interconversions between different snapshot models. Thus, a complete trajectory score for a Markovian process (50) is a product of scores for each snapshot model and each transition in a trajectory (Fig. 1*B*):

$$W(\mathcal{X}) \propto \prod_{t=0}^{T} \mathcal{P}(X_{N_t,t}|D_t) \cdot \prod_{t=0}^{T-1} W(X_{N_{t+1},t+1}|X_{N_t,t}D_{t,t+1}), \quad [3]$$

where t indexes times from 0 until the final time point (T); $\mathcal{P}(X_{N_t,t}|D_t)$ is the snapshot model score (*SI Appendix*, Eq. **S2**); and $W(X_{N_{t+1},t+1}|X_{N_t,t}D_{t,t+1})$ is the transition score. The weights of each trajectory, ($W(\mathcal{X})$), are normalized so that they sum to 1. In general, transition scores can be derived from multiple sources of information, including experimental information or physical models of the macromolecular dynamics, such as molecular dynamics simulations, (51–53) Markov state models, (54–56) or Bayesian metamodeling (57).

Here, we scored the transitions between snapshot models with a simple metric that either allowed or disallowed a transition. A transition was allowed if the subcomplexes present in the first snapshot model were included and had the same subcomplex location assignment in the second snapshot model (*SI Appendix*, Fig. S2). This metric was based on the assumptions that Nups are unlikely to dissociate from the assembling complex and that large subcomplex rearrangements are unlikely.

We then searched for good scoring trajectories. Specifically, we enumerated all connections between snapshot models at adjacent time points, followed by scoring these trajectories according to Eq. **3**. The resulting set of trajectories was represented as a directed acyclic graph. In this representation, each node in the graph is a snapshot model, and each edge represents the transition weight between snapshot models. The complexity of the graph varies with the number of time points (6 for the postmitotic NPC assembly model), the number of snapshot models (91 for the postmitotic NPC assembly model), and the number of edges (425 for the postmitotic NPC assembly model, excluding edges with a weight of 0). These numbers determine the total number of enumerated trajectories (5,184 for the postmitotic NPC assembly model). When enumeration is not feasible, sampling methods such as Markov Chain Monte Carlo could be employed.

In the third step of trajectory modeling, the model is assessed. In general, the uncertainty of the output model results from the actual heterogeneity in the samples used to generate the input information, uncertainty in the input information, and imperfections of the modeling method, which in turn reflect uncertain representation, scoring, and sampling. As for static integrative structure models, we assess the trajectory model in four ways, as follows (13).

First, we analyzed the temporal precision (uncertainty) of the trajectory model. Previously, sampling precision has been defined for static integrative models as the variation among good-scoring structural models between independent realizations of the modeling procedure (58). Here, the temporal precision is the temporal analogue of the sampling precision, and represents the variation among good-scoring trajectories between independent realizations of modeling:

$$P_{\mathrm{temp}} = 1 - \frac{1}{2}\sum_i |W_A(\mathcal{X}^i) - W_B(\mathcal{X}^i)|, \quad [4]$$

where P_{temp} is the Manhattan distance between two trajectory models, and i denotes each possible trajectory (χ^i). Comparisons are made between two independently sampled sets of the same snapshot models (A and B), each of which has a corresponding weight for a given trajectory ($W_A(\mathcal{X}^i)$ and $W_B(\mathcal{X}^i)$, Eq. **3**). P_{temp} reflects the degree of overlap between different trajectory models, taking a value of 1.0 when agreement is perfect and 0.0 when there is no overlap. The metric quantifies the precision (uncertainty) of trajectory weights; thus, the weights should only be interpreted up to P_{temp}.

The only origin of the temporal imprecision of our trajectory model was the sampling uncertainty of the snapshot models because the potential assembly trajectories were enumerated. Accordingly, we used P_{temp} to assess how uncertainty originating from the stochasticity of the sampling of snapshot models propagated into the temporal domain.

Second, after assessing the temporal precision of the trajectory model, we evaluated the trajectory model by comparing it to experimental data used in model construction (*SI Appendix*, Table S2). In addition to examining the agreement between the snapshot models in the highest weighted trajectory and static data, such as the structure of the assembled NPC and ET maps (*SI Appendix*, *Assessment of Snapshot Models*), we also examined how the trajectory model matches the FCS data used to score the model. For each scored Nup (Nup107, Nup93, Nup205, and Nup62), we used the trajectory model to compute the weighted sum of the Nup copy number in each snapshot model at each time point. We compared these computed copy numbers to the experimental copy numbers from FCS data (37).

Third, we evaluated the trajectory model by comparing it to experimental data not used in model construction. Specifically, we intentionally excluded Nup copy number data for Nup188 and Seh1 from model construction. Using the same procedure as described in the previous paragraph, we assessed the trajectory model by comparing the computed copy numbers to the experimental measurements.

Finally, we evaluated the precision (uncertainty) of the trajectory model. The precision of the trajectory model is defined by the variation among good-scoring trajectories:

$$\gamma = \sum_i [W(\mathcal{X}^i)]^2, \quad [5]$$

where i denotes each possible trajectory (χ^i), each of which has a weight of $W(\mathcal{X}^i)$ (Eq. **3**). γ ranges from 1.0 to $\frac{1}{d}$, where d is the number of trajectories. Values approaching 1.0 indicate that the trajectory model can be described by a single trajectory, while values approaching $\frac{1}{d}$ indicate that all trajectories in the trajectory model are weighted approximately equally.

2. Results

2.1. Molecular Model of the Postmitotic NPC Assembly Process.

The modeling began by calculating heterogeneity models for the

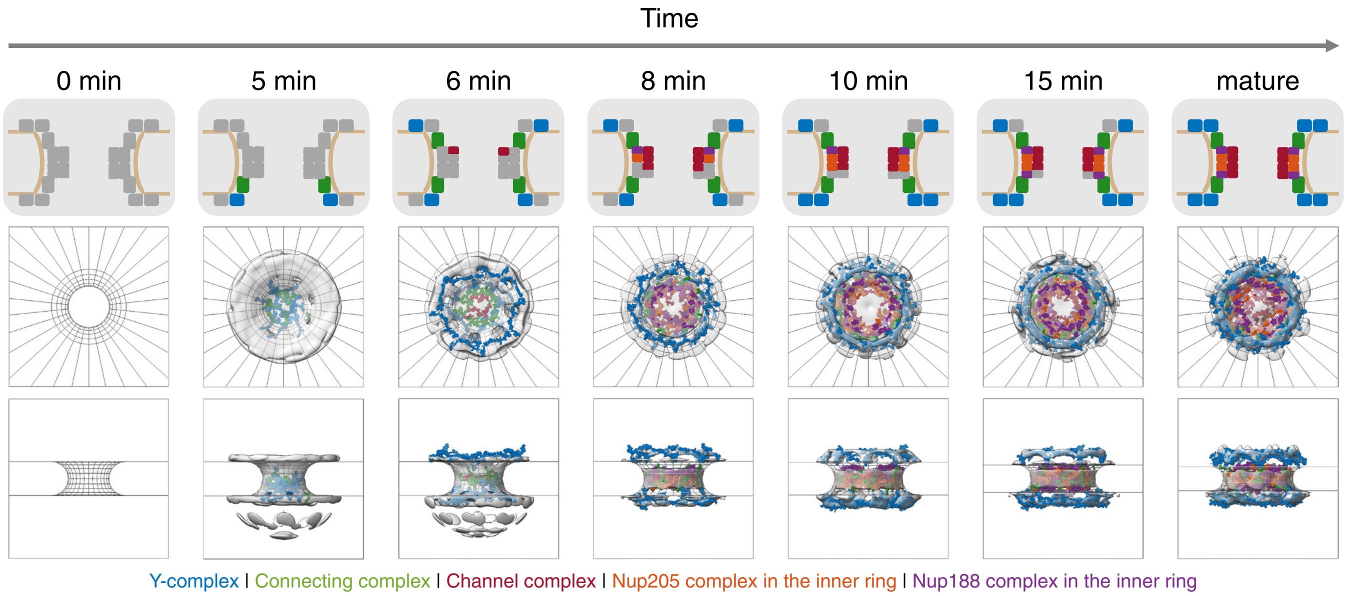

Fig. 2. The pathway model of the postmitotic NPC assembly. The highest weighted trajectory, which accounts for 99.9% of model weight, is shown. For each time point along the assembly trajectory, a a heterogeneity model (*Top* panel) is shown, along with a structural model from the cytoplasm (*Middle* panel) and from within the nuclear envelope (*Bottom* panel). Structural models correspond to the centroid structure of the most populated cluster. ET maps are shown in gray and are compared to structure models corresponding to the Y-complex (blue), connecting complex (green), channel complex (red), Nup205 complex in the inner ring (orange), and Nup188 complex in the inner ring (purple). No experimental data are available at 0 min, corresponding to the completely disassembled NPC in the context of the unsealed nuclear envelope.

most likely Nup combinations at each sampled time point along the assembly process and then computing a snapshot model for each heterogeneity model. A single trajectory consists of a series of snapshot models, with one snapshot model at each time point. Each trajectory can then be scored by multiplying the scores of snapshot models along the trajectory and the scores of transitions between adjacent snapshot models. Thus, the workflow produced a set of weighted trajectories, with the weights summing to one; the larger the weight of a trajectory, the more consistent the trajectory is with the input information.

The trajectory model contained a variety of possible trajectories (*SI Appendix*, Fig. S3); however, the high model precision (γ, Eq. **5**) of 0.999 indicated that the set of alternative trajectories can be represented by the single best-scoring trajectory (Fig. 2). This best-scoring trajectory (the pathway model) is the focus of our analysis as follows.

In the pathway model, eight Y-complexes initiate assembly on the nuclear side of the nuclear envelope (5 min). Next, the FG Nup-containing channel complex (Nup62–Nup58–Nup54) localizes toward the center of the pore, as the first cytoplasmic Y-complex is added (6 min). Two more copies of the channel complex (Nup62–Nup58–Nup54) then join the assembling pore, along with the first components of the inner ring (Nup205–Nup93–Nup155 and Nup93–Nup188–Nup155, 8 min). As assembly proceeds, the inner ring continues to grow, the pore continues to dilate, and the second Y-complex is added to the nuclear side (10 min and 15 min). Eventually, the final copies of the channel complex and cytoplasmic Y-complex complete the pore (mature).

2.2. Assessment of the Postmitotic NPC Assembly Model. A model needs to be assessed before it is interpreted. While the accuracy of a model (the difference between the model and the truth) is often unknown, we can still assess it in at least four ways: estimating the sampling precision, comparing the model to data used to construct it, validating the model against data not used to construct it, and quantifying the precision of the model. These four assessments titrate our confidence in the model; for example, the model precision may be used as a proxy for model accuracy, assuming no systematic errors. Here, each of these assessments is exemplified for the trajectory model of postmitotic NPC assembly.

To estimate the sampling precision of the trajectory model, we quantified the sampling precision of the only stochastic sampling step, namely that of computing static snapshot models, and how that sampling precision impacted the uncertainty of trajectory model weights (58). More specifically, we confirmed that two independently obtained samples of each snapshot model had scores drawn from the same parent distribution (Fig. 3*A* and *SI Appendix*, Fig. S4) and that each structural cluster from two independently obtained samples included models from each sample proportionally to its size (Fig. 3*B* and *SI Appendix*, Figs. S5 and S6). We also found P_{temp} (Eq. **4**) to be 0.9995, indicating that two independent computations of the trajectory model give nearly identical weights to each trajectory.

To compare the trajectory model to data used to construct it, we measured the model's agreement to the mature structure model, the ET maps, and the FCS data. We began by comparing snapshot models at each time point along the pathway model to the mature structure (Fig. 3*B*). The rms deviation (RMSD) to the mature structure model followed the same trend as the sampling precision; snapshot models at early time points (5 min and 6 min) had a higher uncertainty and deviated slightly from the mature structure model, while snapshot models at later time points (8 min onward) were at comparatively higher precision and more closely resembled the mature structure model. Next, we calculated the correlation coefficient between the snapshot models in the pathway model to the ET data used in model construction (Fig. 3*C*). There is good agreement between the snapshot models and the ET maps, with a correlation coefficient of at least 0.60 at each time point. Then, we compared the Nup copy numbers along our trajectory model to experimental

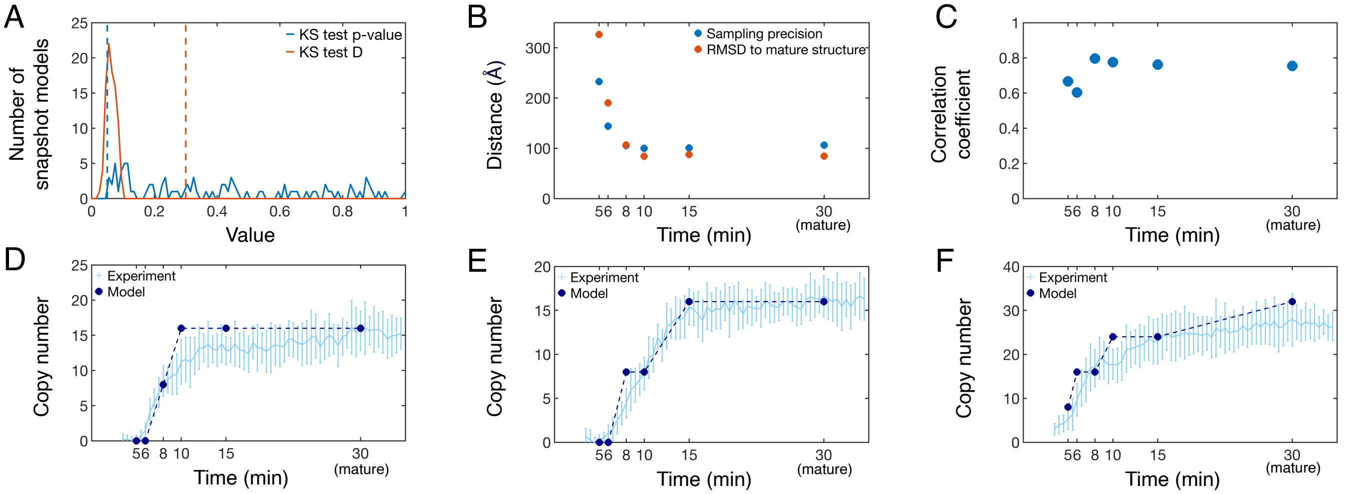

Fig. 3. Validation of the trajectory model of NPC assembly. (*A*) Determining sampling exhaustiveness of snapshot models. For each snapshot model, two samplings are performed and compared via a Kolmogorov–Smirnov two-sample test (59). Sampling is performed until the difference in distribution of scores is not significant (P-value > 0.05) and small in magnitude (Kolmogorov–Smirnov statistic, D, < 0.3). Dashed lines correspond to the cutoffs for Kolmogorov–Smirnov test P-values and Kolmogorov–Smirnov test D. (*B*) Distance-based measures for snapshot models in the pathway model. Both the sampling precision and the average RMSD to the fully mature NPC are shown. (*C*) Correlation coefficient between the snapshot model and the ET map for each snapshot model in the pathway model. (*D*) Comparison of Nup copy number between FCS data [light blue (37)] and the trajectory model (dark blue) for Nup205, which is explicitly included in the scoring function. As the FCS curve plateaus by 30 min, we compare the experimental copy number at 30 min to the mature copy number predicted by the model. Error bars represent SDs over multiple experimental measurements or over weighted trajectories and are smaller than the symbols when not visible. The model SD is undefined for the mature state, which we hold fixed. Comparison to other proteins included in the scoring function are found in *SI Appendix*, Fig. S7. Comparisons of Nup copy number between FCS data [light blue (37)] and the trajectory model (dark blue) for proteins not explicitly included in the scoring function, Nup188 (*E*) and Seh1 (*F*).

copy numbers according to FCS data (Fig. 3*D* and *SI Appendix*, Fig. S7). While reasonable agreement exists for all Nups, the weakest agreement was observed for Nup93 and Nup62, particularly at later time points. These two Nups were constrained to maximum copy numbers observed in the mature structure, (47, 48) which did not fully account for all copies observed *via* FCS. This discrepancy demonstrates how our approach is limited by the accuracy of the input information, in this case the mature structure model.

To validate the trajectory model against the data not used to construct it, we confirmed that our trajectory model is consistent with the time dependence of the Nup188 and Seh1 copy numbers determined by FCS (Fig. 3 *E* and *F*). This test illustrates the predictive power of the trajectory model, although the copy numbers of other members of the same subcomplexes were included in model construction.

To quantify the precision of the trajectory model, we computed its precision, γ, (0.999) and the weight for the most likely trajectory given the data, $W(\chi^1)$, (99.9%). The high sampling precision of the trajectory model ensures that these estimates were not affected by stochasticity of sampling. The high precision indicates that the set of alternative trajectories can be adequately represented by the single, best-scoring trajectory (the pathway model).

2.3. Validation of the Postmitotic NPC Assembly Model with Subsequent Immunofluorescent Images. Perhaps the most convincing test of a model is validating it against data determined after the model was computed. Thus, we validate our model by experimentally confirming its nontrivial prediction that the assembly process of the NPC begins on the nuclear, not cytoplasmic, side of the pore (the "nuclear-side first" mechanism Fig. 2); this prediction emerges from modeling and is not directly evident from either the ET maps or FCS data alone. To validate it, we costained Nup153 and Nup214 in dividing cells (37) and reevaluated their assembly order by immunofluorescence microscopy (*SI Appendix, Immunofluorescent Imaging for Visualizing the Progression of Nuclear Pore Assembly*). Images of these cells identified the order of Nup153 and Nup214 addition to the growing pore: Nup153 joins the assembling NPC before Nup214 (Fig. 4). This experimental observation confirms the nuclear-side first mechanism proposed by the pathway model, given the previous determination that Nup153 and Nup214 asymmetrically bind to the Y-complex of the NPC on the nuclear and cytoplasmic side, respectively (31, 32).

2.4. Temporal Scoring Improves Model Precision. In principle, integrative modeling can improve the accuracy, precision, and completeness of a model by incorporating more information about the modeled process. To illustrate this point concretely,we

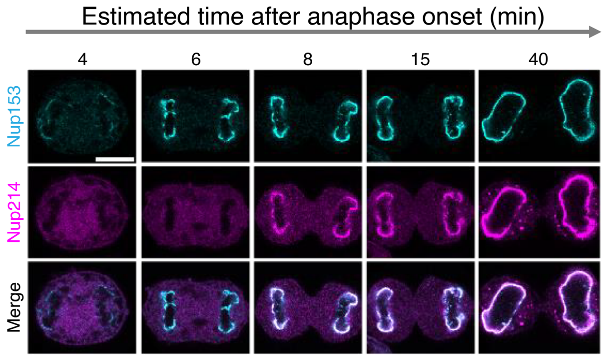

Fig. 4. Validation of the nuclear-side first mechanism predicted by the NPC assembly model. Representative immunofluorescent images of genome-edited Nup214-mEGFP cells labeled with a GFP nanobody and an anti-Nup153 antibody, acquired at various stages after anaphase onset. Division timing was estimated based on the morphology of dividing cells. (Scale bar: 10 μm.)

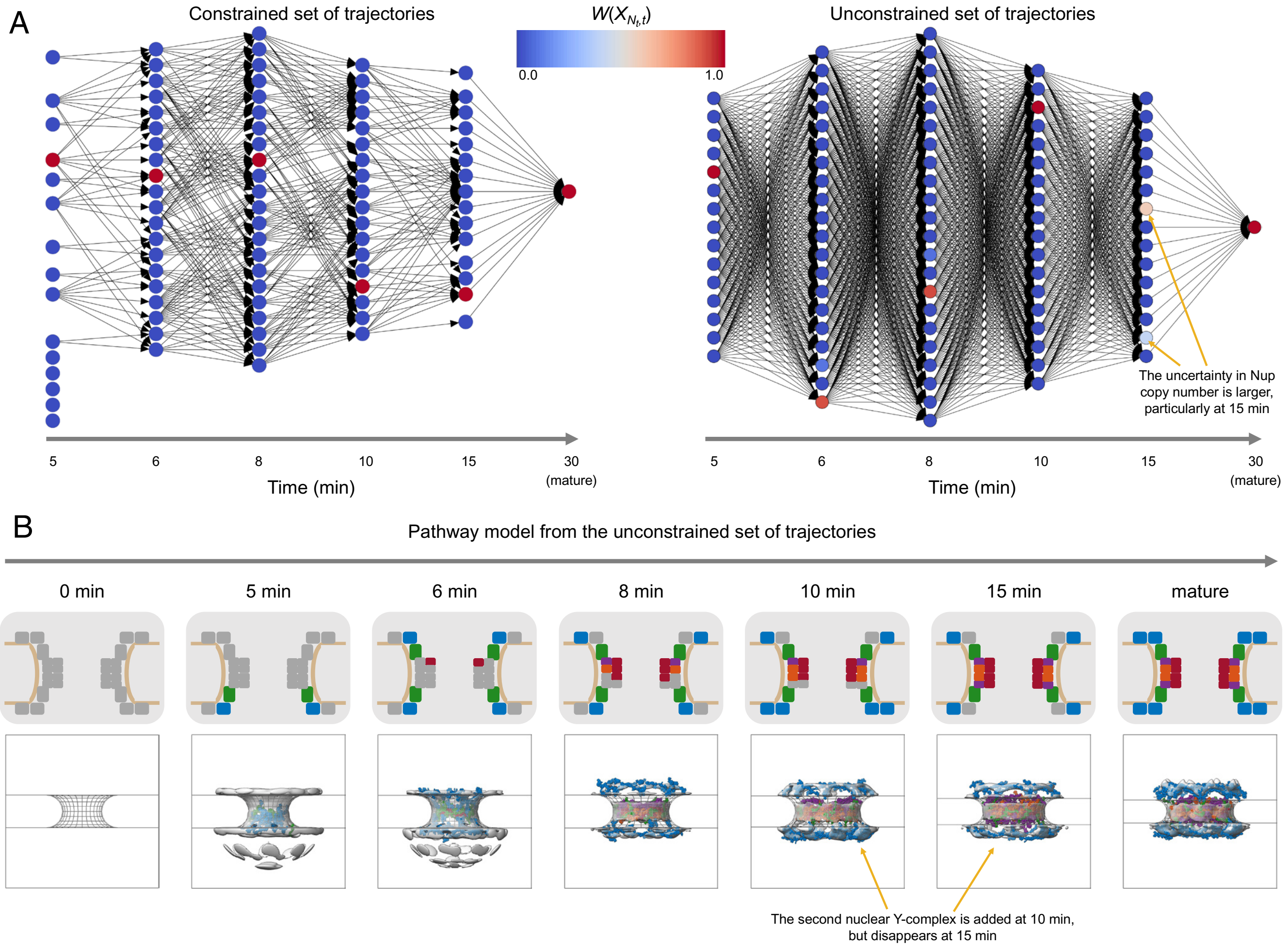

Fig. 5. Temporal scoring terms help improve the precision of the resulting model. (*A*) Directed acyclic graphs of the constrained (*Left*) and unconstrained (*Right*) trajectory models. Each column in the graph corresponds to a different time point in the assembly process (5 min, 6 min, 8 min, 10 min, 15 min, and mature). Each node is shaded according to its weight in the trajectory model ($W(X_{N_t,t})$). (*B*) The most likely trajectory from the unconstrained trajectory model. For each time point along the trajectory, a heterogeneity model (*Top* panel) is shown, along with a structural model from within the nuclear envelope (*Bottom* panel). Experimental electron density maps are shown in gray and are compared to protein models corresponding to the Y-complex (blue), connecting complex (green), channel complex (red), Nup205 complex in the inner ring (orange), and Nup188 complex in the inner ring (purple).

quantified the improvement in the precision of the trajectory model afforded by temporal, in addition to spatial, scoring. The set of possible trajectories without temporal scoring is larger and more varied because it contains transitions between all pairs of neighboring snapshot models, instead of only those that are allowed in the original set of trajectories (*SI Appendix*, Fig. S2).

The precision of a trajectory model is visualized *via* its directed acyclic graph (Fig. 5*A*). These visualizations clearly indicated the greatly reduced trajectory space for the temporally constrained trajectory model compared to the temporally unconstrained trajectory model, reducing the number of possible trajectories from 1,782,000 to 5,184. Correspondingly, the unconstrained trajectory model was less precise, with a γ of 0.36, instead of 0.999 for the constrained trajectory model. The difference in precision of the two trajectory models was also reflected in the weights of the top trajectory (49% and 99.9% for the unconstrained trajectory model and constrained trajectory model, respectively) and the precision of the Nup copy numbers as a function of time (*SI Appendix*, Fig. S8).

Next, we assessed the highest weighted trajectory from unconstrained trajectory model by comparing it to that from the constrained trajectory model. As expected, the highest weighted trajectory in the unconstrained trajectory model included disallowed transitions in the constrained model (Fig. 5*B*). In particular, the second copy of the nuclear Y-complex joined the assembling NPC at 10 min but was then removed at 15 min. This rearrangement seems unlikely, further demonstrating how the transition scores (Eq. **3**) improve the constrained trajectory model.

3. Discussion

We developed integrative spatiotemporal modeling and demonstrated it by modeling the NPC assembly pathway during mitotic exit. A key advantage of an integrative approach is that it improves the accuracy, precision, and completeness of the model because it aims to maximize the amount of input information. This point is illustrated by comparing previous schematic models of NPC assembly with our formal, molecular model of NPC assembly: Previous studies probed the assembly of the NPC with individual techniques, (37, 38, 40) resulting in a coarse schematic of the composition and shape of the assembling NPC as a function of time (39). In contrast, our integrative approach systematically explored "all" Nup trajectories from separate components to the fully assembled mature pore, producing a series of snapshot models with defined molecular architectures that were consistent with all input information.

Incorporating additional types of input information into model construction remains a key challenge in spatiotemporal modeling; using information for model construction, as opposed to only filtering or validation, is more likely to result in models consistent with it. In general, input information can come from experiments, physical theories, statistical analyses, and other prior models: for example, diffusion coefficients, (60) physics-based force fields, (61, 62) metamodeling, (57) previous integrative (63, 64) or machine-learned (65–67) structures of components, protein degradation experiments, (68) binding rate constants, (69) molecular networks, (70) fluorescence microscopies, (71, 72), and electron microscopy (73, 74).

Accordingly, the precision and our confidence in the model of the NPC assembly process could be increased by incorporating additional information into the current method, as illustrated by the following five examples. First, targeted Nups can be depleted through auxin-inducible degradation, (68) allowing us to either validate or improve the model based on this information. Specifically, stalled intermediate states would simplify high-resolution structural characterization, which could either confirm predictions from the original model or inform new snapshot models.

Second, the completeness of snapshot models could be improved by increasing the resolution of the model representation. For instance, deep-learning models (65–67) could predict atomistic structures of Nups and their binding modes; furthermore, the nuclear envelope could be represented by a physics-based membrane force field, (62) instead of the current implicit representation.

Third, the precision of the trajectories could be improved by reflecting additional information in the scoring of dynamic transitions between snapshot models. For example, such information may be provided by molecular dynamics simulations (51, 52) or Markov-state modeling (54–56).

Fourth, our estimates of model precision could be improved by Bayesian approaches (75). For example, the weights of scoring terms could be replaced by Bayes estimators, resulting in a more accurate estimate of model precision that more objectively reflects the uncertainty of nuisance parameters.

Finally, the completeness of modeling the assembly process of the NPC can be improved by expanding comparative modeling of static structures (76) to dynamic processes. In particular, structural models of the mature NPC from multiple organisms (23–32) could be integrated with our model of the NPC assembly pathway to compute assembly pathways across different organisms (57). Such an analysis may be particularly informative about the evolution of the NPC assembly process and thus the evolution of eukaryotes (39).

While the current method is well suited for the snapshot information available about NPC assembly, other integrative spatiotemporal modeling approaches should also be developed. In particular, incorporating experimental measurements into Brownian dynamics simulations could in principle result in a set of trajectories that are by design consistent with the input information; the input into such simulations is a starting configuration for a set of components, their interactions, and their diffusion constants. The resulting models would provide a near-continuous and more realistic description of a dynamic process, compared to the current representation as a series of discrete snapshot models. This modeling process could begin by computing Brownian dynamics trajectories that incorporate physics-based diffusion constants (60) for each molecule, pairwise binding rate constants (69) between each pair of molecules, and binding modes between each pair of molecules derived from comparative, (76) integrative, (11–13) or deep-learning (65–67) methods. Some temporal experimental information could be incorporated by an iterative search for interaction parameters that produce simulated trajectories whose computed properties match the observed ones, similarly to our modeling of molecular transport through the NPC (77). Other temporal experimental information may be better incorporated by guiding possible trajectories through enhanced sampling (78–80) methods. Finally, time-averaged experimental information could be incorporated through biasing potentials such as maximum entropy (81–85) approaches, while time-dependent experimental information could be incorporated through maximum caliber approaches (86, 87).

Our model of the NPC assembly pathway provides several insights. Remarkably, modeling shows that one specific trajectory explains the data significantly better than many alternative trajectories. This result is indicative of a specific assembly pathway as opposed to a more stochastic process with multiple, similarly weighted high-scoring trajectories. Additionally, as discussed previously, (37) the model suggests a specific order of assembly for the four copies of the Y-complex per each spoke: The first copy appears on the nuclear side, followed by the second copy on the cytoplasmic side, the third copy on the nuclear side, and finally the fourth copy on the cytoplasmic side. We were able to confirm this nuclear-side first mechanism with fluorescence imaging experiments. Further, the model suggests that FG Nups in the central channel complex join the assembling NPC before structured inner ring components. This suggestion, combined with the observation that assembly is initiated by FG Nups, Pom121 and Nup153, leads us to hypothesize that the hydrophobic properties of FG Nups may be important for the dilation of the nuclear pore and the recruitment of other NPC components.

Finally, it appears to be more difficult to generalize integrative spatiotemporal modeling methods than to generalize static integrative structural modeling methods. While we developed our current method to model the NPC assembly process, we endeavored to keep the method as generalizable as possible. Its future impact will depend on how many biomolecular processes can be usefully depicted as a linear trajectory through a series of intermediate states, for which some information is available experimentally. Such processes may include gene transcription, (85, 88) exocytosis, (89, 90) viral assembly mechanisms, (91, 92) and ribosome biogenesis (93, 94). The integrative spatiotemporal modeling approach suggested here was implemented in our open sourced IMP package, (95, 96) in an effort to facilitate its application by scientists to a wide range of complex biological phenomena.

4. Materials and Methods

The spatiotemporal modeling method is outlined in the section entitled "Integrative spatiotemporal modeling workflow," with additional details in *SI Appendix*.

Data, Materials, and Software Availability. The spatiotemporal modeling method is implemented in our freely available, open source program *Integrative Modeling Platform* (IMP; https://integrativemodeling.org) (95, 96). The input and output files for modeling of the NPC assembly process are available at https://doi.org/10.5281/zenodo.11129725. The NPC assembly model is also available in the PDB archive for integrative structures (9A8T). Likewise, the tomograms are also available from the Electron Microscopy Public Image Archive (EMPIAR; EMD-3820), and the FCS data are available at Image Data Resource (IDR; idr0115).

ACKNOWLEDGMENTS. We thank Ben Webb and Ignacia Echeverria for help with implementing our code in Integrative Modeling Platform. This work was supported by grants from the Baden Wuerttemberg Foundation (J.E. and A.S.), NIH/NIGMS R01GM083960 (A.S.), NIH/NIGMS P41GM109824 (A.S.), NIH/NIGMS R01GM112108 (A.S.), the European Molecular Biology Laboratory (EMBL; S.O. and J.E.), and the Deutsche Forschungsgemeinschaft (German Research Foundation)–Project-ID 511488495–SFB 1638, P06 (J.E.). A.P.L. was further supported by NIH/NIGMS F32GM150243. S.O. was further supported by the EMBL Interdisciplinary Postdoc Programme under Marie Curie Actions COFUND. S.O. was additionally supported by a Japan Society for the Promotion of Science fellowship (postdoctoral fellowship for research abroad).

Author affiliations: [a]Department of Bioengineering and Therapeutic Sciences, Quantitative Biosciences Institute, University of California, San Francisco, CA 94143; [b]Department of Pharmaceutical Chemistry, Quantitative Biosciences Institute, University of California, San Francisco, CA 94143; and [c]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg 69117, Germany

Author contributions: A.P.L., W.Z., J.O.B.T., S.O., J.E., and A.S. designed research; A.P.L, W.Z., and J.O.B.T. performed research; A.P.L. and J.O.B.T. contributed new reagents/analytic tools; A.P.L., W.Z., J.O.B.T., and S.O. analyzed data; and A.P.L., W.Z., J.O.B.T., S.O., J.E., and A.S. wrote the paper.

The authors declare no competing interest.

1. E. Lerner *et al*., FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *eLife* **10**, 1–69 (2021).
2. J. Livet *et al*., Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
3. T. R. Alderson, L. E. Kay, NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. *Cell* **184**, 577–595 (2021).
4. R. Kaptein, G. Wagner, Integrative methods in structural biology. *J. Biomol. NMR* **73**, 261–263 (2019).
5. M. Karplus, J. A. McCammon, Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646 (2002).
6. A. J. Pak, G. A. Voth, Advances in coarse-grained modeling of macromolecular complexes. *Curr. Opin. Struct. Biol.* **52**, 119–126 (2018).
7. M. R. Jensen, M. Zweckstetter, J. R. Huang, M. Blackledge, Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **114**, 6632–6660 (2014).
8. J. N. Onuchic, Z. Luthey-Schulten, P. G. Wolynes, Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
9. T. J. Lane, D. Shukla, K. A. Beauchamp, V. S. Pande, To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **23**, 58–65 (2013).
10. B. Schuler, W. A. Eaton, Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
11. F. Alber *et al*., Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).
12. M. P. Rout, A. Sali, Principles for integrative structural biology studies. *Cell* **177**, 1384–1403 (2019).
13. A. Sali, From integrative structural biology to cell biology. *J. Biol. Chem.* **296**, 100743 (2021).
14. D. J. Saltzberg *et al*., Using integrative modeling platform to compute, validate, and archive a model of a protein complex structure. *Protein Sci.* **30**, 250–261 (2021).
15. M. Hancock *et al*., Integration of software tools for integrative modeling of biomolecular systems. *J. Struct. Biol.* **214**, 107841 (2022).
16. V. Rantos, K. Karius, J. Kosinski, Integrative structural modeling of macromolecular complexes using Assembline. *Nat. Protoc.* **17**, 152–176 (2022).
17. P. I. Koukos, A. M. Bonvin, Integrative modelling of biomolecular complexes. *J. Mol. Biol.* **432**, 2861–2881 (2020).
18. M. Braitbard, D. Schneidman-Duhovny, N. Kalisman, Integrative structure modeling: Overview and assessment. *Annu. Rev. Biochem.* **88**, 113–135 (2019).
19. S. Bottaro, K. Lindorff-Larsen, Biophysical experiments and biomolecular simulations: A perfect match? *Science* **361**, 355–360 (2018).
20. A. P. Latham, B. Zhang, Improving coarse-grained protein force fields with small-angle X-ray scattering data. *J. Phys. Chem. B* **123**, 1026–1034 (2019).
21. G. Hummer, J. Köfinger, Bayesian ensemble refinement by replica simulations and reweighting. *J. Chem. Phys.* **143**, 243150 (2015).
22. T. Fröhlking, M. Bernetti, G. Bussi, Refinement of molecular dynamics ensembles using experimental data and flexible forward models. *J. Chem. Phys.* **158**, 214120 (2023).
23. F. Alber *et al*., The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701 (2007).
24. S. J. Kim *et al*., Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018).
25. C. W. Akey *et al*., Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* **185**, 361–378.e25 (2022).
26. C. E. Zimmerli *et al*., Nuclear pores dilate and constrict in cellulo. *Science* **374**, eabd9776 (2021).
27. S. Petrovic *et al*., Architecture of the linker-scaffold in the nuclear pore. *Science* **376**, eabm9798 (2022).
28. P. Fontana *et al*., Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science* **376**, eabm9326 (2022).
29. X. Zhu *et al*., Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex. *Science* **376**, eabl8280 (2022).
30. C. J. Bley *et al*., Architecture of the cytoplasmic face of the nuclear pore. *Science* **376**, eabm9129 (2022).
31. S. Mosalaganti *et al*., AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **376**, eabm9506 (2022).
32. D. Singh *et al*., The molecular architecture of the nuclear basket. *Cell* **187**, 5267–5281 (2024).
33. M. Beck, E. Hurt, The nuclear pore complex: Understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* **18**, 73–89 (2017).
34. A. Hoelz, E. W. Debler, G. Blobel, The structure of the nuclear pore complex. *Annu. Rev. Biochem.* **80**, 613–643 (2011).
35. D. H. Lin, A. Hoelz, The structure of the nuclear pore complex (An Update). *Annu. Rev. Biochem.* **88**, 725–783 (2019).
36. D. Russel *et al*., The structural dynamics of macromolecular processes. *Curr. Opin. Cell Biol.* **21**, 97–108 (2009).
37. S. Otsuka *et al*., A quantitative map of nuclear pore assembly reveals two distinct mechanisms. *Nature* **613**, 575–581 (2023).
38. S. Otsuka *et al*., Postmitotic nuclear pore assembly proceeds by radial dilation of small membrane openings. *Nat. Struct. Mol. Biol.* **25**, 21–28 (2018).
39. S. Otsuka, J. Ellenberg, Mechanisms of nuclear pore complex assembly - two different ways of building one molecular machine. *FEBS Lett.* **592**, 475–488 (2018).
40. S. Otsuka *et al*., Nuclear pore assembly proceeds by an inside-out extrusion of the nuclear envelope. *eLife* **5**, 1–23 (2016).
41. W. Antonin, J. Ellenberg, E. Dultz, Nuclear pore complex assembly through the cell cycle: Regulation and membrane organization. *FEBS Lett.* **582**, 2004–2016 (2008).
42. E. Dultz, J. Ellenberg, Live imaging of single nuclear pores reveals unique assembly kinetics and mechanism in interphase. *J. Cell Biol.* **191**, 15–22 (2010).
43. E. Onischenko *et al*., Maturation kinetics of a multiprotein complex revealed by metabolic labeling. *Cell* **183**, 1785–1800.e26 (2020).
44. U. Kutay, R. Jühlen, W. Antonin, Mitotic disassembly and reassembly of nuclear pore complexes. *Trends Cell Biol.* **31**, 1019–1033 (2021).
45. A. Gelman *et al*., *Bayesian Data Analysis* (CRC Press, ed. 3, 2013), pp. 1–667).
46. R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (CRC Press, 2018), pp. 1–469.
47. A. von Appen *et al*., In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015).
48. J. Kosinski *et al*., Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016).
49. N. Go, H. Abe, Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* **20**, 991–1011 (1981).
50. M. A. Pinsky, S. Karlin, *An Introduction to Stochastic Modeling* (Elsevier Inc, Burlington, MA, ed. 4, 2011).
51. K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
52. B. L. Timney *et al*., Simple rules for passive diffusion through the nuclear pore complex. *J. Cell Biol.* **215**, 57–76 (2016).
53. F. Noé, A. Tkatchenko, K. R. Müller, C. Clementi, Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **71**, 361–390 (2020).
54. B. E. Husic, V. S. Pande, Markov state models: From an art to a science. *J. Am. Chem. Soc.* **140**, 2386–2396 (2018).
55. H. Wan, Y. Ge, A. Razavi, V. A. Voelz, Reconciling simulated ensembles of apomyoglobin with experimental hydrogen/deuterium exchange data using Bayesian inference and multiensemble markov state models. *J. Chem. Theory Comput.* **16**, 1333–1348 (2020).
56. T. Hempel *et al*., Independent Markov decomposition: Toward modeling kinetics of biomolecular complexes. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2105230118 (2021).
57. B. Raveh *et al*., Bayesian metamodeling of complex biological systems across varying representations. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104559118 (2021).
58. S. Viswanath, I. E. Chemmama, P. Cimermancic, A. Sali, Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures. *Biophys. J.* **113**, 2344–2353 (2017).
59. S. Siegel, *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York, NY, US, 1956).
60. M. T. Tyn, T. W. Gusek, Prediction of diffusion coefficients of proteins. *Biotechnol. Bioeng.* **35**, 327–338 (1990).
61. A. P. Latham, B. Zhang, Unifying coarse-grained force fields for folded and disordered proteins. *Curr. Opin. Struct. Biol.* **72**, 63–70 (2022).
62. M. Sadeghi, T. R. Weikl, F. Noé, Particle-based membrane model for mesoscopic simulation of cellular dynamics. *J. Chem. Phys.* **148**, 044901 (2018).
63. H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide protein data bank. *Nat. Struct. Biol.* **10**, 980 (2003).
64. B. Vallat *et al*., New system for archiving integrative structures. *Acta Crystallogr. Sec. D Struct Biol.* **77**, 1486–1496 (2021).
65. J. Jumper *et al*., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
66. M. Mirdita *et al*., Colabfold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
67. G. Ahdritz *et al*., Openfold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat. Methods* **21**, 1514–1524 (2024).
68. K. Nishimura, T. Fukagawa, H. Takisawa, T. Kakimoto, M. Kanemaki, An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).

69. M. J. Morelli, P. R. T. Wolde, Reaction brownian dynamics and the effect of spatial fluctuations on the gain of a push-pull network. *J. Chem. Phys.* **129**, 054112 (2008).
70. A. L. Richards, M. Eckhardt, N. J. Krogan, Mass spectrometry-based protein-protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* **17**, e8792 (2021).
71. J. K. Hériché, S. Alexander, J. Ellenberg, Integrating imaging and omics: Computational methods and challenges. *Annu. Rev. Biomed. Data Sci.* **2**, 175–197 (2019).
72. P. J. Thul *et al*., A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
73. L. N. Young, E. Villa, Bringing structure to cell biology with cryo-electron tomography. *Annu. Rev. Biophys.* **9**, 573–595 (2023).
74. C. S. Xu *et al*., An open-access volume electron microscopy atlas of whole cells and tissues. *Nature* **599**, 147–151 (2021).
75. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer, New York, NY, ed. 2, 1985).
76. A. Sali, T. L. Blundell, Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
77. B. Raveh *et al*., Integrative spatiotemporal map of nucleocytoplasmic transport. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2023.12.31.573409 (Accessed 2 January 2024).
78. D. Ray, M. Parrinello, Kinetics from metadynamics: Principles, applications, and outlook. *J. Chem. Theory Comput.* **19**, 5649–5670 (2023).
79. Y. Wang, J. M. L. Ribeiro, P. Tiwary, Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **61**, 139–145 (2020).
80. P. Gkeka *et al*., Machine learning force fields and coarse-grained variables in molecular dynamics: Application to materials and biological systems. *J. Chem. Theory Comput.* **16**, 4757–4775 (2020).
81. J. W. Pitera, J. D. Chodera, On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **8**, 3445–3451 (2012).
82. B. Roux, J. Weare, On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **138**, 084107 (2013).
83. W. Boomsma, J. Ferkinghoff-Borg, K. Lindorff-Larsen, Combining experiments and simulations using the maximum entropy principle. *PLoS Comput. Biol.* **10**, 1–9 (2014).
84. R. Rangan *et al*., Determination of structural ensembles of proteins: Restraining vs. reweighting. *J. Chem. Theory Comput.* **14**, 6632–6641 (2018).
85. X. Lin, Y. Qi, A. P. Latham, B. Zhang, Multiscale modeling of genome organization with maximum entropy optimization. *J. Chem. Phys.* **155**, 010901 (2021).
86. P. D. Dixit *et al*., Perspective: Maximum caliber is a general variational principle for dynamical systems. *J. Chem. Phys.* **148**, 010901 (2018).
87. K. Ghosh, P. D. Dixit, L. Agozzino, K. A. Dill, The maximum caliber variational principle for nonequilibria. *Annu. Rev. Phys. Chem.* **71**, 213–238 (2020).
88. P. Cramer, Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
89. Y. Zhang, F. M. Hughson, Chaperoning snare folding and assembly. *Annu. Rev. Biochem.* **90**, 581–603 (2021).
90. R. Jahn, T. Sudhof, Membrane fusion and exocytosis. *Annu. Rev. Biochem.* **68**, 777–810 (1999).
91. J. D. Perlmutter, M. F. Hagan, Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66**, 217–239 (2015).
92. H. Yao *et al*., Molecular architecture of the SARS-CoV-2 virus. *Cell* **183**, 730–738.e13 (2020).
93. E. Thomson, S. Ferreira-Cerca, E. Hurt, Eukaryotic ribosome biogenesis at a glance. *J. Cell Sci.* **126**, 4815–4821 (2013).
94. L. Jiao *et al*., Ribosome biogenesis in disease: New players and therapeutic targets. *Signal Transduct. Target. Ther.* **8**, 15 (2023).
95. Z. Yang *et al*., UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *J. Struct. Biol.* **179**, 269–278 (2012).
96. D. Russel *et al*., Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, 1–5 (2012).