Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

Biophysical *Journal*
Article

Biophysical Society

NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale

Matthew J. Varga,[1] Yiben Fu,[1] Spencer Loggia,[1] Osman N. Yogurtcu,[2] and Margaret E. Johnson[1,*]
[1]TC Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland and [2]Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland

ABSTRACT Currently, a significant barrier to building predictive models of cellular self-assembly processes is that molecular models cannot capture minutes-long dynamics that couple distinct components with active processes, whereas reaction-diffusion models cannot capture structures of molecular assembly. Here, we introduce the nonequilibrium reaction-diffusion self-assembly simulator (NERDSS), which addresses this spatiotemporal resolution gap. NERDSS integrates efficient reaction-diffusion algorithms into generalized software that operates on user-defined molecules through diffusion, binding and orientation, unbinding, chemical transformations, and spatial localization. By connecting the fast processes of binding with the slow timescales of large-scale assembly, NERDSS integrates molecular resolution with reversible formation of ordered, multisubunit complexes. NERDSS encodes models using rule-based formatting languages to facilitate model portability, usability, and reproducibility. Applying NERDSS to steps in clathrin-mediated endocytosis, we design multicomponent systems that can form lattices in solution or on the membrane, and we predict how stochastic but localized dephosphorylation of membrane lipids can drive lattice disassembly. The NERDSS simulations reveal the spatial constraints on lattice growth and the role of membrane localization and cooperativity in nucleating assembly. By modeling viral lattice assembly and recapitulating oscillations in protein expression levels for a circadian clock model, we illustrate the adaptability of NERDSS. NERDSS simulates user-defined assembly models that were previously inaccessible to existing software tools, with broad applications to predicting self-assembly in vivo and designing high-yield assemblies in vitro.

SIGNIFICANCE A major roadblock in quantitative modeling of cell biology is the need to resolve fast processes over long timescales and individual proteins over large length scales. This is particularly evident in self-assembly processes such as clathrin-mediated endocytosis and viral assembly, which proceed over minutes, building from often weak and short-lived binding events. The NERDSS software presented here uses the reaction-diffusion model to simulate such self-assembly processes, with broad applications to biological processes as they occur in the nonequilibrium cell.

INTRODUCTION

Watching the dynamics of individual molecules in the cell as they function as part of a collective is now possible thanks to revolutions in live-cell microscopy. Despite technical advances in mapping dynamics onto nanoscale structure formation in the cell (1), resolution is still limited by events that are too fast and combine too many distinct components. Computer simulations offer the promise of reproducing dynamics using models designed from the underlying physics and mechanics, providing high-resolution spatial and temporal predictions and exquisite control over components and their interactions. Current tools for capturing cell-scale complexity are a growing companion to cell biology. For example, fields such as cell-signaling (2), developmental biology (3,4), and systems biology (5) have benefitted from a variety of spatial (6,7) and nonspatial tools (8,9) in which interactions and reactions are modeled as events parameterized by rate constants. However, many cell-scale processes involve self-assembly, which is a challenge for modeling because it spans similarly long length and timescales as biochemical signals while also depending fundamentally on molecular structural geometry. Although rate-based approaches have been shown to be highly insightful for modeling self-assembly, previous models lack explicit spatial resolution (8,10–12), apply only to small systems (13,14), or include potentials that prevent quantitative

Submitted December 27, 2019, and accepted for publication May 5, 2020.
*Correspondence: margaret.johnson@jhu.edu
Matthew J. Varga and Yiben Fu contributed equally to this work.
Editor: Padmini Rangamani.
https://doi.org/10.1016/j.bpj.2020.05.002

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

comparison to experiment (15). Here, we present the nonequilibrium reaction-diffusion self-assembly simulator (NERDSS), a higher-resolution rate-based software tool with the addition of user-specified, coarse molecular structure to enable the study of self-assembly at the cell scale.

NERDSS was developed to build self-assembly into the reaction-diffusion (RD) model because of the strength of RD in simulating relatively slow, nonequilibrium dynamics. The name of the software may sound somewhat redundant (particularly NERD). The "nonequilibrium" is included in the name to emphasize how self-assembly can be simulated here in generalized, nonequilibrium systems such as the cell. Additionally, although the RD model is commonly applied to studying systems not at thermodynamic equilibrium (16), it is nonetheless useful for quantifying the pathways to and composition of equilibrium for complex multicomponent systems. The standard computational approach for studying self-assembly, using coarse-grained molecular modeling (17–19), provides access to these pathways and equilibria with more physically detailed models, in which interactions emerge because of distance-dependent energy functions rather than the rate-controlled events of RD models. A coarse molecular model thus naturally can accommodate a range of structures or defects (17,19,20) and capture emergent cooperativity, strain, and localization of components (21,22). However, they are still relatively limited in the length and timescales of dynamics they can access, which does not extend to the cell scale. They are parameterized by potential energy functions, and converting these to the binding free energies (e.g., as done here (21,23)) or rate constants (24–26) for comparison to experiment is nontrivial. Lastly, chemical reactions require covalent bond breaking that is not typically accessible in molecular modeling, preventing systematic and transferable methods for involving enzymatic or ATP-driven reactions ubiquitous in cells. These models are unable to simulate in vivo cell signaling, cytoskeletal dynamics, or clathrin-mediated endocytosis, for example. Our NERDSS software addresses this substantial application gap, and although it lacks the detail of energy-function based models, it is able to uniquely preserve important features of molecular assembly.

NERDSS overcomes technical challenges with adding structure to RD and allowing large, reversible complexes to form in a user-friendly and widely applicable way. The major challenges for applying RD to self-assembly are expanding to multisite, volume-excluding species and capturing orientation-dependent interactions for arbitrary species. For continuum partial differential equation (PDE) approaches (6,7) to RD and lattice-based reaction-diffusion master equation (RDME) approaches (27–29), species have no individual coordinates but are well mixed in the infinitesimal or finite grid spacings, respectively, preventing the resolution of structural assemblies. Capturing assembly kinetics is possible in these methods for complexes with uniquely defined subunits, with recent RDME-based studies of ribosome formation (30) and essential steps in spliceosome assembly (31). Although single-particle RD is considerably more computationally expensive than PDE or RDME, it does track individual coordinates (32–39), often with excluded volume, providing a starting point for multisite species and structure resolution. Recent methods have generated multisite species by linking them using spherically symmetric potentials (14,40), producing simulations that can capture effects of clustering (41) and polymer assembly (40). However, without orientation-dependent interactions, well-defined structural geometries such as lattices and spherical shells will not form. Brownian dynamics methods have been successfully used to capture orientation-dependent, rate-controlled association but are limited to small systems (42,43). Here, we use our recent algorithm for structure-resolved RD of rigid bodies, which includes and accounts for effects of rotational as well as translational diffusion on binding reactions (44). Instead of using potentials, orientational constraints are applied after binding events to maintain accurate solutions to the RD equations of motion and direct comparison to experimental rates. With previous application of this algorithm to clathrin lattice assembly in solution (44), NERDSS generalizes these routines for orienting molecules to apply here to arbitrary user-defined systems, including on the membrane, which can act as a potent driver of self-assembly via dimensionality reduction (45). This approach retains the flexibility and the adaptability of the method to new molecules and structures because binding is parameterized by rates and not customized interaction potentials.

A critical feature of generalized rate-based software for usability, model portability, and construction of conditional and cooperative interactions is the adoption of well-defined formatting languages for model encoding. Rule-based modeling (46), developed for nonspatial models, supports descriptions of species as multisite molecules in which each site can exist in distinct states. Interactions or binding rates can be coupled within species to capture effects of phosphorylation, for example, on subsequent binding events. Formatting languages such as BioNetGen Language (46,47) and Kappa (48,49) implement rule-based modeling, which precludes the need to define and track all possible species and avoids the issue of combinatorial complexity (50). The rule-based format is now available in efficient nonspatial methods such as RuleBender (51) and NFSim (8) and spatial methods such as Virtual Cell (52) and Smoldyn (34), facilitating model comparison and portability. Rule-based encoding translates quite naturally to NERDSS because of the spatial specification of each site. We validate our rule-based encoding here by simulating a model of a circadian clock (53) (and other simple models), compared with the same model simulated in Virtual Cell (6).

To illustrate the capabilities of NERDSS as a spatial self-assembly simulator, we apply it here to steps in clathrin-mediated endocytosis (CME) and virion assembly. These examples highlight several challenges and opportunities for modeling

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

approaches, and both assembly processes have been studied using coarse molecular modeling simulations (19,54–57). CME is an essential pathway used by all eukaryotes for transport across the plasma membrane. It is a stochastic assembly process that depends on membrane mechanics (1,58), enzymatic reactions (59–61), and the stoichiometry of dozens of distinct components (59,62–64), making it a complex system that can benefit from integrated physics-based simulations. Spatial simulations of clathrin, a 600 kDa trimeric protein, assembling in solution (44,54,56,65) and on the membrane (via recruitment by adaptor proteins) (19) have determined assembly pathways, energetic and structural determinants for assembly, and dependence on links to the membrane. The clathrin models developed here for NERDSS illustrate how rate-based models can be adapted and extended to additional components without requiring the substantial expertise of energy-function-based methods (19,66,67) and without sacrificing spatial resolution (45,68–70). To enable nonequilibrium simulations of immature viral lattice formation in vivo, we introduce here a model of the retroviral Gag protein, which assemble a spherical lattice (71), an essential component of the HIV infection and maturation cycle (72). Viral capsid assembly is widely studied using coarse molecular modeling approaches (see (17)), but applications to studying assembly of the precursor lattices that occur within the host cell are more limited (55). In this study, we focus on describing the technical features of the models necessary to produce realistic assemblies, leaving much of the biological implications for future work.

In summary, NERDSS offers a distinct tool that uses the RD model to simulate self-assembly at the cell scale, thus allowing for space- and time-dependent dynamics and kinetics that are immediately comparable to experiment by a broad user base. Below, we first describe the operation of the software and the features we introduced as necessary to move beyond existing tools (Fig. 1). We provide multiple forms of validations, including against fundamental reactions and nonassembly problems such as a circadian clock model, accessible to most RD software. We then present multiple self-assembly applications, with input files provided in the software repository. Models of clathrin lattice assembly are chosen to highlight distinct features of models that can be used to tune the dynamics and stability of the observed structures. We simulate clathrin assembly driven by membrane localization and show how introducing enzymes to the system to control the lipid populations on the membrane drives disassembly. These clathrin models thus capture relaxation from out-of-equilibrium starting states to an equilibrium steady state, with or without active energy consumption. We illustrate the model design process for the self-assembly of Gag retroviral protein monomers into an immature lattice in solution, in which monomers are titrated into solution and have limited lifetimes when unbound. Finally, we discuss the current limitations and most promising future advancements of NERDSS software for realistic cell-scale dynamics.

METHODS

Implementation

NERDSS is written in ANSI/ISO standard C++11 and is available on Linux and macOS. The accompanying GUI is implemented in Java. The source code for both can be found at https://github.com/mjohn218/

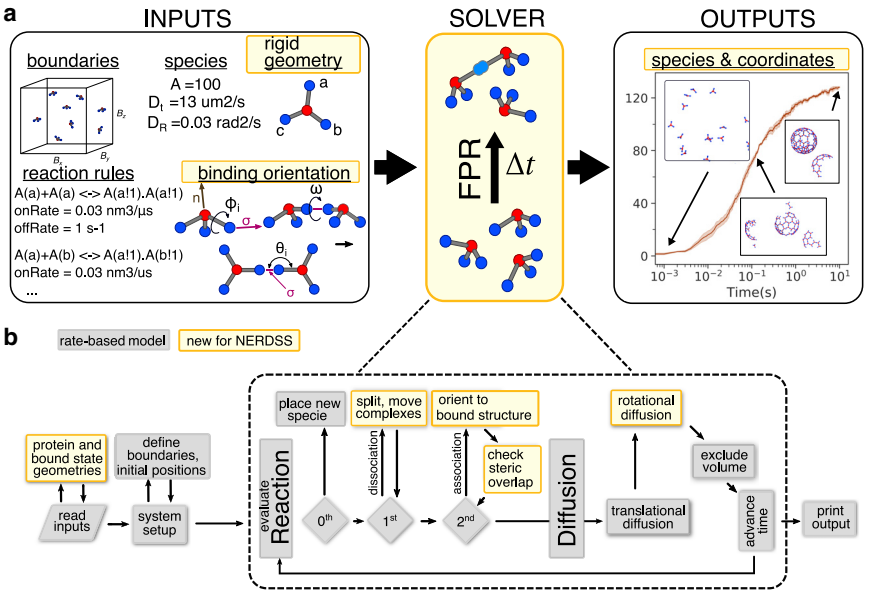


FIGURE 1 NERDSS software overview. (*a*) NERDSS requires additional inputs relative to other rate-based models, highlighted in yellow boxes, with each species modeled as a rigid body with coordinates for a COM (*red*) and its discrete binding sites (*blue*) and a binding orientation for a pair of bound species. These orientation parameters include a binding radius, σ, shown as a purple line between two sites; two angles, θ_1 and θ_2; and three dihedral angles, φ_1, φ_2, and ω (see Supporting Materials and Methods). NERDSS uses free-propagator reweighting (FPR) algorithms to solve the RD model for association and dissociation over time. These algorithms were previously accessible in open source code but without rule-based models and coupling to other reaction types. In addition to coordinates of all species in the simulation as a function of time, NERDSS can track a variety of variables, including species copy numbers, which are formatted for analysis and visualization. Trajectories are output in one of two standard formats, XYZ or PDB, which can be visualized in software such as VMD (90) or Ovito (91). (*b*) An algorithmic flowchart highlights where modifications to time and/or space dependent rate-based models (*gray*) are needed because of the addition of molecular structure and assembly (*yellow boxes*) (Methods). Reactions can be zeroth order: creation; first order: death, transformations, dissociations; second order: binding. For each time step, every molecule either reacts or diffuses. To see this figure in color, go online.

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

NERDSS and is provided under the GNU general public license. A user guide and simulation example systems can be found on the software page. Input files are formatted according to BioNetGen Language (47), such as are used in RuleBender (51) and the NFSim software (8) and is compatible with Virtual Cell (6). Additional features are necessary, however, because of the spatial and structural details used in NERDSS.

Reactions

NERDSS supports the three primary types of reactions: zeroth, first, and second order. The rates of these reactions, other than creation, can depend on the states and/or pre-existing interactions of any participant species' sites.

Zeroth-order reactions

De novo particle creation reactions are treated as a Poisson process. No events occur if a uniform random number (URN) fulfills $URN < \exp(-\lambda)$, where $\lambda = k_0 V \Delta t$, k_0 is the creation rate input in units of M/s, V is the simulation volume (converted from nm^3 to units of M^{-1} by 0.602), and Δt is the timestep. Therefore, N events occur based on $\sum_{n=0}^{N-1}(\exp(-\lambda)\lambda^n/n!) < URN \leq \sum_{n=0}^{N}(\exp(-\lambda)\lambda^n/n!)$.. Created particles are placed within the simulation volume with random coordinates. Zeroth-order reactions can be used to titrate species into the simulation volume.

First-order

First-order reactions are treated as Poisson processes, with a reaction probability per each site of $p_1(\Delta t) = 1 - \exp(-\lambda)$, where $\lambda = k_1 \Delta t$, k_1 is the microscopic rate in units of s^{-1}, and Δt is the time step. The reaction rates can be specified either as microscopic rates or as macroscopic rates. Only one such event can occur per time step per molecule. Four types of first-order reactions are supported:

1) Creation: create a copy of a particle B from another particle A, in which sites required such as (a) are indicated within the parentheses—examples are transcription or translation, e.g., A(a) → A(a) + B(b).
2) Destruction: destroy a molecule and all other molecules in its complex. If the sites (a) are specified as unbound, it will only destroy unbound molecules, e.g., A(a) → NULL.
3) State change: change a state of a site on a molecule, e.g., A(a∼p) → A(a∼u).
4) Dissociation: remove an interaction between the sites of two particles, written as the conjugate back reaction of an association reaction. To retain detailed balance (32), the particles are left such that the dissociating sites are at their binding radius, e.g., A(a!1).B(b!1) → A(a) + B(b).

Second-order reactions

Bimolecular reactions are applied to specific sites between two proteins. Association events can be conditional on the state of the site, e.g., occurring only when the binding site is phosphorylated, or only if the protein is bound through another site as well.

For the free propagator reweighting (FPR) algorithm (32,44,73) used in NERDSS, reaction probabilities are invariant with respect to orientation. Therefore, the reacting particles are "snapped" into place in a predefined geometry to prevent arbitrary structures from forming, with sites always placed at the binding radius σ from one another. Thus, the rigid assemblies that form have strict structures they can form; they do not predict emergent structures. The geometry is defined by a set of vectors and five angles between and within each particle (Fig. 1). If angles and/or orientations are undefined, sites bind to a separation of σ at the orientation they were when the event occurred. Both molecules are translated and rotated into place based on their relative translational and rotational diffusion constants. Thus, smaller complexes will move more to orient, and complexes restricted to the membrane do not rotate out of their membrane-localized orientation. Detailed definitions of these vectors and angles can be found in the Supporting Materials and Methods.

Reaction rates can be defined either as macroscopic rates k_{on} (i.e., rates measured experimentally and dependent on diffusion to contact and energetic barriers), in units of $\mu M^{-1}s^{-1}$, or as microscopic or intrinsic rates k_a, in units of $nm^3/\mu s$ (converts to $\mu M^{-1}s^{-1}$ with multiplication by 0.602). All rates are inputted as three-dimensional (3D) values, where $K_D = k_b/k_a = k_{off}/k_{on}$, where k_b and k_{off} are the corresponding unbinding rates in units s^{-1}. For example, reaction rates in two dimensions (2D) use the 3D value divided by a length scale that by default is set to 2σ (units of nm) but that can be independently specified per reaction using an input parameter. A reaction is identified as 2D when it involves two species that have no diffusion in z (e.g., lipids) or between sites on two complexes that are localized to 2D ($D_z = 0$). Probabilities of each reaction are calculated with the FPR method as previously described and are parameterized by the intrinsic reaction rate k_a, the net diffusion coefficient of the reactants D_{tot}, and the binding radius σ. Conversions between macroscopic rates and the rates needed to calculate the reaction probabilities for each reaction type are performed by the software (see Table S1). Self-binding between molecules such as clathrin with repeated identical sites requires distinctions between binding sites with the same versus distinct labels (see Supporting Materials and Methods). Two types of bimolecular reactions are supported:

1) Association: form an interaction between two sites of two molecules, which, if the association is reversible, has a conjugate first-order dissociation reaction. The resulting complex is then treated as one unit for future propagation, e.g., A(a) + B(b) → A(a!1).B(b!1)
2) State change: change the state of a site on a particle, facilitated by another particle, such as phosphorylation by a kinase. This is thus a composite reaction that combines a bimolecular reaction with an automatic first-order state change of one site, and both reactant species remain unbound. Hence, it is a simplification of Michaelis-Menten, in which reversible binding can trigger a chemical reaction and rapid dissociation, here assuming the chemical reaction is instantaneous; e.g., A(a) + B(b∼u) → A(a) + B(b∼p).

Coupling reactions

NERDSS allows the completion of one reaction (such as dissociation of a complex) to be coupled to another reaction occurring immediately. Thus, one can readily model enzymatic reactions and Michaelis-Menten kinetics. In particular, if an enzyme binds a substrate, the dissociation reaction can be directly coupled to a state change on the substrate, for example, from unphosphorylated to phosphorylated.

Intracomplex binding and defect formation

An important consequence of forming self-assemblies is that molecules can be capable of binding to one another through free sites when they are within the same complex. These intracomplex binding events involve two sites becoming a bound complex but are not truly bimolecular because they do not involve a search to find one another—they are already colocalized. They are thus comparable to conformational rearrangements and are dependent on time (and fundamentally on the details of the atomic interactions) and not on the system size or density. This type of unimolecular or first-order reaction cannot be treated with the same binding probabilities as are used for bimolecular events despite it involving a bond forming between two species. In particular, for the rigid species that we model here, they will have no motion relative to one another ($D_{tot} = 0$), and as noted previously (44), this results in binding probabilities of either one or zero, with the former violating detailed balance and the latter preventing the formation of important stabilizing bonds. Even with flexible molecules, in which intracomplex sites can undergo relative diffusion, constraints on mobility due to the bonded structure would, over most time steps, require adjusted reaction probabilities. We define the binding probability of these intracomplex

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

or loop-closure events thus using a Poisson probability, similar to the first-order reactions described above. Then, we must specify a unimolecular rate, given a bimolecular rate constant. We define this rate such that the equilibrium between the bound and unbound states for a two-step process of loop closure, that is, a bimolecular event and a unimolecular loop closure, is the same as if the protein closed the loop in a single step, forming both bonds at once. The rate is given by

$$k_{close} = k_a C_0 \exp\left(-\Delta G_{coop} / k_B T\right),$$

where C_0 is the standard state concentration of 1 M. For loop closure, the user can thus specify one additional parameter, ΔG_{coop}. By default, $\Delta G_{coop} = 0$, and loop-closure events are typically difficult to reverse, requiring two bond-breaking events to separate the loop back into two pieces. Positive values of this parameter render loops less stable and more reversible or dynamic (Fig. S3). We derive this expression in the Supporting Materials and Methods.

We allow these events to form when the sites are spatially proximate, with the user able to specify a maximum distance to allow these events to occur. The maximum distance is the binding radius, σ, multiplied by a factor, bindRadSameCom, which by default is set to 1.1, which includes only perfectly (with numerical precision errors allowed) aligned contacts. The primary reason to increase it is because for some self-assembly structures (such as the curved clathrin cage), the rigid structures cannot form perfect, defect-less structures (see Fig. S8). Allowing binding between legs that are close together mimics structural flexibility present in the biological molecules. A more sophisticated treatment in the future could couple defects to energetic penalties that impact reaction rates and allow relaxation within the rigid assembly as deviations relative to the user-defined orientation angles. Finally, the software prohibits binding between a pair of molecules if they are already bound through a distinct set of sites. Hence, intracomplex or loop-closure events must be mediated through at least a third molecule (e.g., four molecules in the case of clathrin lattices).

Evaluation of steric overlap

Once association events occur between two sites, their two parent complexes are rotated into place to generate the proper, predefined geometry of the two binding proteins. The exception is intracomplex binding, in which because the binding happens within a single complex, no rotation or movement occurs. For two separate complexes, however, this rotation into place can result in steric overlap between sites that are part of the complex but not part of the binding event. We evaluate overlap after association by measuring distances between all centers of mass (COMs) in the new complex and defining a minimum distance, overlapSepLimit, that will determine steric overlap. If any pair of proteins have COMs less than overlapSepLimit, which by default is 1 nm, the binding event is rejected, and the two complexes instead undergo a diffusion move during the time step. This is the simplest evaluation of steric overlap and could be improved, particularly in cases of defect-forming structures (Fig. S9). Here again, more elaborate schemes could determine whether the full space occupied by a molecule overlaps with other molecules, not just the COM.

A final steric overlap check is then performed between the newly bound complex and all other binding sites of complexes present in the simulation to ensure excluded volume is maintained. If the new complex produces steric clashes with other binding sites in the simulation volume, the binding event is rejected.

Excluded volume

Sites that react with one another have excluded volume that is determined by their binding radius, σ. Specifically, in 3D, a reactive site excludes all partners from entering within a sphere of radius σ, and in 2D, within a disk of radius σ. By default, sites otherwise do not exclude volume with each other but act as point particles, including the COM sites. Therefore, if one desires sites to exclude one another, they must be specified as undergoing a bimolecular reaction, and by setting the rate to zero, they will reflect off one another at σ without ever binding. We note that the addition of reactions inevitably slows down the simulations.

An important consequence of having excluded volume only between specified reactive pairs is that once a pair of sites is in the bound state, they no longer automatically exclude volume from other sites in the system. For proteins that are fully bound (and thus are temporarily nonreactive to bind), they can thus pass through other proteins, which is often physically undesirable. To maintain excluded volume, proteins can be given dummy sites that are present purely to exclude volume with dummy sites on other molecules, at all times. A future alternative will be to have an option for all bimolecular reactions that sites can continue reflecting off their reaction partners, whether they are in the free or bound states.

Boundary effects

By default, the simulation boundaries are reflecting. Rectangular and spherical geometries are implemented, with no restriction on user-defined size. If association events result in a very large lattice, it is possible for the structure to extend beyond the physical boundaries of the simulation volume. These moves are thus rejected. On a related note, if association events result in very large rotational reorientation of a complex, these moves can also be rejected because they result in an unphysical amount of displacement per time step. Binding to lipids or sites restricted to the surface can be defined using explicit lipids or a field-based implicit lipid model (74).

Diffusion constants of complexes

Once proteins bind to form a new complex, their translational and rotational diffusion constants are updated to reflect the larger hydrodynamic radius of the bound complex. Each protein molecule has a user-defined D_t and D_R. Once a bound complex forms, the new transport coefficients are then defined based on all N components of the complex by simply assuming the radii sum:

$$D_{t,x} = \left[\sum_{i=1}^{N} D_{i,x}^{-1}\right]^{-1},$$

$$D_{R,x} = \left[\sum_{i=1}^{N} D_{Ri,x}^{-1/3}\right]^{-3},$$

and the same in y and z. For molecules restricted to the surface, $D_{t,z} = 0$ and all rotational diffusion except (if desired) $D_{R,z}$ are also zero. Correspondingly, any complex that contains these molecules will also have that diffusional component set at zero.

Data output

NERDSS produces restart files that ensure that when simulations are interrupted for any reason, they can be restarted exactly from where they left off, similar to molecular dynamics software. NERDSS also allows restarts to introduce new molecules or reactions into the system using input files formatted the same as the standard inputs. This allows systems to be prepared or equilibrated in one state before, for example, enzymes are introduced.

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

Simulation algorithm outline

1) Initialization. Copies of rigid species are created from the provided molecule templates, given random coordinates, and checked for placement within the simulation volume and overlap with other particles. If two reactive partners are found to be overlapping, one of the particles is reinitialized with new random coordinates, and all particles are rechecked for overlap.
2) Optimization. To optimize evaluation of the two-body (bimolecular) events, the simulation volume is split into N sub-boxes, in which the length of each sub-box edge is at least $R_{max,sys} = \max_m\{\sigma_m + 3\sqrt{6D_m\Delta t} + l_{1m} + l_{2m}\}$, where m loops over all binding reactions, σ is the binding radius, D the total diffusion coefficient of both reactants, and l_{im} is the radius of the reactant protein i. Molecules are assigned to boxes based on the position of their COM. This allows for looping over all possible binding partners only in a molecule's current and neighboring boxes, without excluding any partners with a nonzero reaction probability.
3) Reactions. Within each time step, reaction probabilities are calculated in order from zeroth-, first-, and second-order reactions. Reactions are described in detail above. We note that if a molecule is created or involved in a reaction in any way, it cannot participate in any other reaction during the course of the time step, nor can it diffuse. Hence, only one site per molecule can react per step. Other diffusing molecules will continue to include these molecules in overlap checks to prevent overlap in subsequent steps.
4) Propagation and overlap. After all reaction checks are completed, sites that were within R_{max} of a reaction-partner during the time step are checked for overlap. Because a multicomponent complex only moves as a rigid unit, each complex with reactive sites within is looped over, along with a stored list of all reaction partners. If the newly displaced positions overlap (with displacements sampled from Gaussians for translational and rotational diffusion for both reaction partners), the displacements of complex i and its partner are resampled, and the overlap check restarts. Once a complex's position has been updated, it cannot be moved again.

Simulation details and parameters

Simulation parameters are listed in Table 1, and values are included in figure legends. Transport coefficients are defined using the Einstein-Stokes equation, using each molecule's hydrodynamic radius estimated from its approximate radius of gyration. Unless otherwise noted, $\sigma = 1$ nm for binding.

We describe here the geometry of the molecules from the clathrin and Gag systems and the orientation angles.

Clathrin model

For the flat clathrin molecule, each clathrin trimer has a COM and three leg binding sites rotated 120° relative to one another, and each leg site is 10 nm from the center such that when dimerized, the centers are 21 nm apart, which is comparable to experimental lattices (75). When two clathrin molecules bind to each other, we suppose their leg sites are head to head and on a line with their centers, and, to generate a flat hexagon structure, all their centers and legs should be on the same plane. Thus, the five orientation angles are $[\theta_1, \theta_2, \varphi_1, \varphi_2, \omega] = [\pi, \pi, \text{nan}, \text{nan}, 0]$ ("nan" means no value constraint, and see Fig. 1 and Supporting Materials and Methods for the definition of orientation angles). For binding to adaptor proteins, each flat clathrin includes three more leg sites, which we choose to locate directly beneath the legs ($\pi/4.656$ angle between the long leg and short leg), as shown in Fig. 5 *a*. When the clathrin binds to a linear adaptor protein, the orientation should make the long legs parallel to the membrane, where the adaptor and lipid site are designed to be perpendicular to membrane surface. Therefore, the five orientation angles of adaptor to clathrin (adaptor is molecule 1) should be set as $[\pi/2, 3.656\pi/4.656, \text{nan}, \pi, \pi]$. For the puckered clathrin, the legs are tilted below the plane (5, 10, or 20°) and rotated 120° around the molecule normal, with $l_{leg} = 7.5$ nm. The orientation angles of two bound puckered clathrin is the same as flat clathrin [π, π, nan, nan, 0]. The adaptor binding to the lipid is two linear molecules binding with angles [$\pi/2$, $\pi/2$, nan, nan, π].

TABLE 1 Parameters for NERDSS Simulations

Name	Unit	Meaning
Δt	μs	time step
k_{on}	$\mu M^{-1}s^{-1}$	macroscopic reaction rate
k_{off}	s^{-1}	macroscopic unbinding rate
k_a	$nm^3/\mu s$	microscopic/intrinsic reaction rate
k_b	s^{-1}	microscopic/intrinsic unbinding rate
K_D	μM	dissociation constant
D_t	$nm^2/\mu s$	translational diffusion coefficient/constant
D_R	$rad^2/\mu s$	rotational diffusion coefficient/constant
D_{tot}	$\mu nm^2/\mu s$	net diffusion coefficient of one reactant relative to another
σ	nm	binding radius
R_{max}	nm	largest distance within which binding reactions can occur
ΔG_{coop}	k_BT	adjustment of the free energy during the loop-closure event
f		loop cooperativity factor, $f = \exp(-\Delta G_{coop}/k_BT)$
bindRadSameCom		Multiplier σ, sets max distance between sites to allow the loop-closure event to occur
overlapSepLimit	nm	Max distance between molecule COMs resulting in steric overlap

Gag model

The Gag lattice will form a sphere of radius ~50 nm because we design the diameter of each hexamer to be ~8 nm and the angle between two hexamers is ~9.2°, comparable to experimental lattices (76). Relative to the COM, the Gag has homodimerization sites at [0, 0.5 nm, 0]. The homodimerization angles are $[\theta_1, \theta_2, \varphi_1, \varphi_2, \omega] = [1.49, 1.49, 0, 0, 0]$. The heterodimerization sites are set at [−2 nm, 0, 0] and [2 nm, 0, 0], such that the side of each hexamer is ~4 nm (slightly longer because of the binding radius). The five angles are [2.62, 2.62, π, π, 0]. Our Gag monomer has been designed with binding sites for membrane lipids and RNA binding sites, so future work can effectively probe assembly kinetics as controlled by Gag binding partners and membrane localization.

RESULTS

NERDSS enables self-assembly simulations

NERDSS expands beyond existing single-particle RD software by propagating rigid, user-defined, orientable components that assemble rigid multicomponent structures such as hexagonal clathrin lattices (Fig. 1 *a*). The algorithmic elements of NERDSS that are new for single-particle and spatial RD (Fig. 1 *b*) arise because of this propagation of rigid multisite molecules (44), which requires generalized routines to orient molecules and complexes once they bind, corresponding assessment of steric overlap and defect

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

contacts, integration of rule-based control of each site, and generalized output for visualization and analysis (see Methods). We provide a GUI to facilitate the otherwise cumbersome construction of rigid molecules and bound states. The binding between molecules is independent of their orientation before the event (44), which is true of all single-particle methods, including those with interaction potentials, because of their spherical symmetry (14,40). NERDSS is efficient enough to simulate seconds to minutes of assembly on a single processer (Fig. S1), primarily because of algorithms that support relatively large time steps (32). Localization and assembly on membranes also can proceed efficiently and accurately by using a recent algorithm for implicit lipids, which avoids the need to propagate large populations of individual lipid binding sites (74). Although not yet optimized for parallel or GPU architectures (27,34,40), this is an active development. The significant advantages of NERDSS are the types of biological model systems it makes accessible and the accuracy of the single-particle algorithms used for propagating species, illustrated in the next section.

Validation of NERDSS kinetics and equilibria

We validate the NERDSS software first for several non-self-assembly models to verify fundamental reactions and the rule-based implementation of the RD model either against theory or other simulation methods. Zeroth- and first-order reactions reproduce the exact analytical solutions as expected, and they are not dependent on space (newly created particles are placed randomly in the volume) (Fig. 2 *a*). Bimolecular reactions for single-particles have been previously validated for the FPR algorithms in 3D (32), 2D (73), and transitions from 3D to 2D with flat and curved surfaces (74), but we perform additional verification here (Fig. 2 *b*). The FPR algorithms are derived to preserve detailed balance for reversible binding interactions, ensuring thermodynamic equilibrium is reached when systems are designed to (excluding kinetic traps). Binding to the surface can be mediated through explicit lipid sites, or we have implemented an implicit lipid model that is substantially more accurate than surface adsorption models at reproducing binding kinetics of the explicit lipid simulations (74). This model provides orders of magnitude speedups relative to explicit lipid simulations (74) and couples with all other reaction types (Fig. 2 *c*). Although NERDSS is compatible with arbitrarily curved surfaces (77), the user inputs currently accommodate only spherical geometries (Fig. 2 *c*) or rectangular volumes in which the flat membrane is on one surface. We propagate diffusion on the sphere by exactly sampling displacements on the spherical surface (Supporting Materials and Methods), and all distances between surface-bound particles are measured using geodesics.

To validate the self-assembly simulations, we follow previous work (44) by comparing our results against nonspatial rate-based simulations and by comparing our simulation results as we change the time step of integration (Fig. 2 *d*). Changing the time step is a valuable test for two reasons: first, as the time step decreases, the assumption that each

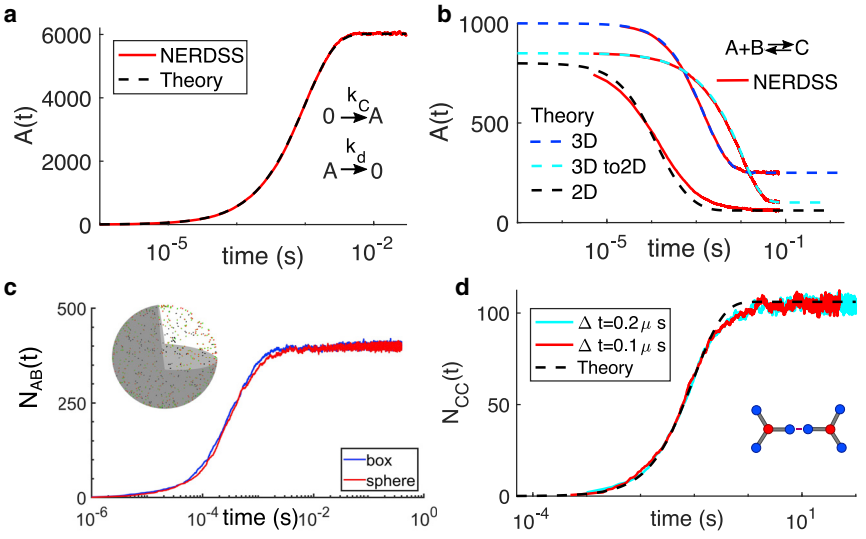


FIGURE 2 Multiple validations of NERDSS algorithms and features. (*a*) Creation and destruction are compared with analytical theory from the corresponding oridinary differential equation (ODE), where $k_C = 0.01$ M/s and $k_d = 1000$ s^{-1}, V = 1 μm^3. (*b*) Bimolecular reactions in different dimensions are shown. 3D and 2D analytical theory curves from the ODE, 3D → 2D theory as the numerical solution to the PDE are shown, solved using Virtual Cell. The kinetics of the 2D single-particle RD solution is theoretically known to differ from the continuum solution (73). For all reactions, $\sigma = 2$ nm. For 3D: V = 0.83 μm^3, $D_A = D_B = 20$ μm^2/s, $k_{on} = 300$ μM^{-1}s^{-1}, $k_{off} = 50$ s^{-1}, $A_0 = B_0 = 2$ μM. For 2D: A = 1 μm^2, $D_A = D_B = 2$ μm^2/s, $k_{on} = 10$ μm^{2}s^{-1}, $k_{off} = 50$ s^{-1}, ($k_a^{2D} = 100$ μm^{2}s^{-1}), $A_0 = B_0 = 800/\mu$m^2. For 3D → 2D: cube of L = 1 μm. $D_A = 40$ μm^2/s, $D_B = 2$ μm^2/s, $k_{on} = 100$ μM^{-1}s^{-1}, $k_{off} = 50$ s^{-1}, $A_0 = 1.412$ μM, $B_0 = 3000$ μm^2. (*c*) Localization and binding on the membrane (as modeled in (45)) with a box geometry and a sphere geometry produce the same equilibrium and kinetics. This model combines all bimolecular reactions tested in part (*b*) and includes diffusion on the curved surface and the implicit lipid model (see Fig. S5 for parameters). (*d*) Clathrin trimer self-assembly, where theory is the analytical solution to the ODE for reversible binding between identical, independent sites, is shown. V = 0.494 μm^3, $N_{trimer,0} = 100$, $k_{on} = 1$ μM^{-1}s^{-1}, $k_{off} = 1$ s^{-1}, $D_{trimer} = 13$ μm^2/s, $D_R = 0.03$ rad^2/s. $<N_{CC}>_{eq} = 104.7$ and 104.8 for $\Delta t = 0.1$, 0.2 μs respectively. The NERDSS simulations capture the spatial lattice formation, and thus, their equilibrium values also depend on the formation of closed loops, where here $f = 5.9e{-}6$, bindRadSameCom = 5 nm, and overlapSepLimit = 7 nm. To see this figure in color, go online.

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

multisite molecule is interacting with only one other site becomes more and more accurate, which improves adherence to solving only two-body problems per step (32). Second, the reaction probabilities depend nonlinearly on the time step, so if inaccuracies existed in diffusion or reactions, they show up quantifiably in the kinetics and equilibria. We performed the same validation for assembly in 2D (Fig. S4). For self-assembly, we must introduce a new loop-closure reaction type (Methods), and we show how the choice of this parameter, which controls the binding free energy of loop-closure events, impacts the equilibrium (Fig. S3). We cannot compare NERDSS self-assembly with other single-particle software because they do not capture orientation-dependent interactions and thus cannot form a structured lattice or cage. It is informative to compare NERDSS to NFSim (8), which lacks spatial resolution but simulates rate-based models that can form multiprotein assemblies via multisite species and rule-based modeling (Figs. S2 and S3). In weak-binding regimes, they agree exactly, but with stronger binding, the structure resolution in NERDSS drives spatially localized interactions and captures several elements of self-assembly not possible with NFSim, including lattice geometries and topologies and their growth patterns.

To test the rule-based implementation with a more complex nonassembly model, we simulated a previously developed minimal model of a circadian genetic oscillator (53) with NERDSS, as well as with PDEs and the stochastic simulation algorithm (via Virtual Cell software (6)), showing quantitative agreement across all models (Fig. 3). Our simulation recapitulates how the expression of two proteins, an activator protein (A) and repressor protein (R), can be coupled to produce robust oscillations of both proteins and their bound complex. For our NERDSS model, when messenger RNA and proteins are created, they are placed adjacent to the molecule creating them. All rates were accelerated by a factor of 3600 relative to the original rates because of computational costs, such that oscillations occurred over timescales of seconds rather than hours. This change made the binding events diffusion limited, meaning they could become sensitive to the spatial distribution of particles. However, because the species mix rapidly and the volume is not too large (Video S5), the kinetics were essentially insensitive to diffusion and the spatial dimension, indicating the resolution offered by NERDSS is not necessary for this model. We note the PDE and nonspatial simulations have significantly faster CPU time.

NERDSS tests clathrin cage assembly designed by four distinct models of clathrin-clathrin interactions

Clathrin, a 600 kDa trimeric protein, assembles into both flat and spherical lattices in vivo and in vitro (1,78,79), although in vivo this process only occurs on the plasma membrane, with the requirement of a host of accessory and adaptor proteins, because clathrin itself does not bind the membrane. Clathrin trimers are observed to assemble in solution in vitro either at low pH or in the presence of clathrin terminal-binding adaptor proteins (80,81). We model distinct assembly conditions by changing binding strengths, cooperativity, and binding geometries of clathrin-only solutions, contrasting how multiple observables vary with distinct model parameterizations at a fixed (1.3 μM) clathrin

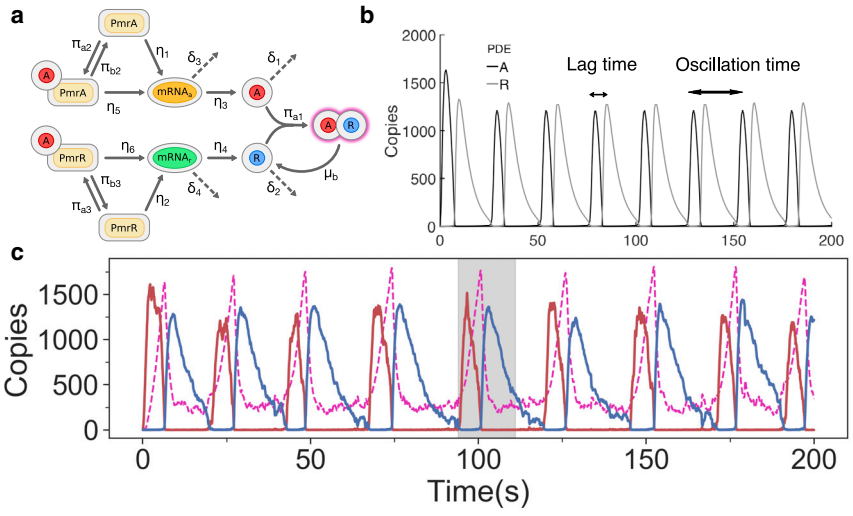


FIGURE 3 NERDSS simulations of a circadian clock model recapitulate oscillatory protein expression. (*a*) The model contains six components, with the activator (A, in *red*) and repressor (R, in *blue*) proteins produced by their corresponding messenger RNA, which are produced by a single copy of each gene (PmrA and PrmR). All species have $D_t = 10\ \mu m^2/s$, and initial copies were zero except for one copy of PrmA and PrmR. Unimolecular reactions (except dissociation) are shown with reactions indicated by η or δ ($\eta_1 = 50\ s^{-1}$, $\eta_2 = 0.01\ s^{-1}$, $\eta_3 = 50\ s^{-1}$, $\eta_4 = 5\ s^{-1}$, $\eta_5 = 500\ s^{-1}$, $\eta_6 = 50\ s^{-1}$; $\delta_1 = 1\ s^{-1}$, $\delta_2 = 0.2\ s^{-1}$, $\delta_3 = 10\ s^{-1}$, $\delta_4 = 0.5\ s^{-1}$) binding and unbinding reactions are shown with rates π ($\pi_{a1} = 1204\ \mu M^{-1}s^{-1}$, $\pi_{a2} = 602\ \mu M^{-1}s^{-1}$, $\pi_{a3} = 602\ \mu M^{-1}s^{-1}$, $\pi_{b2} = 100\ s^{-1}$, $\pi_{b3} = 50\ s^{-1}$, $u_b = 1\ s^{-1}$). For bimolecular association reactions, NERDSS converts the macroscopic rates above to microscopic rates, where $\sigma = 5$ nm for binding PrmA/R and $\sigma = 8$ nm for A + R. One exception is unbinding of the complex A-R (*magenta dashed*). Because unbinding results in only R (A is modeled as degraded instantaneously), there is no possibility for geminate recombination, and thus, the unbinding follows the macroscopic rate u_b. (*b*) Solution using a PDE solver produces regular oscillations, where the calculated oscillation time is 25.1 s. (*c*) The oscillation time of A (and R) from NERDSS is in close agreement with the PDE-based simulations, calculated as 24.5 s. In the Supporting Materials and Methods, we show similar results for a nonspatial model (time = 24.8 s). Simulation results were the same with $\Delta t = 10$ or 50 μs. Because of the constant production and degradation of components in this model, we used Ovito software to produce videos for this model (Video S5) and the Gag model (Video S4). To see this figure in color, go online.

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

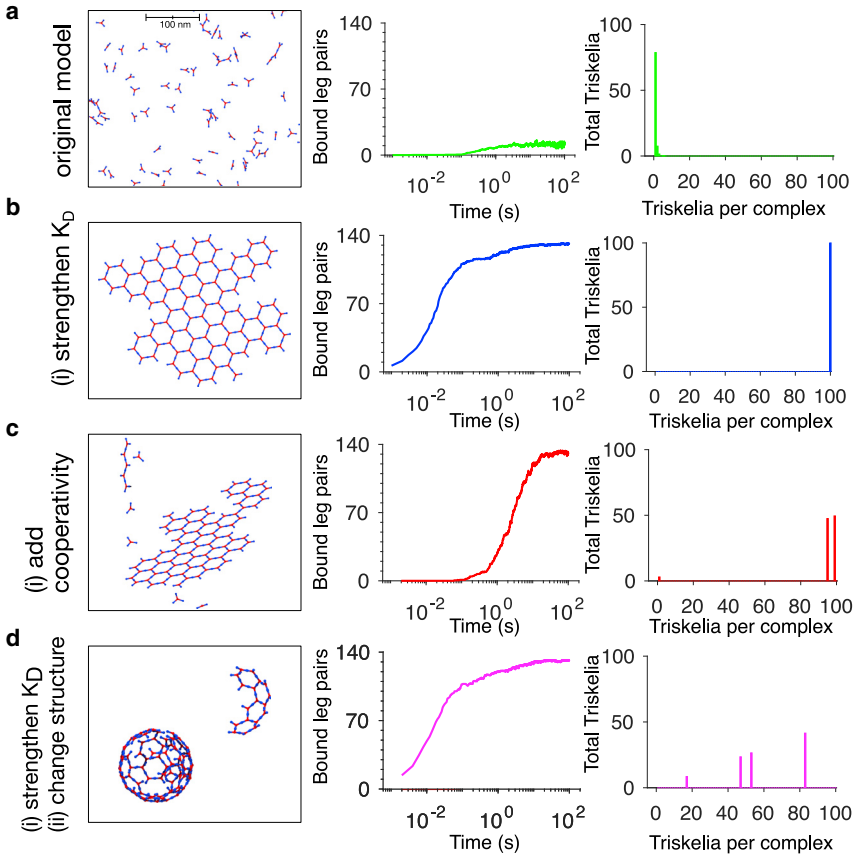


FIGURE 4 NERDSS simulations of designed clathrin lattice assemblies in solution, with 1.3 μM clathrin triskelia. (*a*) For K_D = 100 μM, most clathrin remains monomeric in solution. The second column shows corresponding counts of pairs of bound clathrin trimer legs in time. The third column shows a histogram of the total copies of clathrin triskelia as distributed complexes of increasing size. (*b*) Strengthening K_D to 0.2 μM is shown. (*c*) The cooperative model has binding of monomeric to nonmonomeric clathrin stronger by a factor of 10, and nonmonomeric to nonmonomeric is again 10 times stronger. Lattices can now form despite a weak monomeric K_D = 100 μM. (*d*) By changing geometry of the clathrin monomers (K_D = 0.2 μM), they now assemble into spherical cages with similar binding kinetics but that nucleate smaller complexes (d_3 relative to b_3). Results are averaged over three to five trajectories, and Supporting Videos are generated with VMD (90). Cube with L = 0.494 μm, k_{off} = 1 s^{-1}, and f = 0.001. $N_{trimer,0}$ = 100, D_{trimer} = 13 $\mu m^2/s$, D_R = 0.03 rad^2/s, Δt = 0.2 μs. Flat clathrin molecules: leg length l_{leg} = 10 nm. Puckered clathrin, l_{leg} = 7.5 nm, and bindRadSameCom = 5 nm, overlapSepLimit = 7 nm. To see this figure in color, go online.

concentration. The baseline model has a K_D for leg-leg binding of 100 μM, which is comparable to experimental estimates of trimer dimerization (82). Very little binding is observed in solution at this concentration (Fig. 4 *a*), which is consistent with in vitro studies of clathrin in solution at physiological pH (81). The same results are observed in nonspatial simulations (Fig. S3).

By simply strengthening the K_D, the flat clathrin lattices start to assemble rapidly, to the point that a K_D = 1 μM produces a similar equilibrium in bound legs as 0.2 μM (Figs. 2 *d*, 4 *b*, and S10). As we strengthen K_D, average size of complexes transitions to a giant component at ~10 to 1 μM. With the formation of closed loops in the clathrin lattice, we can introduce positive or negative cooperativity, due to the formation of two simultaneous bound interfaces within the closed loop (Methods and Supporting Materials and Methods). This can thus alter the final equilibrium for a given dimerization K_D, favoring fewer bonds and more remodeling (Figs. S3 and S10), and will also shift the value of k_{on} when nucleation of complexes progresses to complete lattices.

By instead introducing cooperativity in binding between clathrins such that the rate depends on whether a monomer is already bound to another clathrin, flat lattices assemble after a delay, followed by rapid growth (Fig. 4 *c*). This model thus has three on-rates for clathrin-clathrin binding, with strongest binding between clathrins that are already bound to others. This growth model results in a broader distribution of lattice sizes (Fig. 4 *c*; Video S1).

By taking the model of Fig. 4 *b* (K_D = 0.2 μM) and altering the geometry of a clathrin monomer to tilt legs 10° below the plane (100° pucker), spherical cages form (Video S2). The kinetics of leg-leg binding remains very similar to Fig. 4 *b*, but the sizes of the complexes are smaller due to spatial constraints of closed spheres and steric overlap (Fig. 4 *d*). The extent of the tilt controls the radius of the sphere, with this 100° clathrin producing a sphere of R ~45 nm containing ~100 clathrin. A 95° clathrin, in contrast, produce a sphere of R ~92 nm, estimated to contain ~500 trimers, and a 110° degree clathrin produces a sphere of R ~25 containing ~20 trimers (Fig. S7). We note that the rigid clathrin trimer cannot form a perfect spherical lattice, and thus, defects appear even within a single closed hexagon. In the most puckered clathrin (110°), the geometry much better accommodates pentagons, not hexagons (Fig. S7). Biomolecules typically have the flexibility to bend and form contacts by supporting a distribution of structural geometries, and clathrin is known to form cages in solution and on the membrane at a range of curvatures (75). We thus allow sites within a complex to bind one another even when they are not at perfect contact (Methods). The choice of cutoff distances for binding within a complex and for identifying steric overlap (which is more difficult

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

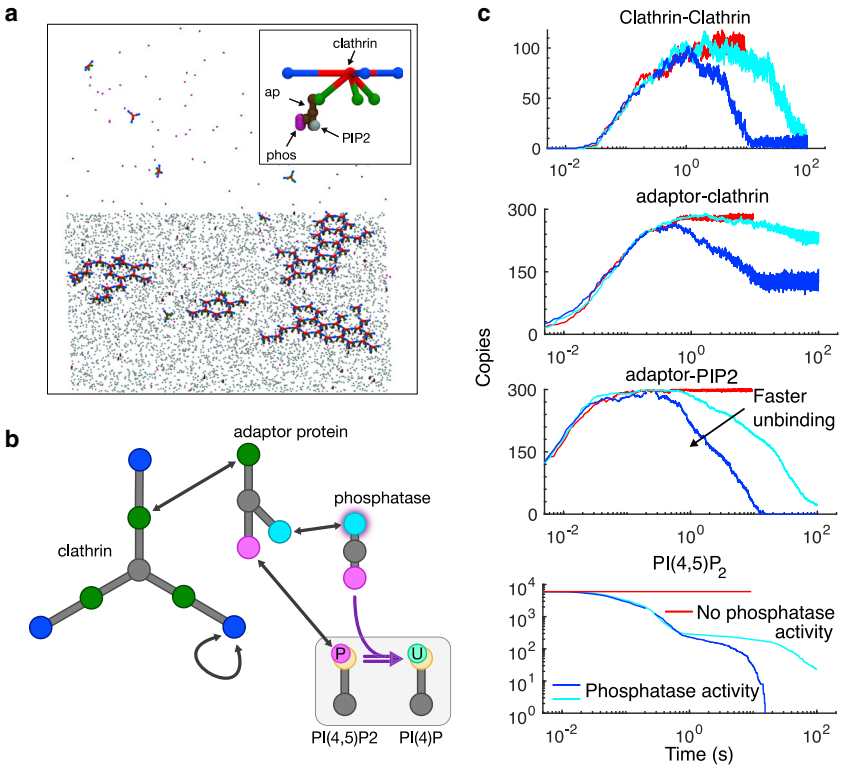


FIGURE 5 Clathrin assembly and disassembly by membrane localization and delocalization. (*a*) Three multivalent solution proteins are included in this model (clathrin, adaptor protein, phosphatase) along with one membrane lipid ($PI(4,5)P_2$). (*b*) Clathrin contains three sites for binding other clathrin and three separate sites for binding to the adaptor. The adaptor protein has a dedicated site for each of the three other components, which is typical of a wide range of clathrin adaptors (83). This allows clustering on the membrane. The $PI(4,5)P_2$ is present comparable to its physiologic value, ~1 mol % of plasma membrane lipids (92). The phosphatase has independent sites for adaptor and lipid binding. (*c*) With no enzymatic activity, the assembly proceeds downhill, directly to an equilibrium steady state (*red curves*). With phosphatase activity turned on, the majority of the lipids are converted by 1 s, except for the population that has been already bound to adaptors. The clathrin lattice thus has time to assemble but is gradually destabilized as determined by the rate of unbinding between adaptor and lipid and adaptor and clathrin, which increases from 1 to 10 s^{-1} between cyan and blue. $N_{cla} = 100$ clathrin trimers, $N_{adaptor} = 300$, $N_{pip2} = 6000$, $N_{phos} = 10$. Rectangular box of size [0.7, 0.7, 0.494] μm, flat and immobile membrane on the bottom. $\Delta t = 0.2$ μs. $k_{off}^{Cla\text{-}Cla} = 10\ s^{-1}$. $k_{on}^{Cla\text{-}adap} = 6\ \mu M^{-1}s^{-1}$ $k_{off}^{Cla\text{-}adap} = 1\ s^{-1}$, $10\ s^{-1}$. $k_{on}^{pip2\text{-}adap} = 6\ \mu M^{-1}s^{-1}$, $k_{off}^{pip2\text{-}adap} = 1\ s^{-1}$, $10\ s^{-1}$. $k_{on}^{phos\text{-}adap} = 0.6\ \mu M^{-1}s^{-1}$, $k_{off}^{phos\text{-}adap} = 1\ s^{-1}$, $k_{on}^{pip2\text{-}phos} = 25\ \mu M^{-1}s^{-1}$, or 0 for "No activity." For all 2D binding interactions, $k_a^{2D} = k_a^{3D}/2\sigma$, k_b is unchanged, $\sigma = 1$ nm, $f = 0.001$, $D_{cla} = 13\ \mu m^2/s$, $D_{R,cla} = 0.03$ rad^2/s. $D_{ap} = 25\ \mu m^2/s$, $D_{R,ap} = 0.5\ rad^2/s$. $D_{pip} = 1\ \mu m^2/s$ ($D_{z,pip} = 0$), $D_{R,pip} = 0$. $D_{syn} = 25\ \mu m^2/s$, $D_{R,syn} = 0.5\ rad^2/s$. To see this figure in color, go online.

with defects) does impact the stability of the cages and whether they prevent unphysical overlap (Figs. S8 and S9); more sophisticated treatments in the future will improve the treatment of defect formation (Methods).

NERDSS probes lattice assembly driven by localization to the membrane

Increasing component concentrations is another natural mechanism to nucleate assembly. In Fig. 5, we show how localization to the 2D membrane surface can, on its own, nucleate clathrin lattices by increasing the effective concentration of clathrin relative to the 3D solution (Video S3). Clathrin does not bind the membrane directly but can be localized to the surface by an adaptor protein. The clathrin-clathrin interaction is set here to a weak value of 100 μM (Fig. 5 *a*; (82)), and thus, minimal binding occurs in solution. Once the clathrin have been localized to the surface, however, they can assemble into a lattice that is stabilized by relatively weak clathrin-clathrin contacts that are favorable because of the small search space available on the 2D surface (45). Although binding to many adaptor proteins is known to stabilize clathrin-clathrin interactions (80), we do not incorporate this cooperativity here to isolate the role of 2D localization in nucleating assembly.

Enzymes can drive lattice disassembly by removing links to the membrane

By removing links between clathrin and the membrane, we can drive the clathrin assembly back into solution where our clathrin lattice is no longer stable. Physiologically, this is achievable by changing the phosphorylation state of the lipid $PI(4,5)P_2$, which is essential for localizing adaptor proteins, and thus clathrin, to the membrane (83). Thus, without altering the clathrin or the adaptor proteins directly, we can drive disassembly of the lattice. We include here 10 copies of a lipid phosphatase (e.g., synaptojanin), which converts $PI(4,5)P_2$ to PI(4)P irreversibly. Critically, the adaptor protein cannot bind to PI(4)P, thus removing its link to the surface. Importantly, the phosphatase synaptojanin can be localized to sites of clathrin-coated structures in vivo through protein interactions with the adaptor protein AP-2[85], thus allowing it to act in 2D to rapidly dephosphorylate unbound $PI(4,5)P_2$. Although in vivo, it has been shown to play an important role in dephosphorylating $PI(4,5)P_2$ after fission from the membrane (59), we demonstrate here how it is capable of driving disassembly of plasma-membrane-bound clathrin-coated lattices, at least in this in vitro-type model.

We find the timescales of the lattice disassembly are sensitive to the binding and unbinding kinetics of all

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

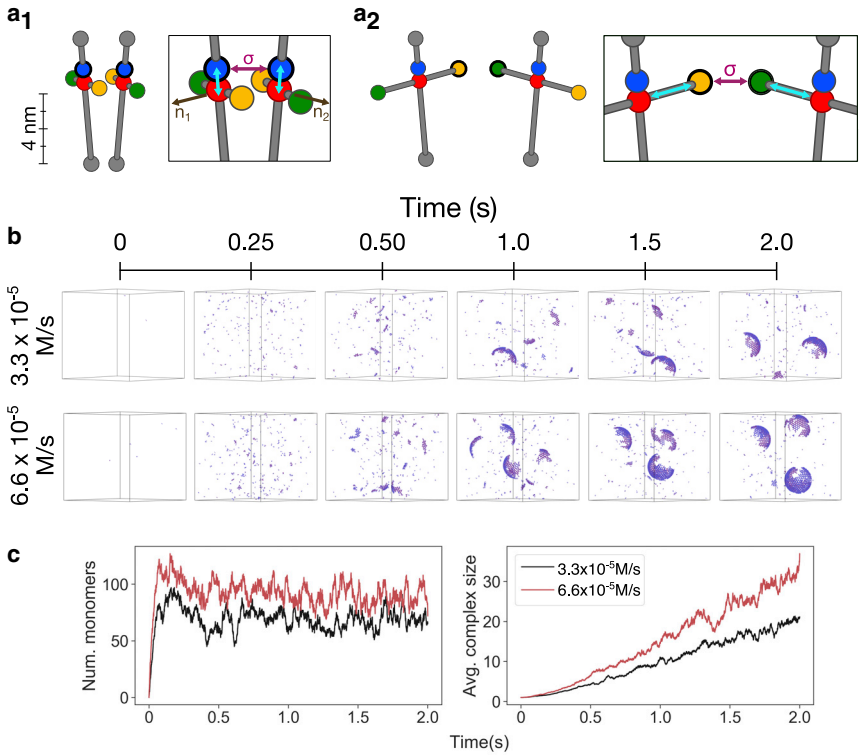


FIGURE 6 Gag monomer assembly model setup and titration experiment. (*a*) Each Gag protein monomer is given three active binding sites and two inactive binding sites, positioned relative to a COM site in red. The blue site is for homodimerization, H, and the green (G1) and orange (G2) sites are for heterodimerization to produce hexamers. The five orientation angles are based on the COM to binding site vectors (*turquoise lines*), the sigma vector (*purple line*), and two molecule normals (*black vectors*) that must not be co-linear with the site vectors to provide an additional dimension (Methods). The orientation of the bound heterodimer similarly requires five angle definitions, and G1 must bind G2 at a *cis* orientation to ensure a hexamer loop can form (inset). (*b*) Trajectories show the titration in of Gag monomers at 3.3×10^{-5} and 6.6×10^{-5} M/s. (*c*) The number of monomers initially grows rapidly, but reaches a relative steady state because of assembly and degradation, which occurs at a rate of 1 s^{-1}. Assembled proteins are protected from degradation, causing a continual growth in sizes of assembled complexes. Box = [1, 1, 1] μm. $\Delta t = 0.1$ μs, $k_{on}^{H\text{-}H} = 3$ $\mu M^{-1}s^{-1}$, $k_{off}^{H\text{-}H} = 1$ s^{-1}. $k_{on}^{G1\text{-}G2} = 0.3$ $\mu M^{-1}s^{-1}$, $k_{off}^{G1\text{-}G2} = 1$ s^{-1}, $D_t = 10$ $\mu m^2/s$, $D_R = 0.05$ rad^2/s. $f = 0.001$, bindRadSameCom = 1.5 nm, overlapSeplimit = 0.7 nm. $k_{deg} = 1$ s^{-1} (only fully monomeric Gag). To see this figure in color, go online.

protein-protein and protein-lipid interactions (Fig. 5). In particular, although the phosphatases rapidly dephosphorylate unbound PI(4,5)P_2, the adaptor proteins protect the subset of lipids they are bound to, and their links are cut on a timescale that is determined by unbinding of the adaptor-lipid interactions. Because clathrin can bind up to three adaptor proteins, its dissociation from the membrane then also depends on dissociation of all three adaptors from the surface. Although this model is not fully parameterized to reproduce experiment, it illustrates how localization, enzymatic activity, and binding kinetics can be tuned to control assembly and disassembly. Equilibrium constants used are within reasonable ranges of experimental measurements for endocytic proteins (collected here (45)).

Model building requires orientations for each pairwise binding reaction: application to viral assembly in solution

To illustrate the model design process, we develop a coarse model of the Gag monomer, a retroviral protein that dimerizes with itself and oligomerizes into a hexagonal lattice on the plasma membrane of infected cells to form the so-called immature virion (72). To form a lattice of hexamers from these approximately linear monomers, two distinct types of dimers are sufficient, a homodimer and heterodimer interaction, with relative orientations defined by five separate angles for each dimer (Fig. 6). The heterodimer forms between head-to-tail monomers, allowing polymer-like growth at a specific hexameric geometry. The lattice forms with sizes of hexameric subunits (~8 nm in diameter) and a spherical curvature (R ~50 nm) that would support an inclusion of ~2500 Gags in the full spherical shell, as observed experimentally (76). The Gag-Gag homodimerization is stronger than other Gag-Gag contacts at an estimated 1–10 μM and appears to form first (84). The weaker Gag-Gag hexameric interactions are known to be autoinhibited before binding to either RNA, negatively charged analogs, or the membrane (72,84); however, here we treat all interactions as constitutive for simplicity. Our Gag monomer has been designed with binding sites for membrane lipids and RNA binding sites, so future work can effectively probe assembly kinetics as controlled by Gag binding partners and membrane localization.

To illustrate the functionality of NERDSS, we initialize the volume with zero Gag monomers. We then titrate them in using a zeroth-order reaction, for which they are placed randomly in the simulation volume. Thus, the concentration slowly increases, favoring slower nucleation and more efficient growth of the Gag into a small number of larger assemblies (Video S4). Monomeric Gag is also degraded, although assembled Gag is not. Thus, the long-time behavior consists of Gag as stable spheroids or as short-lived (~1 s) monomers. As designed, the Gag is able to assemble into large spheroids entirely because of the two specified reactions (Fig. 6). The assembly kinetics depends on the titration speed, which can provide additional knobs to help optimize the model against experimental data.

Please cite this article in press as: Varga et al., NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale, Biophysical Journal (2020), https://doi.org/10.1016/j.bpj.2020.05.002

DISCUSSION

NERDSS has several advantageous features that make it a powerful and immediately useful tool for cell-scale simulations. First, NERDSS is transferrable between distinct systems thanks to its rate-based interaction framework, which avoids the time-consuming parameterization of energy functions often hard coded for specific assembly systems (19). Although protein geometries and orientations of bound complexes in NERDSS are system specific, we provide a GUI to facilitate user design of proteins and their bound states. Second, unlike existing spatial rate-based approaches (33), the built-in molecular structure of NERDSS not only enforces excluded volume of binding sites but evaluates whether steric overlap or volume restrictions would prevent the formation of unphysical assembly structures. Third, it uses the FPR algorithms (32,44,73) to efficiently propagate species while retaining accurate rates of association and dissociation, allowing for current simulations of timescales on the order of seconds to minutes. NERDSS converts carefully between microscopic rates used in FPR and macroscopic rates commonly defined from experiment, with extensive validation (44). Fourth, NERDSS uses a BioNetGen Language-style syntax (46), and models built in other software packages using similar syntax can be ported to NERDSS with minimal alterations, making it available for immediate use. In this combination of structural resolution, efficient propagation and accurate treatment of association rates, and user-friendly input file syntax, NERDSS is a unique tool for simulating equilibrium and nonequilibrium self-assembly and other cell-scale phenomena.

A current limitation of NERDSS is that all species must be rigid bodies, and thus, assembly defects that arise within complexes that cannot perfectly tile surfaces (such as rigid bodies on a sphere) are accommodated approximately by using a distance tolerance. Molecule flexibility (14) is a natural extension beyond rigid molecules that would improve defects and also enable better treatment of disordered regions and genomes. NERDSS also lacks any orientation dependence on binding between complexes before association, which is an approximation that can result in unphysical binding events, particularly between large multicomponent intermediates, as also discussed previously (44). Although we have implemented options to reject these moves, introducing orientation dependence to reaction probabilities (13,85) would be more physically realistic and could allow NERDSS to predict a range of structures rather than the ones predetermined by the defined geometries, although this would require more expense or approximate. NERDSS is designed to be extensible to not only users but to developers expanding functionality. The FPR algorithms support, for example, the implementation of more elaborate physical models that introduce electrostatic interactions between particles (13,32). A particularly exciting future development is integration of NERDSS with continuum membrane models (74) to allow more realistic simulations of vesicle and virion formation dynamics and coupling of assembly to mechanical force generation, which is a key component, for example, in cytoskeletal assembly models (86) and continuum membrane remodeling studies (87–89).

NERDSS is designed to be more user friendly and portable than previous versions of the FPR algorithms (44) with the addition of rule-based modeling, the formatting of input and output files, and the addition of the GUI, all available open source here: github.com/mjohn218/NERDSS. However, it does not match the usability of software such as Virtual Cell in terms of model setup and analysis, including defining cell geometries from experimental data (6,52). It also does not match the level of parallelization or optimization of software such as Lattice Microbes (27). These will be important future software developments to support a broad and diverse user base.

We have introduced self-assembly models here motivated by CME and virion formation in cells as a foundation to extend these models to additional components and experimental comparison. We show how assembly can be nucleated despite weak clathrin-clathrin interactions through cooperativity or through reduction of dimensionality to the membrane (3D to 2D) (45). With NERDSS, we predicted here how the dephosphorylation of essential plasma membrane lipids, which occurs at sites of vesicle budding (59), could de-stabilize clathrin-coated structures and drive their disassembly. Ultimately, NERDSS has encoded as a core feature the ability to resolve relatively fast processes over long timescales and individual proteins over long length scales, simulating otherwise intractable protein assembly dynamics. This provides immediate use in helping overcome the challenges of understanding or designing self-assembling structures in biology.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2020.05.002.

AUTHOR CONTRIBUTIONS

M.J.V., Y.F., and M.E.J. wrote the software. Y.F. implemented the implicit lipid and spherical model to the software. O.N.Y. and S.L. designed and developed the GUI. M.J.V., Y.F., and M.E.J. ran simulations, collected results, and wrote the article.

ACKNOWLEDGMENTS

We thank Danny Evans for setting up the Gag model, Ipsita Saha for providing feedback on Gag simulations, Nomongo Dorjsuren for formatting the output files from the GUI, and Dr. Sikao Guo for software updates.

M.E.J. gratefully acknowledges funding from a National Science Foundation CAREER Award #1753174 and a National Institutes of Health

MIRA Award R35GM133644. We acknowledge use of the MARCC supercomputer at Johns Hopkins, and the XSEDE supercomputer Stampede2 through XRAC MCB150059.

REFERENCES

1. Bucher, D., F. Frey, …, S. Boulant. 2018. Clathrin-adaptor ratio and membrane tension regulate the flat-to-curved transition of the clathrin coat during endocytosis. *Nat. Commun.* 9:1109.
2. Bartocci, E., and P. Lió. 2016. Computational modeling, formal analysis, and tools for systems biology. *PLoS Comput. Biol.* 12:e1004591.
3. Lewis, J. 2008. From signals to patterns: space, time, and mathematics in developmental biology. *Science.* 322:399–403.
4. Jilkine, A., and L. Edelstein-Keshet. 2011. A comparison of mathematical models for polarization of single eukaryotic cells in response to guided cues. *PLoS Comput. Biol.* 7:e1001121.
5. Gunawardena, J. 2014. Models in biology: 'accurate descriptions of our pathetic thinking'. *BMC Biol.* 12:29.
6. Moraru, I. I., J. C. Schaff, …, L. M. Loew. 2008. Virtual Cell modelling and simulation software environment. *IET Syst. Biol.* 2:352–362.
7. Angermann, B. R., F. Klauschen, …, M. Meier-Schellersheim. 2012. Computational modeling of cellular signaling processes embedded into dynamic spatial contexts. *Nat. Methods.* 9:283–289.
8. Sneddon, M. W., J. R. Faeder, and T. Emonet. 2011. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat. Methods.* 8:177–183.
9. Sanft, K. R., S. Wu, …, L. R. Petzold. 2011. StochKit2: software for discrete stochastic simulation of biochemical systems with events. *Bioinformatics.* 27:2457–2458.
10. Sweeney, B., T. Zhang, and R. Schwartz. 2008. Exploring the parameter space of complex self-assembly through virus capsid models. *Biophys. J.* 94:772–783.
11. Zlotnick, A. 1994. To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* 241:59–67.
12. Endres, D., and A. Zlotnick. 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* 83:1217–1230.
13. Zhou, H. X. 1993. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys. J.* 64:1711–1726.
14. Michalski, P. J., and L. M. Loew. 2016. SpringSaLaD: a spatial, particle-based biochemical simulation platform with excluded volume. *Biophys. J.* 110:523–529.
15. Gruenert, G., B. Ibrahim, …, P. Dittrich. 2010. Rule-based spatial modeling with diffusing, geometrically constrained molecules. *BMC Bioinformatics.* 11:307.
16. Turing, A. M. 1953. The chemical basis of morphogenesis. *Philos Trans R Soc B.* 237:37–72.
17. Hagan, M. F. 2014. Modeling viral capsid assembly. *Adv. Chem. Phys.* 155:1–68.
18. Grime, J. M. A., J. F. Dama, …, G. A. Voth. 2016. Coarse-grained simulation reveals key features of HIV-1 capsid self-assembly. *Nat. Commun.* 7:11568.
19. Giani, M., W. K. den Otter, and W. J. Briels. 2017. Early stages of clathrin aggregation at a membrane in coarse-grained simulations. *J. Chem. Phys.* 146:155102.
20. Hall, D. M., and G. M. Grason. 2017. How geometric frustration shapes twisted fibres, inside and out: competing morphologies of chiral filament assembly. *Interface Focus.* 7:20160140.
21. Hagan, M. F., O. M. Elrad, and R. L. Jack. 2011. Mechanisms of kinetic trapping in self-assembly and phase transformation. *J. Chem. Phys.* 135:104115.
22. Perlmutter, J. D., F. Mohajerani, and M. F. Hagan. 2016. Many-molecule encapsulation by an icosahedral shell. *eLife.* 5:e14078.
23. Wilber, A. W., J. P. Doye, …, A. C. Lewis. 2009. Monodisperse self-assembly in a model with protein-like interactions. *J. Chem. Phys.* 131:175102.
24. Saglam, A. S., and L. T. Chong. 2016. Highly efficient computation of the basal kon using direct simulation of protein-protein association with flexible molecular models. *J. Phys. Chem. B.* 120:117–122.
25. Saglam, A. S., and L. T. Chong. 2018. Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chem. Sci. (Camb.).* 10:2360–2372.
26. Martinez, M., N. J. Bruce, …, R. C. Wade. 2015. SDA 7: a modular and parallel implementation of the simulation of diffusional association software. *J. Comput. Chem.* 36:1631–1645.
27. Roberts, E., J. E. Stone, and Z. Luthey-Schulten. 2013. Lattice microbes: high-performance stochastic simulation method for the reaction-diffusion master equation. *J. Comput. Chem.* 34:245–255.
28. Drawert, B., S. Hellander, …, L. Petzold. 2016. A framework for discrete stochastic simulation on 3D moving boundary domains. *J. Chem. Phys.* 145:184113.
29. Drawert, B., S. Engblom, and A. Hellander. 2012. URDME: a modular framework for stochastic simulation of reaction-transport processes in complex geometries. *BMC Syst. Biol.* 6:76.
30. Earnest, T. M., J. Lai, …, Z. Luthey-Schulten. 2015. Toward a whole-cell model of ribosome biogenesis: kinetic modeling of SSU assembly. *Biophys. J.* 109:1117–1135.
31. Ghaemi, Z., J. R. Peterson, …, Z. Luthey-Schulten. 2020. An in-silico human cell model reveals the influence of spatial organization on RNA splicing. *PLoS Comput. Biol.* 16:e1007717.
32. Johnson, M. E., and G. Hummer. 2014. Free-propagator reweighting integrator for single-particle dynamics in reaction-diffusion models of heterogeneous protein-protein interaction systems. *Phys. Rev. X.* 4:031037.
33. van Zon, J. S., and P. R. ten Wolde. 2005. Simulating biochemical networks at the particle level and in time and space: Green''s function reaction dynamics. *Phys. Rev. Lett.* 94:128103.
34. Andrews, S. S. 2017. Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. *Bioinformatics.* 33:710–717.
35. Schoneberg, J., and F. Noé. 2013. ReaDDy–a software for particle-based reaction-diffusion dynamics in crowded cellular environments. *PLoS One.* 8:e74261.
36. Takahashi, K., S. Tanase-Nicola, and P. R. ten Wolde. 2010. Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proc. Natl. Acad. Sci. USA.* 107:2473–2478.
37. Gillespie, D. T., E. Seitaridou, and C. A. Gillespie. 2014. The small-voxel tracking algorithm for simulating chemical reactions among diffusing molecules. *J. Chem. Phys.* 141:234115.
38. Arjunan, S. N. V., and K. Takahashi. 2017. Multi-algorithm particle simulations with spatiocyte. *Methods Mol. Biol.* 1611:219–236.
39. Sokolowski, T. R., J. Paijmans, …, P. R. Ten Wolde. 2019. eGFRD in all dimensions. *J. Chem. Phys.* 150:054108.
40. Hoffmann, M., C. Fröhner, and F. Noé. 2019. ReaDDy 2: fast and flexible software framework for interacting-particle reaction dynamics. *PLoS Comput. Biol.* 15:e1006830.
41. Chattaraj, A., M. Youngstrom, and L. M. Loew. 2019. The interplay of structural and cellular biophysics controls clustering of multivalent molecules. *Biophys. J.* 116:560–572.
42. Antosiewicz, J., and J. A. McCammon. 1995. Electrostatic and hydrodynamic orientational steering effects in enzyme-substrate association. *Biophys. J.* 69:57–65.
43. Vijaykumar, A., T. E. Ouldridge, …, P. G. Bolhuis. 2017. Multiscale simulations of anisotropic particles combining molecular dynamics and Green's function reaction dynamics. *J. Chem. Phys.* 146:114106.

44. Johnson, M. E. 2018. Modeling the self-assembly of protein complexes through a rigid-body rotational reaction-diffusion algorithm. *J. Phys. Chem. B.* 122:11771–11783.

45. Yogurtcu, O. N., and M. E. Johnson. 2018. Cytosolic proteins can exploit membrane localization to trigger functional assembly. *PLoS Comput. Biol.* 14:e1006031.

46. Blinov, M. L., J. R. Faeder, …, W. S. Hlavacek. 2004. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics.* 20:3289–3291.

47. Faeder, J. R., M. L. Blinov, and W. S. Hlavacek. 2009. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol. Biol.* 500:113–167.

48. Deeds, E. J., J. Krivine, …, W. Fontana. 2012. Combinatorial complexity and compositional drift in protein interaction networks. *PLoS One.* 7:e32032.

49. Boutillier, P., M. Maasha, …, W. Fontana. 2018. The Kappa platform for rule-based modeling. *Bioinformatics.* 34:i583–i592.

50. Hlavacek, W. S., J. R. Faeder, …, W. Fontana. 2006. Rules for modeling signal-transduction systems. *Sci. STKE.* 2006:re6.

51. Smith, A. M., W. Xu, …, G. E. Marai. 2012. RuleBender: integrated modeling, simulation and visualization for rule-based intracellular biochemistry. *BMC Bioinformatics.* 13 (*Suppl 8*):S3.

52. Schaff, J. C., D. Vasilescu, …, M. L. Blinov. 2016. Rule-based modeling with virtual cell. *Bioinformatics.* 32:2880–2882.

53. Vilar, J. M., H. Y. Kueh, …, S. Leibler. 2002. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. USA.* 99:5988–5992.

54. Ilie, I. M., W. K. den Otter, and W. J. Briels. 2014. Rotational Brownian dynamics simulations of clathrin cage formation. *J. Chem. Phys.* 141:065101.

55. Ayton, G. S., and G. A. Voth. 2010. Multiscale computer simulation of the immature HIV-1 virion. *Biophys. J.* 99:2757–2765.

56. Schoen, A. P., N. Cordella, …, S. C. Heilshorn. 2013. Dynamic remodelling of disordered protein aggregates is an alternative pathway to achieve robust self-assembly of nanostructures. *Soft Matter.* 9:9137–9145.

57. VanDersarl, J. J., S. Mehraeen, …, N. A. Melosh. 2014. Rheology and simulation of 2-dimensional clathrin protein network assembly. *Soft Matter.* 10:6219–6227.

58. Antonny, B., C. Burd, …, S. Schmid. 2016. Membrane fission by dynamin: what we know and what we need to know. *EMBO J.* 35:2270–2284.

59. He, K., R. Marsland, III, …, T. Kirchhausen. 2017. Dynamics of phosphoinositide conversion in clathrin-mediated endocytic traffic. *Nature.* 552:410–414.

60. Chang-Ileto, B., S. G. Frere, …, G. Di Paolo. 2011. Synaptojanin 1-mediated PI(4,5)P2 hydrolysis is modulated by membrane curvature and facilitates membrane fission. *Dev. Cell.* 20:206–218.

61. Schöneberg, J., M. Lehmann, …, F. Noé. 2017. Lipid-mediated PX-BAR domain recruitment couples local membrane constriction to endocytic vesicle fission. *Nat. Commun.* 8:15873.

62. Schmid, E. M., and H. T. McMahon. 2007. Integrating molecular and network biology to decode endocytosis. *Nature.* 448:883–888.

63. Traub, L. M. 2009. Tickets to ride: selecting cargo for clathrin-regulated internalization. *Nat. Rev. Mol. Cell Biol.* 10:583–596.

64. Sochacki, K. A., A. M. Dickey, …, J. W. Taraska. 2017. Endocytic proteins are partitioned at the edge of the clathrin lattice in mammalian cells. *Nat. Cell Biol.* 19:352–361.

65. den Otter, W. K., M. R. Renes, and W. J. Briels. 2010. Asymmetry as the key to clathrin cage assembly. *Biophys. J.* 99:1231–1238.

66. Cordella, N., T. J. Lampo, …, A. J. Spakowitz. 2015. Membrane indentation triggers clathrin lattice reorganization and fluidization. *Soft Matter.* 11:439–448.

67. Ramanan, V., N. J. Agrawal, …, R. Radhakrishnan. 2011. Systems biology and physical biology of clathrin-mediated endocytosis. *Integr. Biol.* 3:803–815.

68. Holland, D. O., and M. E. Johnson. 2018. Stoichiometric balance of protein copy numbers is measurable and functionally significant in a protein-protein interaction network for yeast endocytosis. *PLoS Comput. Biol.* 14:e1006022.

69. Banerjee, A., A. Berezhkovskii, and R. Nossal. 2012. Stochastic model of clathrin-coated pit assembly. *Biophys. J.* 102:2725–2730.

70. Sorokin, A., K. F. Heil, …, O. Sorokina. 2018. Rule-based modelling provides an extendable framework for comparing candidate mechanisms underpinning clathrin polymerisation. *Sci. Rep.* 8:5658.

71. Wright, E. R., J. B. Schooler, …, G. J. Jensen. 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* 26:2218–2226.

72. Sundquist, W. I., and H. G. Kräusslich. 2012. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* 2:a006924.

73. Yogurtcu, O. N., and M. E. Johnson. 2015. Theory of bi-molecular association dynamics in 2D for accurate model and experimental parameterization of binding rates. *J. Chem. Phys.* 143:084117.

74. Fu, Y., O. N. Yogurtcu, …, M. E. Johnson. 2019. An implicit lipid model for efficient reaction-diffusion simulations of protein binding to surfaces of arbitrary topology. *J. Chem. Phys.* 151:124115.

75. Morris, K. L., J. R. Jones, …, C. J. Smith. 2019. Cryo-EM of multiple cage architectures reveals a universal mode of clathrin self-assembly. *Nat. Struct. Mol. Biol.* 26:890–898.

76. Briggs, J. A., J. D. Riches, …, H. G. Kräusslich. 2009. Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. USA.* 106:11090–11095.

77. Feng, F., and W. S. Klug. 2006. Finite element modeling of lipid bilayer membranes. *J. Comput. Phys.* 220:394–408.

78. Kelly, B. T., S. C. Graham, …, D. J. Owen. 2014. Clathrin adaptors. AP2 controls clathrin polymerization with a membrane-activated switch. *Science.* 345:459–463.

79. Saffarian, S., E. Cocucci, and T. Kirchhausen. 2009. Distinct dynamics of endocytic clathrin-coated pits and coated plaques. *PLoS Biol.* 7:e1000191.

80. Zaremba, S., and J. H. Keen. 1983. Assembly polypeptides from coated vesicles mediate reassembly of unique clathrin coats. *J. Cell Biol.* 97:1339–1347.

81. Pearse, B. M., and R. A. Crowther. 1987. Structure and assembly of coated vesicles. *Annu. Rev. Biophys. Biophys. Chem.* 16:49–68.

82. Wakeham, D. E., C. Y. Chen, …, F. M. Brodsky. 2003. Clathrin self-assembly involves coordinated weak interactions favorable for cellular regulation. *EMBO J.* 22:4980–4990.

83. McMahon, H. T., and E. Boucrot. 2011. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* 12:517–533.

84. Datta, S. A., Z. Zhao, …, A. Rein. 2007. Interactions between HIV-1 Gag molecules in solution: an inositol phosphate-mediated switch. *J. Mol. Biol.* 365:799–811.

85. Shoup, D., G. Lipari, and A. Szabo. 1981. Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints. *Biophys. J.* 36:697–714.

86. Mund, M., J. A. van der Beek, …, J. Ries. 2018. Systematic nanoscale analysis of endocytosis links efficient vesicle formation to patterned actin nucleation. *Cell.* 174:884–896.e17.

87. Rangamani, P., K. K. Mandadap, and G. Oster. 2014. Protein-induced membrane curvature alters local membrane tension. *Biophys. J.* 107:751–762.

88. Wu, Z., M. Su, …, J. Liu. 2018. Membrane shape-mediated wave propagation of cortical protein dynamics. *Nat. Commun.* 9:136.

89. Akamatsu, M., R. Vasan, …, D. G. Drubin. 2020. Principles of self-organization and load adaptation by the actin cytoskeleton during clathrin-mediated endocytosis. *eLife.* 9:e49840.

90. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–38.

91. Stukowski, A. 2010. Visualization and analysis of atomistic simulation data with OVITO-the open visualization tool. *Model Simul. Mater. Sci. Eng.* 18:015012, Published online December 15, 2009.

92. Yoon, Y., P. J. Lee, …, W. Cho. 2011. In situ quantitative imaging of cellular lipids using molecular sensors. *Nat. Chem.* 3:868–874.

Biophysical Journal, Volume *118*

Supplemental Information

NERDSS: A Nonequilibrium Simulator for Multibody Self-Assembly at the Cellular Scale

Matthew J. Varga, Yiben Fu, Spencer Loggia, Osman N. Yogurtcu, and Margaret E. Johnson

Supporting Information
for

NERDSS: a non-equilibrium simulator for multibody self-assembly at the cellular scale

Matthew J. Varga[1], Spencer Loggia[1], Yiben Fu[1], Osman N Yogurtcu[2], and Margaret E. Johnson[1,*]

Supporting Methods

I. Loop closure probability derivation:

Consider a complex A—B, where both protein A and protein B have sites free to bind C. Protein C has two sites that can bind each to A and B, thus being able to form a closed loop. We can write the single step reaction as:

$$A(a).B(b) + C(a,) \rightleftharpoons A(a!1).B(b!2).C(a!1,b!2) \quad \text{Eq. S1}$$

If we treat the association of C to A—B as a single binding event, we can define the K_D due to binding free energy gained by both CA and CB forming, plus an additional free energy gain or loss due to the specifics of the molecules themselves:

$$K_D^{1step} = C_0 \exp\left(\frac{[\Delta G_{CA}+\Delta G_{CB}+\Delta G_{coop}]}{k_B T}\right) = \frac{A(a).B(b)_{eq} * C(a,b)_{eq}}{A(a!1).B(b!2).C(a!1,b!2)_{eq}}, \quad \text{Eq. S2}$$

where C_0 is the standard state concentration of 1M, k_B is Boltzmann's constant, T is the temperature, and ΔG_{CA} is the binding free energy difference of bound minus unbound CA, with the same definition for CB. The free energy does not have to be strictly the sum of the two contributions, hence the addition of the parameter ΔG_{coop}.

In practice, treating the loop closure as a single binding step makes it very difficult and time-consuming to properly evaluate binding probabilities and unbinding probabilities on-the-fly. The rate k_a would change if it were a loop closure relative to a standard binding event. Loop closure events are spatial and not known in advance, so one would have to check for each possible reaction evaluation if the two proteins would spatially orient in a way to allow loop closure. For unbinding, to preserve detailed balance, both links would have to be broken at once, whether A, B, or C unbound.

Because of the algorithmic complexity of single-step closure, we will instead preserve the same free energy difference between the open and closed loop conformations as would occur in a single step closure, by breaking the binding into two steps, a standard bimolecular association with unchanged rates, and a unimolecular loop closure. In BNGL syntax, this looks like:

$$A(a).B(b) + C(a,b) \rightleftharpoons A(a!1).B(b).C(a!1,b) \quad \text{Eq. S3a}$$

$$A(a!1).B(b).C(a!1,b) \rightleftharpoons A(a!1).B(b!2).C(a!1,b!2) \quad \text{Eq. S3b}$$

or

$A(a).B(b) + C(a,b) \rightleftharpoons A(a).B(b!2).C(a,b!2)$ Eq. S4a

$A(a).B(b!2).C(a,b!2) \rightleftharpoons A(a!1).B(b!2).C(a!1,b!2)$ Eq. S4b

For an individual bimolecular association between C and A, we have a standard definition:

$K_D^{CA} = C_0 \exp(\frac{[\Delta G_{CA}]}{k_B T}) = \frac{k_b}{k_f^{CA}}$ Eq. S5

and the same for CB, where k_f and k_b are the binding and unbinding rates for the bimolecular association reaction. If we compare Eq. S5 with Eq. S2, we can state that:

$K_D^{1step} = \frac{k_b k_b \exp(\Delta G_{coop}/k_B T)}{k_f^{CA} k_f^{CB} C_0}$. Eq. S6

We want from our two-step model that:

$K_D^{2step} = \frac{A(a).B(b)_{eq} * C(a,b)_{eq}}{A(a!1).B(b!2).C(a!1,b!2)_{eq}}$. Eq. S7

We assume the first step in the two-step model, the bimolecular association event, is fixed and unchanged, given by:

$\frac{A(a).B(b)_{eq} * C(a,b)_{eq}}{A(a!1).B(b).C(a!1,b)_{eq}} = \frac{k_b}{k_f^{CA}}$. Eq. S8

To recover the desired result of Eq. S7 equivalent to Eq. S6, we must define the second step, the unimolecular step, with equilibrium:

$\frac{A(a!1).B(b).C(a!1,b)_{eq}}{A(a!1).B(b!2).C(a!1,b!2)_{eq}} = \frac{k_b \exp(\Delta G_{coop}/k_B T)}{k_f^{CB} C_0}$. Eq. S9

The forward rate for the unimolecular reaction thus becomes $k_{close} = k_f^{CB} C_0 \exp(-\Delta G_{coop}/k_B T)$. The reaction probability is then evaluated per time-step as the other unimolecular reactions, based on a Poisson process. The same result occurs if CB forms first, followed by CA. Both the on- and off-rates can also be re-scaled by the same constant, since we constrain here only the equilibrium. By default, the scale factor is 1 and k_b is unchanged. We note that in the two-step process, the total A is partitioned amongst four species, A(a).B(b), A(a!1).B(b).C(a!1,b), A(a).B(b!2).C(a,b!2), and A(a!1).B(b!2).C(a!1,b!2), whereas in the one-step process it is only two species (first and last in the list). Hence the equilibrium populations of these species will be slightly different in both models, although the free energy difference between the monomers and closed loops is the same.

In practice, the loop closing frequency using this definition with ΔG_{coop}=0 is much higher than the opening frequency, resulting in loops that rarely disassemble, because the K_D from Eq. S2 is much stronger than the individual binding events. Positive values of ΔG_{coop} thus destabilize the loops, making the open rate more competitive with the close rate, and producing more dynamic and reversible loops. This additional parameter can be specified per reaction pair as the scale factor (loopCoopFactor):

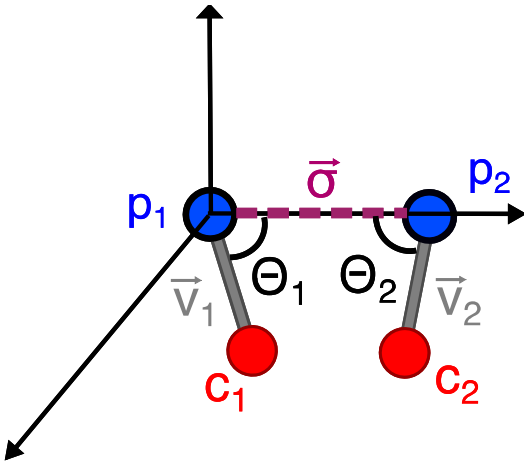
$f = \exp(-\Delta G_{coop}/k_B T)$ Eq. S10

It will be applied to the loop closure rate if those proteins can ultimately form a loop.

II. Orientations of proteins in a bound complex: Angle definitions

Once two molecules have bound to one another, their relative orientations can be specified to 'snap' them into place. This is specified in the model input files via 5 angles: θ_1, θ_2, φ_1, φ_2, and ω, which can be designed using our GUI. Each molecule has two angles defined relative to the binding radius vector (θ_1, φ_1), and ω then defines the orientation of molecule 1 relative to molecule 2. Point particles do not have any orientations to specify, and linear molecules do not have φ values.

When molecules are rotated into position, molecule *i* displaces according to $D_{R,i}/(D_{R,1}+D_{R,2})$. If both molecules have D_R=0, then they displace according to relative D_t values.



Theta Angles: The first two angles, θ_1 and θ_2, determine the orientation [0:π] between the binding radius vector $\vec{\sigma}$ and the site-to-COM vectors

$\overrightarrow{v_1} = p_1 - c_1$ Eq. S11a

$\overrightarrow{v_2} = p_2 - c_2$ Eq. S11b

connecting each molecule center-of-mass (c_1,c_2) to its participating binding site (p_1,p_2).

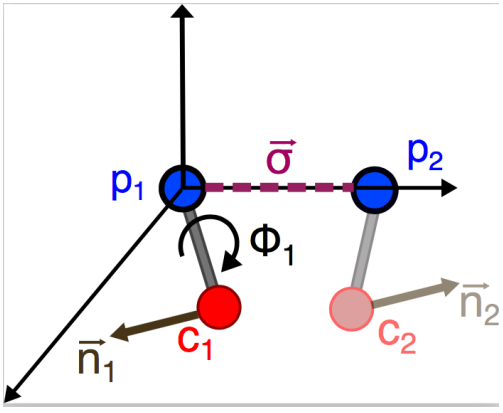
$\theta_1 = \text{acos}\left(\frac{\overrightarrow{v_1}\cdot\overrightarrow{\sigma_1}}{|\overrightarrow{v_1}||\overrightarrow{\sigma_1}|}\right)$ Eq. S12a

and

$\theta_2 = \text{acos}\left(\frac{\overrightarrow{v_2}\cdot\overrightarrow{\sigma_2}}{|\overrightarrow{v_2}||\overrightarrow{\sigma_2}|}\right)$. Eq. S12b

We define the sign of the vector $\vec{\sigma}$ relative to the molecule being moved, such that

$\overrightarrow{\sigma_1} = p_1 - p_2 = -\overrightarrow{\sigma_2}$. Eq. S13



Phi angle: The angles φ_1 and φ_2 are dihedral angles that orient a second axis of each molecule $\overrightarrow{n_1}$ and $\overrightarrow{n_2}$ (not co-linear with the site-to-COM vectors) relative to the binding radius vector $\vec{\sigma}$.

We have that

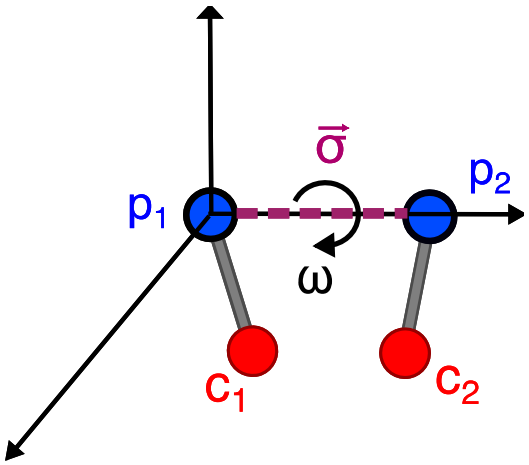
$\varphi_1 = \text{acos}(\hat{t_1} \cdot \hat{t_2})$, Eq. S14a

here $\hat{t_1}$ and $\hat{t_2}$ are the unit normals defined by

$\overrightarrow{t_1} = \overrightarrow{v_1} \times \overrightarrow{\sigma_1}$ Eq. S14b

$\overrightarrow{t_2} = \overrightarrow{v_1} \times \overrightarrow{n_1}$, Eq. S14c

with $\overrightarrow{v_1}$ defined above. The sign of φ_1 is determined by the direction of $\overrightarrow{t_2}$ relative to the right-hand rule of $\overrightarrow{v_1} \times \overrightarrow{t_1}$. The angle φ_2 is defined the same way, with subscripts 2 in the definitions of $\overrightarrow{t_1}$ and $\overrightarrow{t_2}$.



Omega angle: The ω angle is a dihedral angle [-π:π] between the site-to-COM vectors around the binding radius vector σ. This is the only angle that directly constrains the orientations of one molecule to the other (rather than to the binding radius vector $\vec{\sigma}$). We have that

$$\omega = \text{acos}(\hat{t_1} \cdot \hat{t_2}), \qquad \text{Eq. S15a}$$

here $\hat{t_1}$ and $\hat{t_2}$ are the unit normals defined by

$$\overrightarrow{t_1} = \overrightarrow{\sigma_1} \times \overrightarrow{v_1} \qquad \text{Eq. S15b}$$

$$\overrightarrow{t_2} = \overrightarrow{\sigma_1} \times \overrightarrow{v_2}. \qquad \text{Eq. S15c}$$

The sign of ω is determined by the direction of $\overrightarrow{t_2}$ relative to the right-hand rule of $\overrightarrow{\sigma_1} \times \hat{t_1}$. The dihedral is not defined if either $\overrightarrow{v_1}$ or $\overrightarrow{v_2}$ is co-linear with $\vec{\sigma}$. We define a special case that constrains the orientation of molecule 1 relative to molecule 2 when both are co-linear with $\vec{\sigma}$ (e.g. as occurs in the clathrin-clathrin interactions). In that case we use the molecule normals to specify a dihedral, such that $\overrightarrow{t_1} = \overrightarrow{\sigma_1} \times \overrightarrow{n_1}$ and $\overrightarrow{t_2} = \overrightarrow{\sigma_1} \times \overrightarrow{n_2}$.

III. Spherical System

For sphere boundaries, which are treated as reflecting, we must adjust reflection off the curved boundary, diffusion on the curved surface (2D), distances when restricted to the 2D surface, and rigid body motion of complexes on the surface. To calculate the association probability between two particles on the sphere, which is a 2D case, the distance between their sites is calculated as the geodesic distance, not the straight-line distance. All the single particles and complexes on the sphere can diffuse and rotate, but the mathematical description is different from the one in solution.

Diffusion of single point particles on the sphere: Here, a single particle is treated as a point that doesn't have a three-dimensional structure or volume. We suppose that single particles will treat the sphere surface as a planar surface. Therefore, the diffusion on the sphere is the projection of the diffusion on a flat surface.

It is well known that the diffusion on a plane is described by two dimensions, Δx and Δy, which follow the Brownian dynamics:

$$\Delta x = \sqrt{2D_x \Delta t} \cdot \Psi_x, \Delta y = \sqrt{2D_y \Delta t} \cdot \Psi_y \qquad \text{Eq. S16}$$

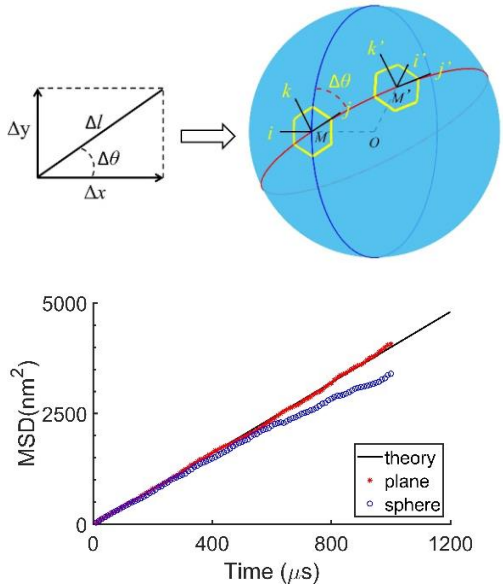
where D_x and D_y are diffusion coefficients, generally $D_x = D_y$, and Ψ_x and Ψ_y are Gauss random numbers, $\langle \Psi \rangle = 0$, $\langle \Psi^2 \rangle = 1$. We transfer them into:

$$\Delta l = \sqrt{(\Delta x)^2 + (\Delta y)^2}, \Delta\theta = \text{acos}(\frac{\Delta x}{\Delta l}) \qquad \text{Eq. S17}$$

where Δl is the displacement distance, and $\Delta\vartheta$ is the movement direction.

The diffusion on the sphere should generate Δl and $\Delta\vartheta$. As shown in the schematic below, M is one point that will diffuse, and O is the center of the sphere, and R is the sphere radius. Firstly, we find the great circle that passes M and the pole points. This great circle is defined as the reference circle (blue circle in the figure). Secondly, rotate this circle around O-M line by $\Delta\vartheta$, we get the new great circle as the trajectory circle (red circle in the figure). Then move M by arc length Δl on the trajectory circle. Through this process, the point M diffuses to

M'. The sign of y is used to decide whether to move up or down the red curve. Our simulation shows that mean squared displacement (MSD) due to diffusion on the sphere matches that expected by theory ($MSD = 4Dt$), up until the trajectories wrap around the sphere. The MSD on the sphere thus drops below the planar diffusion, as the maximum displacement is limited by the geometry.



Top: Illustration of diffusional updates on the sphere. Bottom: For a sphere of R=50nm, D=1nm^2/μs, the mean squared displacement (MSD) on surface matches a plane until the trajectories wrap. 1000 trajectories averaged.

Diffusion of multi-site complexes on the sphere: The diffusion of a rigid complex on the sphere surface is carried out by two steps. First diffuse the center-of-mass (COM) of the complex, then realign the 3D structure of the complex. The diffusion of complex COM is treated as the diffusion of a single particle on the sphere. For example, the point M is the COM of a complex (yellow color in the figure), and following the method above, M moves to M'. To precisely realign the structure, we set up an internal coordinate system of the complex, comprised of three unit-vectors, ***i***, ***j***, ***k***. For example, when the COM of the complex is at M, ***i*** is defined as the normalized vector of $\overrightarrow{OM}$. And ***j*** is defined as the tangent vector of the trajectory circle at point M, and the direction of ***j*** is inclined to M'. ***k*** is defined as $\boldsymbol{i} \times \boldsymbol{j}$. Then every point P in the complex can be expressed of the sum of ***i***, ***j***, ***k***:

$$\overrightarrow{MP} = \alpha \boldsymbol{i} + \beta \boldsymbol{j} + \gamma \boldsymbol{k}. \qquad \text{Eq. S18}$$

When the center-of-mass M moves to M', every point in the complex will also move, such as P moving to P', and the orientation of the internal coordinate system will change into ***i***', ***j***', ***k***'. Since the complex in NERDSS is rigid, we suppose the internal expression of every point in the complex won't change. Thus the new position of P' is:

$$\overrightarrow{M'P'} = \alpha \boldsymbol{i}' + \beta \boldsymbol{j}' + \gamma \boldsymbol{k}'. \qquad \text{Eq. S19}$$

Rotation on the sphere: Due to the constraint of the sphere surface, the complex on the sphere can only rotate around the line of O-M (M is the position of center-of-mass of the complex). Then the rotation is equivalent to a 2-dimentional rotation on a plane.

IV. Reaction rates for molecules with identical, repeated sites

When molecules have identical, repeated sites, such as clathrin (three trimer legs), they must be distinguished by distinct labels, as they have distinct coordinates in space. This labeling results in binding between both self and distinct sites: Cla(c1, c2, c3) can interact with another molecule using all possible pairs, Cla(c1)+Cla(c1)->, or Cla(c1)+Cla(c2)->, or Cla(c2)+Cla(c2)->, etc, with 6 distinct possibilities. The energetics of each of these reactions is (in general) all the same. Importantly, we note this means that the K_D value for distinct reaction sites must be twice as strong as for the self sites, $K_D^{c1c2} = K_D^{c1c1}/2$, i.e. the rates are such that $k_{on}^{c1c2} = 2k_{on}^{c1c1}$.

Only then will the kinetics and equilibrium of the N trimer assembly (excluding all spatial effects), match the behavior of 3*N independent identical sites, with K_D=K_D^{c1c1}. The user must directly specify this increase by a factor of 2, as there are other reactions, e.g. A(a)+A(b)->, where the specific rate is for two truly distinct sites (e.g. actin(barbed)+actin(pointed)->). This modification is performed automatically in NFsim[1] for these reaction types by dividing the user input rates for self sites by 2, and leaving the rates for distinct sites at the input values (we note that this results in a K_D that is half that expected from the input rates). For NERDSS, it is the user's specification when these indices are only distinct due to labeling, and thus the user must introduce the factor of 2 increase to distinct sites in this case. This result was previously validated for a structure-resolved clathrin model[2].

To derive this, we consider a self-binding reaction A(a)+A(a), with equilibrium K_D and initial concentration A_0. The equilibrium value is $A_{eq} = \frac{-K_D}{4} + \frac{\sqrt{K_D^2/4+2A_0K_D}}{2}$. If we now evenly split the sites into *n* distinct labels, we have that $nA_{i0} = A_0$, and at equilibrium, we want that $nA_{ieq} = A_{eq}$, where *i* indexes each of the labels. We know that $A_{i0} = A_{ieq} + 2AA_{iieq} + \sum_{j,j\neq i}^{n} AA_{ijeq}$ and $A_{ieq}^2 = K_D AA_{iieq}$. We then must specify that $A_{ieq}A_{jeq} = \frac{K_D}{2} AA_{ijeq}\ (j \neq i)$ to recover as desired that $A_{ieq} = \frac{1}{n}\left(\frac{-K_D}{4} + \frac{\sqrt{K_D^2/4+2nA_{i0}K_D}}{2}\right) = \frac{A_{eq}}{n}$. We note the equilibrium copy numbers thus generate the distinct products [Cla(c1!1).Cla(c2!1)] twice as numerous as self products [Cla(c1!1).Cla(c1!1)], as order does not matter in listing bound reactants.

Supporting Table

Systems	3D *(A+B)*	$3D \rightarrow 2D$ *(A+B)*	2D *(A+B)*	Self *(A+A)*
Relation to 3D microscopic binding rate	k_a^{3D}	$k_a^{3D\rightarrow 2D} = 2k_a^{3D}$	$k_a^{2D} = k_a^{3D}/h$[a]	$k_a^{self} = 2k_a^{3D}$
Microscopic dissociation rate	k_b	k_b	k_b	k_b
Macroscopic binding rate	$k_{on} = \left(\frac{1}{k_a^{3D}} + \frac{1}{4\pi\sigma D}\right)^{-1}$	$k_{on} = \frac{1}{2}\left(\frac{1}{k_a^{3D\rightarrow 2D}} + \frac{1}{4\pi\sigma D}\right)^{-1}$	$k_{on} = f(k_a^{2D}, \sigma, D, b_\rho)$[b]	$k_{on} = \frac{1}{2}\left(\frac{1}{k_a^{self}} + \frac{1}{4\pi\sigma D}\right)^{-1}$
Macroscopic dissociation rate	$k_{off} = k_{on}\, k_b / k_a^{3D}$	$k_{off} = 2k_{on}\, k_b / k_a^{3D\rightarrow 2D}$	$k_{off} = k_{on}\, k_b / k_a^{2D}$	$k_{off} = 2k_{on} k_b / k_a^{self}$
Equilibrium state	$K_D = k_b / k_a^{3D}$	$K_D = 2k_b / k_a^{3D\rightarrow 2D}$	$K_D = k_b / k_a^{2D}$	$K_D = 2k_b / k_a^{self}$

Table S1. Relationships between microscopic and macroscopic reaction rates. The proper relationship $K_D = k_{off}/k_{on}$ is preserved for all binding reactions. The user always inputs values of k_a^{3D} to the software (or k_{on}). The top row rates are what are input to the Green's function to evaluate binding probabilities, but the software takes care of the conversion from user inputs to these values. [a]The parameter *h* is a lengthscale that by default is set to 2σ. [b]See Ref [3].

Supporting Figures

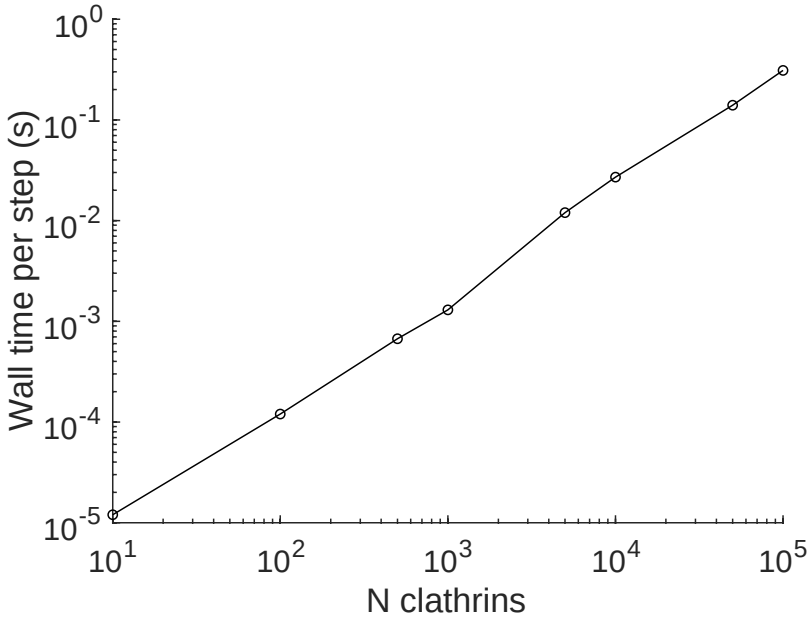


Fig S1: Wall time vs system size for the clathrin assembly is linear. The system here has K_D=1μM, k_{on}=1x10^{6}M^{-1}s^{-1}, k_{off}=1s^{-1}, and a fixed concentration of [Cla]=1.39 μM. The volume is systematically increased with increasing *N*. Exact wall time per system depends on reaction parameters, and diffusion. For smaller diffusion, the system can be partitioned into more sub-volumes, which accelerates pair-wise evaluations. For faster, stronger reactions, fewer sites are left free in solution, generally meaning fewer pairwise evaluations need be evaluated. Timing performed on a Dell workstation running Linux with Intel Core i9-9980XE, with Gnu compiler g++ version 7.4. Times are ~20% faster with Intel compilers.

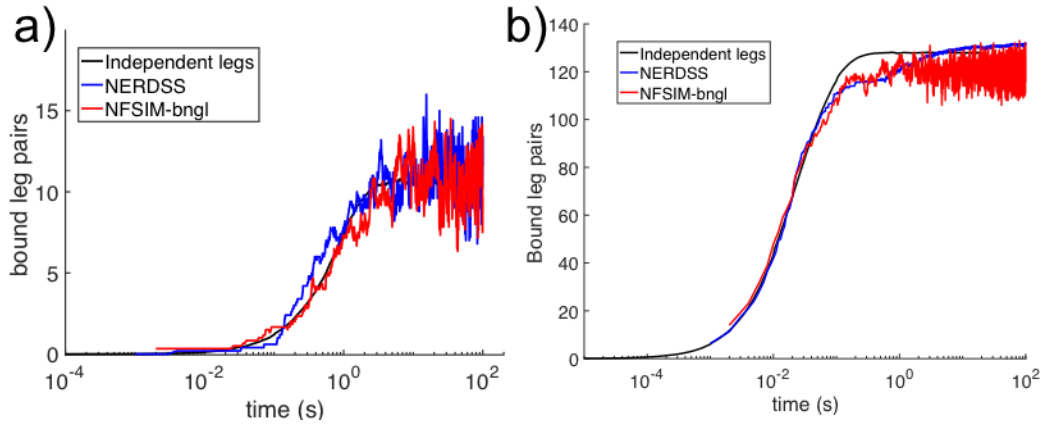


Fig S2: **NERDSS clathrin assembly simulations in solution can be compared with non-spatial stochastic simulations using NFsim software**[1]. a) K_D=100μM, k_{off}=1s^{-1} b) K_D=0.2μM, k_{off}=1s^{-1}.

V=(0.494μm)^3, N_{cla}=100, D_{cla}=13μm^2/s. Blue are NERDSS simulations. Black is the solution with all 300 legs independent—thus no lattice structure forms. Red is solved using the NFSim software package encoded using BNGL syntax. For K_D=100μM, only inter-molecular reactions are specified, that is, only bimolecular association between distinct binding sites. For K_D=0.2μM, intra-molecular reactions are also specified, using the same magnitude of rates as for bimolecular association. Because NFSim is non-spatial, no hexameric structure can be enforced, each clathrin simply has three sites. NFSim is a single trajectory, the NERDSS simulations are averaged over 3-5 trajectories. For NERDSS, f=0.001 (also see Fig S3).

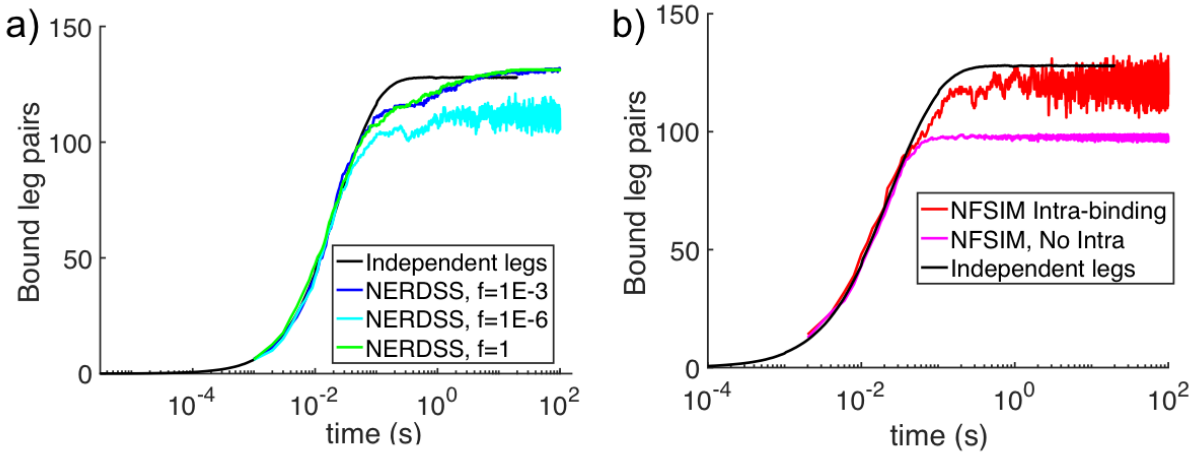


Fig S3. Loop closure events in rate-based models impact total numbers of 'bonds' formed. a) NERDSS simulations of clathrin assembly in solution with varying loop closure factors $f = \exp(-\Delta G_{coop}/k_B T)$. Simulations were performed with leg dimerization strengths of K_D=2 x10^{-7} M, k_{on}=5 x10^6 M^{-1}s^{-1}, k_{off}=1s^{-1}, V=(0.494μm)^3, N_{cla}=100, D_{cla}=13μm^2/s, $D_{R,cla}$=0.03 rad^2/s, Δt=0.2μs. Until the loop factor is comparable to $\exp(-\Delta G_{AB}/k_B T)$, (for this reaction 2 x10^{-7}), loops that form are difficult to break. For comparison, we simulated 300 independent legs, where no lattice forms and all events are dimerization. b) In non-spatial NFSim[1] simulations, the structural geometry is not specified (no hexagonal lattice), but sites that are within the same complex can form intra-molecular bonds, which are effectively loop closure events. For purely inter-molecular, or bimolecular association, the bound pairs plateaus relatively low, and becomes independent of K_D (data not shown), due to formation of a single component, where no internal links are allowed. With internal or intra-molecular events, here defined with the same magnitude rate, more events can occur. For non-spatial models, aside from each clathrin having 3 sites, no geometric or spatial constraints limit where loop closure occurs.

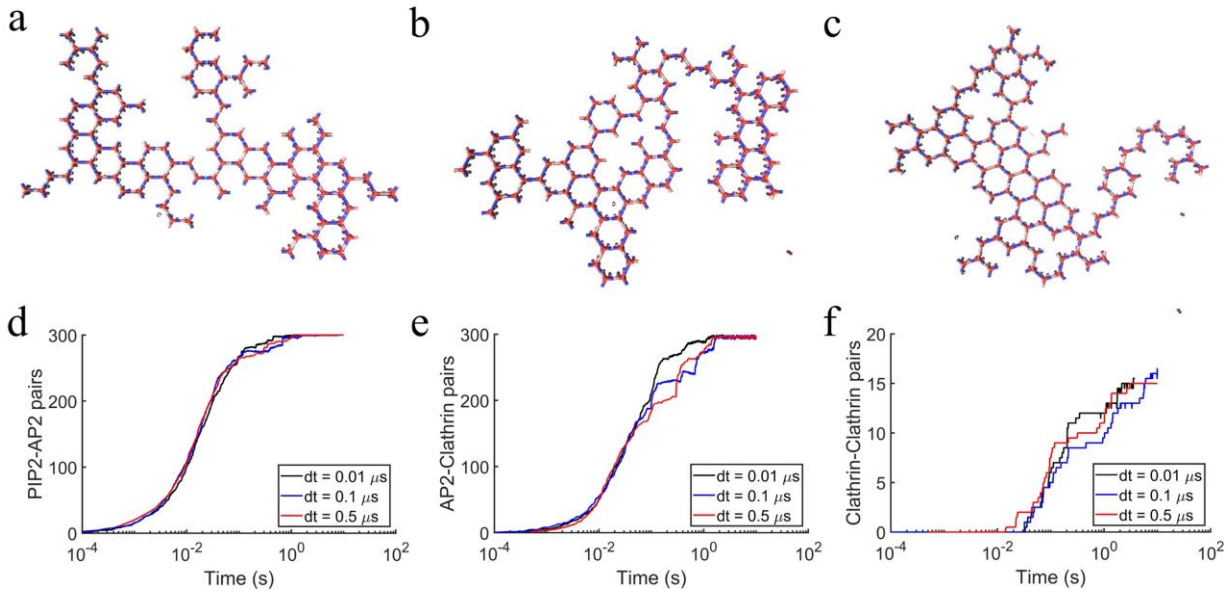


Fig S4. Clathrin assembly on the membrane surface is independent of time-step. Verification that when clathrin forms assemblies in 2D, the kinetics and structures are the same as Δt changes. Example structures from a) Δt=0.01μs, b) Δt=0.1μs, c) Δt=0.5μs. Kinetics of d) recruitment of adaptor protein (labeled AP2) to membrane lipids, e) Binding of clathrin to the adaptor protein, f) binding of clathrin to clathrin. Each clathrin has a reflecting site at its center, to prevent fully bound clathrin from not 'seeing' each other. Each trace is a single trajectory.

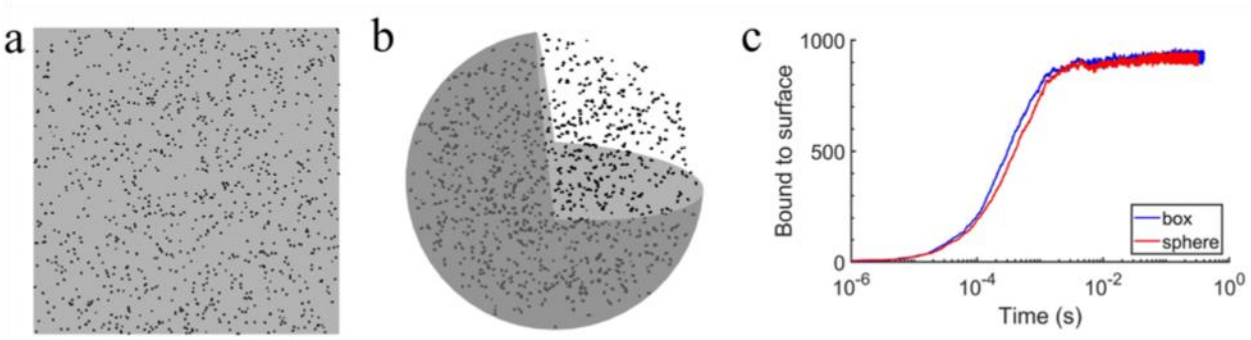


Fig S5. Reversible recruitment of interacting protein pairs to a surface in a spherical system. This model has two species A(a,m) and M(m), and 6 reactions: 1) A(a)+A(a) ⇌A(a!1).A(a!1), 2) A(m)+M(m) ⇌A(m!1).M(m!1), 3) A(a!1).A(a!1,m)+M(m) ⇌A(a!1).A(a!1,m!2).M(m!2), 4) A(a,m!1).M(m!1)+A(a) ⇌A(a!2).A(a!2,m!1).M(m!1), 5) A(a,m!1).M(m!1)+A(a,m!2).M(m!2) ⇌ M(m!1).A(a!3,m!1).A(a!3,m!2).M(m!2) and 6) A(a!1).A(a!1,m!2).M(m!2)+M(m) ⇌ M(m!3).A(a!1,m!4).A(a!1,m!2).M(m!2). The last two reactions are in 2D. The kinetics of binding to the surface of a) a box vs b) a sphere, with the same V/A ratio. Lipids are modeled using the implicit lipid model. c) Number of solution particles bound to the surface agrees well. Sphere R=100nm, Box=[354.4908, 354.4908, 33.333]nm, k_a=8.3056nm^3/μs, k_b=1000s^{-1} for both A+A and A+M binding reactions. σ=1nm. N_A=1000, N_{IL}=2000, D_A=12nm^2/μs, D_{IL}=1nm^2/μs.

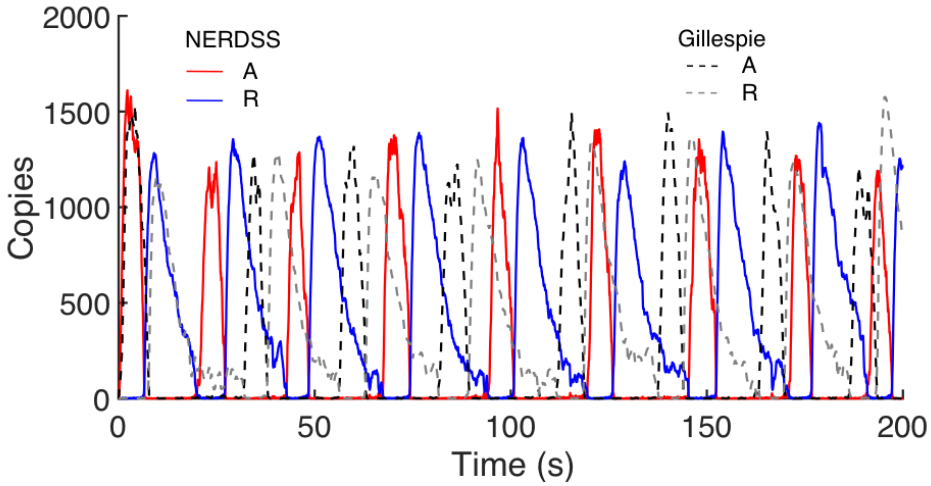


Fig S6: NERDSS simulations of the circadian clock model[4] agree with Virtual Cell[5] simulations. Virtual Cell simulations were solved using either deterministic partial differential equations (Fig 3 main text) (PDE) or stochastic non-spatial algorithms (Gillespie). To quantitatively compare stochastic simulations, which are noisy, we calculate the average period of oscillations for A and R over 200 s. For PDEs (or ODEs—data not shown) the oscillation times are 25.1s and 25.1s, with a lag-time of 6.55s. For Gillespie, averaged over 10, 200s trajectories, the times are 24.8 and 25.1s, with a lag-time of 6.63s. For NERDSS, we have comparable results, indicating that the system remains well-mixed and largely rate-limited, with times of 24.5s and 24.75s for a 200s trajectory, and lag-time of 6.45s. We also ran a set of simulations with increased diffusion from 10 to 20 $\mu m^2/s$, which alters the intrinsic rates for bimolecular reactions, and then measured times of 25.1, 25.2, and 6.4 over 325s. To calculate these times, we performed a discrete Fourier transform of the signals (the signal was zero-padded out to 5000s to increase resolution of the frequencies). The reported oscillation time is the maximum amplitude frequency observed. The lag-time was evaluated using cross-correlation, again reporting the time with maximum amplitude. MATLAB was used. Similar results are obtained by simply finding the peaks, calculating the time-separation between them, and averaging.

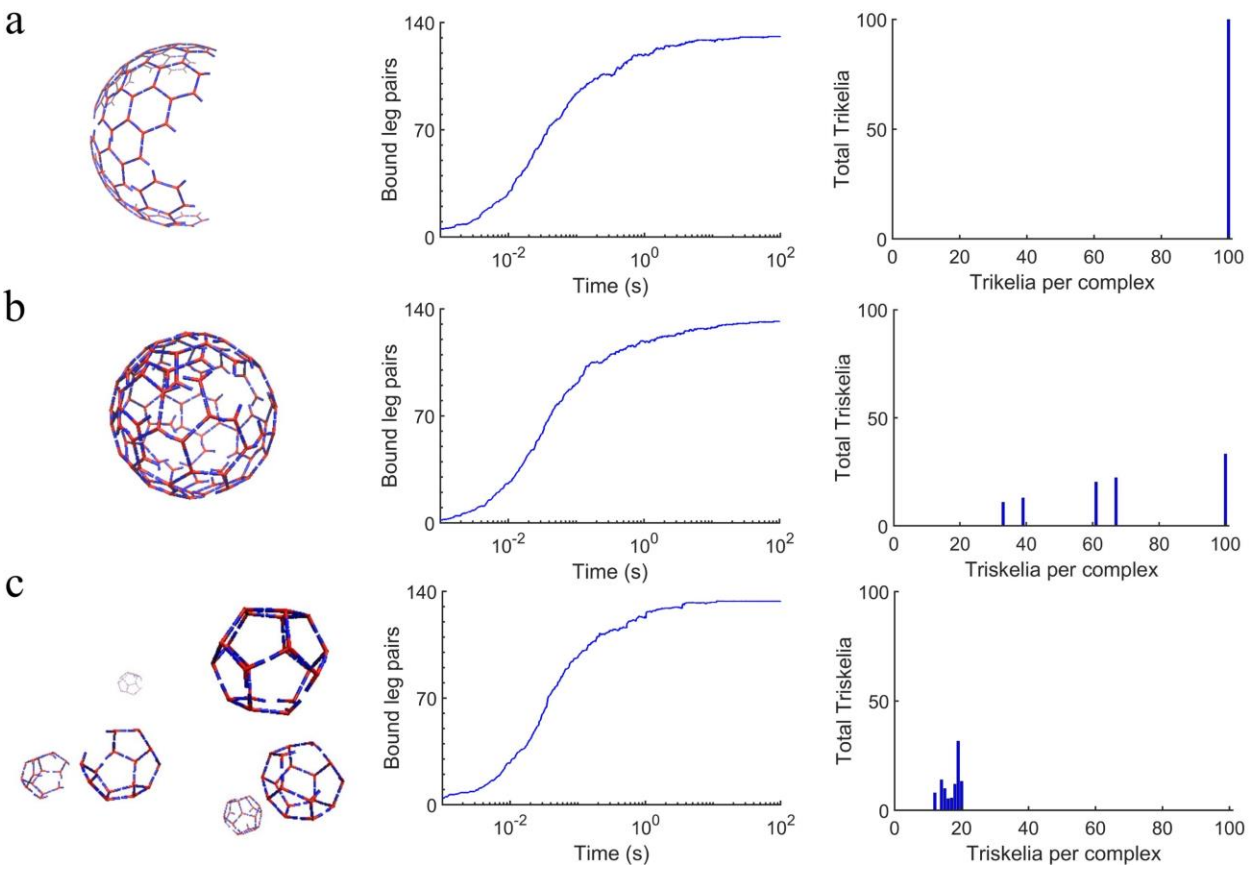


Fig S7. Clathrin assembly with varying pucker angles. Flat planar clathrin has a pucker angle of 90°. Here we increase it to a) 95° b) 100° c) 110°. As the angle increase, the curvature of the cages increases and accommodates fewer trimers. The largest can accommodate ~500 trimers, the smallest ~20, which orient into pentagonal faces not hexagonal faces, despite binding with the same angles. For simulations here and in S8-S10, V=(0.494μm)^3, N_{cla}=100, D_{cla}=13μm^2/s, $D_{R,cla}$=0.03 rad^2/s, Δt=0.2μs. K_D=0.2μM, l_{leg}=7.5 nm, f=0.001, bindRadSamCom = 5nm, overlapSepLimit = 7nm.

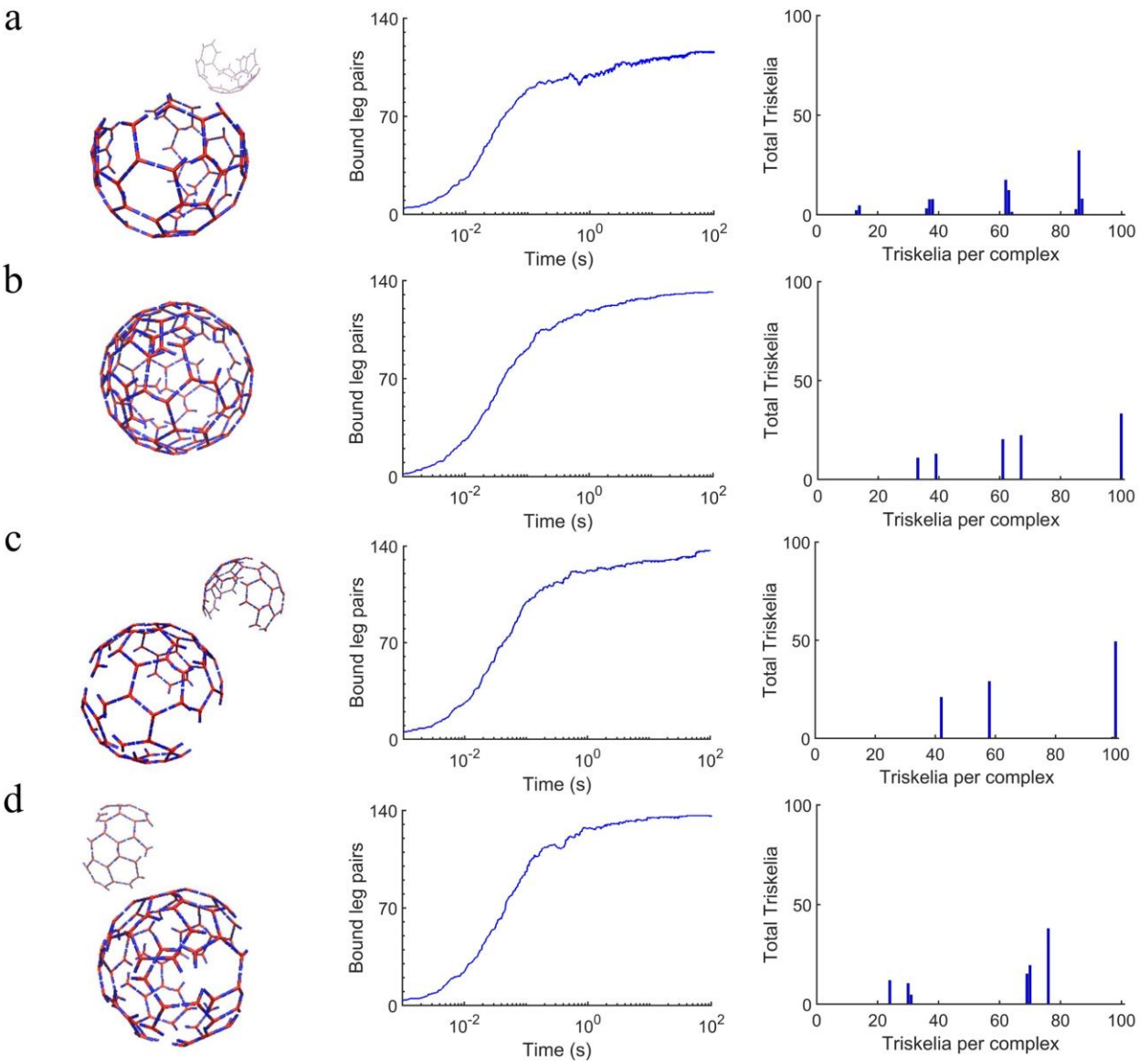


Fig S8. Clathrin cage simulations can form defects that can still stabilized interactions. By allowing free leg sites that are in close by not perfect contact to still form 'bonds', the cages are stabilized further against dissociation. This parameter, bindRadSameCom, is increased to allow bond formation at longer distances, from cutoffs at a) 3nm b) 5nm c) 8nm d) 10nm. When the distances are relatively large compared to the molecule size (l_{leg}=7.5 nm), the bonds are less physical, and prevent recruitment of additional trimers. Pucker is 100°, other parameters match Fig S7.

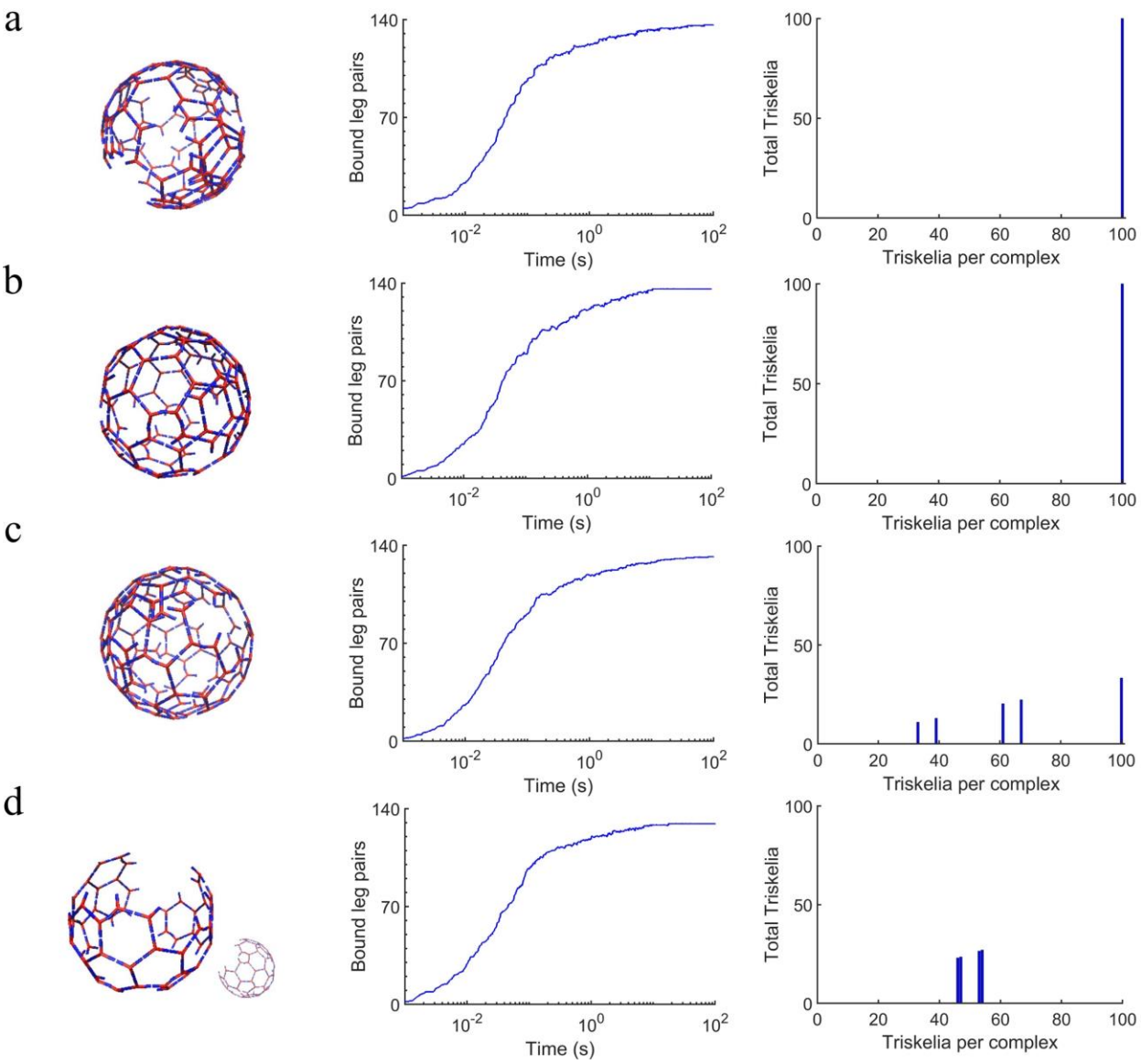


Fig S9. Clathrin cage simulations can form defects, and evaluating steric overlap impacts structure. When binding events occur between clathrin, the events are rejected if components of the two complexes have steric overlap, which is true if any COM-COM distances are less than overlapSepLimit. As this limit is increased from a) 1.5nm b) 5nm c) 7nm d) 10nm, more association moves with any trimer overlap are rejected, resulting in more prevention of defects and fewer trimers per cage. Pucker is 100°, bindRadSameCom=5nm, other parameters match Fig S7.

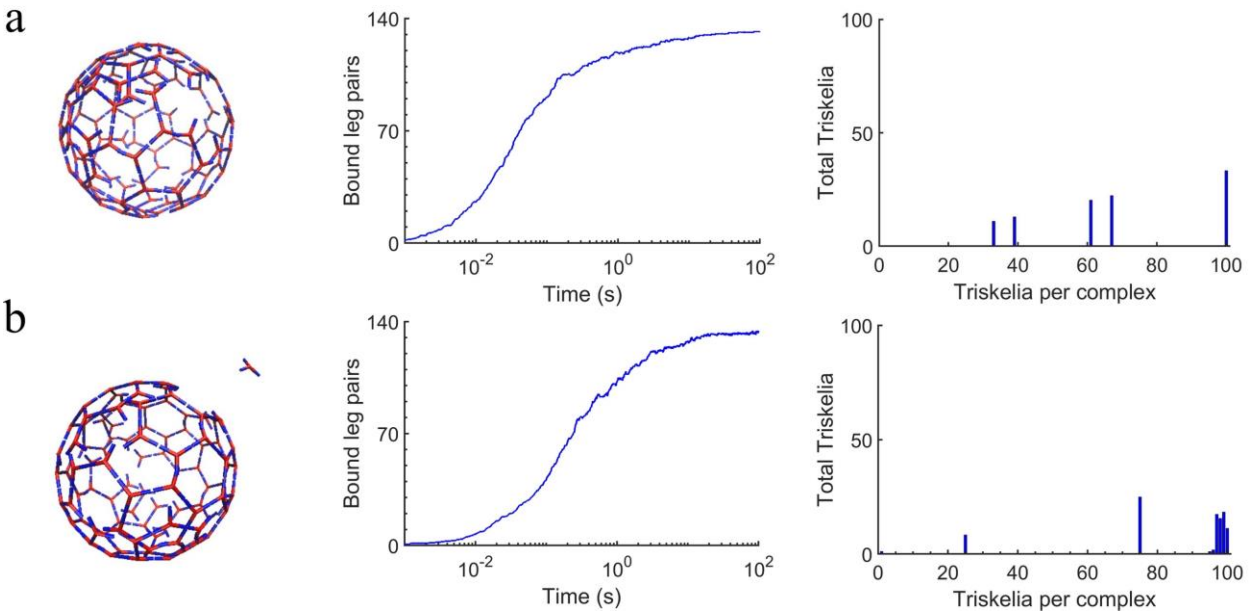


Fig S10. Clathrin cages nucleate once K_D is sufficiently strong. Here we compare results for a) K_D=0.2μM. b) 1μM. Parameters are the same as Fig S7. Here, we note the K_D in (b) is the same as in Fig 2d from the main text. The final equilibrium of bound leg pairs is higher here because we have increased *f* to 0.001, from $5x10^{-6}$, resulting in more stable loops.

SI References

1 Sneddon, M. W., Faeder, J. R. & Emonet, T. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nature methods* **8**, 177-U112 10.1038/Nmeth.1546 (2011).

2 Johnson, M. E. Modeling the Self-Assembly of Protein Complexes through a Rigid-Body Rotational Reaction-Diffusion Algorithm. *J Phys Chem B* **122**, 11771-11783 (2018).

3 Fu, Y. *et al.* An implicit lipid model for efficient reaction-diffusion simulations of protein binding to surfaces of arbitrary topology. *J Chem Phys* **151**, 124115 10.1063/1.5120516 (2019).

4 Vilar, J. M., Kueh, H. Y., Barkai, N. & Leibler, S. Mechanisms of noise-resistance in genetic oscillators. *Proc Natl Acad Sci U S A* **99**, 5988-5992 10.1073/pnas.092133899 (2002).

5 Moraru, I. I. *et al.* Virtual Cell modelling and simulation software environment. *Iet Syst Biol* **2**, 352-362 Doi 10.1049/Iet-Syb:20080102 (2008).