UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Reaction-Diffusion Agent Based Models of Nucleocytoplasmic Transport

Permalink

https://escholarship.org/uc/item/58v3r4dq

Author

Azimi, Mohammad

Publication Date

2013

Peer reviewed|Thesis/dissertation

Reaction-Diffusion Agent Based Models of Nucleocytoplasmic Transport

By

Mohammad Azimi

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy

with University of California San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mohammad R. K. Mofrad, Chair
Professor Donna Hudson
Professor Karsten Weis

Fall 2013

Abstract

Reaction-Diffusion Agent Based Models of Nucleocytoplasmic Transport

by

Mohammad Azimi

Joint Doctor of Philosophy with University of California, San Francisco
in Bioengineering

University of California, Berkeley

Professor Mohammad R. K. Mofrad, Chair

Selective yet efficient transport between the cell nucleus and cytoplasm is critical to cellular function as the nuclear pore complex is a major point of regulation for gene expression, signal transduction, immune response, oncogenesis and viral propagation. The combined physical structure of the pore, biochemical interaction of transport factors with pore components and the presence of the cellular milieu create conditions under which selective and efficient nucleocytoplasmic transport can occur. In this dissertation, I explore the significance of structure and interaction on transport of globular proteins and polymeric mRNA cargo through the crowded pore. In the first part of this work, an agent based modeling software framework is developed and validated for accurately simulating discrete and stochastic reaction-diffusion systems. A simulation environment representing the structure of the nuclear pore complex along with rules for the dynamics of protein movement and interaction was created using *in vivo and in vitro* reported parameters. This setup was then used to perform *in silico* experiments on the role of pore-cargo affinity in optimizing transport efficiency. These experiments demonstrate the pore's sensitivity to cargo affinity in maintaining efficient transport and suggest that a higher affinity binding site at the side of the pore where transport is terminated increases efficiency by reducing futile shuttling of cargo complexes. In the final part of this dissertation, I extend my agent based modeling framework to look at aspects of mRNA export that have remained unaddressed in experimental works. Variations in the number and spacing of transport receptors bound to the mRNA are shown to play a critical role in transport efficiency. In these experiments, a single transport receptor at the 5' end appeared insufficient for facilitating export. Increasing transport receptor coverage along the length of the mRNA improved the chances of successful export. Additionally, it was observed that the presence of a transport receptor near either the 5' or 3' terminus is required for successful export as it likely promotes the emergence of a favorable threading conformation. Finally, it was observed that the use of a single fluorescent tag to track and report mRNA export time, as is standard in current experimental work, is likely to underestimate true transport times. These findings have implications in the design of targeted delivery and export of polymeric molecules into and out the nucleus.

To my son, Zayn Azimi.
Stay curious and seek knowledge and you'll never stop growing.

Table of Contents

Acknowledgements

Section I:

The Nuclear Pore Complex – Structure and Function

Introduction to the Nuclear Pore Complex

The characteristic feature that sets eukaryotic cells apart from prokaryotic cells is the compartmentalization of most of the cell's genetic material in the nucleus – providing increased genomic stability and more efficient genome regulation [1]. This compartmentalization is achieved via the nuclear envelope (NE), a double lipid bilayer that derives its shape and mechanical stability from a two-dimensional meshwork of intermediate filaments termed the nuclear lamina [2]. The lamina is anchored to the envelope via membrane proteins and the nuclear pore complex (NPC). As one of the largest protein complex within a cell, the NPC penetrates the nuclear envelope, establishing a selective yet efficient connection between the cell cytoplasm and nucleoplasm (see Fig. 1). The structure and function of this pore is evolutionarily conserved across distantly related eukaryotes [3].
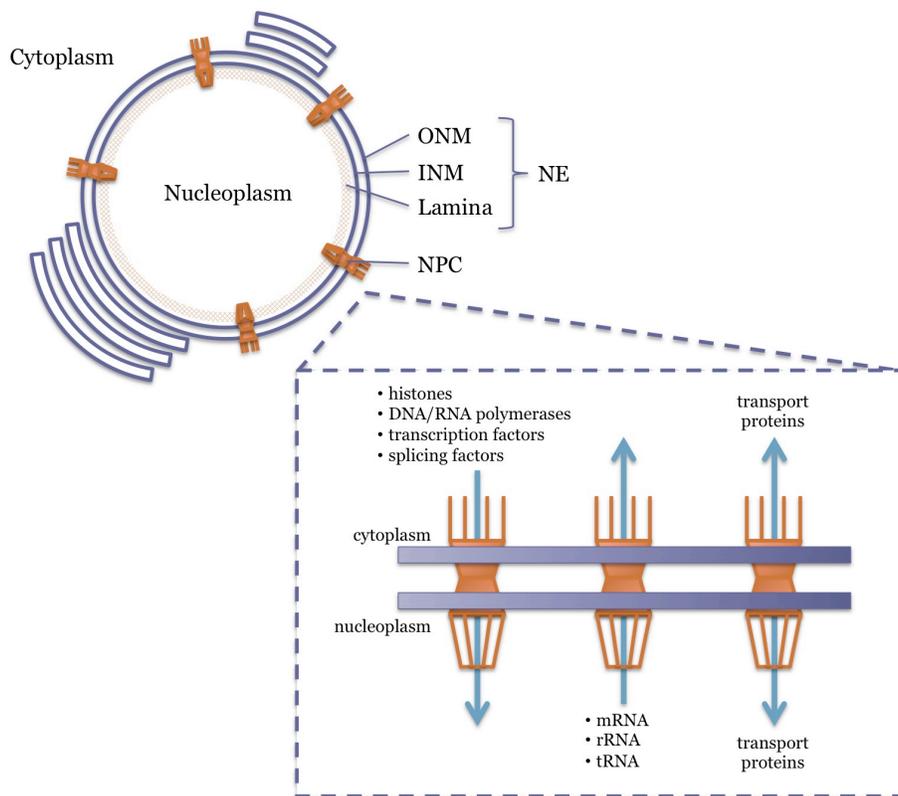
Figure 1. A schematic representation of the cell nucleus and its components. The nucleus is composed of a double lipid bilayer termed the nuclear envelope (NE) – consisting of an outer nuclear membrane (ONM) and inner nuclear membrane (INM) – which derives its shape and structure from the lamina. The NE isolates the majority of the cell's genetic material within the nucleoplasm, while nuclear pore complexes (NPCs) connect the nucleoplasm to the cytoplasm and regulate selective transport between compartments.

Nuclear Pore Function

By connecting the cytoplasmic and nucleoplasmic compartments, the NPC links the two major steps required for gene expression – transcription of DNA, which occurs within the nucleus, to translation of mRNA in the cytoplasm – leading to the synthesis of proteins. Sustaining the process of gene expression requires rapid bidirectional transport of proteins responsible for genomic structure and function (histones, DNA/RNA polymerases, transcription factors and splicing factors) from the cytoplasm to the nucleus, the export of RNAs (mRNA, rRNAs and tRNAs) from the nucleus to the cytoplasm and a continual shuttling back and forth of proteins that play an integral role in transport between the cytoplasm and nucleus [4].

Molecules residing in the cytoplasm and nucleus can be categorized into one of three cargo types based on their size and the presence of signaling sequences or patches that determine their likelihood of transport. The first category of cargo are comprised of molecules that are less than ~40 kDa (~5–9 nm in diameter), termed passive cargo [5-7]. These molecules are capable of passive transport across the NPC, relying solely on diffusive motion to overcome the barrier. Larger molecules (~40 kDA to ~25 MDa or up to ~40 nm in diameter), fall within two categories based on the presence or absence of import/export signals. Cargos that possess either nuclear localization signals (NLS) or nuclear export signals (NES) are capable of binding nuclear transport receptors (NTRs) and traversing the pore; these cargos are termed active or facilitated cargo. The third category of cargo is termed inert cargo as they lack signaling sequences that attract NTRs. As a result of their large size and lacking affinity for NTRs, inert cargo are excluded from entering the pore and traversing the nuclear envelope. A more detailed explanation of NLS/NES interaction with NTRs and subsequent cargo complex formation and transport is provided later in this chapter.

Transport Kinetics

Although the NPC serves as a barrier to large inert cargo, it is capable of selective yet rapid transport of active cargo between compartments. *In vivo* experiments of transport kinetics using the classical import pathway suggest transport rates of approximately 1000 molecules per NPC per second, corresponding to a mass of ~100 MDa per NPC per second [8] – approximately the mass of the pore itself. It may be tempting to extrapolate these transport rates to transport times, ie. 1 ms transport time per molecule; however, single molecule tracking studies of the same pathway have shown that transport events last approximately 10 ms, suggesting that the pore is capable of simultaneous transport of multiple cargos in parallel through a single pore [5]. It should be noted that these transport times vary by cargo size [5,8-12] and reach values of ~100 ms for typical mRNA export and on the order of ~1000 ms for very large mRNA molecules [13-15].

3

Permeabilities for passive cargos have been quantified and observed to decrease logarithmically as a function of the cargo's Stokes radius, ranging from an influx rate of ~1 s^{-1} for small cargos (< 1nm Stokes radius) to ~10^{-4} s^{-1} for larger passive cargos (~2.5 nm Stokes radius) [6]. The permeability of cargos much larger than these in size lacking affinity for transport receptors is negligible (inert cargo), while the permeability of cargos with the capacity to recruit transport receptors or transport receptors themselves are much higher than the permeability of similarly sized inert cargos. Two examples are the transport receptors NTF2 (~30 kDa) and Importin-β (~100 kDa) which have permeabilities of ~1 s^{-1} and 0.4 s^{-1} respectively [16,17].
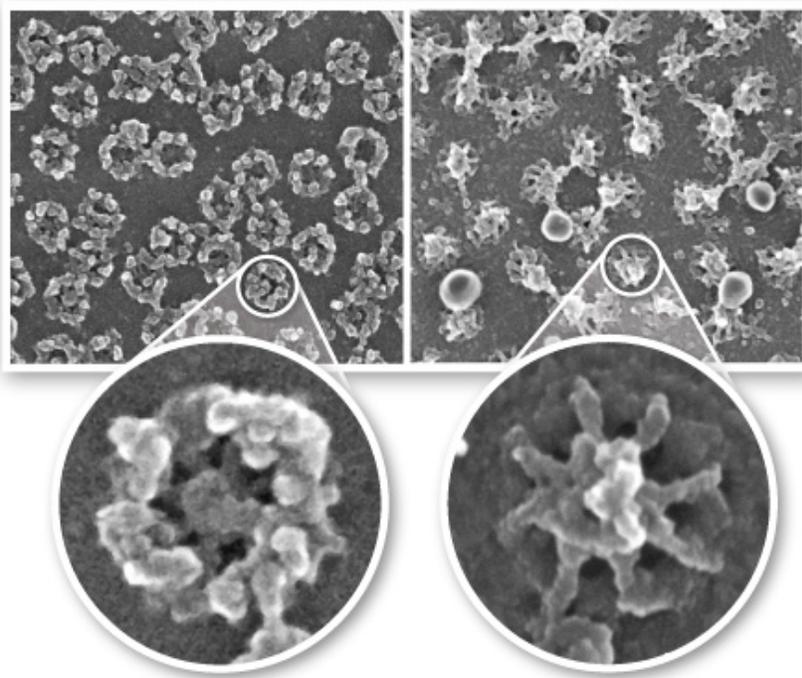
Figure 2. Scanning electron micrographs of *Xenopus* oocytes as seen from outside the nucleus (left) and the nuclear interior (right). Figure adopted from [18].

Finally, the overall kinetics of transport into and out of the nucleus are not only governed by the rate of transport and permeability through each individual pore, but are also a result of the total number of pores present on the surface of a nucleus. The number of NPC per nucleus varies greatly between species and cell cycle and is not simply a function of nucleus size. Across species, *S. cerevisiae* have been observed to have on the order of ~100 pores per nucleus [19], while human nuclei have several thousand pores per nucleus [8,20,21] and an astonishing ~5 x 10^7 pores per nucleus have been reported in a mature *Xenopus* oocyte [22]. Even within species, the number of pores fluctuates over the life of the cell. For example, in *S. Cerevisiae*, it was observed that nucleus contains the greatest number of pores, an average of 142 pores per nucleus, during early mitosis when the nucleus' volume is at its peak (as opposed to 86 pores per nucleus during G1). However, the pore density appeared to peak during S-phase and reached a minimum during late

anaphase, suggesting that the nucleus does not have a fixed pore density and nucleus size does not regulate the number of pores that are present [19]. Scanning electron micrographs of the nuclear envelope from the cytoplasmic and nucleoplasmic sides are shown in Fig. 2 to illustrate the distribution and coarse structure of the pores.

Links to Human Health and Disease

Serving as the physical link between the locations of transcription and translation within the cell, the NPC serves as a major point of regulation for gene expression, signal transduction, immune response, oncogenesis and viral propagation. Subsequently, nucleocytoplasmic transport has been identified as a key player in regulating human health and disease and is an emerging target for new therapies.
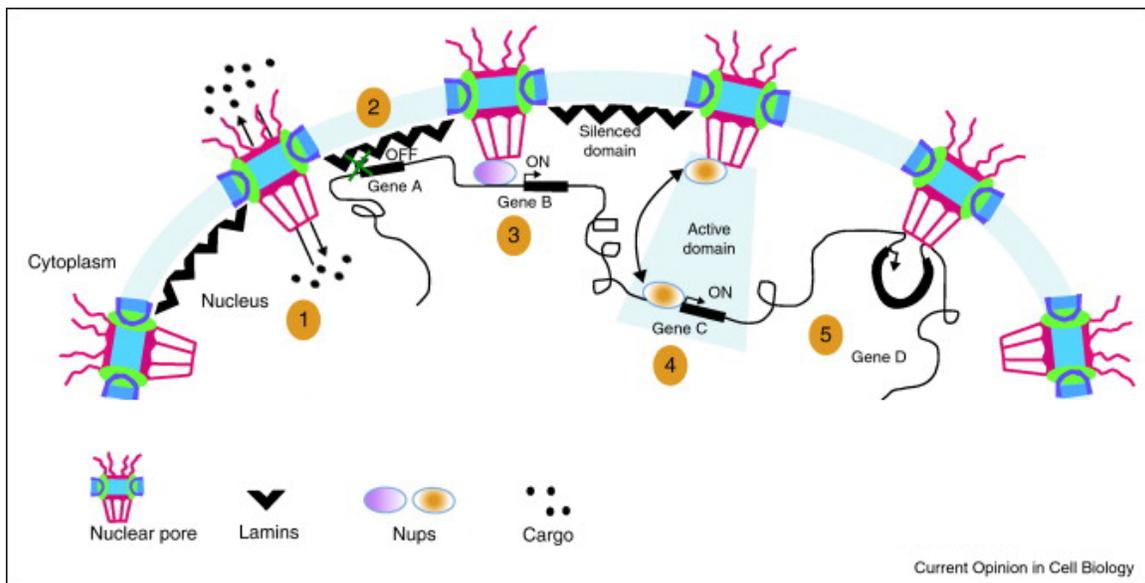
Figure 3. The nuclear pore complex demonstrates a regulatory role that extends beyond import and export of cargo to and from the nucleus (1). Genes localized to the nuclear lamina are typically silenced (2), while genes bound to structural (3) or mobile components (4) of the NPC are actively transcribed. Tethering of gene loops to the nuclear basket has been shown to facilitate 'transcriptional memory' (5). Figure adapted from [23].

Spatial localization and proximity of genes to nuclear pores has been shown to affect gene expression [23,24]. Studies in yeast and higher eukaryotes have shown that heterochromatin – highly condensed, gene-poor and transcriptionally silent regions of chromosomes – preferentially associate with the nuclear periphery and that the binding of genes to the nuclear lamin generally promotes transcriptional silencing whereas localization to the nuclear interior allows for transcription [25-28]. Similarly, the gene-gating hypothesis proposed by Günter Blobel suggested that

NPCs serve as gene-gating organelles that associate preferentially with actively transcribed genes [29]. Recent findings have confirmed that components of the NPC associate preferentially with promoters of active genes both near the pore and within the nucleoplasm [23,30]. Additionally, tethering of genes to NPCs has been reported to facilitate 'transcriptional memory' which results in efficient and repeated transcription of genes within a gene loop [23,31]. This interplay between NPC components and the genome (illustrated in Fig. 3) demonstrates the regulatory role of the pore in gene expression that extends beyond simple import and export events.
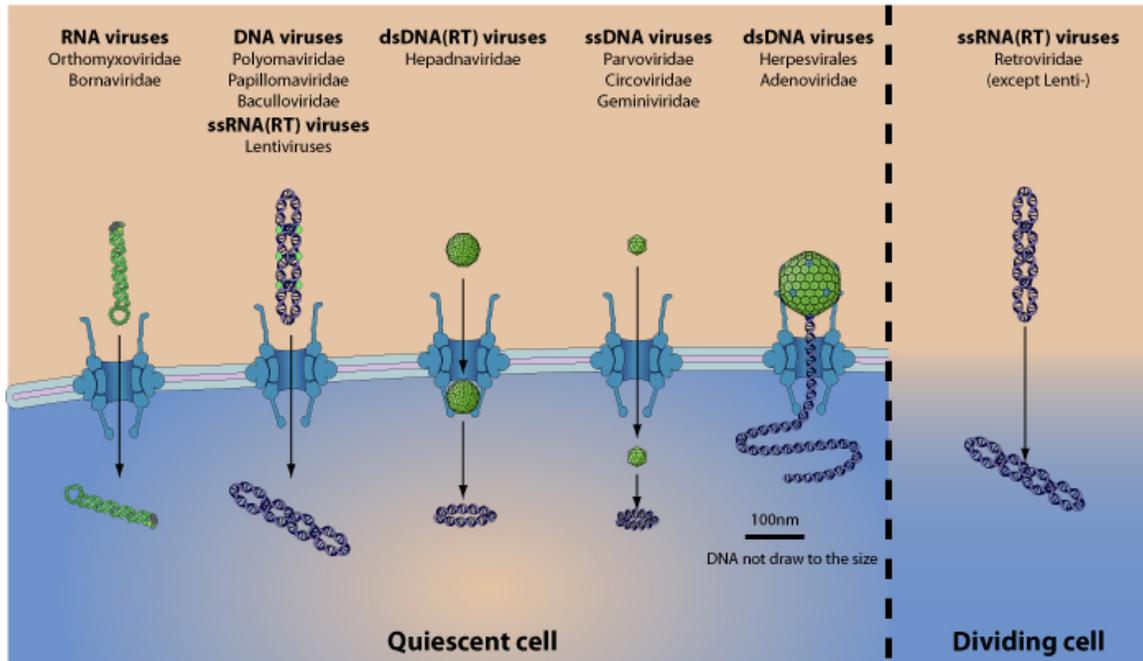
Figure 4. Most viruses infect the cell during the quiescent state and subsequently rely on the NPC for the transfer of viral genetic material to the nucleus. The mechanism of transport varies by virus type, presence of a capsid and size of capsid as shown here. Figure adopted from [32].

While the healthy function of cells requires ongoing signal transduction across the nuclear envelope [33] and interactions between nuclear pores and the genome that aid in gene expression, it has been shown that abnormalities in the composition of the NPC and its components lead to a multitude of human diseases including heart disease [34], reduced immune response [35] and neurodegenerative disorders [36-38]. Additionally, mRNA export dysregulation and aberrant expression of export factors have been linked to breast, lung, ovarian, prostate, cervical, pancreatic, colon, lymphatic and brain cancers [39,40]. Finally, most DNA viruses and some RNA viruses leverage the cell nucleus' DNA replication, transcription and RNA processing machinery for viral replication. Although most retroviruses wait for nuclear envelope disassembly during mitosis to gain access to this nuclear machinery, most other viruses gain access via transport through the nuclear pore

complex. Transport through the nuclear pore complex is associated with various levels of viral capsid disassembly as illustrated in Fig. 4. While some viruses enter the NPC without a capsid by leveraging transport receptors (RNA, DNA, ssRNA(RT) viruses), others traverse the pore with intact capsids (ssDNA) and others bind their capsid to the NPC and release their genetic material into the nucleus (dsDNA). Similarly, export of viral material from the nucleus takes on different forms (ssRNA, RNA, ssDNA and dsDNA) and may involve a capsid or be capsid-free. [32,41] As a result, the NPC presents itself as a point of regulation for import and export of viral material that may be a potential target of therapeutic interventions.

Nuclear Pore Structure

The NPC is a ~100 MDa macromolecular assembly of approximately 30 different proteins termed nucleoporins (Nups). The dimensions of the pore have been determined via cryo-electron tomography to be ~125 nm in diameter near the peripheries and ~55 nm in diameter at the center, with an overall height of ~95 nm [42]. The ~30 proteins that compose the NPC exist in several copies, each in multiples of eight as a result of the pore's eight-fold rotational symmetry. The constitutive Nups compose five distinct regions of the NPC. The first two regions consist of the membrane layer composed of coat nucleoporins and integral membrane proteins of the pore membrane domain (POMs) that anchor the NPC to the nuclear envelope and the scaffold layer composed of adapter nucleoporins that connect the membrane layer to the channel layer and give the pore its structure. The third region is the FG-Nups layer consisting of channel nucleoporins. While the membrane and scaffold layers consist of highly structured proteins, mostly composed of α-helices and β-propellers, the channel layer is composed of nucleoporins containing regions rich in phenylalanine-glycine residues (FG-repeats) that are natively unfolded and are thought to play a central role in the pore's selectivity mechanism. These three regions together comprise the symmetric core of the NPC while the two remaining regions are distinct to the nuclear and cytoplasmic sides of the pore and consist of the cytoplasmic filament and nuclear basket nucleoporins. [3] Several cytoplasmic filament and nuclear basket Nups are FG-repeat containing Nups and interact with cargo in transit. Additionally, some of these Nups serve as catalysts in common import and export pathways and aid in the emergence of transport directionality as discussed later in this chapter.

FG Nucleoporins

As mentioned above, 13 of the Nups that comprise the cytoplasmic, central channel and nuclear basket regions contain natively unfolded regions, rich in phenylalanine and glycine residues (FG-repeats). The regions containing FG-repeats provide hydrophobic affinity for active cargo complexes while forming a barrier to inert

cargos. Analysis of amino acid composition of FG-Nups has revealed that these proteins are significantly more enriched in charged and polar residues that contribute disorder to protein structure [43].

Symmetric nucleoporins

	Yeast	Human
Coat nucleoporins	Seh1	Seh1
	Nup85	Nup75
	Nup120	Nup160
	Sec13	Sec13
	Nup145C	Nup96
	Nup84	Nup107
	Nup133	Nup133
	---	Nup37
	---	Nup43
Adaptor nucleoporins	Nic96	Nup93
	Nup192	Nup205
	Nup188	Nup188
	Nup157 }	Nup155
	Nup170	
	Nup53 }	Nup35
	Nup59	
Channel nucleoporins	Nsp1	Nup62
	Nup57	Nup54
	Nup49 {	Nup58
		Nup45
POMs	NDC1	NDC1
	POM34	---
	POM152	---
	---	Gp210
	---	POM121

Asymmetric nucleoporins

	Yeast	Human
Cytoplasmic filaments	---	Nup358
	Nup82	Nup88
	Nup159	Nup214
	Nup42	CG1
	Gle2	Rae1
	Nup116	
	Nup100 }	Nup98
	Nup145N	
	---	ALADIN
Nuclear basket	Nup1	Nup153
	Nup2	Nup50
	Mlp1 }	TPR
	Mlp2	
	Nup60	---

Figure 5. The structure of the nuclear pore complex consists of a symmetric NPC core and asymmetric cytoplasmic filaments and nuclear basket (left). Each region of the NPC is composed of multiple constitutive nucleoporins – Yeast and Human homologs shown (right). Figure adapted from [44] and [3].

Common motifs found within FG nucleoporins include FxFG and GLFG (where x denotes any residue). Interestingly, most Nups are predominantly composed of either one or the other motif. Similar motifs tend to reside near each other when both FxFG and GLFG appear in the same Nup. In the case of Nsp1, a ~190 amino acid GLFG-like region is followed by a ~430 amino acid FxFG-like region. Nups containing predominantly FxFG motifs or stretches composed of FxFG contain charged amino acids in spacers between motifs that result in dynamic, extended coiled-coil conformations. Conversely, Nups or segments of Nups predominantly composed of GLFG motifs contain relatively uncharged spacers and take on more globular, collapsed coil conformations [45].

These properties of FG-Nups give them the ability to transport cargo complexes by enabling simultaneous interaction with multiple binding partners. Additionally, the unstructured nature of their binding sites allows them to conform to a range of import and export receptors and the hydrophobic patches provide molecular binding and unbinding rates with cargo that are conducive to efficient import and export [43].

Models of Nucleocytoplasmic Transport and Directionality

There is general consensus that the permeability barrier is formed by unstructured and FG-rich regions of nucleoporins that occlude the pore and are responsible for facilitating selectivity and transport. The physical behavior of these nucleoporins remains the topic of much debate and has led to the proposal of multiple contrasting models that attempt to explain the physical behavior responsible for experimentally observed characteristics of Nups and selective transport. Some popular models of transport include (*i*) selective phase, (*ii*) virtual gating, (*iii*) polymer brush, (*iv*) forest and (*v*) reduction of dimensionality, which are illustrated in Fig 6.

The selective phase model (Fig. 6a), which was first proposed by the Görlich group, suggests that FG-repeats of unstructured domains of FG-Nups exhibit interrepeat cohesion that results in the formation of a hydrogel barrier [46]. The hydrophobic affinity between phenylalinines is believed to be crucial for the formation of this gel [47]. The porosity in this meshwork is fine enough to inhibit the transport of inert cargos while being coarse enough to allow entry of passive cargos. Active cargos that bind to transport receptors possess hydrophobic affinity for FG-domains of the Nups and locally "melt" the mesh in order to translocate. The mesh is thought to be self-healing and reseals immediately after a cargo passes [48]. The sieve-like properties of the NPC have been recapitulated in *in vitro*-formed hydrogels of several nucleoporins that were determined to possess either β-amyloid or random coil structures [49-51].

The virtual gating or entropic barrier model (Fig. 6b), which was first proposed by the Rout group, suggests that any pore presents a barrier to transport and that the entropic cost of diffusion across and subsequently the likelihood of transport are dependent on cargo size. To overcome this entropic barrier, active cargo binds FG-Nups. It's suggested that FG-Nups possess an ideal binding energy for transport receptor-bound cargo such that the sum of the entropies of binding and diffusion ($\sim kT$) can be cancelled out by the binding energy (ΔH), flattening the energy landscape (ΔG) [52,53]. Similarly, the polymer brush, entropic brush or reversible collapse model, which was first proposed by the Aebi group, suggests that the pore and the FG-Nups that take on a brush-like configuration within the channel present an entropic barrier to transport. The polymer brush model goes one step further and suggests that transport receptors bind and collapse the unstructured regions of FG-Nups that extend into the channel and that this collapse is reversed by the addition of Ran guanosine triphosphate (RanGTP), a critical factor in transport directionality [54,55].

The forest model, which was first proposed by the Rexach group, suggests that two different gating mechanisms operate at the NPC in distinct locations, which they refer to as Zone 1 and Zone 2. The region in the inner diameter of the pore (Zone 1) is said to take on a hydrogel conformation while the peripheral region (Zone 2) is

said to take on an entropic brush conformation (see Fig. 6c) [56,57]. Through the analysis of 20 different intrinsically disordered protein domains, Yamada et al. observed that the domains could be clustered into two categories, one that had low content of charged residues (2-4%) that took on globular configurations and a second category of domains that contain higher content of charged residues (18-35%) that took on relaxed or extended coil configurations. The distribution of domain configurations was then used to classify FG-Nups into two distinct categories – Nups which mostly contain globular configurations and therefore form "shrubs" near the periphery, and Nups which contain domains that take on globular configurations but are linked to the periphery via extended coil domains, resembling "trees". This bimodal distribution of structure of "shrubs" and "trees" forms a "forest" landscape that leads to the emergence of bimodal function in which very large active cargo are transported via Zone 1, smaller active cargo are transported via Zone 2 and passive cargo can diffuse through either zone, depending on cargo size and channel porosity which may depend on the degree of transport receptor saturation. Interestingly, the Yang group recently demonstrated the presence of a size-dependent radial distribution of cargo transport [11].

The reduction of dimensionality model, which was first proposed by Peters, suggests that cargo efficiently translocate through the NPC via a two-dimensional walk on the pore periphery, which is lined with FG motifs, that are saturated with transport receptors and permanently collapsed – presenting a hydrophobic affinity for cargo complexes (see Fig. 6d) [58,59]. In this model, selectivity is provided by the central channel region, which contains Nups with unstructured, extended and hydrophilic domains that repel inert cargo from entering the channel. Inert molecules are then free to pass through a narrow cylindrical opening in the center of the channel. This model builds on the virtual gating and polymer brush models by suggesting that the saturation of FG-Nups by transport receptors leads to the emergence of a two-dimensional surface rather than a volume occupied by FG-Nups.

(a) Selective phase/hydrogel model
(b) Virtual gate/polymer brush model
(c) Forest model — Transporter, Zone 1, Zone 2
(d) Reduction of dimensionality model
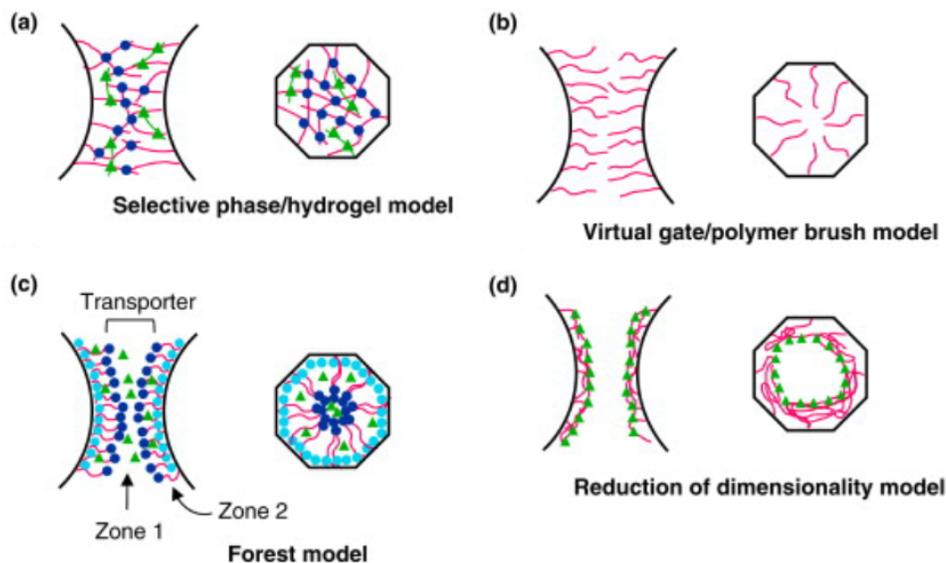
<u>Figure 6.</u> Illustration of several competing models that attempt to explain the mechanism of transport through the nuclear pore complex. These models attempt to resolve the physical behavior of FG-Nups that occlude the pore and facilitate selective transport. The models illustrated represent (a) the selective phase model in which FG-Nups form a cohesive hydrogel that is locally disrupted by transport receptors, (b) the virtual gate or polymer brush model in which the FG-Nups form an entropic barrier to cargo that lack affinity for the Nups, (c) the forest model in which the pore is composed of both hydrogel and entropic brush regions, and (d) the reduction of dimensionality model in which collapsed Nups provide a two-dimensional surface on which cargo can bind to and diffuse along. Adapted from [60].

Although the mechanism governing selective transport is believed to be a result of the physical behavior of nucleoporins, the directionality of transport is primarily governed by the asymmetric distribution of key biochemical transport factors. This was established when the Weis group demonstrated that direction of transport through the pore can be inverted when the gradient of RanGTP is reversed across the nuclear envelope (typically in high concentration in the nucleoplasm and in low concentration in the cytoplasm – discussed in detail below) [61]. The affinity gradient model emerged early on as a possible model for transport [62,63] but was abandoned when it was shown that deletion of asymmetrical Nups (cytoplasmic or nucleoplasmic) did not eliminate cell viability or NPC permeability [64,65]. Nevertheless, recent *in vitro* experiments by the Görlich group show the presence of similar high affinity binding sites for export cargo at the cytoplasmic face that may assist in efficient cargo export by stalling cargo and preventing backflow [50,66]. Our recent computational models of nucleocytoplasmic transport (*Section V*) confirm that the presence of a high-affinity binding site at the transport periphery increases transport efficiency and that the affinity gradient model may be a valid model that contributes to transport directionality.

Nucleocytoplasmic Transport Biochemistry

While the mechanism by which FG-Nups form a selective permeability barrier remains elusive, the biochemical factors involved in facilitating transport are much better understood, resulting in the availability of atomic structures, detailed interaction pathways, steady-state concentrations and kinetic parameters.

Cargo Complex Formation

Large proteins destined for the nucleus (active cargo) typically possess a nuclear localization signal (NLS) – a specific stretch of amino acids that contain basic charged residues, typically a short lysine-rich sequence that occurs in either monopartite or bipartite motifs (classical NLS), or longer arginine-rich sequences (non-classical NLS) [67,68]. Similarly, export cargo contain a nuclear export signal (NES) which typically consists of a leucine-rich or hydrophobic sequence of amino acids [44]. The presence of an NLS/NES gives proteins the ability to recruit and bind karyopherins (Kaps) or nuclear transport receptors (NTRs) that are categorized as either importins or exportins depending on the direction in which they carry NLS/NES containing cargo and may be collectively referred to as transportins (note: although karyopherins may broadly be referred to as transportins, the name transportin itself is used to refer to a specific class of karyopherin as well). While some large proteins are capable of spontaneously transporting across the NPC in a karyopherin-independent manner as a result of their surface hydrophobicity [69], the vast majority of proteins are transported in a karyopherin-dependent manner. The most well studied class of karyopherins belong to the Importin-β (Imp-β) superfamily that are ~100-125 kDa in size that share ~20% sequence homology and a structure composed of ~20 HEAT repeats, with structure and function conserved across species [45]. The binding of NLS/NES-cargo to karyopherins constitutes the formation of a "cargo complex". The process by which cargo complexes are formed in the cell's crowded environment has been determined to be the rate-limiting step in cargo transport rather than the physical translocation between compartments [45].

One of the most well-studied nucleocytoplasmic transport pathways is the classical import pathway in which cargo possessing a classical NLS requires the presence of an adaptor protein Importin-α in order to bind the transport receptor Importin-β and form a cargo complex. It should be noted, however, that proteins can also bind Importin-β directly without the aid of an adaptor protein (see Fig. 7). Interestingly, some viruses have evolved NLS sequences capable of using both classical and multiple non-classical modes of import, giving the virus flexibility over the rate-limiting step in the import pathway and reducing import latency during infection [67].

The formation of a cargo complex provides cargo with hydrophobic affinity for the FG-motifs that occupy the pore channel – a feature that all models of NCT agree upon and require. It has been demonstrated that karyopherins possess multiple FG binding sites on their surfaces [70]. For example, experimental evidence confirms that Imp-β contains two FG binding sites near the N terminus (between HEAT repeats 5 and 6, another between HEAT repeats 6 and 7) as well as two FG binding sites near the C terminus (between HEAT repeats 14 and 15, another between HEAT repeats 15 and 16) [71-73]. Interestingly, computational methods have also been used to confirm and identify FG binding sites – with up to six additional FG binding sites having been identified on Imp-β using molecular dynamics simulations [74].
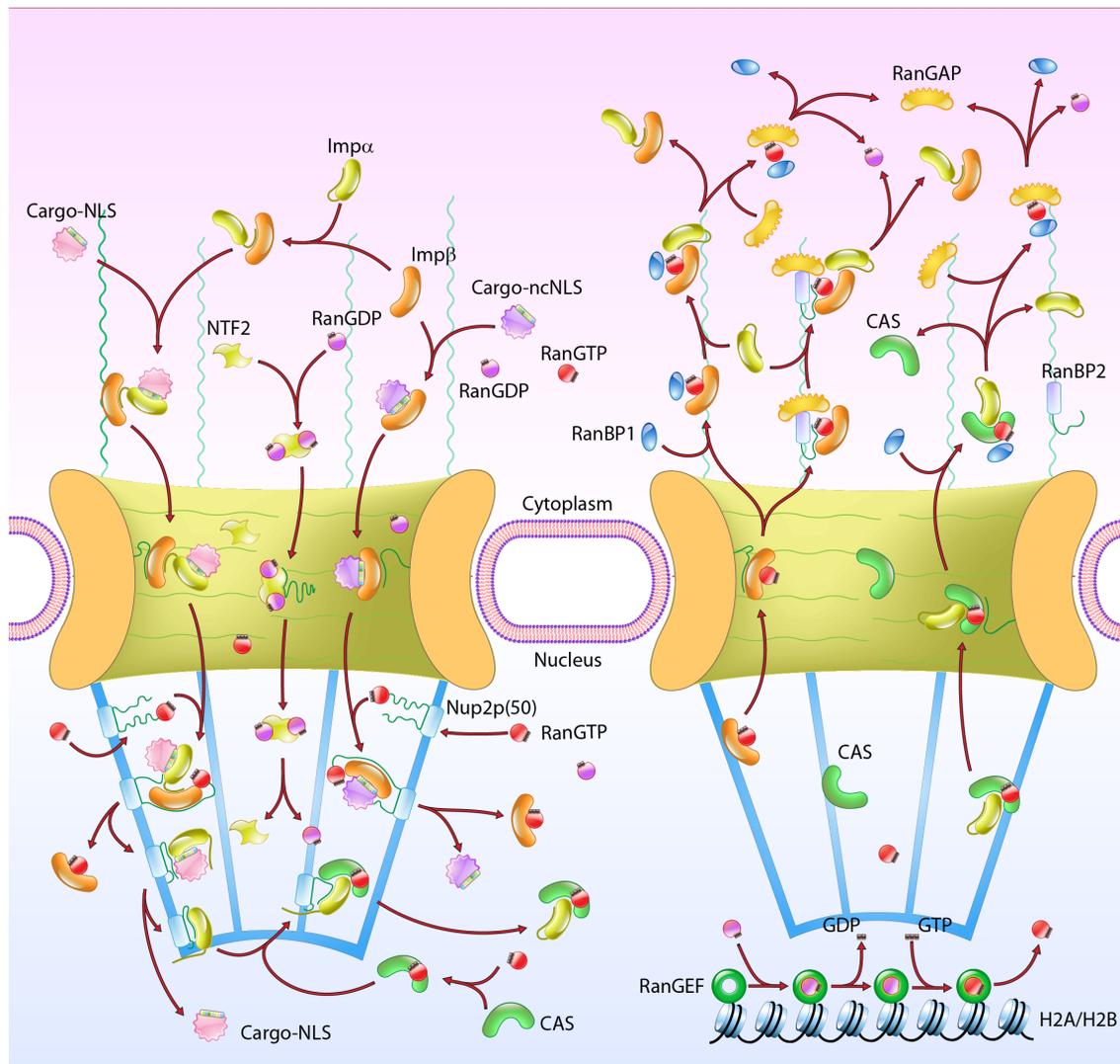
Figure 7. An illustration of the transport pathway for import cargo with cargo complex formation in the cytoplasm (classical and non-classical pathway), transport through the channel, cargo-complex disassembly at the basket, transport receptor recycling from nucleus to cytoplasm along with maintenance of the RanGTP

concentration gradient across the pore via both cytoplasmic (RanGAP/RanBP2) and nucleoplasmic (RanGEF) factors. Adapted from [44].

Transport Directionality and Cargo Complex Disassembly

As mentioned above, the primary determinant of transport directionality is the asymmetric distribution of biochemical factors involved in facilitating transport. Although transport through the pore itself does not require energy input, maintaining efficient transport directionality requires energy via the Ras related GTPase, Ran, which acts as a switch to terminate transport and prevent futile shuttling of cargo between cytoplasm and nucleus. Ran is a ~25 kDa protein that exists in two forms, GDP-bound (guanosine-5'-diphosphate) or GTP-bound (guanosine-5'-triphosphate). The GDP-bound state is predominant in the cytoplasm while the GTP-bound state is predominant in the nucleus. The high concentration of RanGTP (and subsequently low concentration of RanGDP) in the nucleus is maintained by the localization of Ran Guanine Nucleotide Exchange Factor (RanGEF) RCC1 to the nucleus, which is responsible for stimulating the release of GDP from Ran, allowing Ran to bind GTP. Conversely, the low concentration of RanGTP (and subsequently high concentration of RanGDP) in the cytoplasm is maintained by the localization of GTPase Activating Protein (RanGAP) and RanBP2, which are involved in the hydrolysis of RanGTP to RanGDP. Finally, Nuclear Transport Factor 2 (NTF2) is responsible for binding RanGDP in the cytoplasm and recycling it to the nucleoplasm. Maintaining the Ran gradient is essential for transport directionality [61] as well as other cellular functions such as spindle and nuclear envelope assembly [75,76]. The pathway responsible for maintaining the Ran gradient is illustrated in Fig. 7 alongside the nuclear import and transport protein shuttling pathways.

Upon encountering and binding RanGTP, Importin-β releases bound cargo, disassembling the cargo complex. The disassembly of import cargo complexes occurs primarily in the nucleoplasm as a result of the high RanGTP concentration, while the mechanism for RanGTP hydrolysis via RanGAP/RanBP2 in the cytoplasm ensures that RanGTP is not readily available in the cytoplasm to prematurely disassemble import cargo complexes and disrupt transport efficiency. RanGTP mediated cargo complex disassembly is believed to be the result of two factors, (*i*) overlap between cargo binding sites and RanGTP binding sites on Importin-β (N-terminal binding site and the acidic loop within HEAT repeat 8), resulting in steric interference between RanGTP and the cargo, and (*ii*) a conformational change in the structure of Importin-β in which its helical pitch increases after binding RanGTP, resulting in an open conformation that no longer "grips" the cargo tightly [77,78]. The binding of RanGTP to Importin-β and subsequent disassembly of the cargo complex not only contributes to transport directionality but also ensures efficiency by preventing futile cycling of cargo complexes between nucleoplasmic and cytoplasmic compartments. This phenomenon was observed in our computational

model when RanGTP concentrations were decreased and a subsequent decrease in transport was observed (*Section V*). In addition to the RanGTP mediated cargo disassembly, Nup50 (Nup2p) coordinates import cargo complex disassembly for the classical import pathway at the nuclear basket. Nup50 accomplishes this by interfering with the binding between NLS-cargo and Importin-α, further accelerating import cargo complex disassembly [79,80].

Transport Receptor Recycling

Following cargo complex disassembly, transport receptors must be recycled to their respective compartment of origin in order to form new cargo complexes and maintain cargo transport. Karyopherins of the Importin-β superfamily maintain their affinity for FG-Nups even after they bind RanGTP and dislodge their cargo. This allows the karyopherins to readily shuttle back to the cytoplasm where RanGTP undergoes hydrolysis and detaches from the karyopherin, allowing it bind new cargos. Importin-α, on the other hand, requires the aid of an exportin termed Cellular Apoptosis Susceptibility protein (CAS) in conjunction with RanGTP to translocate back to the cytoplasm. Nup50 (Nup2p), which was mentioned previously as having a role in cargo complex disassembly, also aids in the assembly of the Importin-α/CAS/RanGTP complex that is necessary for Importin-α recycling to the cytoplasm [79-81]. Similar to the Importin-β/RanGTP complex that is recycled to the cytoplasm, the Importin-α/CAS/RanGTP complex relies on RanGAP/RanBP2 mediated RanGTP hydrolysis to terminate export of the karyopherin and to detach it from the export complex, making it available for another round of cargo import. The transport receptor recycling pathway for Importin-α and Importin-β is demonstrated in Fig. 7 (right side).

RNA Export Pathways and Kinetics

Aside from transport receptors that are recycled to the cytoplasm, much of what is exported from the nucleus consists of ribonucleic acids (RNA) in the form of transfer RNA (tRNA), microRNA (miRNA), small nuclear RNA (snRNA), ribosomal RNA (rRNA) and messenger RNA (mRNA). Many of these molecules (tRNAs, miRNAs, snRNAs, rRNAs and some mRNAs) follow a karyopherin mediated export pathway that is similar to that of nucleocytoplasmic import and rely on the Ran cycle for directionality. More specifically, tRNAs recruit Exportin-t (Los1) of the karyopherin superfamily in complex with RanGTP to facilitate export, while miRNAs recruit Exportin-5 (Msn5) in complex with RanGTP. The remaining RNAs (snRNAs, rRNAs and some mRNAs) as well as many viral RNAs such as HIV-1 make use of Exportin-1 (Xpo1, a.k.a. CRM1) in complex with RanGTP to achieve nucleocytoplasmic export [82,83]. These exportins rely on the RanGAP/RanBP2-mediated hydrolysis of RanGTP to provide RNAs with export directionality (similar to RanGTP, which acts

as a molecular ratchet for protein import). Conversely, the bulk of mRNA is exported from the nucleus via the NXF1/NXT1-mediated pathway. The dominant transport pathway for each type of RNA is illustrated in Fig. 8.
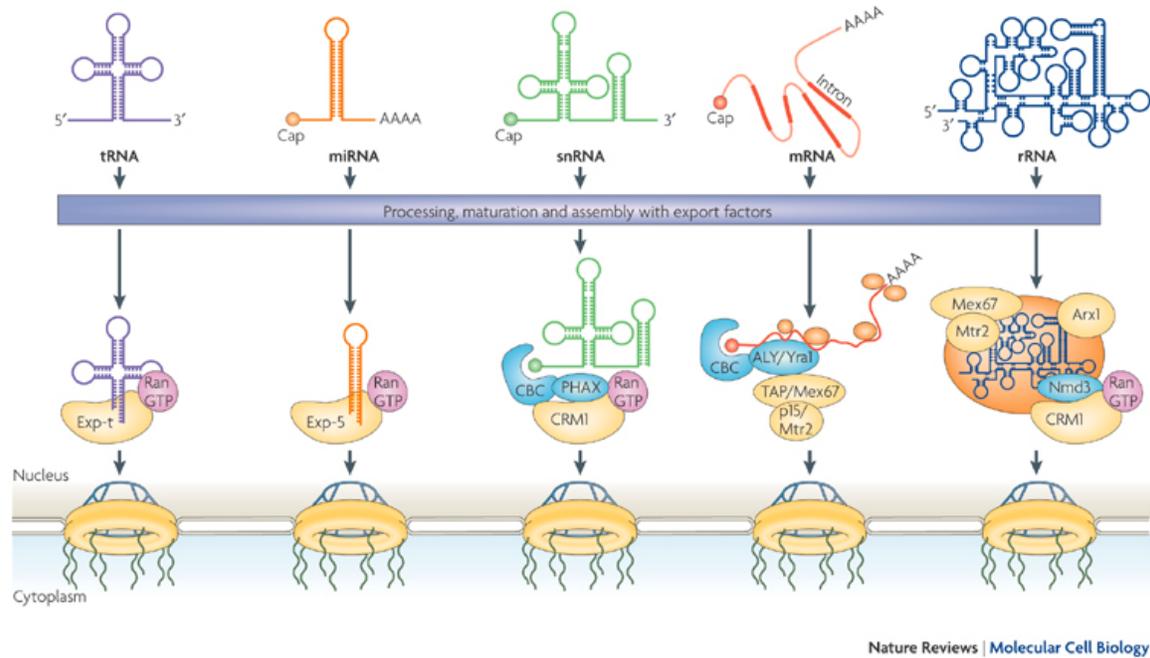
Figure 8. Major RNA export pathways shown for tRNA, miRNA, snRNA, mRNA and rRNA. The bulk of mRNA are exported via an NXF1/NXT1-mediated pathway, while other RNA are exported via transport receptors that belong to the karyopherin superfamily in complex with RanGTP. Figure adopted from [82].

Following transcription, mRNA carry genetic information from the cell's DNA to the ribosomes that translate this message to functioning proteins, primarily via the nuclear pore complex (NPC) and the NXF1/NXT1 pathway. A nascent pre-mRNA must undergo multiple steps before it can be successfully recruited to the NPC and exported from the nucleus. A newly transcribed pre-mRNA undergoes four processing steps prior to being considered a mature mRNA: 1) capping of the 5'-terminal end with 7-methylguanylate, 2) splicing to remove non-coding intron regions, 3) 3'-terminal end cleavage and 4) polyadenylation [84]. During transcription and prior to transport, a maturing mRNA must undergo mRNP assembly in which a series of proteins bind it. Recruitment of proteins to the mRNA is dynamic, with multiple proteins binding, recruiting other proteins, and subsequently unbinding in an effort to produce an export-competent mRNP. The transcription elongation-mRNA export (TREX) complex is a critical factor in producing export-competent mRNA, as it recruits the export factor NXF1/NXT1 (Mex67/Mtr2) heterodimer via a set of core proteins termed the transcription elongation (THO) complex and the associated protein Aly/REF (Yra1). These are subsequently removed from the mRNP prior to export, once NXF1/NXT1 are successfully bound [85]. Once the mRNA is processed and export receptors are

16

recruited, the mRNP is considered export-competent and is recruited to the NPC, where it translocates to the cytoplasm via a series of binding and unbinding events with phenylalanine-glycine (FG) Nups.

Unlike mRNA export pathways in which exportins of the karyopherin superfamily are recruited, bulk mRNA export via the NXF1/NXT1-mediated pathway relies on a different set of export factors to regulate transport directionality. Specifically, the superfamily 2 DEAD-box ATPase DDX19 (Dbp5) is responsible for remodeling mRNPs as they translocate through the NPC and reach the cytoplasmic filaments. DDX19 is localized to the cytoplasmic Nup124 (Nup159) and is activated by the export factor Gle1 and its cofactor inositol hexakisphosphate (IP6), which remodel the mRNP, free transport receptors to return to the nucleus for a successive round of transport, and prevent the return of the mRNP into the nucleus [86].

As was seen with the model of import directionality, in which a RanGTP concentration gradient mediates directional transport, the presence of DDX19 at the cytoplasmic side mediates directional transport of mRNAs. However, the size of a typical mRNA (~2.2 kb) is larger than that of a typical protein that undergoes nucleocytoplasmic import, and much larger mRNAs such as the Balbiani ring mRNA (~37 kb) are known to translocate across the NPC. Subsequently, a model for NXF1/NXT1-mediated nucleocytoplasmic export of mRNA was proposed in which transport receptors are believed to bind multiple sites along the length of an mRNP. The mRNP is exported and transport receptors are sequentially exposed to DDX19, Gle1 and IP6 in the cytoplasm, resulting in local remodeling at each site, preventing the mRNP from locally diffusing back into the NPC, as sites with detached transport receptors will lack affinity for FG-Nups – creating a virtual "Brownian ratchet" in which mRNPs are "ratcheted" through the NPC [87].

Although the mRNA ratcheting model appears feasible, to date, the number and distribution of transport receptors that are required to bind an mRNP for efficient transport remains unknown. Binding of the TREX complex, a factor involved in transport receptor recruitment, has been observed at the 5'-terminal end of mRNA [88]. Coincidentally, it has been reported that transport of very large Balbiani ring mRNA occurs with the 5'-terminal end leading through the pore as seen in electron microscopy (EM) experiments [89-91]. mRNA export orientation could be the result of transport receptors bound at the 5' end and a lack of receptors elsewhere. However, other groups have observed the binding of transport receptor recruiting factors such as REF along the length of the mRNA at the site of exon junction complexes (EJCs) [92-94], suggesting that transport receptors are distributed at multiple sites along an mRNA and lending support to the mRNA ratcheting model. Nevertheless, this leaves unanswered questions as to why there exists a preference for the 5' end leading orientation of export if multiple other transport receptors are present. One could speculate that reasons for this may be due to DNA being read from 3' to 5' end during transcription and complementary mRNA being produced in 5' to 3' end. It is possible that the 5' end simply recruits transport receptors before other sites have a chance to. Another possibility is that the lack of an EJC near the 3'

and the presence of a poly-A tail leaves a segment of the mRNA lacking affinity for the pore. Although it would still be possible for the mRNA to initiate binding to the pore via a transport receptor located somewhere along the length of the mRNA, this would likely be a less favorable orientation for transport than threading through the pore via a leading 5' end orientation. Although the idea that transport receptors bind EJCs seems promising, one must consider that many lower eukaryotes lack the density of introns and subsequently EJCs as seen in higher eukaryotes (Fig. 9) [95]. This could suggest that even though the NXF1/NXT1-mediated export pathway is present in yeast via the homologs Mex67/Mtr2, the mechanism by which mRNPs mature and become export-competent may be slightly different. These questions remain to be answered and an understanding of the number and distribution of transport receptors that bind an export competent mRNP will provide significant insight into mRNA export dynamics with potentially far-reaching impacts in understanding human health and developing successful interventions for human disease.
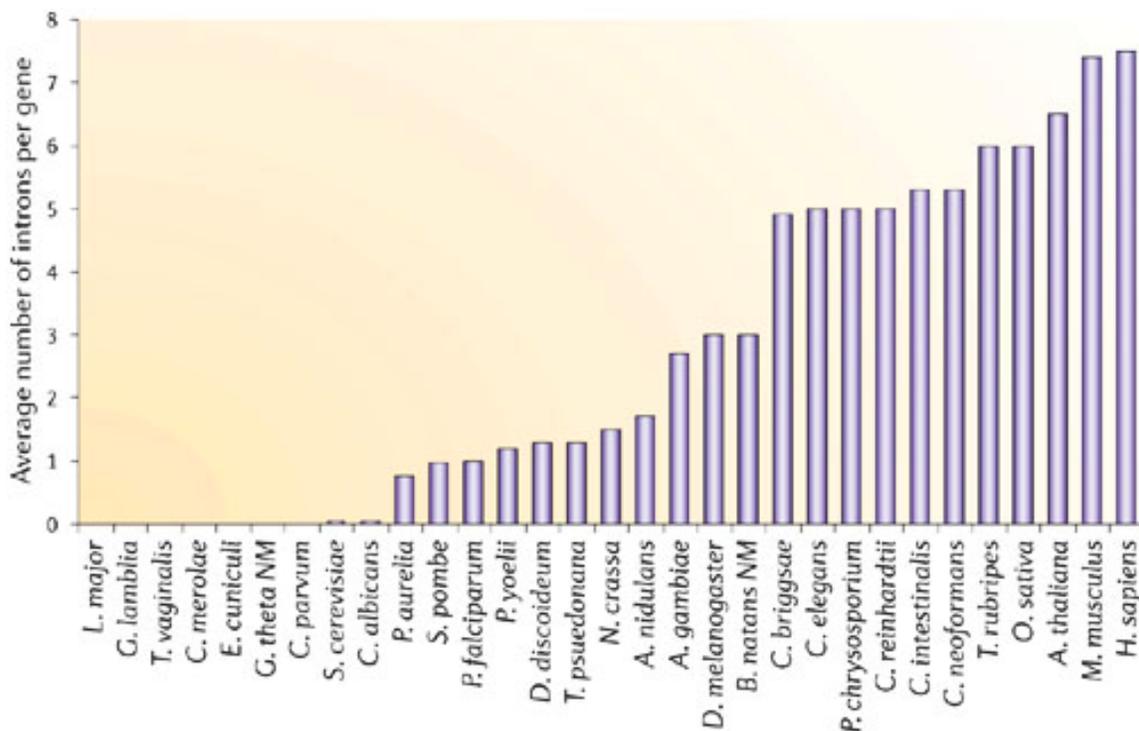
Figure 9. Average number of introns per gene across eukaryotic species. Figure adopted from [95].

In recent years, new molecular techniques have been developed for single molecule tracking of mRNA export, which provide further insight into export kinetics. In 2010, Grünwald and colleagues used the MS2-GFP system [96] to track the movement of β-actin mRNA (~3.3 kb) through the NPC, with temporal resolution of

20 milliseconds and spatial precision of 26 nm. They observed that, rather than transport through the central pore (5-20 ms in duration), docking to and release from the NPC are the rate limiting steps (each ~80 ms in duration) [14]. In 2012, the Kubitscheck group used light sheet fluorescence microscopy to track single export events of native mRNPs with hrp36 fluorescent tags with 20 millisecond temporal resolution and 10 nm spatial resolution [13]. In these experiments, an average transport time of 65 ms was observed for mRNA of all lengths that may be present within the nucleus. In this work, a fluorescent hrp36 protein that natively binds mRNA prior to export was introduced to the system. Since the protein binds mRNA at multiple sites, fluorescent hrp36 was added at sufficiently low concentration to ensure that each mRNA was tagged only once and to avoid tagging multiple proteins within the imaging plane. Subsequently, the group reports an average transport time of 65ms for the majority of mRNA, with transport times of up to several seconds for what are likely very large mRNAs. They further observe mRNA arrest strictly at the nuclear face of the pore that typically lasts ~50 ms.

The discrepancy between the transport times of these two methods (180 ms compared to 65 ms) could be due to the fact that in the MS2-GFP system, a specific mRNA of known length was tracked (β-actin mRNA, ~3.3 kb) whereas the hrp36 protein can bind mRNA of any length (average length of mRNA reported as ~2.2 kb). To further complicate matters, longer mRNA have a greater number of hrp36 binding sites, suggesting that larger mRNAs are more likely to be tagged. However, the typical length mRNAs tracked in the fluorescent hrp36 system also depends on the expression levels for each mRNA. Taken together, this suggests a level of uncertainty in the actual length of mRNAs tracked. Nevertheless, the fluorescent hrp36 system allows for tracking of native mRNA without introduction of multimers of RNA stem loops that bind MS2 coat proteins, which likely alter the size and dynamics of the mRNA molecule [96]. It is likely that the addition of stem loops and the coat protein is a significant source of discrepancy between the observed kinetics of the two experiments, and likely explain why mRNA export was delayed at the cytoplasmic surface in the MS2-GFP system and not the fluorescent hrp36 system [15]. Interestingly, despite the discrepancy in transport kinetics between these works, both single molecule tracking studies as well as the previous EM studies report the observation of a rate-limiting step at the nuclear basket during transport. In the case of the EM experiments, this was thought to be a result of the 5'-end leading orientation of export and the time required for such a large polymer to be oriented to satisfy this condition. It's not clear whether the observance of this rate-limiting step for smaller mRNAs is the result of the polymer searching for a favorable export orientation or whether there is a quality control mechanism or some unidentified processing step present at the nuclear basket that results in delayed entry and transport. Answering this question requires significant improvements in microscopy techniques. Alternatively, the effective computational modeling of the system can provide further insight.

Finally, it should be noted that the above discussion only addresses nuclear export of mRNA via CRM1 and NXF1/NXT1-mediated pathways via the NPC. In 2012,

Speese and colleagues identified an alternate pathway for nuclear export of DFz2C mRNPs via budding of the nuclear envelope analogous to export pathways used by viruses such as herpes [97,98]. There remain many questions as to why these molecules are exported via nuclear budding as opposed to via the NPC (the authors don't exclude the possibility that some DFz2C mRNP could be exported via NPCs). Possible explanations for an alternative transport pathway such as size limits, co-export of multiple mRNPs, preserving mRNP structure for transcriptional repression during export and pathways presenting alternate regulation have been suggested [99]. This alternate pathway remains a topic of interest as more details become available.

Computational Models of Nucleocytoplasmic Transport

Much has been learned about the nuclear pore complex structure and function as a result of novel experiments over the past few decades. Improvements and innovations in imaging will surely allow researchers to overcome limitations in spatiotemporal resolution and resolve many of the questions that remain unanswered. Additionally, over the past decade we have seen many applications of computational modeling, simulation and visualization that aid in the understanding of the structure, function, kinetics and mechanisms of the nuclear pore complex and transport pathways. While many early applications related to systems biology and modeling NCT pathways and coarse-grained models of structure, recent years have seen the application of computation to biophysical modeling of all atomic representations of entire nucleoporin structures and their dynamics. The continuation of Moore's Law combined with advances in parallel computing and graphics processing unit (GPU) computing have enabled these discoveries and will allow us to explore larger and more complex systems in the future [100,101]. Advances in spatiotemporal resolution on the experimental side coupled with increasing system size and complexity on the computational side will continue to bridge the gap between experiment and simulation. Here we review some computational models of the NPC and NCT from recent years that span a wide range of spatial and temporal scales.

Computational Models of Nucleoporin Structure and Interactions

There have been significant efforts by the Sali and Schulten labs over recent years to leverage computational methods for determining nucleoporin structure and interactions. In 2004, Devos and colleagues used multiple computational techniques to predict a structure for Nup84, a core building block of the NPC that is comprised of a complex of seven proteins. In order to do this, they used protein threading software to assign structure folds based on similarity to already resolved protein structures. Although this method did not provide the exact structure of the proteins that make up Nup84, it helped determine regions of the Nups comprised of α-solenoid and β-propeller domains, resulting in a coarse-grained view of the structure. The same group recently demonstrated that these methods for protein structure prediction produce reliable representations of structure by comparing their computationally predicted structures to recently resolved crystal structures that became available after their initial publication (shown in Fig. 10) [102]. This approach was later extended beyond the Nup84 complex to assign folds to ~95% of residues that compose the nuclear pore complex [103].

As an increasing number of crystal structures for proteins relating to nucleocytoplasmic transport became available, the application of molecular

21

dynamics (MD) simulations to studying protein-protein interactions involved in transport became more feasible. In these MD simulations, proteins can be represented by their constitutive atoms (all-atom simulations) or by beads representing a grouping of atoms (coarse-grained simulations). The trajectory of the protein and its interactions over time are determined by solving Newton's equations of motion for each particle, taking into account the forces between a particle and all other particles in the system, including solvent. The computational cost associated with performing all-atom simulations typically results in limitations to system size (size and number of proteins/atoms considered) as well as limitations in simulation temporal length. Coarse-graining methods as well as implicit solvation methods in which the solvent is treated as a continuous medium rather than explicitly modeling all water molecules have been developed to allow modeling of larger systems for longer time scales. Isgro and Schulten used this modeling framework to identify several FG-Nup binding sites on the surface of the transport receptors Importin-β, NTF2 and CAS and were able to confirm experimentally determined FG-Nup binding sites and further identified several additional binding sites that were previously undetected [74,104,105]. In these all-atomic simulations, the authors modeled short peptides of the FG-Nup sequence and ran simulations for durations of ~250 nanoseconds as a result of the computational cost associated with modeling such large systems. They further incorporated computational techniques to assess the significance of novel binding sites by performing sequence alignment from transport receptors of several distant species and assessing evolutionary conservation of potential binding sites.
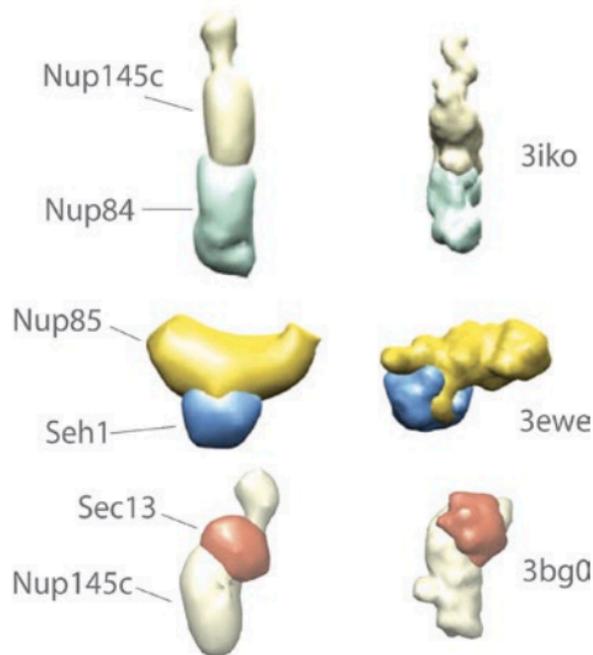
Figure 10. Comparison of protein structures for the constituent proteins of the Nup84 complex generated using computational protein structure prediction

22

methods (left) and their respective crystal structures (right). Figure adapted from [102].

Computational Models of Nuclear Pore Complex Structure and Dynamics

In 2007, Alber and colleagues developed an elegant method to weigh experimental information about nucleoporins such as composition, size, shape, stoichiometry and localization from proteomic, affinity purification, ultracentrifugation and electron microscopy methods to determine a likely arrangement of the nucleoporins that produces the NPC structure [106,107]. Given all the experimental information, generating a likely structure was seen as an optimization problem, well suited for computation. In their model, each nucleoporin was represented by a set of connected spheres that accounted for molecular mass and Stokes radius. Information from experiments was used to create a scoring function or "goodness of fit" for potential configurations of the bead-representation of Nups. The authors then used modeling methods that included molecular dynamics (MD) with simulated annealing (SA) and conjugate gradient (CG) minimization to sample the solution space of configurations that produced an optimum value based on the scoring function, leading to a likely solution to detailed nucleoporin configuration that produces the observed NPC structure. A schematic of the process is shown in Fig. 11.
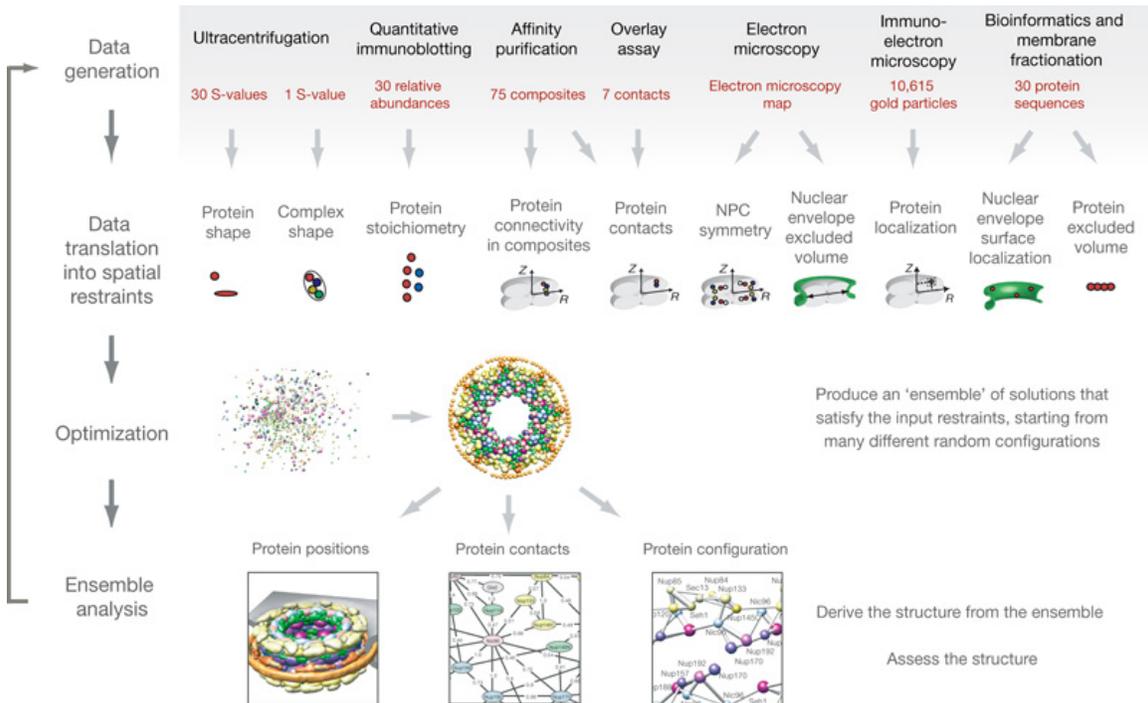
Figure 11. Schematic showing the three components to the optimization problem developed by Alber et al. to computationally determine the structure of macromolecular assemblies. The three components consist of (*i*) data generation from experiments, (*ii*) definition of a scoring function consisting of spatial restraints

and (*iii*) optimization of the scoring function using molecular dynamics (MD) with simulated annealing (SA) and conjugate gradient (CG) minimization to sample the solution space followed analysis of potential structures. Figure adopted from [107].

Other groups have used computational methods to look at the mechanical behavior of the NPC rather than structural composition. Wolf and Mofrad performed normal mode analysis (NMA) using the finite element method (FEM) to determine the natural mode shapes and frequencies of the NPC under vibration [108]. Using their FEM model, the authors demonstrated that the eightfold rotational symmetry of the NPC is ideal for maximizing the bending stiffness of the structure and concluded that the mechanical properties of the NPC accommodate transport of very large molecules via their mechanical compliance in the necessary flexural modes. Following this work, in 2009, Lezon and colleagues developed elastic network models (ENMs) which are conceptually similar to FEMs to also perform normal mode analysis on the NPC structure [109]. Interestingly, the authors observed flexural modes that would assist in the transport of large cargo and concluded that the eightfold rotational symmetry presented ideal dynamics that might have been evolutionarily selected for.

Computational Systems Biology Models of Nucleocytoplasmic Transport

The early applications of computational modeling to nucleocytoplasmic transport took the form of compartmentalized ordinary differential equation (ODE) models. With these mass action approximation models, experimentally determined kinetic rates are used to compute deterministic changes in species concentrations over time. These models make the assumption that the distribution of reactants is continuous in a well-mixed system in which reactant concentrations are sufficiently high and rapidly diffusing to avoid the emergence of concentration gradients and stochastic behavior. Although these models sacrifice significant spatial detail, they are computationally efficient to solve and can span length (entire cell) and time (thousands of seconds) scales that are comparable to experimental measurements and are an excellent resource for performing sensitivity analysis in complex systems [110].

Görlich and colleagues demonstrated one of the first applications of ODE modeling to the NCT pathway when they analyzed the RanGTPase system which is tightly coupled to multiple nucleocytoplasmic import and export pathways to provide directionality [16]. This model was used to predict that the steepness of the RanGTP concentration gradient across the nuclear envelope was limited by the availability of RanGAP in the cytoplasm, leakage of RanGTP across pores and the ability for NTF2 to import Ran into the nucleus and not limited by the availability of RanGEF in the nucleus. Interestingly, the model was able to predict that affinity measurements made *in vitro* were likely one to two orders of magnitude too high to reproduce transport rates observed *in vivo*.

Riddick and Macara later applied the two-compartment ODE model to the complete classical import pathway and performed sensitivity analysis to identify additional factors that may enhance or limit nucleocytoplasmic transport [17]. Their model correctly demonstrated that Importin-α, Ran and NTF2 are limiting factors, as had been confirmed by previous experiments. Additionally, the model predicted that increases in Importin-β and RanGEF surprisingly inhibit import. These predictions from the model were subsequently validated experimentally. In their subsequent work, the authors expanded the simulations to a three-compartment model in which the pore itself was treated as a compartment in which transport receptors had affinities for nucleoporins [111]. This model was then used to understand the role of Importin-α in the classical transport pathway and were able to demonstrate that the use of the adapter protein, Importin-α, reduced overall import efficiency but provided the cell with increased dynamic range for control of import rates, likely having evolved to give the cell the flexibility to control transport rates under different cellular conditions.

Computational Biophysical Models of Nucleocytoplasmic Transport Dynamics

Molecular dynamics methods have also been used to model the collective behavior of Nups to get an understanding of their dynamic structure and how it contributes to facilitating selective transport. Miao and Schulten used a combination of coarse-grained and all-atomic MD simulations to determine how the dynamics of fully extended, individually grafted Nups differed from an array of fully extended, neighboring Nups grafted on a surface. Their simulations showed that an individual Nup rapidly collapses into a globular conformation while the 5 X 5 array of Nups grafted on a surface formed a collective brush-like structure. These observations were considered to be supportive of both the virtual gate and the reduction of dimensionality models and did not support the Nup conformation proposed by the selective phase (hydrogel) model. Additional simulations were run using the grafted Nup configuration in which most hydrophobic phenylalanine residues were mutated to serines. The result of these simulations was still a brush-like configuration, albeit less collapsed than the wild-type Nups (brush height of ~83 Å for wild-type vs. brush height of ~114 Å for mutant). Additionally, when the wild-type brush-like configuration was probed with the transport receptor NTF2, the transport receptor was observed to enter the brush-like structure, while in the mutant configuration, the transport receptor remained near the brush surface. Finally, both structures were also probed with an inert molecule, which failed to enter either the wild-type or mutant brush – suggesting that the observed simulation-derived conformation of the wild-type Nups is capable of reproducing both transport facilitating and excluding properties of the NPC [7,112]. Notably, the computational cost of these simulations was so great that runs were limited to 4 microseconds in duration for the course-grained representation and to 10 nanoseconds in duration for the all-atomic representation.

In 2011, Moussavi-Baygi and colleagues developed a two-dimensional coarse-grained Brownian dynamics (BD) model of nucleocytoplasmic transport in which they considered all the Nups in a planar segment of the NPC and their interactions with single import cargo complexes for durations of up to 8 milliseconds [113,114]. Using this BD model, they were able to simulate the conformation that nucleoporins adopted in the confined geometry of the central channel and its peripheries, which was supportive of the virtual-gate model. They further probed this structure with cargo complexes of varying size to determine the effect of cargo size on transport time, transport path and nucleoporin interaction dynamics. Mean transport times of 2.6 milliseconds were recorded for cargos of 15 nm in diameter with an inverse Gaussian distribution about the mean. Transport time for cargo ranging in size from 9 nm to 15 nm was observed to be ~2.6 milliseconds, regardless of size, while cargo larger than 15 nm in diameter had longer transport times (i.e. a 30 nm cargo had a transport time of 7.1 milliseconds). Interestingly, the authors observed that FG-binding sites on cargo complexes were saturated throughout most of the transport process, with binding site saturation peaking in the central channel >97% of the time. Size dependent average trajectories of cargo are shown in Fig. 12.
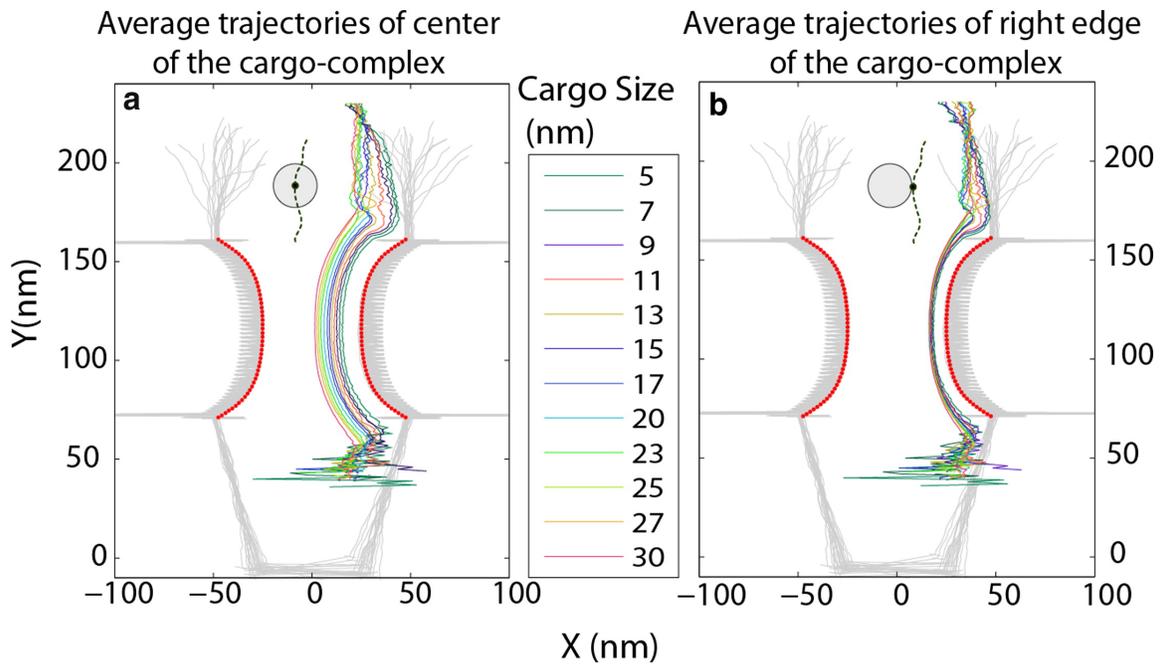
Figure 12. Cargo complex import trajectories averaged over 50 independent replicate coarse-grained Brownian dynamics simulations for each cargo size ranging from 5 nm to 30 nm in diameter. Figure adopted from [114].

Mincer and Simon developed a three-dimensional model of the pore by extending a filament bending dynamics model [115] to account for repulsive terms leading to polymer self-avoidance as well as the various modes of FG-FG and FG-karyopherin binding [116]. Their model was able to reproduce many experimentally observed phenomena such as size exclusion cutoffs for active and passive transport. Notably,

the authors reported that their simulation results of FG-Nup conformation are most consistent with a hybrid of both the selective phase and virtual gate models. In this example, the authors were able to extend the model to three dimensions and achieve millisecond time scales by using an even further coarse-grained model than the previously mentioned Brownian dynamics model. More recently, Tagliazucchi and colleagues developed a single chain mean field (SCMF) theory model similar to coarse-grained MD in which they extended the three-dimensional model of the NPC to account for the non-uniformity among nucleoporin sequences in terms of electrostatic charge and hydrophobicity – a complex factor that was unaccounted for in previous models [117]. Accounting for this additional detail, the authors were able to demonstrate that the conformation of nucleoporins is most consistent with a hybrid of both the selective phase and virtual gate models and also suggest that the specific distribution of charge in the NPC may be optimized for facilitating the transport of the negatively charged cargo complex.

Afterword

Our understanding of nuclear pore complex structure and function has far reaching implications ranging from insight into biological phenomena such as gene expression to treating human diseases such as cancer and viral infection. Experimental approaches that contribute to our understanding of pore functions and mechanisms by which import and export events occur have been limited by the smallest spatiotemporal scales that these methods can explore, with researchers continually pushing the envelope of experimental techniques to peer into the ever more intricate details of this complex structure. Computational techniques spanning a wide range of spatial and temporal scales have been adopted to gain a better understanding of the pore at scales that have at times been unreachable with purely experimental methods. Early models of computational models of transport consisted of deterministic ODE models that achieved substantially long spatiotemporal scales (macro-scale models) but downplay the effects of spatial inhomogeneity due to diffusion, stochastic effects and steric effects. More recently biophysical models have been applied to protein and multi-protein size scales at limited temporal resolutions (nano-scale models). With advances in computing and the application of modern coarse-graining approaches we've reached increased spatial resolutions to system sizes containing the entire NPC and temporal scales in the microseconds (micro-scale models). As experimental techniques continue to push to new limits of spatiotemporal resolution, as is the case in recent single molecule export studies of mRNA, we find that current micro-scale models are inadequate for simulating these very large systems under observation that span several micrometers in length and transport events that are on average tens to hundreds of milliseconds and in extreme cases several seconds in duration (meso-scale). Subsequently, there exists a need for a modeling framework that can efficiently model phenomena at the meso-scale with sufficient accuracy that accounts for factors such as spatial inhomogeneity in the system, steric effects within the pore that play a major role in selectivity and the stochastic effects that are inherent to such a system. In the next section, I introduce agent based modeling (ABM) as a framework, along with a methodology for using ABM to model reaction-diffusion systems at the meso-scale.

References

1. Markov AV (2005) On the origin of the eukaryotic cell. Paleontological Journal 39: 109-116.

2. Alberts B (2002) Molecular biology of the cell. New York: Garland Science.

3. Hoelz A, Debler EW, Blobel G (2011) The structure of the nuclear pore complex. Annu Rev Biochem 80: 613-643.

4. Cooper GM, Hausman RE (2007) The cell : a molecular approach. Washington, D.C.: ASM Press.

5. Yang W, Gelles J, Musser SM (2004) Imaging of single-molecule translocation through nuclear pore complexes. Proc Natl Acad Sci U S A 101: 12887-12892.

6. Mohr D, Frey S, Fischer T, Guttler T, Gorlich D (2009) Characterisation of the passive permeability barrier of nuclear pore complexes. EMBO J 28: 2541-2553.

7. Miao L, Schulten K (2009) Transport-related structures and processes of the nuclear pore complex studied through molecular dynamics. Structure 17: 449-459.

8. Ribbeck K, Gorlich D (2001) Kinetic analysis of translocation through nuclear pore complexes. EMBO J 20: 1320-1330.

9. Kubitscheck U, Grunwald D, Hoekstra A, Rohleder D, Kues T, et al. (2005) Nuclear transport of single molecules: dwell times at the nuclear pore complex. J Cell Biol 168: 233-243.

10. Herrmann M, Neuberth N, Wissler J, Perez J, Gradl D, et al. (2009) Near-field optical study of protein transport kinetics at a single nuclear pore. Nano Lett 9: 3330-3336.

11. Ma J, Yang W (2010) Three-dimensional distribution of transient interactions in the nuclear pore complex obtained from single-molecule snapshots. Proc Natl Acad Sci U S A 107: 7305-7310.

12. Dange T, Grunwald D, Grunwald A, Peters R, Kubitscheck U (2008) Autonomy and robustness of translocation through the nuclear pore complex: a single-molecule study. J Cell Biol 183: 77-86.

13. Siebrasse JP, Kaminski T, Kubitscheck U (2012) Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy. Proc Natl Acad Sci U S A 109: 9426-9431.

14. Grunwald D, Singer RH (2010) In vivo imaging of labelled endogenous beta-actin mRNA during nucleocytoplasmic transport. Nature 467: 604-607.

15. Kaminski T, Siebrasse JP, Kubitscheck U (2013) A single molecule view on Dbp5 and mRNA at the nuclear pore. Nucleus 4: 8-13.

16. Gorlich D, Seewald MJ, Ribbeck K (2003) Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation. EMBO J 22: 1088-1100.

17. Riddick G, Macara IG (2005) A systems analysis of importin-{alpha}-{beta} mediated nuclear protein import. J Cell Biol 168: 1027-1038.

18. Center GSL (2012) Membranes Organize Cellular Complexity. LearnGenetics.

19. Winey M, Yarar D, Giddings TH, Jr., Mastronarde DN (1997) Nuclear pore complex number and distribution throughout the Saccharomyces cerevisiae cell cycle by three-dimensional reconstruction from electron micrographs of nuclear envelopes. Mol Biol Cell 8: 2119-2132.

20. Maul GG, Maul HM, Scogna JE, Lieberman MW, Stein GS, et al. (1972) Time sequence of nuclear pore formation in phytohemagglutinin-stimulated lymphocytes and in HeLa cells during the cell cycle. J Cell Biol 55: 433-447.

21. Ori A, Banterle N, Iskar M, Andres-Pons A, Escher C, et al. (2013) Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. Mol Syst Biol 9: 648.

22. Cordes VC, Reidenbach S, Franke WW (1995) High content of a nuclear pore complex protein in cytoplasmic annulate lamellae of Xenopus oocytes. Eur J Cell Biol 68: 240-255.

23. Arib G, Akhtar A (2011) Multiple facets of nuclear periphery in gene expression control. Curr Opin Cell Biol 23: 346-353.

24. Strambio-De-Castillia C, Niepel M, Rout MP (2010) The nuclear pore complex: bridging nuclear transport and gene regulation. Nat Rev Mol Cell Biol 11: 490-501.

25. Cockell M, Gasser SM (1999) Nuclear compartments and gene regulation. Curr Opin Genet Dev 9: 199-205.

26. Zink D, Amaral MD, Englmann A, Lang S, Clarke LA, et al. (2004) Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. J Cell Biol 166: 815-825.

27. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, et al. (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453: 948-951.

28. Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, et al. (2006) Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nat Genet 38: 1005-1014.

29. Blobel G (1985) Gene gating: a hypothesis. Proc Natl Acad Sci U S A 82: 8527-8529.

30. Schmid M, Arib G, Laemmli C, Nishikawa J, Durussel T, et al. (2006) Nup-PI: the nucleopore-promoter interaction of genes in yeast. Mol Cell 21: 379-391.

31. Brickner JH, Walter P (2004) Gene recruitment of the activated INO1 locus to the nuclear membrane. PLoS Biol 2: e342.

32. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, et al. (2012) ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 40: W597-603.

33. Parry G (2013) Assessing the function of the plant nuclear pore complex and the search for specificity. J Exp Bot 64: 833-845.

34. Zhang X, Chen S, Yoo S, Chakrabarti S, Zhang T, et al. (2008) Mutation in nuclear pore component NUP155 leads to atrial fibrillation and early sudden cardiac death. Cell 135: 1017-1027.

35. Faria AM, Levay A, Wang Y, Kamphorst AO, Rosa ML, et al. (2006) The nucleoporin Nup96 is required for proper expression of interferon-regulated proteins and functions. Immunity 24: 295-304.

36. Basel-Vanagaite L, Muncher L, Straussberg R, Pasmanik-Chor M, Yahav M, et al. (2006) Mutated nup62 causes autosomal recessive infantile bilateral striatal necrosis. Ann Neurol 60: 214-222.

37. Kinoshita Y, Ito H, Hirano A, Fujita K, Wate R, et al. (2009) Nuclear contour irregularity and abnormal transporter protein distribution in anterior horn cells in amyotrophic lateral sclerosis. J Neuropathol Exp Neurol 68: 1184-1192.

38. Sheffield LG, Miskiewicz HB, Tannenbaum LB, Mirra SS (2006) Nuclear pore complex proteins in Alzheimer disease. J Neuropathol Exp Neurol 65: 45-54.

39. Culjkovic-Kraljacic B, Borden KL (2013) Aiding and abetting cancer: mRNA export and the nuclear pore. Trends Cell Biol.

40. Funasaka T, Wong RW (2011) The role of nuclear pore complex in tumor microenvironment and metastasis. Cancer Metastasis Rev 30: 239-251.

41. Whittaker GR, Kann M, Helenius A (2000) Viral entry into the nucleus. Annu Rev Cell Dev Biol 16: 627-651.

42. Frenkiel-Krispin D, Maco B, Aebi U, Medalia O (2010) Structural analysis of a metazoan nuclear pore complex reveals a fused concentric ring architecture. J Mol Biol 395: 578-586.

43. Denning DP, Patel SS, Uversky V, Fink AL, Rexach M (2003) Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. Proc Natl Acad Sci U S A 100: 2450-2455.

44. Jamali T, Jamali Y, Mehrbod M, Mofrad MR (2011) Nuclear pore complex: biochemistry and biophysics of nucleocytoplasmic transport in health and disease. Int Rev Cell Mol Biol 287: 233-286.

45. Aitchison JD, Rout MP (2012) The yeast nuclear pore complex and transport through it. Genetics 190: 855-883.

46. Ribbeck K, Gorlich D (2002) The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. EMBO J 21: 2664-2671.

47. Frey S, Richter RP, Gorlich D (2006) FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. Science 314: 815-817.

48. Frey S, Gorlich D (2009) FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties. EMBO J 28: 2554-2567.

49. Frey S, Gorlich D (2007) A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. Cell 130: 512-523.

50. Labokha AA, Gradmann S, Frey S, Hulsmann BB, Urlaub H, et al. (2013) Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. EMBO J 32: 204-218.

51. Ader C, Frey S, Maas W, Schmidt HB, Gorlich D, et al. (2010) Amyloid-like interactions within nucleoporin FG hydrogels. Proc Natl Acad Sci U S A 107: 6281-6285.

52. Rout MP, Aitchison JD, Magnasco MO, Chait BT (2003) Virtual gating and nuclear transport: the hole picture. Trends Cell Biol 13: 622-628.

53. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, et al. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. J Cell Biol 148: 635-651.

54. Lim RY, Fahrenkrog B, Koser J, Schwarz-Herion K, Deng J, et al. (2007) Nanomechanical basis of selective gating by the nuclear pore complex. Science 318: 640-643.

55. Lim RY, Deng J (2009) Interaction forces and reversible collapse of a polymer brush-gated nanopore. ACS Nano 3: 2911-2918.

56. Patel SS, Belmont BJ, Sante JM, Rexach MF (2007) Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. Cell 129: 83-96.

57. Yamada J, Phillips JL, Patel S, Goldfien G, Calestagne-Morelli A, et al. (2010) A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. Mol Cell Proteomics 9: 2205-2224.

58. Peters R (2005) Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. Traffic 6: 421-427.

59. Peters R (2009) Translocation through the nuclear pore: Kaps pave the way. Bioessays 31: 466-477.

60. Walde S, Kehlenbach RH (2010) The Part and the Whole: functions of nucleoporins in nucleocytoplasmic transport. Trends Cell Biol 20: 461-469.

61. Nachury MV, Weis K (1999) The direction of transport through the nuclear pore can be inverted. Proc Natl Acad Sci U S A 96: 9622-9627.

62. Ben-Efraim I, Gerace L (2001) Gradient of increasing affinity of importin beta for nucleoporins along the pathway of nuclear import. J Cell Biol 152: 411-417.

63. Pyhtila B, Rexach M (2003) A gradient of affinity for the karyopherin Kap95p along the yeast nuclear pore complex. J Biol Chem 278: 42699-42709.

64. Walther TC, Pickersgill HS, Cordes VC, Goldberg MW, Allen TD, et al. (2002) The cytoplasmic filaments of the nuclear pore complex are dispensable for selective nuclear protein import. J Cell Biol 158: 63-77.

65. Strawn LA, Shen T, Shulga N, Goldfarb DS, Wente SR (2004) Minimal nuclear pore complexes define FG repeat domains essential for transport. Nat Cell Biol 6: 197-206.

66. Antonin W (2013) Don't get stuck in the pore. EMBO J 32: 173-175.

67. Cherezova L, Burnside KL, Rose TM (2011) Conservation of complex nuclear localization signals utilizing classical and non-classical nuclear import pathways in LANA homologs of KSHV and RFHV. PLoS One 6: e18920.

68. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, et al. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. J Biol Chem 282: 5101-5105.

69. Kumeta M, Yamaguchi H, Yoshimura SH, Takeyasu K (2012) Karyopherin-independent spontaneous transport of amphiphilic proteins through the nuclear pore. J Cell Sci 125: 4979-4984.

70. Kutay U, Izaurralde E, Bischoff FR, Mattaj IW, Gorlich D (1997) Dominant-negative mutants of importin-beta block multiple pathways of import and export through the nuclear pore complex. EMBO J 16: 1153-1163.

71. Bayliss R, Littlewood T, Stewart M (2000) Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. Cell 102: 99-108.

72. Bayliss R, Littlewood T, Strawn LA, Wente SR, Stewart M (2002) GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta. J Biol Chem 277: 50597-50606.

73. Bednenko J, Cingolani G, Gerace L (2003) Importin beta contains a COOH-terminal nucleoporin binding region important for nuclear transport. J Cell Biol 162: 391-401.

74. Isgro TA, Schulten K (2005) Binding dynamics of isolated nucleoporin repeat regions to importin-beta. Structure 13: 1869-1879.

75. Fu J, Bian M, Jiang Q, Zhang C (2007) Roles of Aurora kinases in mitosis and tumorigenesis. Mol Cancer Res 5: 1-10.

76. Clarke PR, Zhang C (2008) Spatial and temporal coordination of mitosis by Ran GTPase. Nat Rev Mol Cell Biol 9: 464-477.

77. Cook A, Bono F, Jinek M, Conti E (2007) Structural biology of nucleocytoplasmic transport. Annu Rev Biochem 76: 647-671.

78. Zachariae U, Grubmuller H (2008) Importin-beta: structural and dynamic determinants of a molecular spring. Structure 16: 906-915.

79. Matsuura Y, Stewart M (2005) Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling. EMBO J 24: 3681-3689.

80. Sun C, Fu G, Ciziene D, Stewart M, Musser SM (2013) Choreography of importin-alpha/CAS complex assembly and disassembly at nuclear pores. Proc Natl Acad Sci U S A 110: E1584-1593.

81. Harreman MT, Cohen PE, Hodel MR, Truscott GJ, Corbett AH, et al. (2003) Characterization of the auto-inhibitory sequence within the N-terminal domain of importin alpha. J Biol Chem 278: 21361-21369.

82. Kohler A, Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol 8: 761-773.

83. Rodriguez MS, Dargemont C, Stutz F (2004) Nuclear export of RNA. Biol Cell 96: 639-655.

84. Carmody SR, Wente SR (2009) mRNA nuclear export at a glance. J Cell Sci 122: 1933-1937.

85. Natalizio BJ, Wente SR (2013) Postage for the messenger: designating routes for nuclear mRNA export. Trends Cell Biol.

86. Montpetit B, Thomsen ND, Helmke KJ, Seeliger MA, Berger JM, et al. (2011) A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. Nature 472: 238-242.

87. Stewart M (2007) Ratcheting mRNA out of the nucleus. Mol Cell 25: 327-330.

88. Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, et al. (2006) Human mRNA export machinery recruited to the 5' end of mRNA. Cell 127: 1389-1400.

89. Mehlin H, Daneholt B, Skoglund U (1992) Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. Cell 69: 605-613.

90. Mehlin H, Daneholt B, Skoglund U (1995) Structural interaction between the nuclear pore complex and a specific translocating RNP particle. J Cell Biol 129: 1205-1216.

91. Visa N, Izaurralde E, Ferreira J, Daneholt B, Mattaj IW (1996) A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. J Cell Biol 133: 5-14.

92. Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, et al. (2000) Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. Mol Cell 6: 673-682.

93. Le Hir H, Izaurralde E, Maquat LE, Moore MJ (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J 19: 6860-6869.

94. Le Hir H, Gatfield D, Izaurralde E, Moore MJ (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J 20: 4987-4997.

95. Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet 7: 211-221.

96. Querido E, Chartrand P (2008) Using fluorescent proteins to study mRNA trafficking in living cells. Methods Cell Biol 85: 273-292.

97. Speese SD, Ashley J, Jokhi V, Nunnari J, Barria R, et al. (2012) Nuclear envelope budding enables large ribonucleoprotein particle export during synaptic Wnt signaling. Cell 149: 832-846.

98. Hatch EM, Hetzer MW (2012) RNP export by nuclear envelope budding. Cell 149: 733-735.

99. Montpetit B, Weis K (2012) Cell biology. An alternative route for nuclear mRNP export by membrane budding. Science 336: 809-810.

100. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE (2012) Biomolecular simulation: a computational microscope for molecular biology. Annu Rev Biophys 41: 429-452.

101. Moore GE (1998) Cramming more components onto integrated circuits (Reprinted from Electronics, pg 114-117, April 19, 1965). Proceedings of the Ieee 86: 82-85.

102. Fernandez-Martinez J, Phillips J, Sekedat MD, Diaz-Avalos R, Velazquez-Muriel J, et al. (2012) Structure-function mapping of a heptameric module in the nuclear pore complex. J Cell Biol 196: 419-434.

103. Devos D, Dokudovskaya S, Williams R, Alber F, Eswar N, et al. (2006) Simple fold composition and modular architecture of the nuclear pore complex. Proc Natl Acad Sci U S A 103: 2172-2177.

104. Isgro TA, Schulten K (2007) Association of nuclear pore FG-repeat domains to NTF2 import and export complexes. J Mol Biol 366: 330-345.

105. Isgro TA, Schulten K (2007) Cse1p-binding dynamics reveal a binding pattern for FG-repeat nucleoporins on transport receptors. Structure 15: 977-991.

106. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, et al. (2007) The molecular architecture of the nuclear pore complex. Nature 450: 695-701.

107. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, et al. (2007) Determining the architectures of macromolecular assemblies. Nature 450: 683-694.

108. Wolf C, Mofrad MR (2008) On the octagonal structure of the nuclear pore complex: insights from coarse-grained models. Biophys J 95: 2073-2085.

109. Lezon TR, Sali A, Bahar I (2009) Global motions of the nuclear pore complex: insights from elastic network models. PLoS Comput Biol 5: e1000496.

110. Chen WW, Niepel M, Sorger PK (2010) Classic and contemporary approaches to modeling biochemical reactions. Genes Dev 24: 1861-1875.

111. Riddick G, Macara IG (2007) The adapter importin-alpha provides flexible control of nuclear import at the expense of efficiency. Mol Syst Biol 3: 118.

112. Miao L, Schulten K (2010) Probing a structural model of the nuclear pore complex channel through molecular dynamics. Biophys J 98: 1658-1667.

113. Moussavi-Baygi R, Jamali Y, Karimi R, Mofrad MR (2011) Brownian dynamics simulation of nucleocytoplasmic transport: a coarse-grained model for the functional state of the nuclear pore complex. PLoS Comput Biol 7: e1002049.

114. Moussavi-Baygi R, Jamali Y, Karimi R, Mofrad MR (2011) Biophysical coarse-grained modeling provides insights into transport through the nuclear pore complex. Biophys J 100: 1410-1419.

115. Alberts JB (2009) Biophysically realistic filament bending dynamics in agent-based biological simulation. PLoS One 4: e4748.

116. Mincer JS, Simon SM (2011) Simulations of nuclear pore transport yield mechanistic insights and quantitative predictions. Proc Natl Acad Sci U S A 108: E351-358.

117. Tagliazucchi M, Peleg O, Kroger M, Rabin Y, Szleifer I (2013) Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex. Proc Natl Acad Sci U S A 110: 3363-3368.

Section II:
Specific Aims and Hypotheses

Aim 1: A major component of this work was the design, implementation and validation of the molecular ABM framework as well as determining correct methods for translating physical properties into ABM event probabilities as outlined in Section III. While the molecular ABM framework was designed to be applicable to a broad range of problems and systems, the initial design direction was heavily influenced by the need to model cellular structures and processes of interest, namely transport across the cell cytoskeleton and transport through the nuclear pore complex.

Aim 2: The first sub-cellular transport phenomena of interest related to the role of actin filament structure in modifying cytoskeletal transport directionality. I hypothesized that the parallel geometric structure of filaments at the very leading edge of the cell (filopodia) and their contrast to the dense branched geometric structure of filaments that are adjacent but more internal (lamellipodia) may contribute to the super-diffusive behavior of molecules at the leading edge of the cell. The basic molecular ABM diffusion model was sufficient for testing this hypothesis as shown in Section IV.

Aim 3: There are many questions and hypotheses relating to the nuclear pore complex and nucleocytoplasmic transport that remain unanswered as a result of limitations in both *in vivo* and *in silico* methods. In this work, I pose and test the hypothesis that a high-affinity binding site for cargo complexes at the terminal side of the pore is advantageous to cargo transport. Furthermore, I aim to quantify the contribution of this high-affinity binding site and a broad range of affinity gradients to nucleocytoplasmic transport rates in an effort to determine the optimum design qualities of selective yet efficient nanopores. To accomplish these aims and test these hypotheses, an extension to the molecular ABM framework was required to account for binding and unbinding events as shown in Section V.

Aim 4: Looking beyond protein transport, I aim to model the export of mRNA molecules through the nuclear pore complex in an effort to understand the effects of nuclear transport receptor density on transport rate and efficiency. Furthermore, I hypothesize that the use of *in silico* single molecule tracking of mRNA export to report transport times overestimates transport efficiency and underestimates transport times. To accomplish these aims and test these hypotheses, an extension to the molecular ABM framework was required to model the dynamics of the mRNA molecule represented as a freely jointed chain as shown in section VI.

Section III:

Molecular Agent Based Modeling

Introduction to Agent Based Modeling

Agent based modeling (ABM), also referred to as multi-agent system (MAS) or individual based modeling, is a modeling paradigm consisting of a collection of discrete and autonomous entities, termed "agents", that possess a set of governing rules that allow them to assess their nearby environment and make decisions, resulting in the emergence of complex behavior. ABMs are a natural extension of cellular automata [1,2] with notable differences being that cellular automata are limited to lattice geometries, operations are performed on lattice sites (termed "cells"), the rules for updating a cell are typically the same for all cells within the lattice and updates to all cells are made synchronously. These differences make ABMs well suited for simulating a wide range of systems. Here I introduce the principle components of an ABM, typical applications, capabilities/limitations and existing modeling frameworks for performing simulations. I then introduce the application of ABM to modeling reaction-diffusion systems, which I refer to as molecular agent based modeling (MABM). I will discuss the data structures and algorithms that were developed in my MABM framework as well as the derivation of methods for accurately representing physical properties within the model.

Principles of Agent Based Modeling

At the core of ABMs are the agents that comprise the system and are a representation of real world entities. A typical model can consist of multiple types of agents, each in multiple copies and with distinct properties that can be modified throughout the course of simulation either independently or via interaction with other agents. The agents reside in an environment that they can navigate and can take the form of a discrete lattice, continuous space or a series of interconnected nodes that comprise a network. Furthermore, the environment can be limited to a single dimension or be comprised of multiple dimensions depending on the system being modeled and level of detail desired. Discrete lattice environments can be comprised of triangular, square or hexagonal elements in 2D or tetrahedronal, cubic or rhombic dodecahedronal in 3D. In a continuous environment, neighbors are defined by a distance cutoff. In network environments, neighbors are determined by graph edges that connect two nodes. In a discrete lattice environment, two popular neighborhood types exist – the von Neumann and Moore neighborhoods. The von Neumann neighborhood consists of neighboring cells that share an edge with a central cell; for example, in a 2D square lattice, the four cells orthogonally surrounding a cell comprise its neighbors. The von Neumann neighbor can be expanded to the extended von Neumann neighborhood in which cells sharing an edge with the first set of von Neumann neighbors also belong to the central cell's neighborhood (see Fig. 1). In the Moore neighborhood, all cells in contact with (not necessarily sharing an edge) a central cell belong to its neighborhood; for example, in a 2D square lattice, the eight cells surrounding the central cell belong to its Moore neighborhood. The choice of environment and neighborhood type is problem

specific and has implications for computational cost, design of governing rules and ultimately, model correctness. The selection of a discrete lattice environment leads to computational efficiency in very large systems comprised of many agents by limiting the search for neighboring agents at the cost of changes to degrees of freedom in movement.
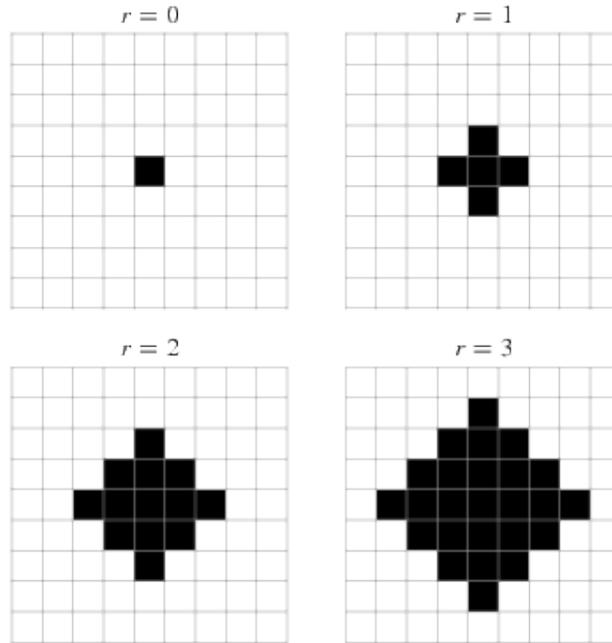
Figure 1. Illustration of various extended von Neumann neighborhoods on a 2D square lattice. r=1 represents the standard von Neumann neighborhood with other values of r representing extensions of it. [*Weisstein, Eric W. "von Neumann Neighborhood." From MathWorld*]

Finally, governing rules determine the actions of agents and lead to emergent behavior. Governing rules dictate events that agents undergo and are typically assigned a likelihood of occurrence, referred to as event probability. Governing rules can determine the actions or state of an agent based on interaction with (*i*) neighboring agents, (*ii*) interactions with the environment, (*iii*) both *i* and *ii*, or (*iv*) the agent on its own without consideration for *i* or *ii*, at any given moment. These rules can be defined using natural language rather than mathematical representations, making ABMs more intuitive and simpler to adopt and communicate.

Applications of Agent Based Modeling

Agent based models have traditionally been applied to macro-scale phenomena involving human agents in which the system is modified to understand how various outputs could be optimized. Examples of this include simulating and optimizing

traffic flow, flow of occupants out of a crowded room during an evacuation, and customer flow management. Other applications to simulating and optimizing flow include models of computer networks in which ABMs can be used to test different scenarios, identify points of weakness and help develop more efficient protocols. ABMs have also been applied to financial markets to both understand the dynamics of price fluctuations and to identify emergent patterns and human/machine trading habits that can be taken advantage of for financial gain. Further applications of ABM at the macro-scale involve models of ecology and climate. [3-8]

Agent based models have also seen application in areas relating to biology and human health. The field of epidemiology has made significant use of ABMs to model the spread of infectious disease. ABMs are well suited for accounting for key factors in disease propagation, namely time, space and human-human interaction. Further applications to human health involve population models of disease and associated interventions to assess efficacy [9]. As previously mentioned, ABMs have traditionally been applied to macro-scale, population-level systems. However, in recent years, ABMs have increasingly been applied to modeling cell-scale phenomena and more recently, sub-cellular phenomena involving the diffusion, binding and unbinding of molecular species involved in signaling pathways [10-16] – systems that have historically been modeled using deterministic mathematical models which usually consist of a series of coupled ordinary differential equations (ODEs). While traditional ABM software frameworks have been designed for applications involving macro-scale problems, there is a need for specialized frameworks that can accurately simulate sub-cellular phenomena. Recent applications of ABM to sub-cellular biological problems use governing rules that have not been validated to reproduce accurate physical phenomena such as diffusion, steric effects, binding and unbinding. Below, I outline other methods for modeling biochemical pathways before reviewing some of the existing general purpose and specialized software frameworks for ABM and then detail my implementation of a molecular agent based modeling framework and methods for correctly simulating physical phenomena.

Computational Modeling Paradigms for Representing Biochemical Pathways

Before discussing ABM software frameworks, a brief explanation of the application of traditional ODE models to simulating biochemical pathways and other less commonly used models is provided with commentary about where ABMs may be more appropriate and more informative.

Traditional deterministic models of biochemical pathways are based on the law of mass action, which describes the kinetics of the system in terms reaction rates that depend on a kinetic constant multiplied by the concentration of reacting species. Given a set of interacting species, each with rate equations dependent on the concentration of other species, there emerges a network of species with time-

varying concentration that are interdependent on the rate of change of concentration of the other species in the system. If initial conditions such as concentrations and kinetic rates are provided, the time evolution of the concentrations of the species in the network can be determined by solving a series of ODEs iteratively to determine both the time evolution of the system as well as steady state concentrations of the species. [17]

There are a number of assumptions made about the system when applying the law of mass action, primarily that the number of molecules is evenly distributed (continuous distribution rather than discrete). This assumption is valid for systems in which the number of species, rate of diffusion and lifetime of these species is sufficiently high such that species that are being consumed or produced redistribute through the system rapidly so that the assumption of a continuous distribution remains valid and the system is termed "well-mixed". The Kuramoto length, l_k (Eq. 1), can be used to test the validity of the well-mixed system assumption for each species in the system given its diffusion coefficient D, and average lifetime of the reactant molecule τ [18]. In the case that $l_k \gg L$, where L represents the length of the container, it can be assumed that a local fluctuation in species concentration will quickly distribute across the system and the well-mixed assumption holds true. However, when $l_k \ll L$, local fluctuations take a long time to redistribute throughout the system and the species is no longer considered continuously distributed. Partial differential equations (PDEs) can be used to account for inhomogeneities in species distribution within a system. PDEs are more complex to implement and as a result, most models that want to account for differences in species localization consist of models containing multiple compartments in which the well-mixed assumption is upheld in each compartment and species are allowed to move between compartments with some rate corresponding to their flux [17].

$$l_k = \sqrt{D\tau} \qquad (1)$$

Many biological systems of interest contain species that are present in discrete numbers; this can occur when species concentrations are very low or when the system size being considered is sufficiently small that the number of molecules modeled is very few, and the continuum assumption breaks down. In such systems, it's likely that natural fluctuations in the number of species are significant enough that deterministic models such as ODEs and PDEs cannot correctly capture the behavior of the system. In a system where there exist on average n molecules of a species, the typical stochastic fluctuation about the average is on the order of $n^{\frac{1}{2}}$ [17]. In the case that the number of molecules in the system is greater than 10^2-10^3, a deterministic solution is sufficient for capturing the behavior of the system. However, when the number of molecules of a species falls below this limit, stochastic fluctuations cannot be ignored as they can result in significant changes to system dynamics and push the system to extreme states that a deterministic model may fail to reach.

More complex modeling paradigms have been developed to account for the natural stochasticity in reaction-diffusion systems. The chemical master equation (CME) is a deterministic ODE that describes the evolution of the probability density function for a Markov process in a well-mixed system of discrete molecules [19,20]. In the case that the well-mixed assumption doesn't hold true, the reaction diffusion master equation (RDME), an extension of CME for inhomogeneous systems in which the system is divided into sub-volumes, each of which are assumed to be well-mixed, can be used to model a discrete and stochastic system. The improvement in model correctness associated with using these advanced methods comes at the expense of increased computational cost. Furthermore, although CMEs and RDMEs can more accurately capture the stochastic behavior of a system, they fail to capture specific physical properties that affect the dynamics of a system, such as steric repulsion between species in crowded environments or confined geometries.
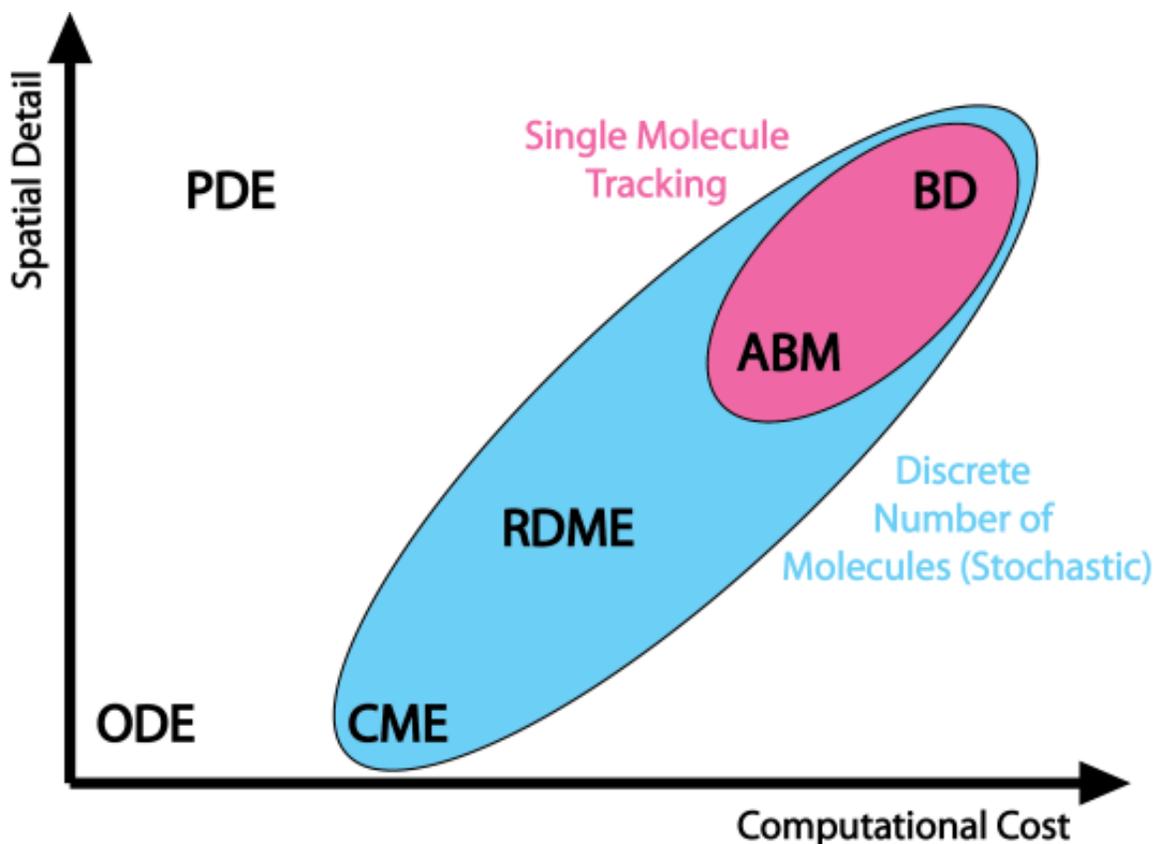
Figure 2. Relative computational cost and the spatial detail between different modeling paradigms and on-lattice molecular ABM. Modeling paradigms enclosed by the blue ellipse are capable of accounting for stochastic effects while modeling paradigms enclosed by the pink ellipse are capable of accounting for steric effects and can be used to track individual molecules throughout a simulation with greater local precision.

In spite of their limitations, compartmentalized ODE models that ignore the effects of stochasticity due to discreteness, inhomogeneities in the system and steric effects

between species in confined and crowded environments have been applied to studying transport through the nuclear pore complex (NPC) as outlined in *Section I* [21]. There exist more physically accurate modeling paradigms such as Brownian dynamics (BD) and molecular dynamics (MD) that can better capture the dynamics of molecules in such systems. However, the sheer size of the nuclear pore complex and the computational cost of BD and MD simulations has limited their applications to modeling nucleocytoplasmic transport (NCT) to smaller subsets of the system or very short time scales as discussed in *Section I*. As a result ODEs, although not ideal, have continued to see application in modeling the NCT pathway as they are capable of simulating large length scales spanning multiple NPCs and time scales on the order of thousands of seconds, which are comparable to the length and time scales achieved through *in vivo* experiments. This highlights the need for a meso-scale modeling paradigm that can explore longer spatiotemporal scales with sufficient consideration for physical phenomena such as diffusion and steric effects that are critical in the confined geometry of the pore. I believe that on-lattice ABMs are well suited to fill this gap at the meso-scale. Furthermore, there are many problems in biology at this scale that can benefit from the availability of an ABM software framework designed specifically for modeling reaction-diffusion systems at the molecular scale. In the next section I will summarize the qualities of existing frameworks and their limitations that prompted the development of a new framework for molecular ABM. The qualities of different modeling paradigms in comparison to on-lattice molecular ABM is illustrated in Fig. 2.

Computational Frameworks for Developing Agent Based Models

There are a substantial amount of software resources dedicated purely to agent based modeling as shown in Table 1. The most popular of these software frameworks are Swarm, MASON and NetLogo. Much of the popularity of Swarm and MASON can be attributed to the open source nature of these frameworks that allows modelers to adapt the software to best suit their needs but its adoption is limited to advanced users. In the case of NetLogo, the ease of use makes the modeling framework attractive for those unfamiliar with modeling and/or programming while limiting its application to specific application areas that it was designed for. The availability of multiple problem-specific software frameworks for deterministic ODE models (Virtual Cell, Cell Designer, CellWare, Jarnac, etc.) have contributed to the popularity and use of ODE models in systems biology modeling, often-times by researchers that lack programming/modeling experience. The availability of such a specific ABM modeling framework that has been tested to correctly reproduce physical phenomena of interest at the meso-scale will contribute the following to the field of computational biology: (*i*) standardized methods for translating physical properties to modeling parameters that will improve reproducibility of results, (*ii*) data structures and algorithms that are optimized to efficiently simulate systems with biochemical agent objects and governing rules and (*iii*) lower barrier to adoption of a modeling paradigm that can provide new insights to biological

problems that have traditionally only been approached using ODE models. Below I describe significant aspects of my implementation of the molecular ABM software framework.

Table 1. Comparison of popular agent based modeling toolkits and their attributes and capabilities.

Framework	License	Application Areas	Modeling Language	Programming Skill Required	3D
AnyLogic	Proprietary	General	Java	Low	Yes
Breve	Open Source	General, AI	Steve	High	Yes
Mason	Open Source	General	Java	High	Yes
MASS	Proprietary	General, Social	Java	High	No
NetLogo	Free	Social, Natural	NetLogo	Low	Yes
Repast	Open Source	Social	Java	Moderate	Yes
SeSAm	Open Source	General	Visual	Moderate	Plug-in
Swarm	Open Source	General	Java	High	Yes

Data Structures and Algorithms for Molecular ABM

The following data structures and algorithms were implemented in FORTRAN90 primarily for its computational efficiency as a compiled language with support across multiple platforms and highly efficient open source implementations of the compiler and numerical/parallelization libraries. Secondly, FORTRAN's learning curve makes it easy for others to contribute to the implementation of the software framework and future development. Significant efforts were made to port the framework to Python but resulted in computational cost that was two orders of magnitude greater for identical simulations. Implementation in Cython reduced computational cost but still resulted in simulation times that were an order of magnitude greater than the FORTRAN implementation.

The software framework consists of four main classes of objects consisting of *element*, *particle*, *reaction* and *polymer*. Collections of objects of these classes are used to form the system and simulate reaction-diffusion phenomena of small molecules and larger polymers. Each class has properties, typically relating to physical properties of the object or its current state that are set at run-time and may vary throughout a simulation.

An ABM simulation consists of an environment in which agents can navigate and interact in. In our framework, this environment consists of an array of objects of type *element*. The fields of an *element* are listed in Table 2 and are basic descriptors of an *element's* position and spatial orientation with regard to other *element* objects. Further descriptors are provided to identify the particles within any given *element* object.

Table 2. Fields belonging to an object of class *element* and their associated data types.

Field	Data Type	Description
X	real (scalar)	X coordinate of an element object.
Y	real (scalar)	Y coordinate of an element object.
Z	real (scalar)	Z coordinate of an element object.
neighbors	integer array (numNeighbors)	Array containing index of neighboring elements.
numParticles	integer (scalar)	Number of particles present in the element.
mapParticles	integer array (maxParticles x 2)	Mapping of particles present in this element to global array of particles.
sumVolume	real (scalar)	Sum of the volumes of all particles within this element.
elementType	integer (scalar)	The type of the current element. May have an effect on how particles behave.

Agents representing molecules or polymers can navigate the constructed environment and are represented as objects or collections of objects of the *particle* class. The fields of an object of class *particle* are provided in Table 3 and are descriptors for the physical properties such as the volume a molecule occupies, its diffusion coefficient within different regions, and an initial concentration of the molecule and whether or not that initial concentration should be maintained throughout the simulation. Other descriptors are used to define the polymer's properties during post-simulation visualization as well as identifying objects of the *polymer* class that comprise objects of the *polymer* superclass.

A three-dimensional array of particles is used to track all *particles* within the system with the first dimension specifying *particle* type, the second dimension indexing *particle* objects of a specific type. The third dimension contains information that identifies which *element* object each *particle* object is located within. Although it appears redundant, the ability to look up all particles within an *element* and the ability to do the reverse-lookup of the *element* that a specific *particle* type belongs to contributes significantly to reducing simulation time at an insignificant cost of increased memory usage. This redundant tracking of objects enables: (*i*) movement

events are only applied to *particle* objects rather than scanning across *elements* to identify *particles* to move, (*ii*) reactions involving two *particles* don't require searching across all *particles* to identify neighboring reactants (only neighbors are searched, an operation of less than $O(n)$ rather than $O(n^2)$). The algorithms that define these events are discussed in further detail below.

Table 3. Fields belonging to an object of class *particle* and their associated data types.

Field	Data Type	Description
numType	integer (scalar)	Internal identifier of particle object.
name	character array (50)	Human readable name of particle object.
partVolume	real (scalar)	Fractional volume of an element that a particle occupies.
initConc	integer array (numElementTypes)	Initial number of particles in total across elements of specific type.
flagConstConc	logical array (numElementTypes)	Flag to identify if number of particles will remain constant throughout simulation.
diffC	real array (numElementTypes)	Probability of movement of this particle into element of specific type.
vizRGB	real array (4)	Particle radius and RGB values for visualization purposes.
isPolymer	logical (scalar)	Identifies whether this particle object is used to construct polymers.

Agents are capable of interacting with other agents or individually transforming to form new agent types to simulate reactions or binding/unbinding events. The methods by which neighboring agents are selected to interact or individually transformed are hardcoded and described in the algorithms below. However, user-defined rules must be provided to inform the model of which agent types are capable of interacting or transforming and the likelihoods by which these events occur. These rules are treated as objects of class *reaction* with their associated fields shown in Table 4. The fields of the *reaction* class are descriptors for how many inputs and outputs constitute a specific reaction, the *particle* classes involved in a specific reaction and the likelihood with which a reaction event will occur.

Table 4. Fields belonging to an object of class *reaction* and their associated data types.

Field	Data Type	Description
numRxnType	integer (scalar)	Internal identifier of reaction object.
numInputs	integer (scalar)	Number of inputs to reaction (either 1 or 2)
rxnInputs	integer array (2)	Array of particle objects that are inputs to the reaction.
K	real (scalar)	Probability with which inputs are replaced by outputs.
numOutputs	integer (scalar)	Number of outputs of reaction (either 1 or 2)
rxnOutputs	integer array (2)	Array of particle objects that are outputs of the reaction.

Table 5. Fields belonging to an object of class *polymer* and their associated data types.

Field	Data Type	Description
numPolyType	integer (scalar)	Internal identifier of polymer object.
polyName	character array (50)	Human readable name of polymer object.
polyLen	integer (scalar)	Length of polymer object.
monomerTypes	integer array (maxPolymerLen)	Sequence of particle objects that constitute a polymer object.
polyCount	integer (scalar)	Number of copies of polymer object in system.

Collections of objects of the *particle* class can be used to define objects of the *polymer* superclass. The fields comprising an object of the *polymer* class are provided in Table 5 and are descriptors such as the length of the polymer chain, the

individual *particles* that construct a *polymer* and their copy number. Another three-dimensional array *polyNeighbor*, keeps track of each *particle* object within a *polymer* and the *particles* bound to it upstream and downstream by recording the *element* object that they belong to as well as their mapping within the array of *particles* to facilitate efficient *movement* operations on *polymer* objects.

There are two main event types in this molecular ABM framework, consisting of *movement* and *reaction* (it should be noted that "reactions" in this model consist entirely of binding and un-binding of molecules – typically proteins). In the case of *movement*, the ABM rules governing agents must be set in a manner that correctly accounts for the physics of diffusion. This means that a single *particle* must undergo normal diffusion at a rate corresponding to its diffusion coefficient while a particle in a crowded environment must undergo anomalous diffusion. Furthermore, the size of *particles* in a crowded environment should determine the extent by which their effective diffusive behavior is corralled and the number of *particles* capable of occupying an *element* should similarly be governed by the size of those *particles* in order to reproduce steric effects. The algorithm used in the *movement* subroutine is outlined in Fig. 3 below followed by derivation of the relationship between diffusion coefficient and simulation movement probabilities. Further details regarding variations on the algorithm to reproduce steric effects and associated computational cost as well as validation of the algorithm are discussed in Section IV of the text with application to cytoskeletal diffusion. Objects of the *polymer* class behave as freely jointed or ideal chains, the simplest representation of a polymer that can convey information about its dynamics and a sufficient representation of long chain polymers that exist in biological systems, such as RNA and DNA. Agents belonging to a *polymer* object are restricted in their movement on the lattice in order to preserve the connectivity of the adjacent agents and subsequently the polymer as a whole. Any agent within the *polymer* object that doesn't represent a terminal end (at position *n* in the chain of agents) is free to move into a lattice site that is the common neighbor of agents that are directly upstream (*n*+1) and downstream (*n*-1) of said agent. The destination site must be a common neighbor of both upstream and downstream connecting agents in order to preserve polymer connectivity. Agents representing terminal ends of a *polymer* object (at position *n* or *N* in the chain of agents) are free to move to a lattice site that is in the neighborhood of their adjacent neighbor (either position 2 or *N*-1), excluding the single neighbor lattice that is directly opposing the agent's current lattice site. By introducing this exclusion, all possible movements for polymer agents in the model are restricted to ones that are into lattice sites diagonal from the lattice site that an agent occupies and are of equal length allowing for the use of a single movement probability for all agents of a single *polymer* object. Validation of accurate *polymer* dynamics is presented in Section VI.

Reaction events may involve a single agent as a reactant or two neighboring agents as reactants. In the case of a single reactant (ie. agent type *A*), every agent of that type in the simulation experiences the *reaction* event with a fixed probability corresponding to the reaction rate constant. If an agent is selected to undergo the

reaction event (ie. the probability of *reaction* event is greater than a randomly generated number), the reactant agent is removed from the simulation and product agent(s) of a different type (ie. agent type *B* or agent types *B* and *C*) are added to the simulation, residing in the same element. In the case where two reactants are involved (ie. agent types *A* and *B*), every agent of the first reactant type (ie. agent type *A*) in the simulation is checked for a neighboring agent of the second reactant type (ie. agent type *B*). If an agent of the second reactant type is present, the *reaction* event may occur with a fixed probability corresponding to the reaction rate constant. If the *reaction* is determined to occur (ie. the probability of *reaction* event is greater than a randomly generated number) the reactant agents are removed from the simulation and product agent(s) of a different type are added to the simulation, residing in the same elements. The removal from or addition to of an agent to the simulation following a *reaction* event requires that both the array of *particles* and array of *elements* be updated to reflect the changes. The algorithm used in the *reaction* subroutine is outlined in Fig. 4 below followed by derivation of the relationship between reaction rate constants and simulation event probabilities. Validation of accurate *reaction* dynamics and comparison with deterministic solutions is presented in Section V.

It should be noted that the selection of polymer *particles* should be randomized in order to avoid simulation artifacts such as preferential movement of one end of the polymer. In the case of single *particles* (non-polymer), randomization of *particle* selection should be randomized across types and objects of types in very crowded systems in order to avoid simulation artifacts. Similarly, the selection of *reaction* events should in all cases be randomized to avoid biasing the system towards artificial equilibriums or outcomes. FORTRAN90 implementations of the *movement* and *reaction* subroutines are provided in Appendix A and Appendix B respectively.
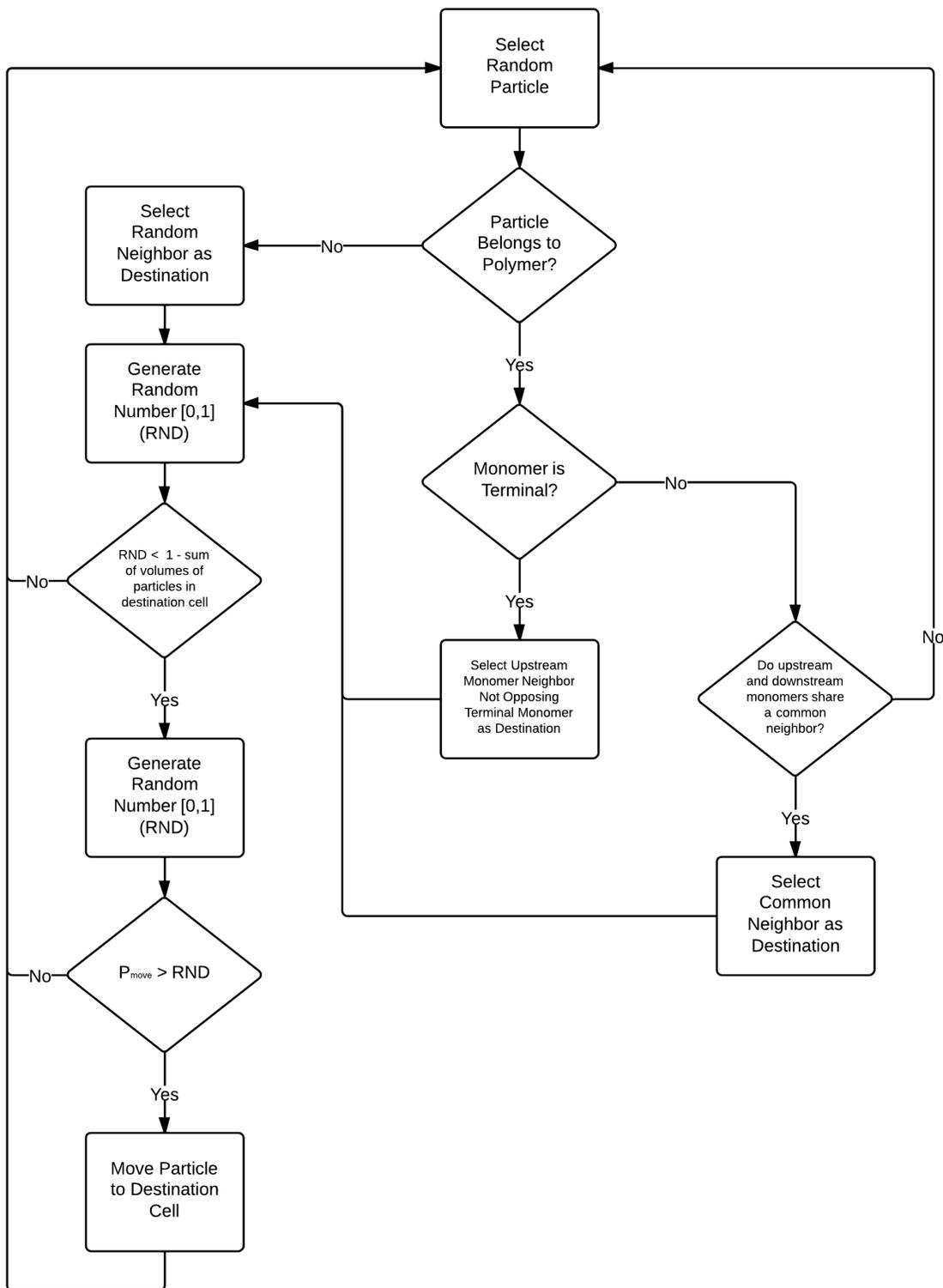
Figure 3. Flowchart of the particle object movement algorithm used to simulate diffusion of small molecules and long chain polymers. See Appendix A for a FORTRAN90 implementation.
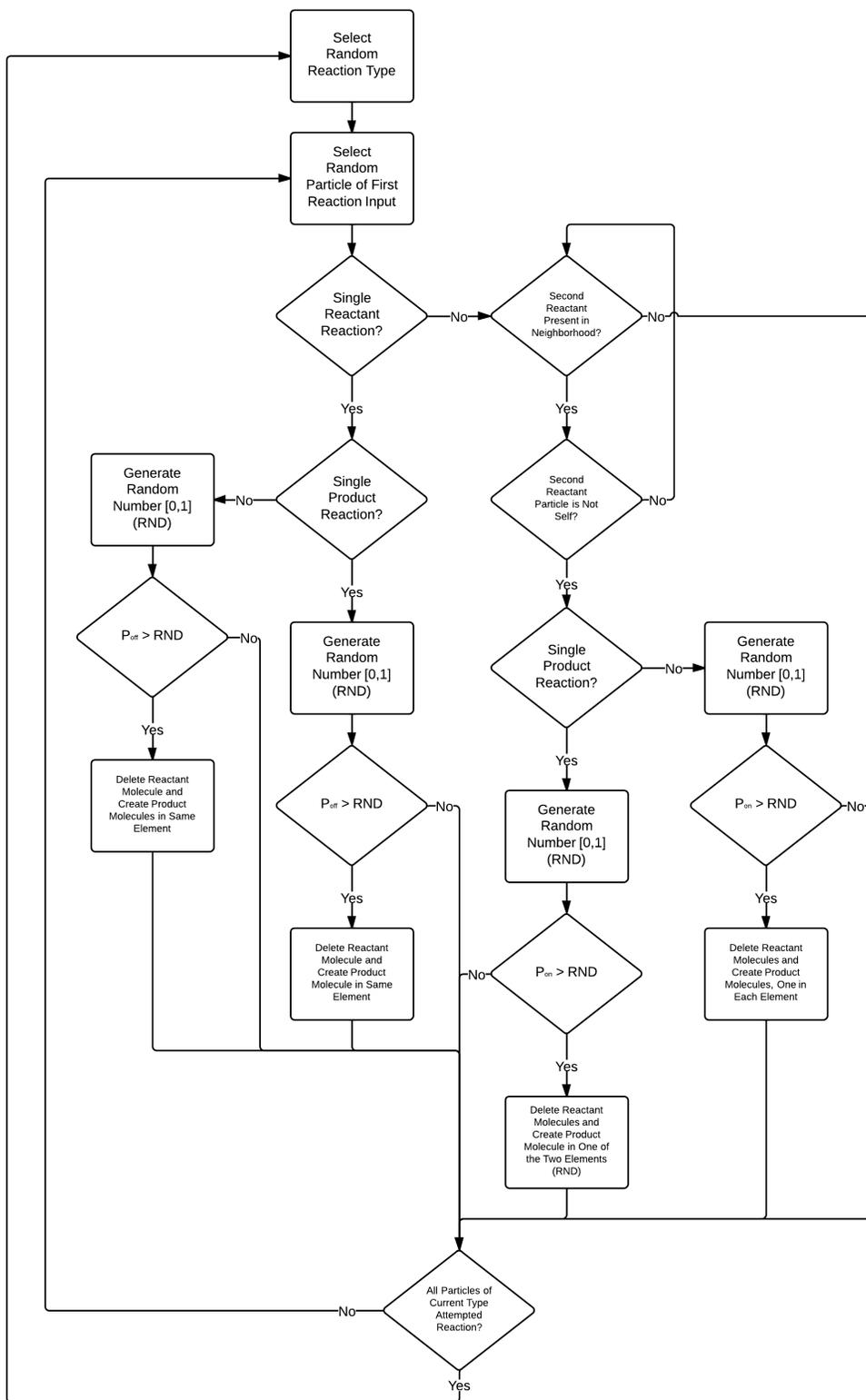
<u>Figure 4.</u> Flowchart of the reaction event used to simulate binding and unbinding events in the molecular ABM. See Appendix B for a FORTRAN90 implementation.

Translating Physical Properties to ABM Event Probabilities

When simulating reaction-diffusion systems, a range of physical properties such as diffusion coefficients and reaction rate constants can be represented by event likelihoods. In the molecular ABM, the event likelihoods considered are P_{move}, P_{on} and P_{off} as shown in Fig. 3 and Fig. 4. As expected, an agent representing a molecule with a higher diffusion coefficient than another molecule would also have a higher probability of movement and likewise for reactions. Translating diffusion coefficients and reaction rate constants to event probabilities should be done in a way that recapitulates real-world agent behavior. Given a real-world agent with a specific diffusion coefficient, a probability of movement should exist that produces the same mean square displacement over time observed for the agent in the simulation. Likewise for a reaction event, a probability of binding/un-binding exists that can produce the same steady-state concentration of their real-world constituents over the same duration of time. Below I will outline the detailed derivation of the equations used to translate diffusion coefficients and reaction rate constants to probabilities that can be used as model parameters. Validation that simulations run using probabilities output from these equations recapitulate the starting values (diffusion coefficients and reaction rate constants) is provided in Section IV and Section V.

Translating Diffusion Coefficients to Agent Movement Event Probabilities

Given a one-dimensional lattice (as shown in Fig. 5) with lattice length L, containing a single agent at site X_n at time t, we can determine the probability of finding the agent at site X_n at the next time step $t+\Delta t$ using Eq. 2.

$$P_{n,t+\Delta t} = P_{n,t} - P_{n,t}T_R - P_{n,t}T_L + P_{n-\Delta L,t}T_R + P_{n+\Delta L,t}T_L \qquad (2)$$
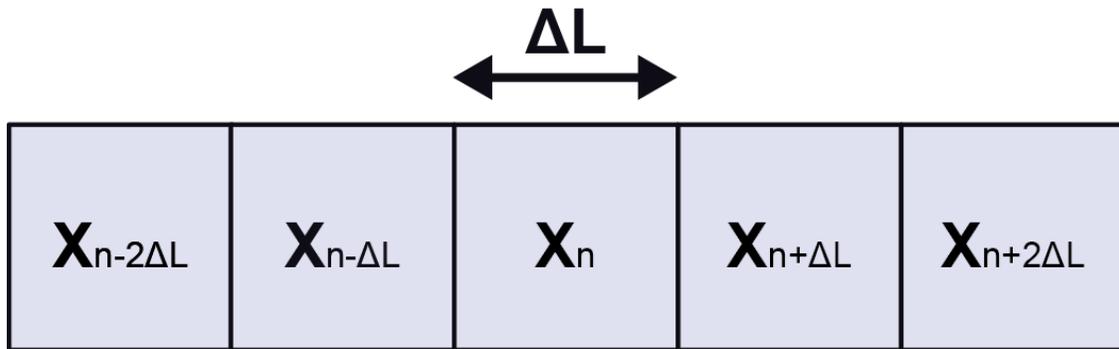
Figure 5. Discretized one-dimensional space with square lattices of length ΔL. This spatial configuration is used to relate movement probability to diffusion coefficient.

Where $P_{n,t}$, $P_{n-\Delta L,t}$ and $P_{n+\Delta L,t}$ represent the probability of finding the particle at position X_n, $X_{n-\Delta L}$ and $X_{n+\Delta L}$ respectively at time t; T_R and T_L represent the probability of the particle moving to the right or left respectively. Note that unless otherwise noted, all probability terms represent the probability of the respective cell being occupied. Equation 1 relays that the probability of a particle being found at X_n at time $t+\Delta t$ can be determined based on the probability that the particle was initially in that position and remained there (first term) less the probability that the particle started in that position and moved to either the right or left cells (second and third term) plus the probability that the particle was initially to the left or right of that cell and moved to the right or left respectively (fourth and fifth term).

Taylor expansion of the terms in Eq. 1 gives:

$$P_{n,t+\Delta t} = P_{n,t} + \Delta t\, \frac{\partial P_{n,t}}{\partial t} + \delta\left(\Delta t^2\right) \tag{3}$$

$$P_{n\pm\Delta L,t} = P_{n,t} \pm \Delta L\, \frac{\partial P_{n,t}}{\partial x} + \frac{1}{2}\Delta L^2\, \frac{\partial^2 P_{n,t}}{\partial x^2} \pm \frac{1}{6}\Delta L^3\, \frac{\partial^3 P_{n,t}}{\partial x^3} + \delta\left(\Delta L^4\right) \tag{4}$$

Assuming that agents have equal probability of moving to the left or right in the one-dimensional lattice (random diffusion, $T_R = T_L = T$), we can simplify Eq. 2 and substitute in the Taylor expansion terms from Eq. 3 and Eq. 4. to get the below equation:

$$P_{n,t} + \Delta t\, \frac{\partial P_{n,t}}{\partial t} = P_{n,t}(1-2T) + 2T(P_{n,t} + \frac{\Delta L^2}{2}\frac{\partial^2 P_{n,t}}{\partial x^2}) \tag{5}$$

Rearranging terms in Eq. 5 gives us a form of the equation (Eq. 6) that is analogous to Fick's second law that demonstrates how diffusion causes the spatial distribution of particle concentrations to change with time (Eq. 7).

$$\frac{\partial P_{n,t}}{\partial t} = \frac{T \cdot (\Delta L)^2}{\Delta t}\frac{\partial^2 P_{n,t}}{\partial x^2} \tag{6}$$

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{7}$$

The definition of $P_{n,t}$, the likelihood of a lattice site to be occupied by agents, is a natural probabilistic description of concentration on a discrete lattice which allows us to define diffusion in terms of movement probability, lattice size and timestep length by combining Eq. 6 and Eq. 7:

$$D = \frac{T \cdot (\Delta L)^2}{\Delta t} \tag{8}$$

As a sanity check we can verify that the right hand side of the Eq. 8 produces units that are in agreement with the units of diffusion coefficient. Eq. 8 establishes the relationship between an agent's real-world diffusion coefficient and its equivalent simulation-based movement probability. The validity of this relationship is demonstrated in Section IV via direct comparison of ABM simulations with simulations of diffusion in Langevin dynamics.

Translating Kinetic Off-Rate Constants to Agent Transformation or Unbinding Event Probabilities

For the case where an agent undergoes transformation to another agent type (Eq. 9) or the unbinding of a complex agent into its constituent agents (Eq. 10), we can use the rate equation (Eq. 11) to determine the change in agent concentration over time.

$$A \xrightarrow{k_{off}} B \qquad\qquad (9)$$

$$A \xrightarrow{k_{off}} B + C \qquad\qquad (10)$$

$$r = \frac{d[A]}{dt} = -k_{off}[A] \qquad (11)$$

While Eq. 11 is expressed in terms of concentration, we can define an analogous description of concentration in terms of number of discrete agents as is shown in Eq. 12:

$$[A] = \frac{N_A}{V \cdot N_{Avagadro}} \qquad (12)$$

Where *[A]* is the continuous concentration of agents of type A, N_A is the discrete number of agents of type A, *V* is the volume of the system and $N_{Avagadro}$ is Avagadro's number. Substituting this definition of concentration in our ABM back into the rate equation (Eq. 11) and simplifying produces the following relationship:

$$\frac{dN_A}{dt} = -k_{off}N_A \qquad (13)$$

In our ABM simulation the change in number of discrete agents of type A is defined as:

$$\Delta N_A = -P_{off} \cdot N_A \qquad (14)$$

Where P_{off} is the likelihood of a transformation or unbinding event of an agent. Substituting Eq. 14 into the discrete form of Eq. 13 and rearranging the terms

provides the relationship between real-world kinetic off rate and simulation-based event probability (Eq. 15).

$$P_{off} = k_{off} \Delta t \qquad (15)$$

Again, as a sanity check we can verify that the right hand side of the Eq. 15 produces a unitless value that is in agreement with the units of event probability. The validity of this relationship is demonstrated in Section V via direct comparison of ABM simulations with the respective deterministic ODE solution.

Translating Kinetic On-Rate Constants to Agent Binding Event Probabilities

For the case where two agents undergo a binding event (Eq. 16) or the transformation of a pair of agents into a different pair of agents (Eq. 17), we can use the rate equation (Eq. 18) to determine the change in agent concentration over time.

$$A + B \xrightarrow{k_{on}} C \qquad (16)$$

$$A + B \xrightarrow{k_{on}} C + D \qquad (17)$$
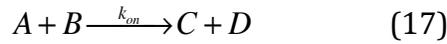
$$r = \frac{d[C]}{dt} = k_{on}[A][B] \qquad (18)$$

Using our previous description of discrete "concentration" in a lattice cell from Eq. 12, we can rewrite Eq. 18 as:

$$\frac{dN_C}{dt} = k_{on} \frac{N_A N_B}{V \cdot N_{Avagadro}} \qquad (19)$$

In the ABM the change in number of agents that undergo binding or joint transformation of can be defined in an analogous manner as Eq. 14 by considering the number of interacting agents of type A and B that are neighboring each other in the system (N_{AB}) and the likelihood of their interaction (P_{on}):

$$\Delta N_C = P_{on} \cdot N_{AB} \qquad (20)$$

Additionally, Eq. 19 can be approximated for the case of discrete number of molecules and time to be:

$$\Delta N_C = k_{on} \frac{N_A N_B \Delta t}{V \cdot N_{Avagadro}} \qquad (21)$$

Setting Eq. 20 and Eq. 21 equal to one another and solving for the probability of interaction (P_{on}) produces:

$$P_{on} = k_{on} \frac{N_A N_B \Delta t}{N_{AB} V \cdot N_{Avagadro}} \qquad (22)$$

The number of agent types of A and B neighboring one another (N_{AB}) is related to the probability of agents A and B neighboring each other at any time (P_{AB}) and the number of discrete lattice sites in the system (N_{cells}):

$$N_{AB} = P_{AB} N_{cells} \qquad (23)$$

The probability of finding two agents of type A and B neighboring each other in the system is related to the number of agents of each type, the number of cells in the system and how many neighbors a cell is defined to have ($N_{neighbors}$):

$$P_{AB} = \frac{N_A}{N_{cells}} \frac{N_B}{N_{cells} - 1} N_{neighbors} \qquad (24)$$

Plugging Eq. 24 in to Eq. 23 and substituting back in to Eq. 22 gives us the overall relationship between real-world kinetic on rate and simulation-based event probability (Eq. 25).

$$P_{on} = k_{on} \frac{\Delta t}{V_{cell} N_{neighbors} N_{Avagadro}} \qquad (25)$$

Where V_{cell} is the volume of the system divided by the number of cells the system is composed of. It should be noted that Eq. 25 is valid for the case where the two interacting agents are of different types. In the case where two interacting agents are of the same type, the event probability is halved. Again, as a sanity check we can verify that the right hand side of the Eq. 25 produces a unitless value that is in agreement with the units of event probability. The validity of this relationship is demonstrated in Section V via direct comparison of ABM simulations with the respective deterministic ODE solution.

In the case where we allow more than one agent to occupy a single lattice site at once, a correction factor α should be added to $N_{neighbors}$ in order to account for the increased number of neighborhoods. This correction factor is a fractional value that depends on the volumes that agent A and B occupy (V_A and V_B respectively) and approaches the value of one in the limit of very small agent volumes (Eq. 26).

$$\alpha \approx 1 - \frac{V_A + V_B}{V_{cell}} \qquad (26)$$

Selection of Simulation Timestep Size

The typical process for determining the largest possible timestep to use for a simulation depends on the slowest occurring event in the system, whether that be movement, transformation, binding or unbinding. As a first step, one should select the fastest diffusing molecule (largest diffusion coefficient) that will be modeled and plug the value in to Eq. 8, specifying a movement probability of one and the lattice size used in order to solve for the largest possible timestep that could be used to model the movement events. The next step is to pick the fastest first-order rate constant (highest k_{off}) and plug the value in to Eq. 15, specifying a movement probability of one in order to solve for the largest possible timestep that could be used to model first-order events. The next step is to pick the fastest second-order rate constant (highest k_{on}) and plug the value in to Eq. 25, specifying a movement probability of one in order to solve for the largest possible timestep that could be used to model second-order events (determine the appropriate cell volume and number of neighbors based on lattice size, dimensionality and neighborhood configuration). The final step is to then select from the three derived timesteps, the timestep that is the smallest – this is the overall fastest occurring event in the system and will occur with a likelihood of one every timestep throughout the simulation (if the appropriate conditions are met) – all other events will occur with lower likelihoods. As an aside, the inclusion of this event in the system is a main determinant of computational cost since it is limiting the selection of a larger timestep size.

References

1. Von Neumann J, Burks AW (1966) Theory of self-reproducing automata. Urbana,: University of Illinois Press. xix, 388 p. p.
2. Wolfram S (1983) Statistical-Mechanics of Cellular Automata. Reviews of Modern Physics 55: 601-644.
3. Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 99 Suppl 3: 7280-7287.
4. Brown DG, Robinson DT (2011) Ecology and Society: Effects of Heterogeneity in Residential Preferences on an Agent-Based Model of Urban Sprawl. Ecology and Society.
5. Eppstein MJ, Grover DK, Marshall JS, Rizzo DM (2011) An agent-based model to study market penetration of plug-in hybrid electric vehicles. Energy Policy 39: 3789-3802.
6. Farmer JD, Foley D (2009) The economy needs agent-based modelling. Nature 460: 685-686.
7. Helbing D, Farkas I, Vicsek T (2000) Simulating dynamical features of escape panic. Nature 407: 487-490.
8. Li H, Tang M, Shang W, Wang S (2013) Securities Transaction Tax and Stock Market Behavior in an Agent-Based Financial Market Model. Procedia Computer Science 18: 1764-1773.
9. Auchincloss AH, Diez Roux AV (2008) A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. Am J Epidemiol 168: 1-8.
10. Fallahi-Sichani M, Flynn JL, Linderman JJ, Kirschner DE (2012) Differential risk of tuberculosis reactivation among anti-TNF therapies is due to drug binding kinetics and permeability. J Immunol 188: 3169-3178.
11. Miller J, Parker M, Bourret RB, Giddings MC (2010) An agent-based model of signal transduction in bacterial chemotaxis. PLoS One 5: e9454.
12. Zhang XH, Tapia M, Webb JB, Huang YH, Miao S (2008) Molecular signatures of two cattail species, Typha domingensis and Typha latifolia (Typhaceae), in South Florida. Mol Phylogenet Evol 49: 368-376.
13. Gooch K, Reinhardt J, Krakauer D (2013) Complex Matrix Remodeling and Durotaxis Can Emerge from Simple Rules for Cell-Matrix Interaction in Agent-Based Models. J Biomech Eng.
14. Nadell CD, Foster KR, Xavier JB (2010) Emergence of spatial structure in cell groups and the evolution of cooperation. PLoS Comput Biol 6: e1000716.
15. Bauer AL, Beauchemin CA, Perelson AS (2009) Agent-based modeling of host-pathogen systems: The successes and challenges. Inf Sci (Ny) 179: 1379-1389.
16. Bailey AM, Lawrence MB, Shang H, Katz AJ, Peirce SM (2009) Agent-based model of therapeutic adipose-derived stromal cell trafficking during ischemia predicts ability to roll on P-selectin. PLoS Comput Biol 5: e1000294.
17. Chen WW, Niepel M, Sorger PK (2010) Classic and contemporary approaches to modeling biochemical reactions. Genes Dev 24: 1861-1875.

18. Grima R (2008) Multiscale modeling of biological pattern formation. Curr Top Dev Biol 81: 435-460.
19. Dobrzynski M, Rodriguez JV, Kaandorp JA, Blom JG (2007) Computational methods for diffusion-influenced biochemical reactions. Bioinformatics 23: 1969-1977.
20. Van Kampen N (2011) Stochastic Processes in Physics and Chemistry: North Holland.
21. Riddick G, Macara IG (2007) The adapter importin-alpha provides flexible control of nuclear import at the expense of efficiency. Mol Syst Biol 3: 118.

Section IV:

Agent Based Modeling of Diffusion in the Cell Cytoskeleton

Accounting for Diffusion in Agent Based Models of Reaction-Diffusion Systems with Application to Cytoskeletal Diffusion

Abstract

Diffusion plays a key role in many biochemical reaction systems seen in nature. Scenarios where diffusion behavior is critical can be seen in the cell and subcellular compartments where molecular crowding limits the interaction between particles. We investigate the application of a computational method for modeling the diffusion of molecules and macromolecules in three-dimensional solutions using agent based modeling. This method allows for realistic modeling of a system of particles with different properties such as size, diffusion coefficients, and affinity as well as the environment properties such as viscosity and geometry. Simulations using these movement probabilities yield behavior that mimics natural diffusion. Using this modeling framework, we simulate the effects of molecular crowding on effective diffusion and have validated the results of our model using Langevin dynamics simulations. Furthermore, we investigate an extension of this framework where single discrete cells can contain multiple particles of varying size in an effort to highlight errors that can arise from discretization that lead to the unnatural behavior of particles undergoing diffusion. Subsequently, we explore various algorithms that differ in how they handle the movement of multiple particles per cell and suggest an algorithm that properly accommodates multiple particles of various sizes per cell that can replicate the natural behavior of these particles diffusing. Finally, we use the present modeling framework to investigate the effect of structural geometry on the directionality of diffusion in the cell cytoskeleton with the observation that parallel orientation in the structural geometry of actin filaments of filopodia and the branched structure of lamellipodia can give directionality to diffusion at the interface of filopodia-lamellipodia interface.

AUTHOR SUMMARY

Computational modeling of biological processes plays a key role in embodying various sources of experimental information into a comprehensive understanding of cellular behavior. These models give us the ability to perform in silico experiments to test hypotheses, better understand how pathway components and parameters affect the system and potentially lead to new experiments. Different modeling methodologies cover a wide range of spatiotemporal scales with each methodology being best suited to a particular class of problems based on the assumptions made and computational costs incurred. Biochemical models range from computationally inexpensive but spatially undetailed macro-scale continuum models to computationally expensive and spatially detailed all-atomistic models. Agent based modeling is emerging as a highly capable method for simulating biochemical

systems at the meso-scale range of the modeling spectrum with application to modeling cellular pathways. Agent based models of cellular pathways have traditionally relied on arbitrary movement rules and rates for simulating diffusion of biochemical components despite the fact that diffusion plays an important role in determining how the system evolves through time as a result of phenomena such as emergence of spatial gradients or molecular crowding. Subsequently, we explore methodologies for the accurate modeling of diffusion in agent based models.

Introduction

Diffusion is a key driver of many biological processes in living systems where ions and molecules move down concentration gradients as a result of their thermal motion within solutions. This phenomenon can be modeled using various computational techniques that consume varying degrees of computational resources correlated with the degree of molecular detail provided by the model. Of specific interest are modeling techniques that account for diffusion and reaction of molecules in biological systems.

Current methods for modeling reaction-diffusion systems generally rely on ordinary differential equation (ODE) models in which the system is assumed to be well-mixed and molecules of interest exist in high numbers, satisfying the continuum assumption [1-3]. These models ignore both the spatial detail and the stochastic behavior observed in natural systems. Other techniques with applications to modeling cellular pathways include partial differential equation (PDE), chemical master equation (CME) and reaction-diffusion master equation (RDME) models that are capable of accounting for spatial varying levels of spatial detail and stochasticity at the cost of increased computational time. These techniques are well-suited for modeling a range of biological phenomena (ODE/PDE methods are ideal for metabolic network models, CME/RDME methods are ideal for gene expression models), with each technique limited by spatial, stochastic and computational cost constraints [1-5]. On the other end of the modeling spectrum are more accurate Brownian dynamics (BD) and Langevin dynamics (LD) models that explicitly account for the diffusion and interaction of individual molecules with the ability to track these individual molecules and assess the effects of spatial and environmental properties that result in the emergence of phenomena such as molecular crowding. These models have additional computational costs associated with them, resulting in limitations to the simulation time and length scales. Recently, agent based models (ABM) have been applied to simulating reaction-diffusion systems [6-9] and have the potential to bridge the gap between spatiotemporally detailed but computationally expensive BD/LD methods and the less detailed but computationally inexpensive ODE/PDE/CME/RDME methods.

Agent Based Models

Agent based modeling is a robust computational technique used to simulate the spatiotemporal actions and interactions of real-world entities, referred to as "agents" in an effort to extract their combined effect on the system as a whole. Both space and time are discretized in an agent based model, giving these autonomous agents the ability to move and interact with other agents and their environment at each time step over a given duration. Simple behavioral rules govern the movement and interaction of each individual entity in an effort to re-create or predict more complex behavior of multiple entities. Such a model attempts to simulate the

emergence of complex phenomena that may not be apparent when simply considering individual entities. Agent based modeling has seen applications in a broad range of fields ranging from artificial intelligence and gaming to modeling emergent social behavior such as the spread of disease and outcomes of financial markets [10-14]. In their simplest form, these agent based models consist of a mesh of "cells" that make up the discretized space that agents occupy. The agents occupy these cells and are typically only aware of other agents within their "neighborhood"; in the simplest form a neighborhood consists of adjacent cells. Agents are given the ability to move into adjacent cells and to interact with other agents with some probability in conjunction with governing rules that define what movement and interactions are possible.

In a physical system we can attribute the diffusion of a particle in solvent to the instantaneous imbalance of the combined forces exerted by collisions of the particle with the much smaller solvent molecules surrounding it which are moving due to random thermal motion. In an agent based model the same movement of this particle due to collisions with much smaller solvent molecules can be implicitly modeled by correlating the diffusion coefficient of the particle in the specific solvent to some movement probability for that particle. Furthermore, in a physical system, steric effects prevent two particles from coming closer than a certain distance from one another or occupying the same position. This type of behavior can also be replicated with an agent based model using governing rules that limit the number of particles per discretized space. As a result of these simplifications, the process of modeling particles diffusing throughout a space does not require computationally intensive method for simultaneously calculating velocities of particles and the effects of repulsive and attractive forces of these particles on other particles within the system (as seen in BD/LD models). Rather, we can describe diffusion and interaction in terms of natural language based on simple observations such as: *different particles move throughout space in a random manner*, *these movements are related to particle size*, and *two particles tend to disfavor occupying the same space*. These descriptions based on natural language can be translated into simple logic rules that govern the behavior of the system. Although ABMs seem ideal for modeling reaction-diffusion systems, existing ABM frameworks do not consider the accuracy of particle movement algorithms. Furthermore, particle movement probabilities are oftentimes selected arbitrarily by the modeler without validating that the molecules' movement behavior represents realistic diffusion rates. Subsequently, agent based modeling of biochemical systems can benefit from validated movement algorithms and movement probability selection criteria.

We have outlined an approach for establishing the logic rules that govern particle diffusion along with methods for translating key parameters such as diffusion coefficients that have continuous and deterministic values into probabilities that can be used as inputs to a discrete and stochastic agent based model. Additionally, we validate these methods with single-particle and multi-particle simulations where normal diffusion is modeled. Furthermore, we investigate the effects of molecular crowding and high concentrations of macromolecules in the simulation volume as is

seen in the cell cytoplasm along with their effect on effective diffusion coefficients, comparing our results with Brownian dynamics simulations. We then investigate the effect of allowing multiple particles to reside in a discrete cell of finite volume and quantify and discuss advantages and disadvantages of various approaches of enforcing finite cell volumes. Finally, we apply the ABM framework to investigate the role of geometry on the directionality of diffusion and show how specific geometries can promote diffusion in a particular direction while other geometries hinder the movement of macromolecules in a particular direction as seen in the filopodia and lamellipodia regions of the cell cytoplasm with regard to the diffusion of G-actin.

Results & Discussion

Relating Diffusion Coefficients to ABM Movement Probabilities

Given a lattice with discretization length of ΔL, the relationship between diffusion coefficient (D) and movement probability (T) for a fixed timestep of Δt is shown in Eq. (1) (see Methods for derivation).

$$D = \frac{T \cdot (\Delta L)^2}{\Delta t}$$ (1)

$$\Delta t \rightarrow 0, \Delta L \rightarrow 0$$

The relationship established in Eq. (1) allows us to take diffusion coefficient values that are meaningful in a continuous and deterministic framework and apply them to a discrete and stochastic agent based model via movement probabilities. The relationship between mean square displacement and time can be used to validate the relationship derived in Eq. (1) for simulating diffusion via ABM. This means that movement probability associated with the diffusion coefficient being modeled should result in displacement behavior and rate that would be seen in a physical system. The mean square displacement $\langle r^2 \rangle$ of a particle diffusing due to Brownian motion is proportional to the time elapsed through the following relationship [15]:

$$\langle r^2 \rangle = qDt^\alpha$$ (2)

Where q is the numerical constant which depends on dimensionality, q = 2, 4 or 6 for dimensionalities of 1, 2 or 3 respectively and D is the diffusion coefficient and t is time. The exponent α is the anomalous diffusion exponent where $\alpha=1$ for normal diffusion while all other values of α represent anomalous diffusion. This means that for normal diffusion, there is a linear relationship between the mean square displacement of a particle and time. If we were to plot the calculated mean square displacement versus time in our simulation, the linearity of this plot would demonstrate whether simulated diffusion is normal or anomalous and the slope of this plot would be related to our diffusion coefficient as described by Eq. (2).

To validate our model for the diffusion of a single particle, we simulated a macromolecule with a Stokes radius of $r=5x10^{-7}$ *cm* (*diameter*=10^{-6} *cm*) that was free to diffuse in a solvent in three dimensional space with a diffusion coefficient of $D=10^{-6}$ *cm²/s*. We ran our model for 500,000 time steps with a minimum sampling size of 300 independent runs which resulted in a linear relationship between mean square displacement and time ($\alpha=0.9992$, $R^2=0.9999$) which implies that the model successfully reproduces normal diffusion behavior. Furthermore, when modeling multiple particles with different diffusion coefficients in very low concentration (crowding effects are negligible), it was observed that using Eq. (1) to derive movement probabilities for each particle type produced the same linear relationship

between the measured mean square displacement and time (normal diffusion was observed) with different slopes for each particle type that corresponded to the diffusion coefficients being modeled. The deviation from linearity in this case was on the same order as that of the single particle and is most attributed to the stochastic nature of the model and the sampling size used. Additionally, it should be noted that as Δt and ΔL become larger (more coarse grained models) the error in the simulation also increases as a result of the approximations made in Eq. (1). However, this discretization error is typically negligible when compared to the variations resulting from the stochasticity of the model and more importantly, such a change in discretization will result in the loss of detailed spatial and temporal information.

Crowding effects on Diffusion and Multiparticle Occupation of Cells

The relationship between crowding due to increased concentration and the effective diffusion at low time scales is shown in Eq. (3) (see Methods for derivation).

$$D(C) = (1 - C \cdot V_{element} N_A) \cdot D(0) \qquad (3)$$

Where $D(C)$ is the effective diffusion coefficient as a function of concentration and $D(0)$ is the diffusion coefficient of a particle in a low concentration system and C, $V_{element}$ and N_A represent concentration of crowding molecule, volume of a discrete element and Avagadro's number respectively and the product of these three terms is equivalent to the probability of finding any discrete cell to be occupied by a molecule ($P^{occ} = CV_{element}N_A$). In order to determine how higher concentrations affect particle diffusion we performed Langevin dynamics simulations utilizing the shifted force form of the Lennard-Jones potential energy function that assessed the effective diffusion coefficient of a particle as the concentration of particles in the system was increased. The analytical relationship shown in Eq. (3) is in agreement with the computational result from Langevin dynamics simulations shown in Figure 1 (circle points). These results can be compared with the result of the two different diffusion algorithms, all-neighbor attempt and single-neighbor attempt, used in the agent based model as shown in Figure 1 (square and triangular points respectively). The single-neighbor attempt algorithm results are in agreement with both the Langevin dynamics simulation as well as the analytical relationship, showing that as concentration of particles in the system increases, the effective diffusion coefficient decreases linearly. The all-neighbor attempt algorithm that searches for neighboring vacant cells results in unnaturally higher effective diffusion coefficients.

This higher effective diffusion coefficient of the all-neighbor attempt algorithm can be attributed to the algorithm simulating "intelligent particles" that search for vacancies rather than the behavior of "non-intelligent particles" that diffuse randomly due to Brownian motion. The behavior exhibited by the single-neighbor attempt algorithm is considered best suited for modeling the diffusion of passive

"non-intelligent" molecules and macro-molecules such as proteins involved in reaction-diffusion systems. Furthermore, this phenomenon is critical for modeling macromolecular crowding and its direct effects on intracellular diffusion as well as reaction kinetics in intracellular environments. Alternatively, the all-neighbor attempt algorithm would be better suited for intelligent agents that can sense the environment around them using means other than collisions. The implementation of a single-neighbor movement algorithm is a very computationally efficient way of providing detailed spatial information for diffusing particles while enforcing steric repulsion and simulating molecular crowding.
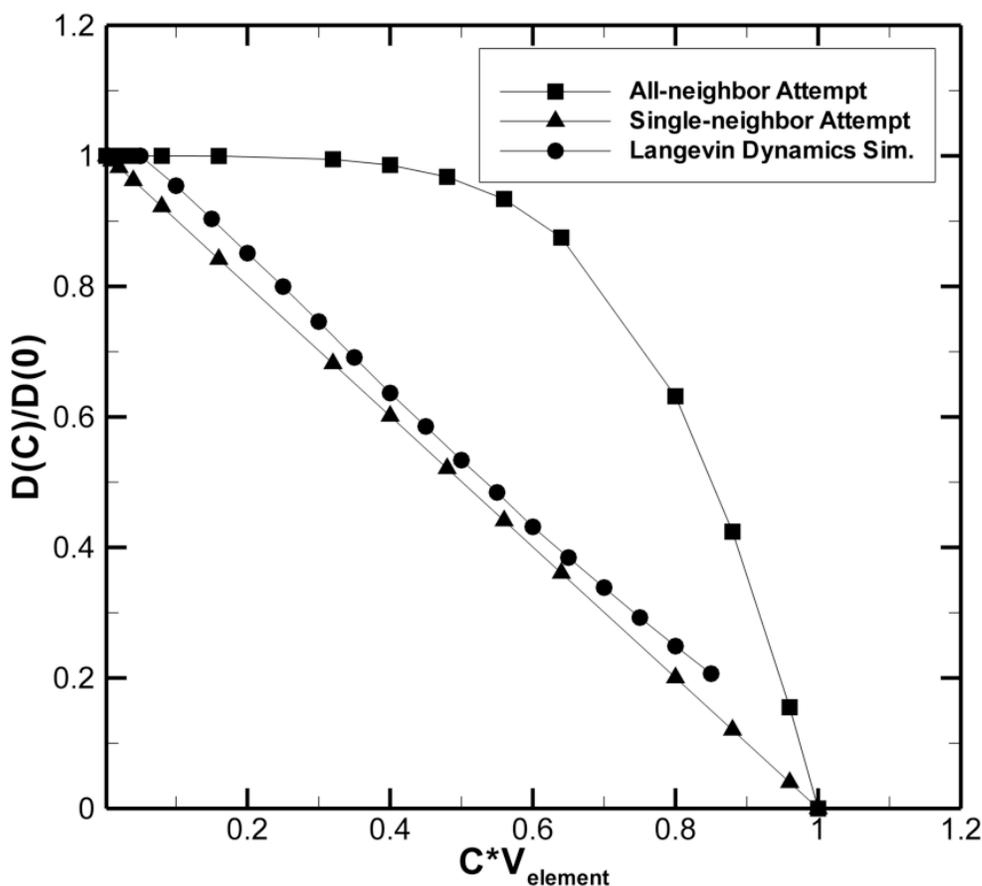
Figure 1. Effective diffusion coefficient versus normalized free particle concentration (volume density) for two agent based model algorithms and a Langevin dynamics simulation for comparison. The graph shows the single-neighbor attempt algorithm to best represent diffusion at higher concentrations as the effective diffusion of this algorithm decreases linearly with increased concentration as does the Langevin dynamics model. As the graph shows, the single-neighbor attempt and Langevin dynamics simulation exhibit the same negative linear slope with a slight difference in offsets resulting from the different

definition of particle volume between the two modeling techniques. Higher concentration data points for Langevin dynamics have been omitted as the volume definition of particles leads to volume overlap at this concentration.

The issue of crowding effects becomes more complex when considering systems with particles of varying size. In the simplest case where the model allows for only a single particle per cell, discretization errors can arise from small molecules saturating the available vacancies and reducing the effective diffusion coefficient, when in reality the volume density of the system has not been changed significantly. This error arises from discretization and the simplification that the smallest particles occupy the same volume as that of the largest particles. This issue can be overcome by introducing an additional layer of complexity in the agent based model where multiple particles are allowed to occupy a single cell. In this framework each particle is given a volume value, typically a fraction of the discretized cell's volume which it occupies, ranging from 0 to 1. As this framework is adopted, multiple particles are allowed to diffuse into a single cell and steric repulsion between particles is no longer intrinsically observed as it was with single particles per cell, raising concerns about individual cells' volume limits being exceeded at high concentrations of particles.

The most intuitive method for ensuring that the number of particles per cell does not exceed the cell's volume is to simply enforce that the movement of any particle into a destination cell will not surpass that cell's volume limit. Although this method is seemingly straightforward and adds minimal computational cost (see flowchart, Figure 2.a), it results in the emergence of artificially high diffusion for particles of smaller size and artificially lower diffusion for larger particles (see Figure 3). In addition, as Figure 3 depicts, the effective diffusion of particles obeying the volume limit (VL) method is subject to artificial limitations resulting in the stair-step behavior. For example, in a concentrated environment where the cell's fraction volume cannot exceed 1, no cell can contain more than a single particle of volume fraction greater than 0.5. This means that when a three dimensional discretized space of 1000 elements has 1000 particles of size 0.51 fractional volume or larger (Figure 3b), no particles in the system will diffuse since any movement will result in the volume limit being exceeded.

One method for rectifying the problem of artificially higher diffusion for smaller particles in the volume limit method is by adding a probability term based on cell capacity to the movement logic. In the combined reduced probability and volume limit method (RP + VL) shown in Figure 3, as a cell's occupied volume increases, the probability of movement into that cell decreases. This adds additional computational time due to the random number generation required each time the cell's volume is not exceeded and a move is attempted (see flowchart, Figure 2.b) but has the added benefit that it better matches the true behavior of the diffusion of multiple particles at physiologically relevant concentrations. However, as depicted in Figure 3, this method of reduced probability combined with the volume limit is only effective at accurately modeling concentrated systems with smaller particles.

Finally, we can best match the actual diffusion behavior by removing the volume limit and simply reducing the probability of movement based on the fraction of a cell's occupied volume (RP method shown in Figure 3). This method is the most computationally intensive of the three as it requires a random number generation for every attempted move, regardless of whether a cell is occupied or empty (see flowchart, Figure 2.c). This method is the suggested method when investigating systems with molecular crowding as it best conforms to the expected behavior of multiple particles in a concentrated environment which can be attributed to the steric repulsion that would prevent multiple particles from occupying the same position in space at a particular time. The error at high concentrations with larger particle sizes in the reduced probability method (RP) models each cell as an elastic box capable of briefly exceeding the cell's maximum volume. However, as shown in Figure 3a, the error in this method is less than 5% at physiologically relevant crowding volumes of 10% - 40% excluded volume [16-17].
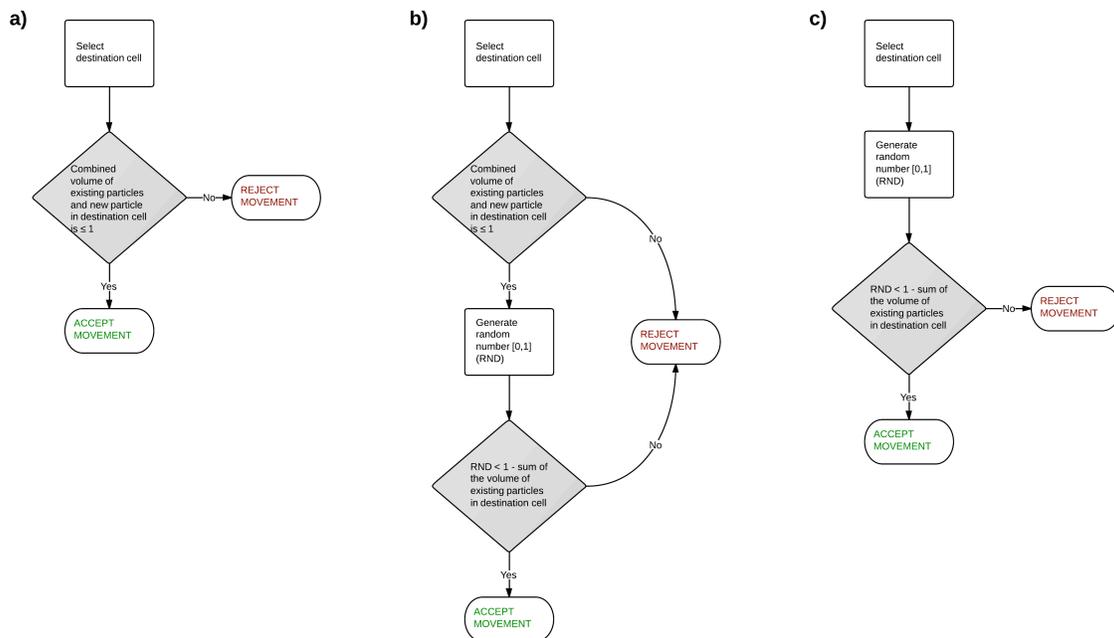
<u>Figure 2.</u> Flowcharts showing the algorithm of three various methods for simulating steric repulsion of multiple particles per cell. a) the Volume Limit (VL) method is the most computationally efficient, b) followed by the Reduced Probability + Volume Limit (RP + VL) method and c) the Reduced Probability (RP) method being the least computationally efficient. The degree of accuracy for which each method models steric repulsion is illustrated in Figure 3.
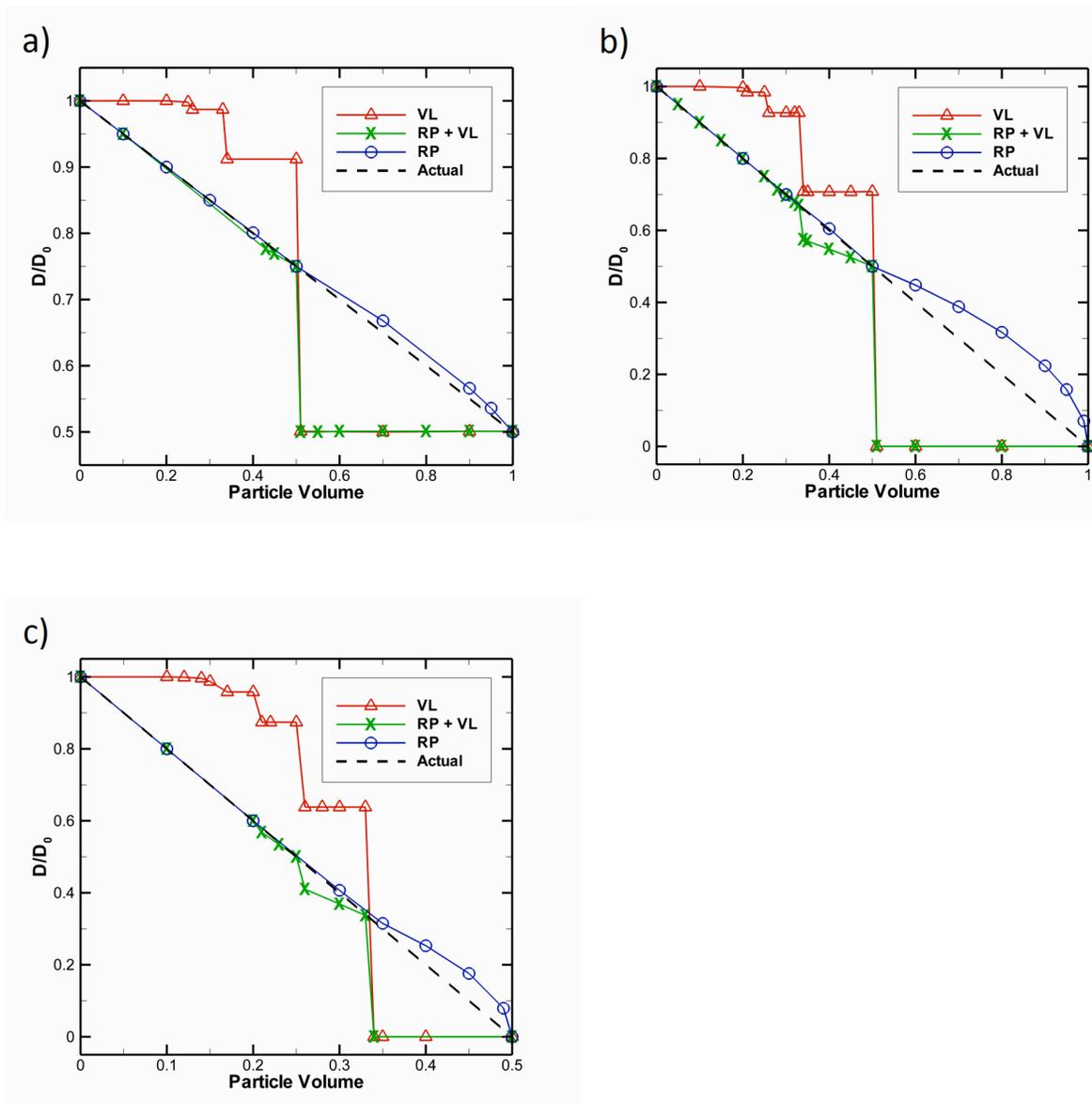
Figure 3. In systems of high concentration a) 500 particles in a system with 1000 cells, b) 1000 particles in a system with 1000 cells and c) 2000 particles in a system with 1000 cells, it can be seen that three different methods for handling the movement of multiple particles per cell result in significantly different behavior. The volume limit method (VL) is the most computationally efficient by simply limiting the movement of particles that would result in the fraction of occupied volume of a cell exceeding 1. However, it is also the least accurate when dealing with crowded environments. The combined reduced probability and volume limit method (RP + VL) is slightly less computationally efficient but is much more representative of crowded diffusion when the particles are of smaller volume. The reduced probability method (RP) is the least computationally efficient of the three but best represents the crowded diffusion for most particle sizes. Additionally, the system with 500 particles in 1000 cells deviates the least from actual when using

74

the RP method while the more crowded systems deviate more, confirming that the RP method can accurately model physiologically relevant concentrations.

Geometry Effects on Diffusion

To demonstrate a biolgoical application of the proposed agent based diffusion method, we have modeled the effect of structural geometry on diffusion directionality. In this model, we show the effect of quasi-random versus parallel structural geometries of filaments similar to what is seen in the structure of cell actin filaments in the form of lamellipodia versus filopodia [18-19]. Actin dynamics are thought to play a key role in cell motility [20-23]. Additionally, it has been shown that the flow of actin monomers in the lamellipodia cannot be explained by diffusion alone and may involve some form of active transport [24]. Moreover, due to the parallel orientation of actin filaments in the filopodia, and their longer length as a result of inhibition of the capping process, the actin monomers required for polymerization of actin filaments of the filopodia must travel a greater distance to where they are needed [19,25-27]. In this model, we investigate how the structural geometry and orientation of these filaments affects the directionality of diffusion of the monomers.
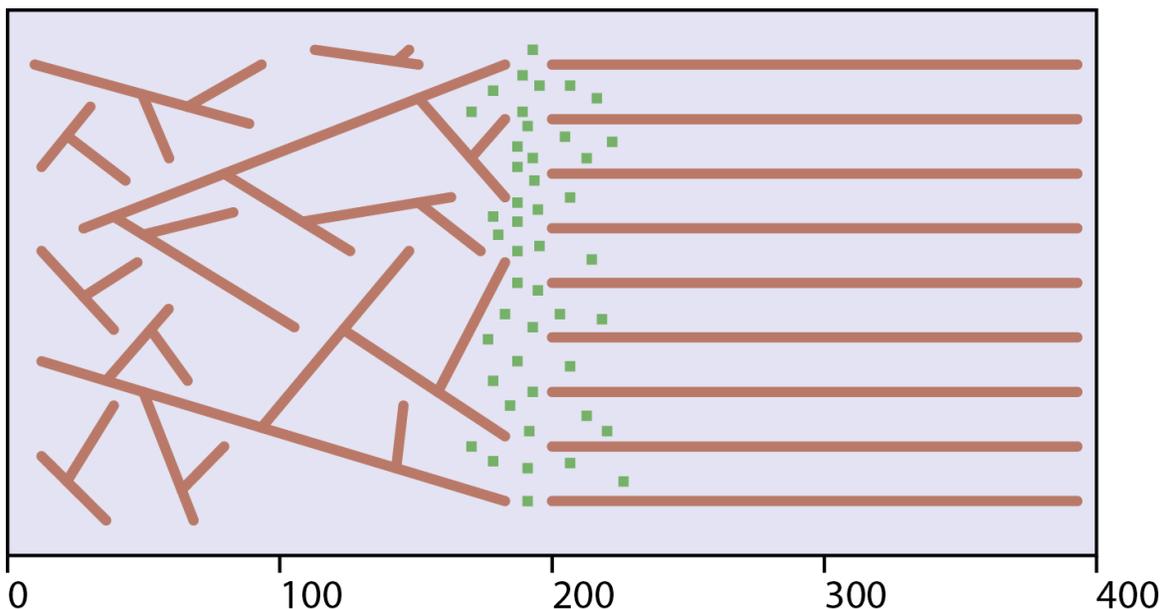
Figure 4. Representatitve cross-sectional illustration of the xy-plane of the three-dimensional simulation box of size $L_X=400nm$, $L_Y=200nm$, and $L_Z=100nm$ with periodic boundary conditions in the y-direction only. The parallel filaments in the right half of the box ($x>200nm$) represent the filopodia while filaments in the left half ($x<200nm$) represent the lamellipodia in the cell. The green particles represent the freely diffusing actin monomers which are distributed in three-dimensional space near $x=200nm$.

75

The model environment consists of a simulation box of size $L_X=400nm$, $L_Y=200nm$, and $L_Z=100nm$ with periodic boundary conditions in the y-direction only. Figure 4 shows an illustration of a representative cross-sectional snapshot of the xy-plane of the simulation box with the right half containing parallel filaments and the left half containing a uniform density of filaments oriented at 68±2 degrees from one another in three dimensional space (prior to discretization) [18,23,28-29]. This configuration was chosen not only to investigate geometry effects on diffusion but more specifically to model actin dynamics at the lamellipodia and filopodia interface.
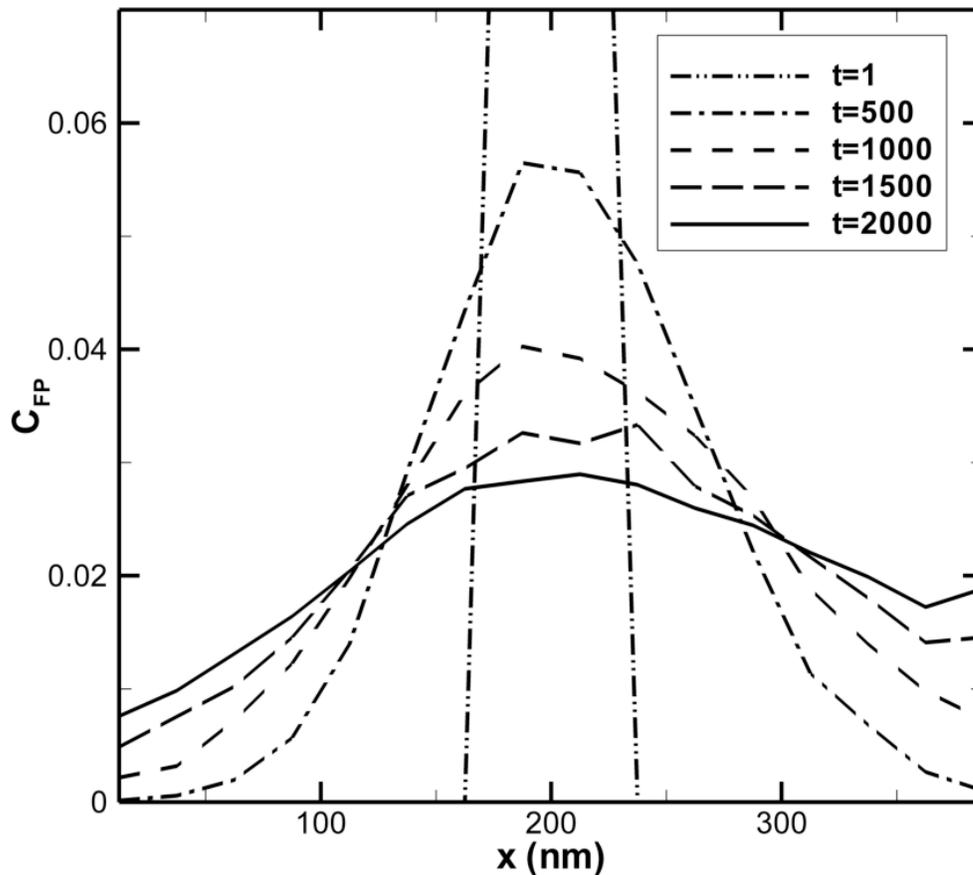
Figure 5. Normalized free particle concentration as a function of position for snapshots of time ranging from 1 to 2000 time steps with a time step increment of 74μs for a fixed actin filament volume density of 0.25 averaged over ten runs. Initially at t=1 the distribution of particles is uniform whereas at each subsequent time step shown, the filopodia region (*x>200nm*) is seen to have a higher free particle concentration than the lamellipodia region (*x<200nm*).

Actin filaments were initially generated in a non-discretized 3D environment with continuous filaments spaced a uniform distance apart and oriented parallel to one another within the filopodia region and conversely, filaments positioned randomly, with uniform density and oriented 68±2 degrees from each other in the lamellipodia region. The continuous actin polymers were then discretized into individual particles representing pairs of g-actin monomers fixed in space that occupy the full volume of each cell. It should be noted that the total number of particles (g-actin monomer pairs) in the lamellipodia and filopodia are equal to avoid obstacle concentration effects. Agent based modeling was used to investigate the effect of various actin filament densities on the directionality of free actin diffusion. As Figure illustrates, free particles diffuse more easily in the direction of the filopodia ($x>200nm$) as opposed to the direction of the lamellipodia ($x<200nm$). The simulation was run using characteristic values for the g-actin monomer diameter and diffusion coefficient of $\Delta L=5nm$ and $D=5.65\mu m^2/s$ [19,30-31] in a three dimensional space with a movement probability of $T=1$ given that only one particle type is diffusing in the simulation and the movement probability is maximized in order to maximize computational efficiency. Subsequently, the time step of the simulation can be determined to be $\Delta t=74\mu s$ using the relationship established in Eq. (1).

In order to show the time progression of concentration differences between the filopodia region and lamellipodia region at different fixed actin filament volume densities, we calculate the ratio of the center of mass of diffusing particles in the filopodia region to that of the lamellipodia. Eq. (4) shows the method used for calculating the center of mass for each region where R represents the center of mass and N_i represents the number of freely diffusing actin monomers at position x_i.

$$R = \frac{\sum_i x_i N_i}{\sum_i N_i} \qquad (4)$$

Figure shows the ratio between the center of mass of particles diffused in the filopodia to that of the lamellipodia as a function of time for different fixed actin filament volume densities using simulation parameters of $\Delta L=5nm$ and $D=5.65\mu m^2/s$ in a three dimensional space with a movement probability of $T=1$ given that only one particle type is diffusing in the simulation. The general trend seen from these results is that there is an initial peak in the tendency of particles to diffuse into the filopodia region (region with parallel filaments) for all fixed particle densities greater than zero. Subsequently, this peak diminishes over time (ratio decreases towards 1) as particles reach the x-direction extremes and begin to distribute uniformly throughout space.

In addition, it can be seen that as the density of fixed actin filaments is increased, the tendency of particles to diffuse towards the region of parallel filaments is only increased. This is most likely a result of random filaments generating a longer path

77

that must be taken from the center of the simulation box to the left extreme whereas the parallel filaments generate the shortest possible distance that can be taken from the center of the simulation box to the right extreme which is effectively a reduction of dimensionality.
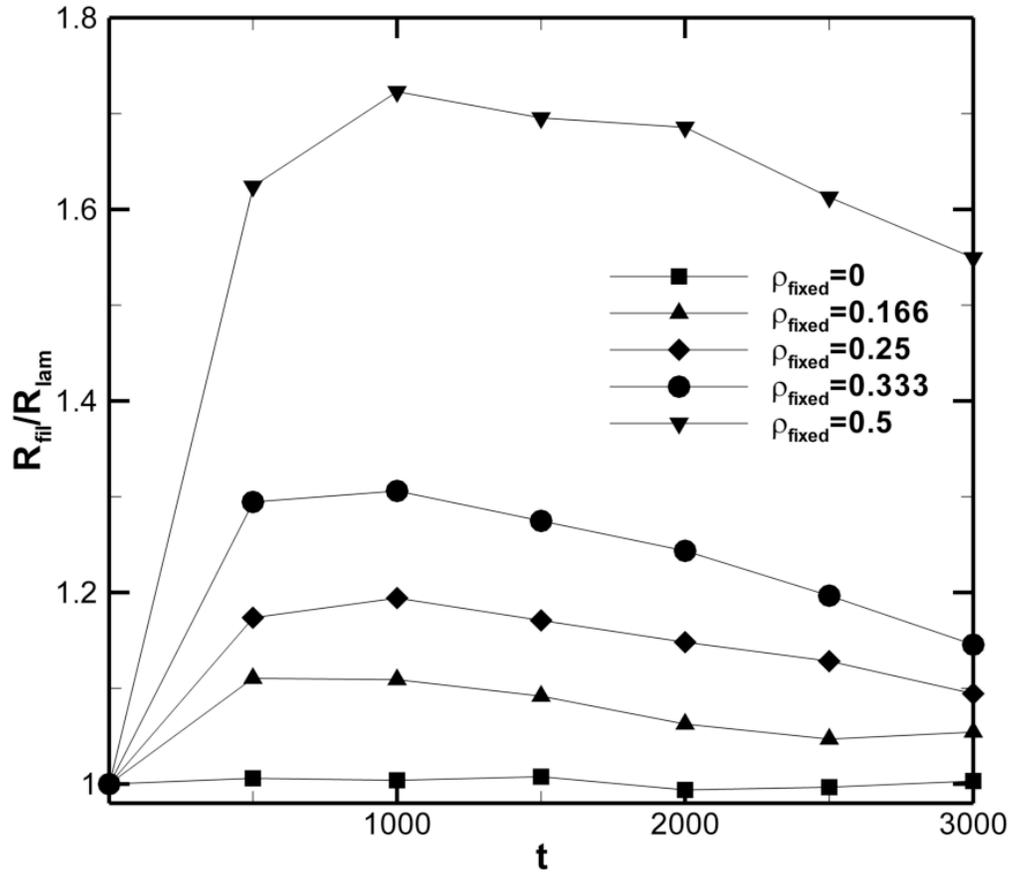
Figure 6. Ratio between the center of mass of particles diffused in the filopodia to that of the lamellipodia as a function of time for different fixed actin filament volume densities. There is a tendency for particles to diffuse towards the filopodia region as a result of the geometry of filaments in each region. This phenomenon is only amplified as the density of fixed actin filaments is increased.

Methods

Relating Diffusion Coefficients to ABM Movement Probabilities

Fick's second law relates the effect of diffusion on the concentration field of particles over time [32]. We can express this relationship in terms of the probability of discretized cells being occupied rather than concentration, Eq. (5), by considering the relationship between concentration and the probability of a cell being occupied by an agent, Eq. (6).

$$\frac{\partial P^{occ}}{\partial t} = D \frac{\partial^2 P^{occ}}{\partial x^2} \tag{5}$$

$$C = \frac{N_{particles}}{N_{cell} V_{element} N_A} = \frac{P^{occ}}{V_{element} N_A} \tag{6}$$

Note that the variables P^{occ}, D, t, and x in Eq. (5) represent probability of finding an occupied cell, diffusion coefficient, time, and position respectively while the variables C, $N_{particles}$, N_{cell}, $V_{element}$, N_A, and P^{occ} in Eq. (6) represent concentration, number of particles, number of cells, volume of each element, Avagadro's number and the probability of finding an occupied cell respectively.

The diffusion term in Eq. (5) is a factor dependent on temperature of the solvent, size and shape of the particle, and viscosity of the solvent that quantifies the ratio of Brownian forces to drag forces. Factors such as force and velocity are not explicitly calculated in a simple agent based model and the coarse discretization of space that limits the direction of movement would make such calculations meaningless. Rather, in an agent based model, diffusion can be simulated by assigning a probability of movement to each particle agent. The relation between movement probability and a physically meaningful diffusion coefficient is derived below.

We consider a one-dimensional lattice with discretized segments of length ΔL as shown in Figure to derive the relationship between a physical diffusion coefficient and a movement probability to be used in our agent based model. We can define the probability of finding a single particle at position X_n at time $t+\Delta t$ as:

$$P_{n,t+\Delta t} = P_{n,t} - P_{n,t}T_R - P_{n,t}T_L + P_{n-\Delta L,t}T_R + P_{n+\Delta L,t}T_L \tag{7}$$

Where $P_{n,t}$, $P_{n-\Delta L,t}$ and $P_{n+\Delta L,t}$ represent the probability of finding the particle at position X_n, $X_{n-\Delta L}$ and $X_{n+\Delta L}$ respectively at time t; T_R and T_L represent the probability of the particle moving to the right or left respectively. Note that unless otherwise noted, all probability terms represent the probability of the respective cell being occupied. Eq. (7) states that the probability of a particle being found at X_n at time $t+\Delta t$ can be determined based on the probability that the particle was initially in that position and remained there (*first term*) less the probability that the particle started

79

in that position and moved to either the right or left cells (*second and third term*) plus the probability that the particle was initially to the left or right of that cell and moved to the right or left respectively (*fourth and fifth term*).
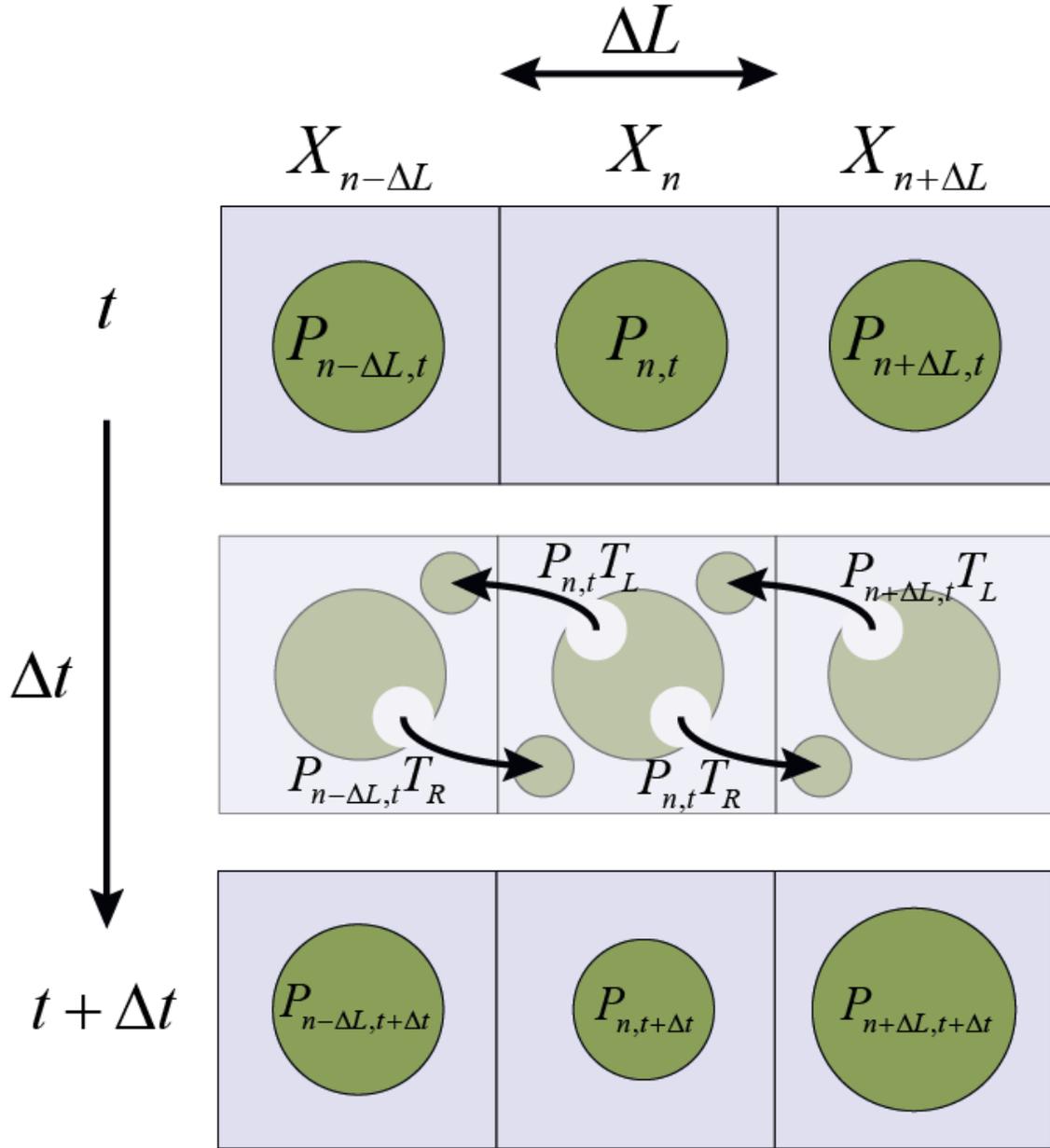
<u>Figure 7.</u> Discretized one-dimensional space with square lattices of length ΔL depicting how probabilities of particles existing in a cell at time t combined with movement probabilities result in a change in the probability of a particle occupying a cell at time $t+\Delta t$ as outlined by Eq. 7. Note that the circles in each cell do not represent individual particles; rather they qualitatively represent probabilities of a particle residing in that cell.

Taylor expansion of the terms in Eq. (7) as $\Delta t, \Delta L \to 0$ gives the following relationship:

$$\frac{\partial P_{n,t}}{\partial t} = \frac{2T}{2} \frac{(\Delta L)^2}{\Delta t} \frac{\partial^2 P_{n,t}}{\partial x^2} \qquad (8)$$

This assumes that movement probability in both directions are equal ($T_L + T_R = 2T$). Eq. (8) relates how the transition probability affects the spatial distribution of particles with time, similar to Eq. (5). Thus we can relate diffusion (D) to movement probability (T), using discretization length (ΔL) and time (Δt) as previously shown in Eq. (1) for diffusion of a particle on a discrete cubic lattice.

Crowding effects on Movement Probability

The computational model and analytical solutions described thus far pertain to the diffusion of a single particle in a discretized space. In addition, it is necessary to validate the model behavior in high particle concentrations to ensure that the model behaves in accordance with physical phenomena. Using the same approach used to determine movement probability for a single particle in Eq. (1), we can analytically derive the effective diffusion coefficient for a high concentration, multiple particle system in a stochastic agent based model. Given the same one-dimensional discretized environment from Figure , we can modify Eq. (7) to now incorporate the effect of multiple particles.

$$P_{n,t+\Delta t} = P_{n,t} - P_{n,t}T_R P_{n+\Delta L,t}^{vac} - P_{n,t}T_L P_{n-\Delta L,t}^{vac} + P_{n-\Delta L,t}T_R P_{n,t}^{vac} + P_{n+\Delta L,t}T_L P_{n,t}^{vac} \quad (9)$$

Where P^{vac} represents the probability of finding the given cell to be vacant. Eq. (9) states that the probability of a particle being found at X_n at time $t+\Delta t$ can be determined based on the probability that the particle was initially in that position and remained there less the probability that the particle started in that position and moved to either the right or left cells plus the probability that the particle was initially to the left or right of that cell and moved to the right or left respectively if cell X_n was vacant plus the probability of the particle in cell X_n attempting to move to the right or left into an occupied cell resulting in the particle remaining in cell X_n. Taylor series expansion of the time and position varying terms along with the relationship that $P^{vac} = 1\text{-}P^{occ}$ gives the solution shown in Eq. (10).

$$\frac{\partial P_{n,t}}{\partial t} = (1 - P^{occ}) \frac{2T}{2} \frac{(\Delta L)^2}{\Delta t} \frac{\partial^2 P_{n,t}}{\partial x^2} \qquad (10)$$

Note that Eq. (10) is similar to the relationship derived for the single particle concentration field and has the same relationship relating movement probability to diffusion coefficient as the single particle with an additional term related to the

probability of cells being occupied by particles. As shown in Eq. (6), this probability of cells being occupied by particles is directly related to the concentration of the system ($C=P^{occ}/V_{element}N_A$).

Model Details

Our agent based model consists of a three-dimensional discretized space that can be bounded or unbounded in which various types of agents diffuse by moving between neighboring cells of cubic shape with a given movement probability that corresponds to the particle's respective diffusion coefficient (Eq. (1)). In this model, we incorporate a von Neumann neighborhood consisting of the six cells orthogonally surrounding an agent in 3D space. Agents in this model can only interact with other agents within their von Neumann neighborhood and can only move in the direction of von Neumann neighborhood cells.

At higher concentrations, two methods for particle movement consisting of an *all-neighbor attempt* and a *single-neighbor attempt* algorithm are assessed. These movement methods differ in that an all-neighbor attempt is an intelligent agent movement procedure in which all von-Neumann neighborhoods are searched at random until an empty cell is found for the agent to move to while in the single-neighbor attempt a von Neumann neighbor is selected at random, disregarding whether it is occupied or vacant. If the cell is occupied the movement is rejected and the agent remains in its current cell while movement into a vacant cell is accepted with some probability correlated to the diffusion coefficient. In this second method, rejected movements represent a collision between two particles resulting in both particles remaining in their respective cells. Figure also shows the process for single-neighbor attempt movements.

The model was developed using object oriented FORTRAN to maximize computational efficiency. Agent based modeling benefits significantly from object oriented programming since the concept of an object is similar to the concept of an agent. Moreover, agent based modeling is very computationally efficient for large systems and long time scales when compared to modeling techniques such as Langevin dynamics as shown in Figure at the cost of reduced spatial and temporal detail. Additionally, the discrete nature of the model makes it an ideal candidate for parallelization and distributed computing, resulting in further computational efficiency [33].
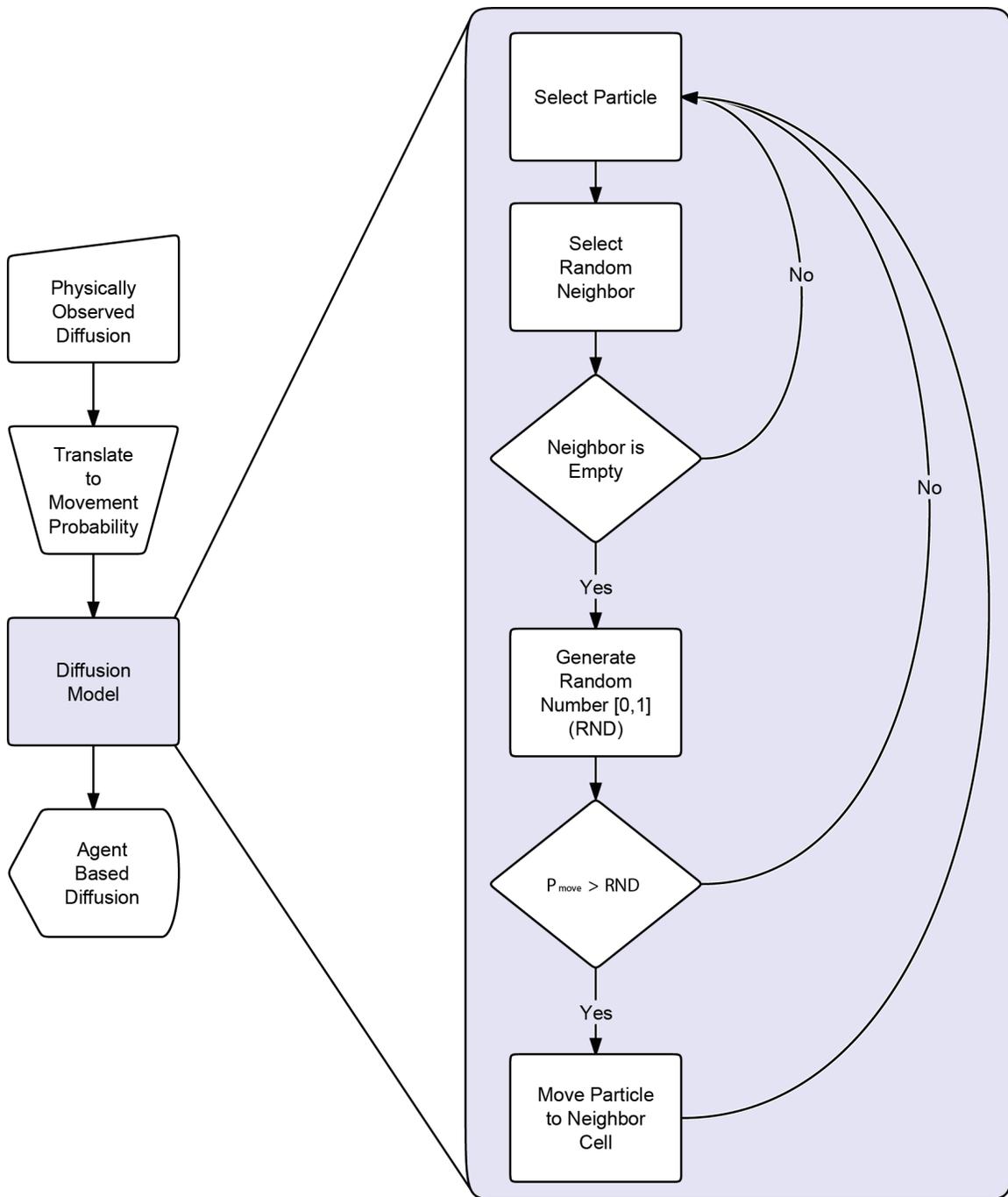
<u>Figure 8.</u> Method for modeling diffusion using physically observed diffusion coefficients (translated to movement probabilities) as an input in an agent based model. Additional details regarding the movement algorithm (specifically the single-neighbor attempt) are illustrated.

	Langevin dynamics	ABM	
# of particles	400		
Diffusion Coefficient	7.1×10^{-7} cm^2/s		
L of simulation box	52 nm		
Simulation time	32 ms		
Δt	0.16 ns	37.5 ns	
# of time steps	~10^8	~10^6	
σ	5 nm	N/A	
Cutoff	1.1 σ*	N/A	
Discretization	N/A	(4 nm)^3	13 x 13 x 13
Particle diameter	5 nm	4 nm	
Volume fraction occ.	0.186	0.182	
COMPUTE TIME	**1.75 CPU hours**	**0.004 CPU hours**	

Figure 9. Benchmark showing compute time between a Langevin dynamics simulation and an agent based model of 400 particles diffusing in the same size simulation box and both simulations having the same fraction of volume occupied. The agent based model used a discretized simulation box of 13 x 13 x 13 with cubic cells of 4nm in length. The Langevin dynamics model incorporated the *Cichocki and Hinsen* method [34] for simulating hard-spheres and utilized further optimizations such as neighbor lists to maximize computational efficiency. *The Langevin dynamics simulation cutoff was set lower than typical to simulate hard spheres to match that of the agent based model. Typical Langevin dynamics models use higher cutoffs resulting in additional computation time.

References

1. Dobrzynski M, Rodriguez JV, Kaandorp JA, Blom JG (2007) Computational methods for diffusion-influenced biochemical reactions. Bioinformatics 23: 1969-1977.

2. Gillespie DT (2007) Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry 58: 35-55.

3. Ridgway D, Broderick G, Ellison MJ (2006) Accommodating space, time and randomness in network simulation. Curr Opin Biotechnol 17: 493-498.

4. Takahashi K, Arjunan SN, Tomita M (2005) Space in systems biology of signaling pathways--towards intracellular molecular crowding in silico. FEBS Lett 579: 1783-1788.

5. Grima R, Schnell S (2008) Modelling reaction kinetics inside cells. Essays in Biochemistry: Systems Biology, Vol 45 45: 41-56.

6. Bonchev D, Thomas S, Apte A, Kier LB (2010) Cellular automata modelling of biomolecular networks dynamics. SAR QSAR Environ Res 21: 77-102.

7. Devillers J, Devillers H, Decourtye A, Aupinel P (2010) Internet resources for agent-based modelling. SAR QSAR Environ Res 21: 337-350.

8. Dong X, Foteinou PT, Calvano SE, Lowry SF, Androulakis IP (2010) Agent-based modeling of endotoxin-induced acute inflammatory response in human blood leukocytes. Plos One 5: e9249.

9. Miller J, Parker M, Bourret RB, Giddings MC (2010) An agent-based model of signal transduction in bacterial chemotaxis. Plos One 5: e9454.

10. Berry BJL, Kiel LD, Elliott E (2002) Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling. Proceedings of the National Academy of Sciences of the United States of America 99: 7187-7188.

11. Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 99 Suppl 3: 7280-7287.

12. Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, et al. (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310: 987-991.

13. Helbing D, Farkas I, Vicsek T (2000) Simulating dynamical features of escape panic. Nature 407: 487-490.

14. Tesfatsion L (2006) Chapter 16 Agent-Based Computational Economics: A Constructive Approach to Economic Theory. In: Tesfatsion L, Judd KL, editors. Handbook of Computational Economics: Elsevier. pp. 831-880.

15. Einstein A (1905) Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annalen der Physik 322: 549-560.

16. Ellis RJ (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. Current Opinion in Structural Biology 11: 114-119.

17. Rivas G, Ferrone F, Herzfeld J (2004) Life in a crowded world. EMBO Rep 5: 23-27.

18. Haviv L, Brill-Karniely Y, Mahaffy R, Backouche F, Ben-Shaul A, et al. (2006) Reconstitution of the transition from lamellipodium to filopodium in a membrane-free system. Proc Natl Acad Sci U S A 103: 4906-4911.

19. Schafer DA (2004) Cell biology - Barbed ends rule. Nature 430: 734-735.

20. Cojoc D, Difato F, Ferrari E, Shahapure RB, Laishram J, et al. (2007) Properties of the Force Exerted by Filopodia and Lamellipodia and the Involvement of Cytoskeletal Components. Plos One 2: -.

21. Koestler SA, Auinger S, Vinzenz M, Rottner K, Small JV (2008) Differentially oriented populations of actin filaments generated in lamellipodia collaborate in pushing and pausing at the cell front. Nature Cell Biology 10: 306-U333.

22. Mitchison TJ, Cramer LP (1996) Actin-based cell motility and cell locomotion. Cell 84: 371-379.

23. Pollard TD, Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. Cell 112: 453-465.

24. Zicha D, Dobbie IM, Holt MR, Monypenny J, Soong DYH, et al. (2003) Rapid actin transport during cell protrusion. Science 300: 142-145.

25. Atilgan E, Wirtz D, Sun SX (2005) Morphology of the lamellipodium and organization of actin filaments at the leading edge of crawling cells. Biophys J 89: 3589-3602.

26. Fletcher DA, Mullins RD (2010) Cell mechanics and the cytoskeleton. Nature 463: 485-492.

27. Mogilner A, Rubinstein B (2005) The physics of filopodial protrusion. Biophys J 89: 782-795.

28. Schaub S, Meister JJ, Verkhovsky AB (2007) Analysis of actin filament network organization in lamellipodia by comparing experimental and simulated images. J Cell Sci 120: 1491-1500.

29. Small JV, Auinger S, Nemethova M, Koestler S, Goldie KN, et al. (2008) Unravelling the structure of the lamellipodium. Journal of Microscopy-Oxford 231: 479-485.

30. Gardel ML, Valentine MT, Crocker JC, Bausch AR, Weitz DA (2003) Microrheology of entangled F-actin solutions. Physical Review Letters 91: -.

31. Koestler SA, Rottner K, Lai F, Block J, Vinzenz M, et al. (2009) F- and G-Actin Concentrations in Lamellipodia of Moving Cells. Plos One 4: -.

32. Fick A (1995) On Liquid Diffusion (Reprinted from the London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Vol 10, Pg 30, 1855). Journal of Membrane Science 100: 33-38.

33. Scheutz M, Schermerhorn P (2006) Adaptive algorithms for the dynamic distribution and parallel execution of agent-based models. Journal of Parallel and Distributed Computing 66: 1037-1051.

34. Strating P (1999) Brownian dynamics simulation of a hard-sphere suspension. Physical Review E 59: 2175-2187.

Section V:

Higher Affinity Importinβ Binding at the Nuclear Basket Increases Nucleocytoplasmic Import

Higher Affinity Importinβ Binding at the Nuclear Basket Increases Nucleocytoplasmic Import

Abstract

Several *in vitro* studies have shown the presence of an affinity gradient in nuclear pore complex proteins for the import receptor Importinβ, at least partially contributing to nucleocytoplasmic transport, while others have historically argued against the presence of such a gradient. Nonetheless, the existence of an affinity gradient has remained an uncharacterized contributing factor. To shed light on the affinity gradient theory and better characterize how the existence of such an affinity gradient between the nuclear pore and the import receptor may influence the nucleocytoplasmic traffic, we have developed a general-purpose agent based modeling (ABM) framework that features a new method for relating rate constants to molecular binding and unbinding probabilities, and used our ABM approach to quantify the effects of a wide range of forward and reverse nucleoporin-Importinβ affinity gradients. Our results indicate that transport through the nuclear pore complex is maximized with an effective macroscopic affinity gradient of 2000μM, 200μM and 10μM in the cytoplasmic, central channel and nuclear basket respectively. The transport rate at this gradient is approximately 10% higher than the transport rate for a comparable pore lacking any affinity gradient, which has a peak transport rate when all nucleoporins have an affinity of 200μM for Importinβ. Furthermore, this optimal ratio of affinity gradients is representative of the ratio of affinities reported for the yeast nuclear pore complex – suggesting that the affinity gradient seen *in vitro* is highly optimized.

Introduction

The Nuclear Pore Complex and Nucleocytoplasmic Transport

Spatial segregation of genetic material in the nucleus from the cytoplasm gives eukaryotes the ability to highly regulate gene expression and DNA replication. By regulating import to and export from the nucleus, the nuclear pore complex (NPC) plays a critical role in cell physiology – enabling rapid yet selective bi-directional flow of material into and out of the nucleus to maintain cellular functions. For example, RNA synthesized in the nucleus must be shuttled to the cytoplasm for protein synthesis while proteins involved in the transcription of these RNAs must simultaneously be shuttled into the nucleus. While small molecules (diameter of <9 nm) passively diffuse across the channel, larger molecules rely on an active (energy dependent) mechanism for transport. This pore's selectivity for active cargo transport is achieved through a combination of structural features and biochemical pathways that lead to a still elusive transport mechanism for which there exist several competing hypotheses.

The NPC itself is a ~60MDa (in yeast) to ~125MDa (in vertebrates) macromolecular assembly composed of multiple copies of ~30 different proteins termed nucleoporins (Nups) that are evolutionarily conserved across eukaryotes and embedded in the nuclear membrane with eight-fold radial symmetry [1-4]. These Nups form eight cytoplasmic filaments that protrude from the nuclear envelope into the cytoplasm and another eight that project into the nucleus and are bound by a ring at their distal end to form a basket. The pore is anchored to the nuclear envelope by a membrane layer that surrounds the scaffold layer (Fig. 1). This scaffold layer provides structure and serves as an anchor for Nups that contain both structured domains as well as highly unstructured domains – rich in phenylalanine-glycine repeats –that are believed to be principally responsible for selective transport (FG-Nups). The FG-rich regions of these Nups present an affinity for hydrophobic patches present on transport receptor proteins involved in shuttling cargo across the nuclear envelope. The mechanism by which these Nups regulate transport remains a topic of much debate and has lead to the proposal of several competing models such as Brownian affinity gating [2,5,6], selective phase [7-11], affinity gradient [12,13], reversible collapse [14-16] and reduction of dimensionality [17], among others. Additional details of NPC structure and function can be found in our recent review [18].

While these models aim to explain the pore's selectivity, other models were proposed to resolve contributors to transport directionality. The affinity gradient model emerged as a plausible explanation as a result of *in vitro* measurements in yeast and vertebrae demonstrating the presence of an increasing affinity gradient in Nups moving from the cytoplasm to the nucleus – suggesting that this affinity gradient provided transport directionality to cargo bound to transport receptors [12,13]. Other studies found that the source of directionality of import complexes was a result of the steep RanGTP gradient present across the pore with high

concentration of RanGTP in the nucleus, stemming from the presence of the nucleotide exchange factor RanGEF (Ran Guanine nucleotide Exchange Factor) and low concentrations of RanGTP in the cytoplasm due to hydrolysis via RanGAP (GTPase activating protein). These studies showed that reversal of the RanGTP gradient across the pore resulted in the reversal of transport directionality in spite of any affinity gradient [19]. Nevertheless, the contribution of the observed affinity gradient or a partial affinity gradient to transport efficiency remains unexplored with traditional experimental methods. More specifically, it is unclear whether (*i*) the presence of a Nup-Impβ affinity gradient affects nucleocytoplasmic transport rate and whether (*ii*) there exists an affinity gradient that can optimize transport rate beyond that of the reported *in vitro* gradient. To answer these questions, we have developed an agent based model to perform *in silico* measurements of Impβ translocation across the NPC.
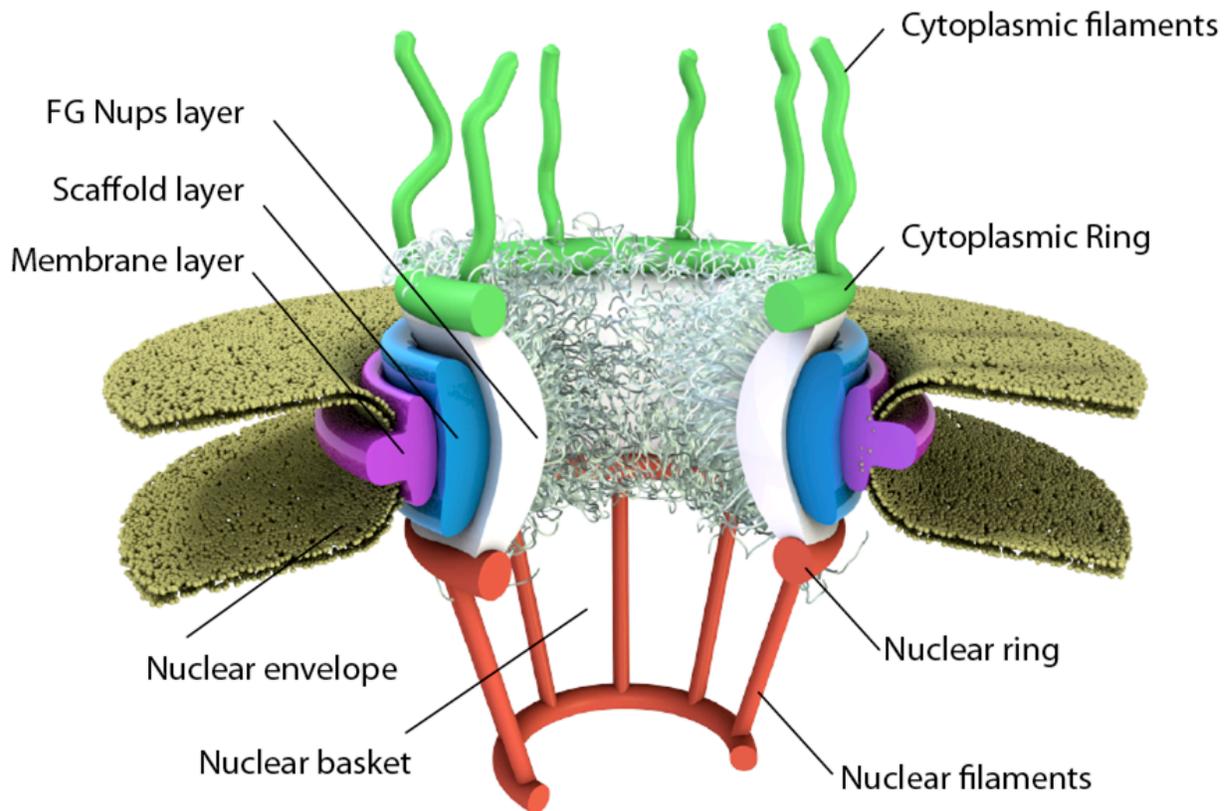
Figure 1. Schematic of the nuclear pore complex. The pore is anchored to the nuclear envelope by a membrane layer that surrounds the scaffold layer. This scaffold layer provides structure and serves as an anchor for Nups that contain both structured domains as well as highly unstructured domains that are thought to form a barrier that excludes non-interacting molecules while allowing for selective transport of others. This central channel exhibits eight-fold rotational symmetry and has eight cytoplasmic filaments as well as eight nuclear filaments protruding into the

cytoplasm and nucleoplasm respectively. The nuclear filaments are bound via a ring, resulting in a basket structure.

Agent Based Modeling

Agent based modeling (ABM) is a robust computational technique used to simulate the spatiotemporal actions and interactions of real-world entities or "agents", in an effort to extract their combined effect on the system as a whole. Both space and time can be discretized in an ABM, giving these autonomous agents the ability to move and interact with other agents and their environment at each timestep over a given duration. Simple behavioral rules govern the movement and interaction of each individual entity in an effort to reproduce or predict more complex behaviors of multiple entities. Such a model attempts to simulate the emergence of complex phenomena that may not be apparent when simply considering individual entities. Agent based modeling has seen applications in a broad range of fields ranging from artificial intelligence and gaming to modeling emergent social behavior such as the spread of disease and outcomes of financial markets [20-23]. In their simplest form, on-lattice agent based models consist of a mesh of "cells" that make up the discretized space that agents occupy. The agents occupy these cells and are typically only aware of other agents within their "neighborhood"; in the simplest form a neighborhood consists of adjacent cells. Agents are given the ability to move into adjacent cells and to interact with other agents with some probability in conjunction with governing rules that define what movement and interactions are possible (Fig. 2). On-lattice agent based models have previously been applied to biological systems involving diffusion, binding and unbinding [24-26]; establishing methods for event probability selection – relating diffusion and rate constants to event probability – will improve model accuracy and enable quantitative analysis of results from these models [27,28].

In the present work we develop a method for relating real world rate constants to molecular binding and unbinding probabilities within the agent based model. We then build upon our ABM framework [27] to explore the role of an affinity gradient between Nups and the nuclear transport factor, Impβ in nucleocytoplasmic import efficiency. We model the system using affinity gradients derived from *in vitro* experiments and compare these to NPCs lacking affinity gradients as well as a wide range of forward and reverse affinity gradients in order to address the following questions: (*i*) Does the presence of a Nup-Impβ affinity gradient affect transport rate? (*ii*) Does there exist an affinity gradient that can optimize transport rate beyond that of the reported *in vitro* gradient?

To answer these questions, simulations were carried out using a computationally efficient, spatiotemporally detailed, three-dimensional agent-based model developed specifically for modeling molecular diffusion, binding and unbinding events with consideration for physical factors such as molecular crowding and steric

repulsion. In addition to movement and interaction rules, event probabilities govern system dynamics in the agent-based model. Methods for accurate selection of movement, binding and unbinding probabilities to best represent actual diffusion coefficients and kinetic rate constants can build confidence in the output of agent based models and deductions from these models. The procedure for relating real world rate constants to molecular binding and unbinding probabilities is detailed in the Materials and Methods.

Figure 2. Simplified representation of the agent based model. Abstract cartoon representation of the nuclear pore structure environment (not to scale) projected onto a simplified, 2-dimensional, on-lattice ABM with agents representing proteins that move within the system and interact with other agents within their von-Neumann neighborhood. The actual model consists of a three-dimensional representation of the NPC structure and physiologically relevant concentrations of biochemical factors and channel dimensions. In our model, the purple region representing the cytoplasmic periphery is treated as a compartmentalized volume containing non-interacting Nup and Impβ-interacting FG-Nup agents. Similarly, central channel (blue) and nuclear basket (green) regions are represented by compartmentalized volumes, containing both non-interacting and interacting Nup agents at physiologically meaningful concentrations. Grey regions of the diagram represent the scaffold and nuclear envelope regions of the model that are impermeable to diffusing species.

Materials & Methods

Probability Selection for Molecular Movement on an ABM Lattice

In our previous work we proposed a method for movement probability selection based on molecular diffusion coefficient along with algorithms for realistic consideration of crowding and steric repulsion [27] that was also used in the current model:

$$P_{move} = \frac{D \cdot \Delta t}{\Delta L^2}$$ (1)

Here, movement probability of an agent is determined by its diffusion coefficient (D), simulation timestep (Δt) and lattice discretization length (ΔL). We implemented the *reduced probability (RP)* method to account for the steric effects of multiple agents occupying individual lattice sites [27].

Probability Selection for Molecular Binding and Unbinding Events on an ABM Lattice

The simpler case of the molecular unbinding event, which is representative of a first-order unimolecular reaction, can be modeled in the ABM using an unbinding probability, for which derivation of the relationship between kinetic rate constant and probability is trivial. The reversible binding of two molecules *A* and *B* is given in Eq. (2), followed by the rate law for the unbinding event as a function of number of bound molecules within the volume of interest (V):

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB$$ (2)

$$\frac{\partial}{\partial t}\left(\frac{N_{AB}}{V \cdot N_{Avagadro}}\right) = -k_{off}\frac{N_{AB}}{V \cdot N_{Avagadro}}$$ (3)

Change in the number of bound molecules (∂N_{AB}) is a function of elapsed time (∂t), kinetic rate constant (k_{off}) and initial number of bound molecules (N_{AB}) (Eq. (3)). Subsequently, the probability that two bound molecules become unbound is independent of interaction with other molecules; this unbinding probability (P_{off}) is shown in Eq. (6) in the limit of very small Δt.

$$\partial N_{AB} = -\partial t \cdot k_{off} N_{AB} \qquad (4)$$

$$\Delta N_{AB} = -P_{off} N_{AB} \qquad (5)$$

$$P_{off} = k_{off} \Delta t \qquad (6)$$

For molecular binding of two molecules, representative of a second order reaction between adjacent molecules on a lattice - factors such as number of lattice neighbors and lattice size must be considered. This relationship can be derived from the second order rate law as a function of number of unbound molecules within the volume of interest as shown in Eq. (7).

$$\frac{\partial}{\partial t}\left(\frac{N_{AB}}{V \cdot N_{Avagadro}}\right) = k_{on} \frac{N_A}{V \cdot N_{Avagadro}} \frac{N_B}{V \cdot N_{Avagadro}} \qquad (7)$$

Similar to the case of unbinding, change in number of bound molecules can be expressed as a function of binding events in terms of number of unbound molecules (N_A and N_B) in the volume and their binding kinetic rate constant (k_{on}) (Eq. (8)) as well as a function of binding probability (P_{on}) and the probability of finding neighboring binding molecules within the lattice system ($P_{A-B-neighboring}$), Eq. (9).

$$\partial N_{AB} = \frac{k_{on} N_A N_B \partial t}{V \cdot N_{Avagadro}} \qquad (8)$$

$$\Delta N_{AB} = P_{on} P_{A-B-neighboring} \qquad (9)$$

The likelihood of finding two unbound agents, *A* and *B*, neighboring each other on the lattice ($P_{A-B-neighboring}$) is a function of the number of unbound *A* molecules (N_A), number of unbound *B* molecules (N_B), number of lattice cells (N_{cells}), where number of lattice cells in the system is much larger than the number of unbound molecules and the number of lattice neighbors each cell has ($N_{neighbors}$), Eq. (10).

$$P_{A-B-neighboring} = \frac{N_A N_B}{N_{cells}} N_{neighbors} \qquad (10)$$

Subsequently, the probability of a binding event between two neighboring molecules (P_{on}) can be derived by solving for the likelihood of neighboring binding molecules in the system ($P_{A-B-neighboring}$) and combining with Eqs. (8) and (9) as shown in Eq. (11).

$$P_{on} = \frac{k_{on}\Delta t}{V/N_{cells} \cdot N_{neighbors} \cdot N_{Avagadro}} \tag{11}$$

The general form of the probability of binding provided in Eq. (11) is valid for two molecules of different types. For binding events consisting of two molecules of the same type, the probability is reduced by half. Furthermore, Eq. (11) represents the case where the lattice is restricted to containing a single molecule per cell. In the case where multiple smaller molecules can occupy a single cell ($V_A + V_B \leq 1$), a correction factor (α) must be added to the number of neighbors since two unbound agents within the same cell can bind one another. Eq. (12) provides an approximation for the correction factor as a function of the sum of molecular volumes and cell volume.

$$\alpha \approx 1 - \frac{V_A + V_B}{V_{cell}} \tag{12}$$

It should be noted that timestep selection is governed by the smallest of all time scales associated with diffusion or molecular interaction. In other words, the simulation timestep should be selected in a manner so that movement or binding/unbinding event probability does not exceed a value of one.

Comparison of On-Lattice ABM Method to Deterministic ODE Solution

The agent based modeling framework used was an extension of the framework described in our previous publication, which explored methods for accounting for diffusion in agent based models of reaction-diffusion. The combined Reduced-Probability & Volume-Limit (RP+VL) method was used to govern diffusion behavior [27]. Binding events occurred between neighboring agents or agents within the same lattice point with probabilities as defined in Eq. (11), while unbinding events occurred with probabilities as defined in Eq. (6). Binding and unbinding rules were executed in random order for each agent type at each timestep to avoid the possibility of biasing a particular agent type to a specific bound or unbound state.

In order to validate these methods, we used the binding relation given in Eq. (11) to relate rate constants to event probabilities by modeling a system consisting of an initial concentration of 3mM (2000 molecules in the well-mixed volume) molecules of type A that undergo a irreversible binding event, $A + A \rightarrow B$. We compare the time-course data of the model using a deterministic ordinary differential equation (ODE) solver to that of the stochastic agent based model solution in Fig. 3 for multiple rate constants. The ABM solution reproduces the average behavior of the ODE solution without the unnatural smoothness that is seen in the deterministic model.
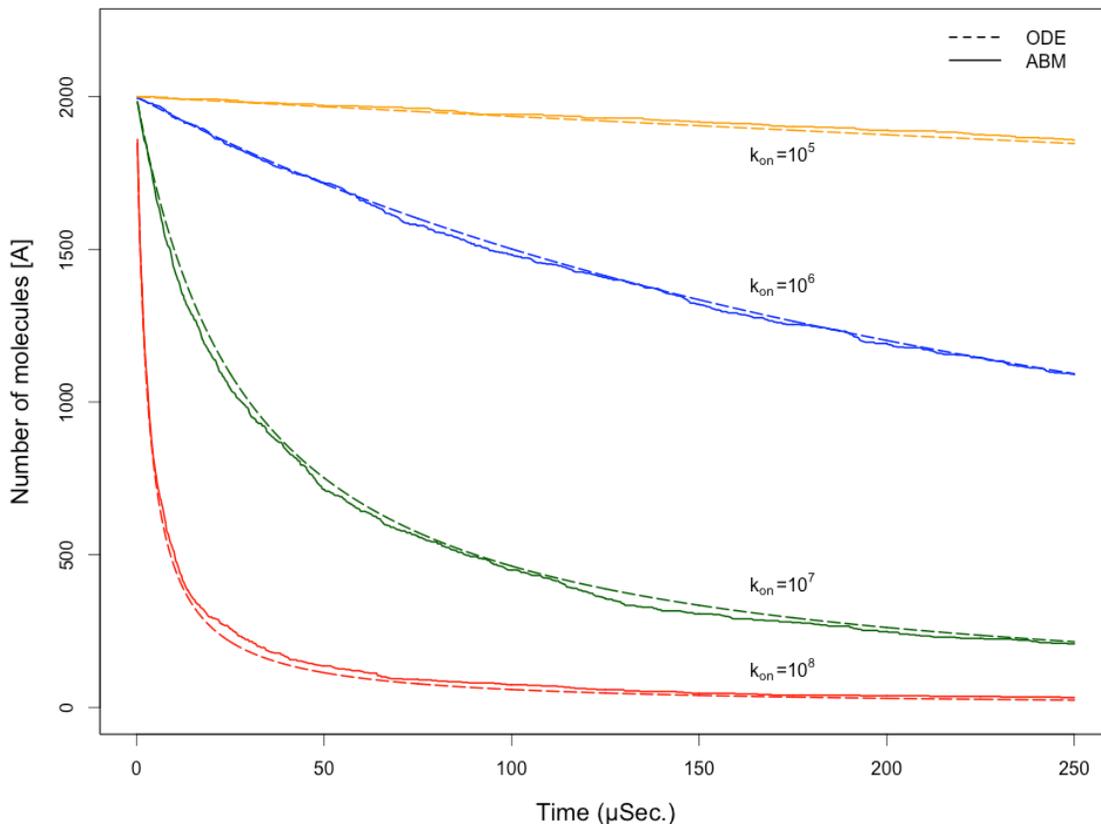
Figure 3. Comparison of ABM and ODE time course data. Comparison of time-course data from an agent based model of molecular binding to that of the numerical solution of the ordinary differential equation for the same event. Probability selection using the relationship in Eq. 11 produces similar behavior to that of the numerical solution in a well-mixed system at multiple rate constants with the addition of stochasticity that is expected from natural systems.

ABM System and Simulation Details

The model environment consists of a 42,108 element, three-dimensional lattice comprised of elements with dimensions of 5nm x 5nm x 5nm. The lattice size was selected to accommodate the volume associated with the Stokes radius of the largest single-agent species in the system, in this case Impβ (Nups being represented by a collection of multiple agents). Additionally, the model allowed for multiple agents of the same or different species type to occupy the same lattice element at any given time, so long as the available volume of a lattice element was not exceeded by agents diffusing into it. Discrete lattice elements belong to one of six region types, cytoplasmic, nuclear membrane, nucleoplasm, cytoplasmic filament periphery, central channel or nuclear basket. The cytoplasmic region contains Impβ molecules at a steady state concentration of 2.5μM while the nucleoplasm contains RanGTP

97

molecules at a steady state concentration of 1μM throughout the simulation [29]. The 35nm thick nuclear membrane which partitions the two compartments is impermeable to all agent types and contains a single nuclear pore with diameter of 30nm at the center and 50nm at the peripheries. The cytoplasmic filament periphery consists of a 50nm diameter region that extends 30nm into the cytoplasm while the nuclear basket is composed of a basket shaped region that extends 55nm into the nucleoplasm [4,30]. The cytoplasmic periphery, central channel and nuclear basket each contain 24, 80 and 32 agents respectively, which represent the distribution of FG Nups [31]. In addition to these FG agents, non-FG agents are added to the channel to represent regions of the Nups that lack affinity for Impβ but play a role in sterically repelling molecules, with the sum of the volume of these Nups corresponding to experimentally reported volumes [31]. Impβ and RanGTP agents are free to diffuse throughout the system while FG agents and non-FG agents are restricted to movement within their respective pore regions in order to maintain the permeability barrier. Impβ and RanGTP agents bind with an on rate of 9.6×10^4 μM^{-1}sec^{-1} [29] while Impβ binds and unbinds FG-Nup agents with a dissociation constant that was varied from 200nM to 2mM [32].

For simulations where an influx rate was reported (units: Sec^{-1}), the system was initially brought to steady state. Subsequently, an *in silico* microinjection of inert molecules ranging in size from a Stokes radius of 0.531 nm to 2.819 nm was administered for validation of the model's ability to reproduce the size exclusion properties of the nuclear pore as demonstrated in prior experiments [33]. In addition, the same microinjection was performed using Impβ to compare the model's influx rate for the karyopherin to what had been reported in experiments from other groups [29]. Simulations were run for a length of 25 to 75 seconds using a timestep of 2.5×10^{-6} seconds with 100 independent stochastic ABM simulations performed for each configuration. Time course concentration data for each configuration was averaged over the 100 independent simulations and data points were fitted to Eq. (13) for comparison with experimental values reported by Mohr and colleagues [33].

$$c(t) = c_{max}(1 - e^{k \cdot t}) \qquad\qquad (13)$$

Where *c* represents concentration of microinjected species and *k* represents the influx rate of a given species into the nucleus, which follows first-order kinetics [33].

For simulations where Impβ transport rate was reported (units: molecules/Sec), simulations were run for a length of 2.5 seconds using a timestep of 2.5×10^{-6} seconds. The system was allowed to reach steady state in the first 1 second and linear regression was performed on the remaining 1.5 seconds of simulation data to quantify the number of Impβ transported to the nucleoplasm as a function of time. The transport rate for each set of 100 simulations was averaged and reported along with standard error.

Importinβ Multivalency and Nup-Impβ Affinity

The transport receptor Impβ has been shown to contain multiple hydrophobic patches that serve as FG binding sites and are believed to play a critical role in the shuttling of Impβ and cargo transport [34]. It has been shown experimentally that Impβ contains two FG binding sites near the N terminus (between HEAT repeats 5 and 6, another between HEAT repeats 6 and 7) as well as two FG binding sites near the C terminus (between HEAT repeats 14 and 15, another between HEAT repeats 15 and 16) [35-37]. Furthermore, Isgro and Schulten identified up to six additional FG binding sites on Impβ using computational methods [38]. Subsequently, experimentally determined macroscopic affinities (or multivalent affinities) are the result of combined microscopic affinities (or monovalent affinities) between FG repeats and the multiple binding sites on Impβ [32].

In our agent-based model, FG-Nups and Impβ bind and unbind with a single probability that corresponds to the macroscopic affinity and represents the combined effect of microscopic affinities – a simplification that is common to other models with coarse granularity [39-41]. This simplification is further justified when considering that: (*i*) the high concentration of FG repeats that a Nup presents to Impβ (~150mM) results in the receptor strongly tending toward the fully bound state [32], and (*ii*) consideration that the on-rate for association of Impβ with FG-Nups has been approximated to be 10^7-10^8 M^{-1}s^{-1} [12], which suggests that the receptor reaches a fully bound state within the span of the timestep used in our simulations (2.5×10^{-6} seconds). Finally, coarse-grained Brownian dynamics simulations of cargo transport through the pore have confirmed that once Impβ is hydrophobically engaged with the pore, its monovalent binding sites become fully saturated at the timescales used in this agent based model [42,43].

Model Validation

In silico experiments were performed to determine the model's ability to recapitulate experimentally-determined, size-dependent permeabilities for passive cargos as well as for Impβ [29,33]. These simulations served as a control to validate the model and associated algorithms' ability to simulate selective transport. Following a simulated microinjection of non-interacting species in the cytoplasm, the *in silico* pore is observed to inhibit the influx of larger species while allowing smaller species to diffuse through the pore (shown in Fig. 4). Influx rates of non-interacting species with Stokes radii of ~1nm are on the order of 0.1s^{-1} while larger species with Stokes radii of >2.5nm have influx rates of less than 0.001s^{-1}. As expected, a reduced influx rate was not observed for larger species that had affinity for the FG-Nups. To test this behavior, experiments similar to those performed for non-interacting species were repeated, replacing the non-interacting species with 2.5μM labeled Impβ – in addition to the steady-state concentration of unlabeled Impβ. Influx of Impβ into the nucleus was measured for 100 simulations, averaged

and fit to Eq. 13. Influx rates of 0.367s⁻¹ and 0.391s⁻¹ were observed for a pore with uniform Nup-Impβ affinity of 200μM and a pore with the optimal Nup-Impβ affinity gradient respectively. These values are comparable with experimentally measured influx rates of $0.4s^{-1}$ for Impβ [29].
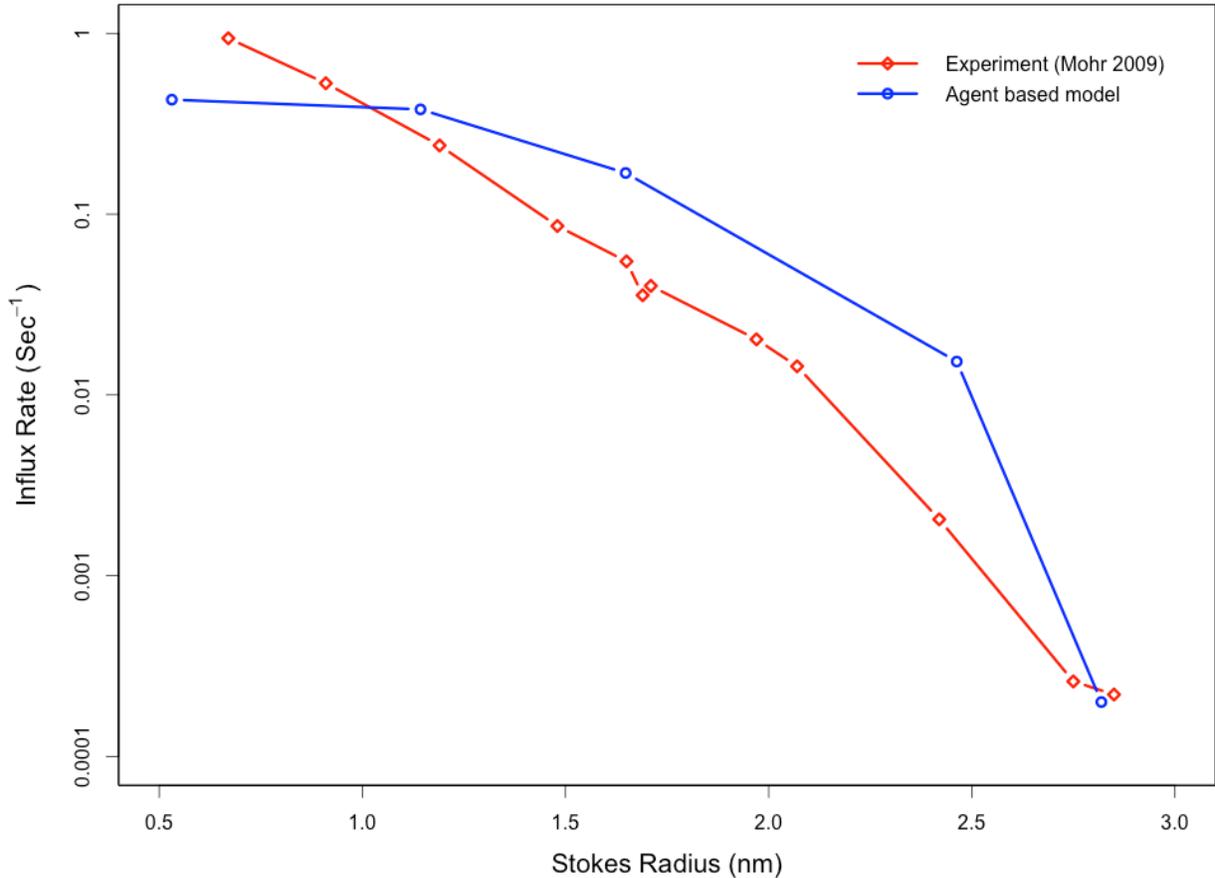
Figure 4. The agent based model recapitulates the experimentally observed size-dependent permeabilities of passive cargos through the nuclear pore. Following a simulated microinjection of non-interacting species in the cytoplasm, the *in silico* pore is observed to inhibit the influx of larger species while allowing smaller species to diffuse through the pore. This is in agreement with previous experimental observations.

Finally, the model's sensitivity to various simulation parameters was explored. Previous studies have proposed that import receptors are the rate limiting species in the import pathway [44,45]. Using the model, we observed a ~50% increase in Impβ shuttling when cytoplasmic Impβ steady-state concentration was increased two-fold. Conversely, there was no significant increase in Impβ shuttling as nuclear RanGTP concentrations were increased up to ten-fold (shown in Fig. 9). These findings are discussed in further detail below. In addition to assessing sensitivity to biochemical species, we assessed the model's sensitivity to the simulation timestep

used. Cutting the simulation timestep in half (from $2.5\text{x}10^{-6}$ seconds to $1.25\text{x}10^{-6}$ seconds) resulted in an average change in absolute Impβ shuttling rates of $5.5\pm3.4\%$ for a range of Nup-Impβ affinities. Nevertheless, the overall relationship between Nup-Impβ affinity and Impβ transport rate were preserved, regardless of the selected timestep.

Results & Discussion

Transport in the Absence of an Affinity Gradient

Initially, Nup-Impβ affinity was kept homogenous throughout the NPC's three regions – cytoplasmic periphery, central channel and nuclear basket – to investigate how a uniform binding affinity affects transport rates. The results of these simulations display a biphasic behavior in transport rate as channel affinity is increased from 100nM to 4mM, with peak transport rate of 86.241±1.68 Impβ translocations per second observed for a channel with a dissociation constant of 200μM (Fig. 5). Recent Brownian dynamics models of single-cargo transport have shown a similar biphasic behavior in transport time as a function of Nup-cargo-complex affinity [42]. It is worth mentioning that since Impβ is the only transportin considered in this model, relative transport rates are of more interest than absolute transport rates (absolute transport rates of Impβ may differ when considering the effects of competing transport receptors).
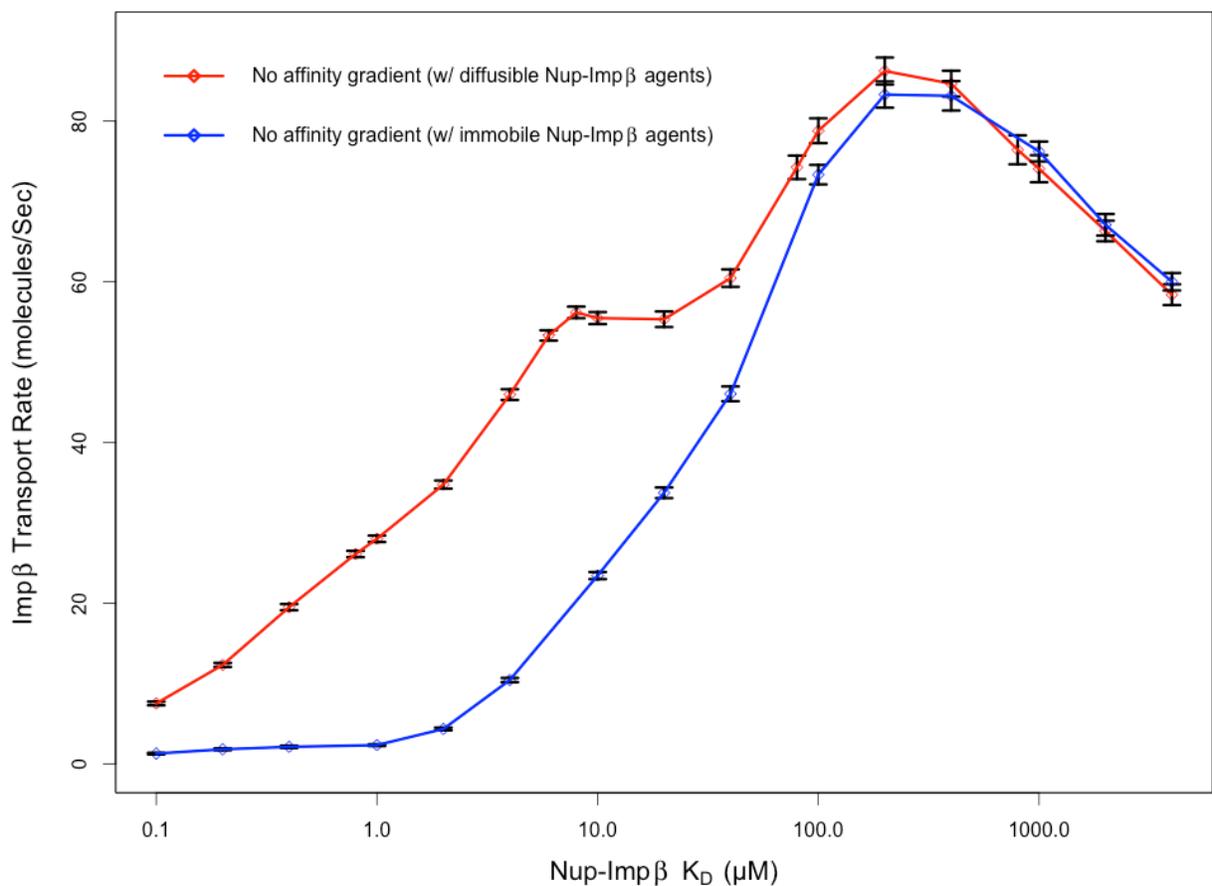
Figure 5. Impβ transport rate through a pore with Nups of uniform affinity. Transport rates for ABM simulations of Impβ through a nuclear pore containing Nups with uniform affinity (no gradient). Nup-Impβ affinity is varied from 100nM to

4mM. The transport rate exhibits biphasic behavior as a function of affinity. At very high affinities (low K_D), Impβ is tightly bound to Nups, resulting in slow transport rates as the Nups become saturated. At very low affinities, Impβ isn't able to bind Nups as efficiently, reducing its resident time at the pore periphery and subsequently excluding it from the pore interior as a result of steric effects. Peak transport of 86.24±1.68 transports per second were observed at a Nup-Impβ affinity of 200μM. Pores containing Nup bound Impβ agents that are capable of diffusing locally exhibit increased transport rate compared to simulation configurations where Impβ becomes immobile once bound to an FG-Nup.

The presence of an optimal binding affinity for transport through a uniform channel is not surprising. As expected, very low dissociation constants will result in increased binding time between molecules, resulting in reduced mobility. Conversely, high dissociation constants will result in a reduction in the time that Impβ is bound to the channel, similar to a molecule with no binding affinity, relying on diffusion alone to traverse the Nup-obstructed channel. It was observed that the optimal dissociation constant of 200μM for the uniform affinity pore was the same as the optimal Nup-Impβ affinity within the central channel of the pore with affinity gradient (Fig. 6 and Fig. 7). Coincidentally, this is also the region with the highest relative FG motif density [31].

The saddle point observed in the plot of Impβ transport rate as a function of Nup-Impβ affinity is due to the motility of Impβ when bound to high affinity Nups (Fig. 5 – No affinity gradient w/ diffusible Nup-Impβ agents). To confirm this, we repeated the simulations but configured the system so that once an Impβ agent bound to an FG-Nup agent, the complex had a movement probability of zero. In this configuration, the Impβ transport rate decreased sharply as Nup-Impβ affinity was increased past 100μM (Fig. 5 – No affinity gradient w/ immobile Nup-Impβ agents). Our assumption that Nup-Impβ complexes are locally diffusive is based on the presence of hydrophobic FG-pockets, which bind Impβ on unstructured regions of the Nups. Furthermore, Brownian dynamics models demonstrate that cargo complexes continue to exhibit diffusive movement when bound to Nups lining the nuclear pore, albeit at a lower rate than their unbound state [42]. The combination of an increase in time bound between Nup and Impβ at higher affinities and the complex's local mobility leads to an increase in likelihood that the Impβ overcomes the Nup-dense central channel. However, the motility of Nup-bound Impβ agents has no effect on our primary parameter of interest – the Nup-Impβ binding affinity at which peak transport is observed.
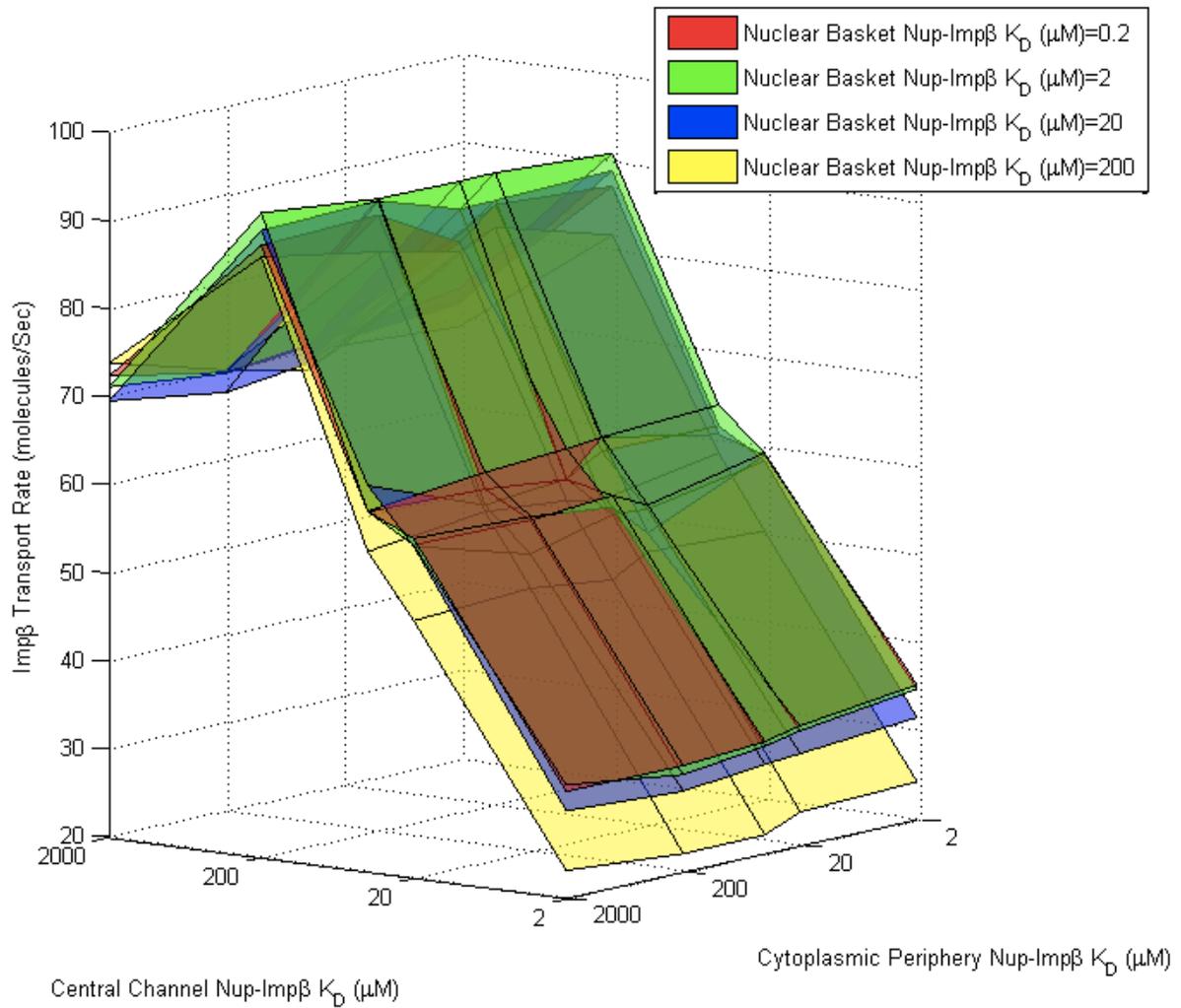
Figure 6. Transport rate as a function of cytoplasmic, central channel, and nuclear basket Nup-Impβ affinity. Impβ transport rate (z-axis) as a function of cytoplasmic (x-axis) and central channel (y-axis) Nup-Impβ affinity ranging from 2μM to 2mM. The four three-dimensional surfaces represent a range of nuclear basket affinities ranging from 0.2μM to 200μM. Transport rates appear to be least sensitive to cytoplasmic affinities and most sensitive to central channel and nuclear basket affinities. Varying central channel affinities results in biphasic behavior with maximum transport at $K_D \approx 200$μM. Transport rates appear to increase as nuclear basket affinity is increased up to $K_D \approx 10$μM and don't appear to show significant increase at higher affinities.
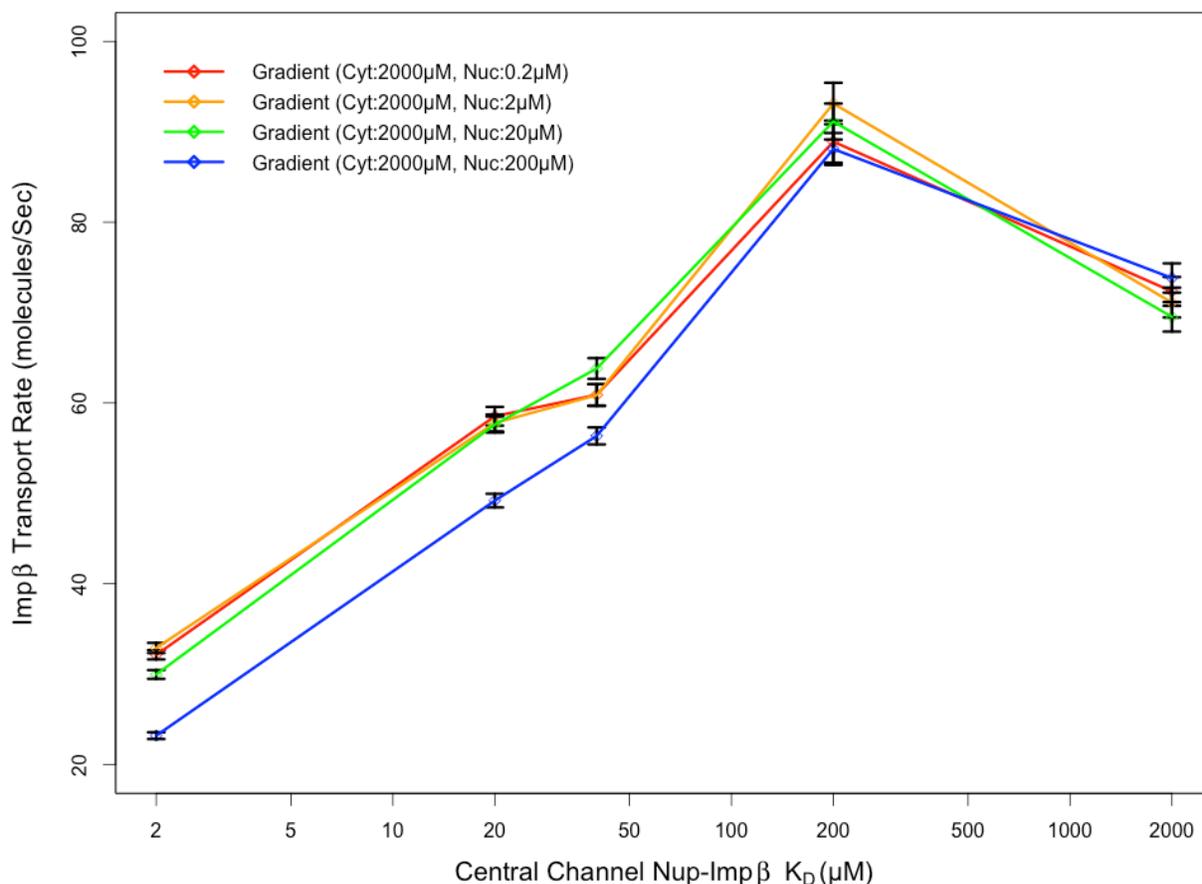
Figure 7. Transport rate as a function of central channel affinity. Impβ transport rate appears most sensitive to central channel affinity, regardless of nuclear basket affinity, with a peak transport rate when Nup-Impβ affinities are on the order of 100μM. (Cyt: Cytoplasmic periphery, Nuc: Nuclear basket)

Simulation-Derived Transport Rates for Experimentally Measured Affinity Gradients

The model was subsequently modified so that the affinity of the Nups in each region for Impβ matched that of the affinity reported in two previous experiments as outlined in Table 1.

Using affinity gradients measured *in* vitro in yeast and vertebrates, our model predicts transport rates that are approximately an order of magnitude lower than the peak transport rate observed for pores with uniform 200μM affinity. The presence of this affinity gradient brought about minimal gains in transport rate when compared to pores lacking a Nup-Impβ affinity gradient in the same nano-molar affinity range. Low transport rates at these experimentally derived affinities

105

can be attributed to the very slow off-rate between Impβ and Nups that act to hinder transport and are in contrast to simulations where significantly higher transport rates are seen in pores containing Nups with micro- to milli-molar affinities. It has previously been suggested that affinities derived from *in vitro* experiments are too tight to account for experimentally observed transport rates, with the simplest reason being that the associated off-rates are much slower than the observed transport times of *in vivo* cargo [10,44,46,47]. More recently, it has been shown that *in vivo* affinities between Nup and Impβ are likely much lower than *in vitro* measurements claim due to the presence of non-specific competitors in the cell milieu which are generally not considered in *in vitro* studies [32,48,49].

Table 1. Summary of affinity gradients and *in silico* derived transport rates. Summary of *in vitro* affinity gradients for yeast and vertebrates and the agent based model derived transport rate corresponding to each affinity gradient. The model optimum affinity gradient is included for comparison. The ratio of gradients between yeast and model optimum are comparable while the magnitude of the individual affinities is approximately 1000 times weaker in the model optimum.

Source	Cytoplasmic Nup Affinity (µM)	Central Channel Nup Affinity (µM)	Nuclear Basket Nup Affinity (µM)	*in silico* Transport Rate (Sec^{-1})	Ref.
Yeast	1.5	0.2	0.01	12.23 ±0.27	[13]
Vertebrates	0.2	0.1	0.01	7.58 ±0.21	[12]
Model Optimum	2000	200	10	94.73 ±1.92	

Nucleocytoplasmic Transport Sensitivity to Pore Affinity Gradient

To quantify the contribution of each region's affinity to Impβ transport rate, we varied the *in silico* affinity of FG-Nup agents for Impβ in each of the three regions independently. The range of affinities explored spanned 2µM to 2mM in the cytoplasmic periphery and central channel and 200nM to 200µM in the nuclear basket (Fig. 6). This range allowed us to explore transport rates for moderate and steep gradients in both forward and reverse affinity gradients.

As our results indicate, the transport rate appears mostly insensitive to the affinity of FG-Nups in the cytoplasmic periphery for Impβ, especially when compared to that of the other two regions. Varying the affinity of FG-Nups for Impβ in the central channel of the pore resulted in a clear biphasic behavior in transport rates observed, with very low and high affinity Nups hindering the transport of Impβ irrespective of affinity in the other two regions. The contrast between the effects of affinity in the cytoplasmic region compared to the central channel can likely be attributed to the

difference in Nup density in each region; the higher Nup concentration within the central channel [31] makes affinity a much more critical parameter. Very low affinities prevent Impβ from binding FG agents and traversing the channel as Impβ molecules are sterically repelled by the Nups – while very high affinities result in longer binding and reduced mobility of Impβ molecules. Our data suggest that a FG-Impβ K_D of 200μM in the central channel is ideal for transport irrespective of cytoplasmic and nuclear basket affinities. Transport rates were observed to peak at cytoplasmic FG-Impβ K_D of 2mM, although this was not much higher (less than one standard error) than transport rates observed with K_D ranging from 2μM to 4mM. The relationship between cytoplasmic periphery Nup-Impβ affinity and transport rate is depicted in more detail in Fig. 8.
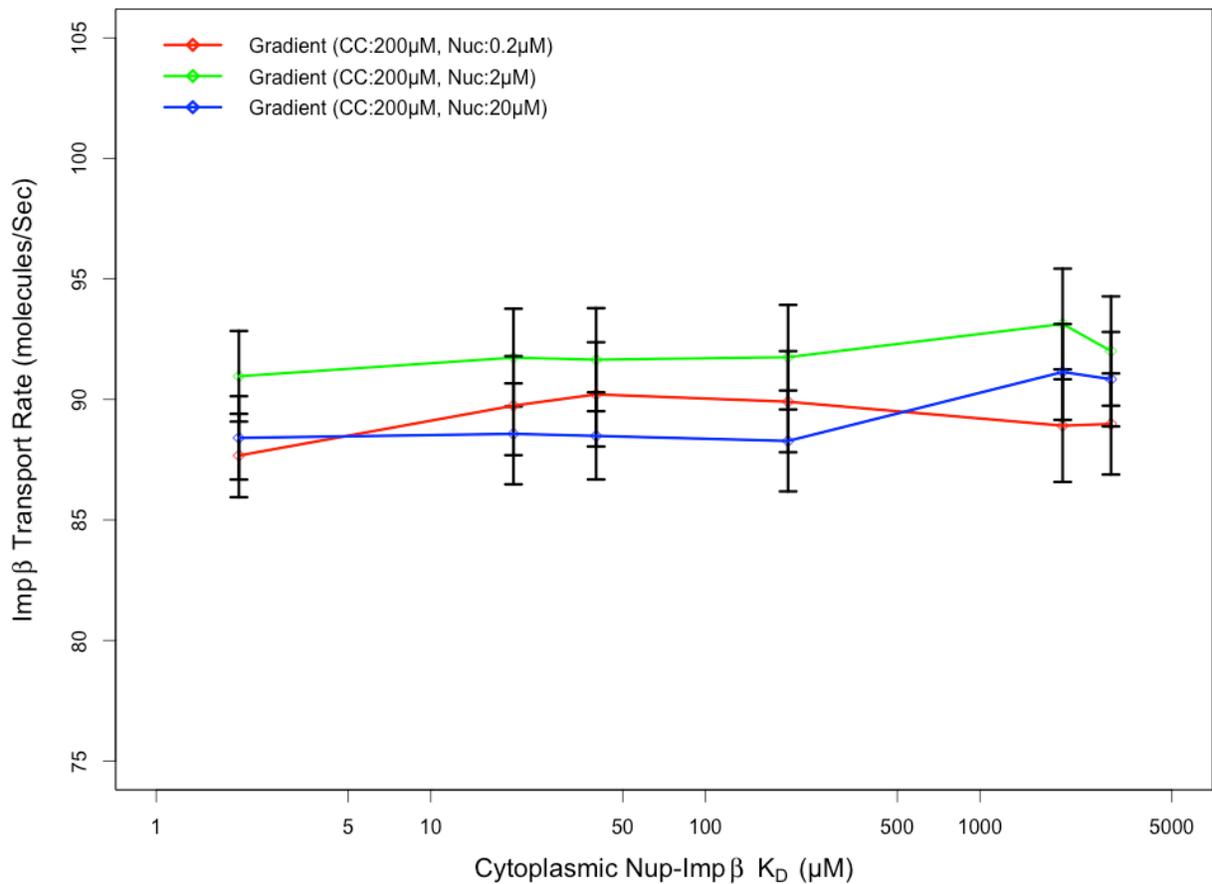
Figure 8. Transport rate as a function of cytoplasmic affinity. Transport rate appears insensitive to cytoplasmic Nup-Impβ affinity as opposed to central channel and nuclear basket Nup-Impβ affinity. An increase or decrease to affinity in the cytoplasmic region by an order of magnitude results in a change in transport rate that is within a standard error. (CC: Central Channel, Nuc: Nuclear basket)

Affinity of nuclear basket FG-Nups for Impβ also plays a critical role in determining transport rate. Our model indicates that as this affinity is increased, the transport

107

rate is also observed to increase up to a limit that is dependent on the availability of RanGTP. As shown in Fig. 9, for the standard case of nuclear RanGTP concentration of 1μM, the transport of Impβ increases as nuclear basket affinity is increased, up to an affinity of 2μM. A decrease in Impβ transport rates is observed at affinities higher than 2μM (lower K_D), under standard RanGTP concentrations. When nuclear RanGTP concentrations were increased, up to ten-fold, the absolute transport rate was not observed to change significantly. Conversely, when Impβ concentrations were increased by as little as two-fold, there was a consistent ~50% increase in Impβ shuttling into the nucleus observed across all nuclear basket affinity configurations. This is in agreement with previous studies that propose that import receptors, rather than Ran, are the limiting species in the import pathway [44,45]. Although an increase in nuclear RanGTP concentration had little effect on Impβ shuttling, a decrease in nuclear RanGTP concentration was shown to reduce Impβ shuttling from cytoplasm to nucleus. Nevertheless, a nearly ten-fold decrease in nuclear RanGTP concentration only resulted in a ~17% average decrease in Impβ shuttling across nuclear basket affinity configurations shown in Fig. 9 – suggesting that import is more sensitive to the availability of transport receptors than the availability of Ran.

Our results suggest that the peak transport rate for Impβ shuttling are produced with a Nup-Impβ affinity gradient consisting of 2000μM in the cytoplasmic fibril region, 200μM in the central channel and 10μM in the nuclear basket which produced a transport rate of 94.73 ±1.92 transports per second. This produces a regional Nup-Impβ affinity ratio of 1:20:200 (nuclear : central channel : cytoplasmic). Despite the difference in magnitude, the *in vitro* measurements of yeast affinity gradient exhibit a similar affinity gradient ratio of 1:20:150. It would be speculative to state that competitors in the cellular milieu would decrease *in vitro* affinities in a linear manner that would result in μM range affinities with the same gradient ratio. However, recent experiments by the Rout group showing the effective Kd after the addition of 0.1mg/mL of lysate as competitor hint at such a relationship [32]. Furthermore, the difference between the *in vitro* affinity gradient ratio and the *in silico* derived optimum is restricted to the cytoplasmic affinity (150 vs 200 respectively) which, as stated previously, is the least significant determinant of overall transport rate among the three regions. Simulation using a 1500μM cytoplasmic affinity instead of the optimum 2000μM to reproduce the affinity gradient ratio seen *in vitro* resulted in less than a 1% reduction in overall transport rate.

Finally, as indicated by these simulations, transport rates are relatively insensitive to the Nup-Impβ affinity in the cytoplasm. Varying the cytoplasmic affinity from the optimum 2000μM to 200μM, eliminating the steep gradient between the cytoplasmic and central channel Nups, results in only a 2% decrease in transport rate, well within the standard error of our measurements. As previously mentioned, this is not the case for central channel and nuclear basket Nups. This observation supports the notion that a continuous affinity gradient isn't necessary for efficient transport [32]; rather, the majority of the pore can be composed of Nups with a

moderate affinity for Impβ ($K_D \approx 200\mu M$) combined with high affinity Nups in the nuclear basket ($K_D \approx 10\mu M$) to achieve transport rates comparable to computationally derived optimum values.
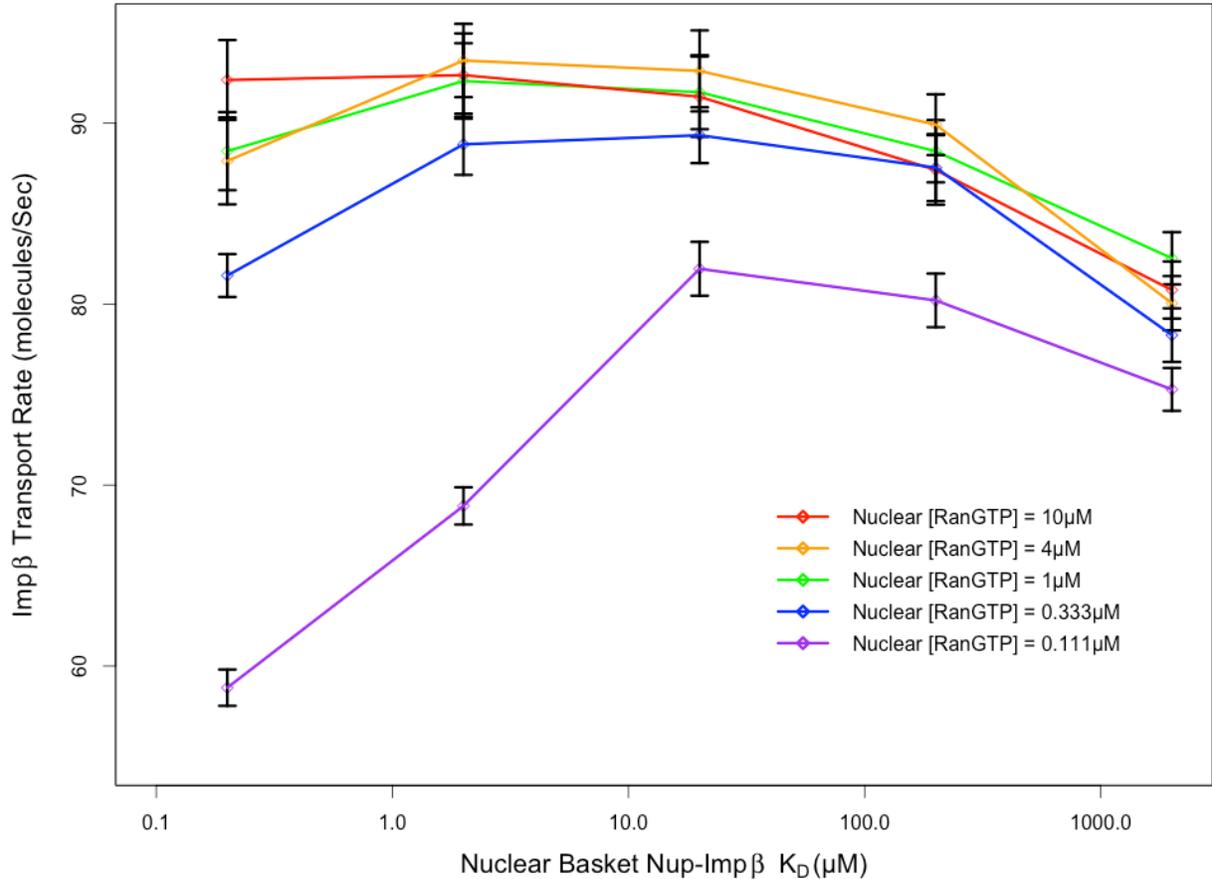
Figure 9. Transport rate as a function of nuclear basket affinity. Transport rates are very sensitive to nuclear basket Nup-Impβ affinity, with maximum transport rates emerging in the presence of a high affinity target for Impβ in the nuclear basket. Transport rates peak at an affinity of ~2μM with a slight decrease in transport rate as affinities are increased beyond that. This peak in transport rate doesn't appear to be due to a lack of RanGTP to terminate transport at the nuclear periphery of the pore since there aren't significant changes to transport rate under very high nuclear RanGTP concentrations. Conversely, when nuclear RanGTP concentrations are much lower than physiological values, the effect on transport rate is more noticeable.

Interestingly, the idea of maximizing transport rate by inhibiting diffusion at a terminal side of the pore is not limited to the Impβ import pathway. Hydrogels composed of Nup214 and Nup358 (which are located on the farthest cytoplasmic side of the pore) have been shown to selectively inhibit the diffusion of CRM1, an export receptor of the same karyopherin-β family as Impβ, while allowing the import receptor Impβ to diffuse across rapidly [50,51]. Such a high affinity target for

export receptors at the cytoplasmic side of the pore and likewise, as our model shows, a high affinity target for import receptors at the nuclear side of the pore can improve transport efficiency by reducing backflow of export and import complexes, and increasing the likelihood that they interact with RanGAP and RanGTP respectively.

Conclusion

The presence of an affinity gradient, or at the very least, a high affinity target within the nuclear basket for import cargo complexes has been observed but remains contested. The effect of such affinity gradients on transport rates was not previously explored in detail. We developed a coarse-grained, biophysical model of the nuclear pore complex translocating Impβ under various forward and reverse affinity gradients. Our results are in agreement with previous reports that the affinity gradient within the nuclear pore is not essential for cargo transport [2,52]; rather an affinity gradient, or more specifically, a high affinity target within the nuclear basket can increase overall transport rates. In fact, the reversal of the affinity gradient, shown in Fig. 6, illustrates that net movement of the Impβ molecules continues in the same direction, albeit at a lower rate, indicating that the presence of a RanGTP concentration gradient is more influential than the contribution of an affinity gradient. Nevertheless, our results reveal that the slope of the affinity gradient that maximizes transport through the pore is very similar to the slope of the affinity gradient measured *in vitro*, albeit at much lower affinity values (μM vs. nM). These lower affinity values are in agreement with the range of affinities reported in recent experimental findings, suggesting that competitors present *in vivo* reduce the effective affinity gradient. These findings could have additional implications for the design and optimization of highly efficient artificial nanopores. These modeling techniques can be used to further assess the role of nucleoporin density and distribution along with channel geometry on transport efficiency and selectivity in an effort to optimize the design and function of artificial nanopores.

References

1. Reichelt R, Holzenburg A, Buhle EL, Jr., Jarnik M, Engel A, et al. (1990) Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. J Cell Biol 110: 883-894.

2. Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, et al. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. J Cell Biol 148: 635-651.

3. Rout MP, Blobel G (1993) Isolation of the yeast nuclear pore complex. J Cell Biol 123: 771-783.

4. Loschberger A, van de Linde S, Dabauvalle MC, Rieger B, Heilemann M, et al. (2012) Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution. J Cell Sci 125: 570-575.

5. Rout MP, Aitchison JD, Magnasco MO, Chait BT (2003) Virtual gating and nuclear transport: the hole picture. Trends Cell Biol 13: 622-628.

6. Zilman A, Di Talia S, Chait BT, Rout MP, Magnasco MO (2007) Efficiency, selectivity, and robustness of nucleocytoplasmic transport. PLoS Comput Biol 3: e125.

7. Frey S, Gorlich D (2007) A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. Cell 130: 512-523.

8. Frey S, Richter RP, Gorlich D (2006) FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. Science 314: 815-817.

9. Petri M, Frey S, Menzel A, Gorlich D, Techert S (2012) Structural characterization of nanoscale meshworks within a nucleoporin FG hydrogel. Biomacromolecules 13: 1882-1889.

10. Ribbeck K, Gorlich D (2001) Kinetic analysis of translocation through nuclear pore complexes. EMBO J 20: 1320-1330.

11. Ribbeck K, Gorlich D (2002) The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. EMBO J 21: 2664-2671.

12. Ben-Efraim I, Gerace L (2001) Gradient of increasing affinity of importin beta for nucleoporins along the pathway of nuclear import. J Cell Biol 152: 411-417.

13. Pyhtila B, Rexach M (2003) A gradient of affinity for the karyopherin Kap95p along the yeast nuclear pore complex. J Biol Chem 278: 42699-42709.

14. Lim RY, Fahrenkrog B, Koser J, Schwarz-Herion K, Deng J, et al. (2007) Nanomechanical basis of selective gating by the nuclear pore complex. Science 318: 640-643.

15. Lim RY, Huang NP, Koser J, Deng J, Lau KH, et al. (2006) Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. Proc Natl Acad Sci U S A 103: 9512-9517.

16. Schoch RL, Kapinos LE, Lim RY (2012) Nuclear transport receptor binding avidity triggers a self-healing collapse transition in FG-nucleoporin molecular brushes. Proc Natl Acad Sci U S A.

17. Peters R (2005) Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. Traffic 6: 421-427.

18. Jamali T, Jamali Y, Mehrbod M, Mofrad MR (2011) Nuclear pore complex: biochemistry and biophysics of nucleocytoplasmic transport in health and disease. Int Rev Cell Mol Biol 287: 233-286.

19. Nachury MV, Weis K (1999) The direction of transport through the nuclear pore can be inverted. Proc Natl Acad Sci U S A 96: 9622-9627.

20. Berry BJL, Kiel LD, Elliott E (2002) Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling. Proceedings of the National Academy of Sciences of the United States of America 99: 7187-7188.

21. Bonabeau E (2002) Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America 99: 7280-7287.

22. Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, et al. (2005) Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. Science 310: 987-991.

23. Helbing D, Farkas I, Vicsek T (2000) Simulating dynamical features of escape panic. Nature 407: 487-490.

24. Dong X, Foteinou PT, Calvano SE, Lowry SF, Androulakis IP (2010) Agent-based modeling of endotoxin-induced acute inflammatory response in human blood leukocytes. PLoS One 5: e9249.

25. Bonchev D, Thomas S, Apte A, Kier LB (2010) Cellular automata modelling of biomolecular networks dynamics. SAR QSAR Environ Res 21: 77-102.

26. Devillers J, Devillers H, Decourtye A, Aupinel P (2010) Internet resources for agent-based modelling. SAR QSAR Environ Res 21: 337-350.

27. Azimi M, Jamali Y, Mofrad MR (2011) Accounting for diffusion in agent based models of reaction-diffusion systems with application to cytoskeletal diffusion. PLoS One 6: e25306.

28. Jamali Y, Jamali T, Mofrad MRK (In Press) An Agent Based Model of Integrin Clustering: Exploring the Role of Ligand Clustering, Integrin Homo-Oligomerization, Integrin-Ligand Affinity, Membrane Crowdedness and Ligand Mobility. Journal of Computational Physics.

113

29. Riddick G, Macara IG (2005) A systems analysis of importin-{alpha}-{beta} mediated nuclear protein import. J Cell Biol 168: 1027-1038.

30. Adam SA (2001) The nuclear pore complex. Genome Biol 2: REVIEWS0007.

31. Yamada J, Phillips JL, Patel S, Goldfien G, Calestagne-Morelli A, et al. (2010) A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. Mol Cell Proteomics 9: 2205-2224.

32. Tetenbaum-Novatt J, Hough LE, Mironska R, McKenney AS, Rout MP (2012) Nucleocytoplasmic transport: a role for nonspecific competition in karyopherin-nucleoporin interactions. Mol Cell Proteomics 11: 31-46.

33. Mohr D, Frey S, Fischer T, Guttler T, Gorlich D (2009) Characterisation of the passive permeability barrier of nuclear pore complexes. EMBO J 28: 2541-2553.

34. Kutay U, Izaurralde E, Bischoff FR, Mattaj IW, Gorlich D (1997) Dominant-negative mutants of importin-beta block multiple pathways of import and export through the nuclear pore complex. EMBO J 16: 1153-1163.

35. Bayliss R, Littlewood T, Stewart M (2000) Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. Cell 102: 99-108.

36. Bayliss R, Littlewood T, Strawn LA, Wente SR, Stewart M (2002) GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta. J Biol Chem 277: 50597-50606.

37. Bednenko J, Cingolani G, Gerace L (2003) Importin beta contains a COOH-terminal nucleoporin binding region important for nuclear transport. J Cell Biol 162: 391-401.

38. Isgro TA, Schulten K (2005) Binding dynamics of isolated nucleoporin repeat regions to importin-beta. Structure 13: 1869-1879.

39. Riddick G, Macara IG (2007) The adapter importin-alpha provides flexible control of nuclear import at the expense of efficiency. Mol Syst Biol 3: 118.

40. Kolodkin AN, Bruggeman FJ, Plant N, Mone MJ, Bakker BM, et al. (2010) Design principles of nuclear receptor signaling: how complex networking improves signal transduction. Mol Syst Biol 6: 446.

41. Cangiani A, Natalini R (2010) A spatial model of cellular molecular trafficking including active transport along microtubules. J Theor Biol 267: 614-625.

42. Moussavi-Baygi R, Jamali Y, Karimi R, Mofrad MR (2011) Brownian dynamics simulation of nucleocytoplasmic transport: a coarse-grained model for the functional state of the nuclear pore complex. PLoS Comput Biol 7: e1002049.

43. Moussavi-Baygi R, Jamali Y, Karimi R, Mofrad MR (2011) Biophysical coarse-grained modeling provides insights into transport through the nuclear pore complex. Biophys J 100: 1410-1419.

114

44. Timney BL, Tetenbaum-Novatt J, Agate DS, Williams R, Zhang W, et al. (2006) Simple kinetic relationships and nonspecific competition govern nuclear import rates in vivo. J Cell Biol 175: 579-593.

45. Yang W, Musser SM (2006) Nuclear import time and transport efficiency depend on importin beta concentration. J Cell Biol 174: 951-961.

46. Kubitscheck U, Grunwald D, Hoekstra A, Rohleder D, Kues T, et al. (2005) Nuclear transport of single molecules: dwell times at the nuclear pore complex. J Cell Biol 168: 233-243.

47. Yang W, Gelles J, Musser SM (2004) Imaging of single-molecule translocation through nuclear pore complexes. Proc Natl Acad Sci U S A 101: 12887-12892.

48. Cardarelli F, Bizzarri R, Serresi M, Albertazzi L, Beltram F (2009) Probing nuclear localization signal-importin alpha binding equilibria in living cells. J Biol Chem 284: 36638-36646.

49. Cardarelli F, Tosti L, Serresi M, Beltram F, Bizzarri R (2012) Fluorescent recovery after photobleaching (FRAP) analysis of nuclear export rates identifies intrinsic features of nucleocytoplasmic transport. J Biol Chem 287: 5554-5561.

50. Labokha AA, Gradmann S, Frey S, Hulsmann BB, Urlaub H, et al. (2013) Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. EMBO J 32: 204-218.

51. Antonin W (2013) Don't get stuck in the pore. EMBO J 32: 173-175.

52. Strawn LA, Shen T, Shulga N, Goldfarb DS, Wente SR (2004) Minimal nuclear pore complexes define FG repeat domains essential for transport. Nat Cell Biol 6: 197-206.

Section VI:

An Agent Based Model of mRNA Nuclear Export Kinetics

Introduction

RNA Export Pathways and Kinetics

Aside from transport receptors that are recycled to the cytoplasm, much of what is exported from the nucleus consists of ribonucleic acids (RNA) in the form of transfer RNA (tRNA), microRNA (miRNA), small nuclear RNA (snRNA), ribosomal RNA (rRNA) and messenger RNA (mRNA). Many of these molecules (tRNAs, miRNAs, snRNAs, rRNAs and some mRNAs) follow a karyopherin-mediated export pathway that is similar to that of nucleocytoplasmic import, relying on the Ran cycle for directionality. More specifically, tRNAs recruit Exportin-t (Los1) of the karyopherin superfamily in complex with RanGTP to facilitate export, while miRNAs recruit Exportin-5 (Msn5) in complex with RanGTP. The remaining RNAs (snRNAs, rRNAs, and some mRNAs), as well as many viral RNAs such as HIV-1, make use of Exportin-1 (Xpo1, a.k.a. CRM1) in complex with RanGTP to achieve nucleocytoplasmic export [1,2]. These exportins rely on the RanGAP/RanBP2-mediated hydrolysis of RanGTP to provide RNAs with export directionality (similar to RanGTP, which acts as a molecular ratchet for protein import). Conversely, the bulk of mRNA is exported from the nucleus via the NXF1/NXT1-mediated pathway.

Following transcription, mRNA carries genetic information from the cell's DNA to the ribosomes that translate this message to functioning proteins, primarily via the nuclear pore complex (NPC) and the NXF1/NXT1 pathway. A nascent pre-mRNA must undergo multiple co-occuring steps before it can be successfully recruited to the NPC and exported from the nucleus. A newly transcribed pre-mRNA undergoes four processing steps prior to being considered a mature mRNA: 1) capping of the 5'-terminal end with 7-methylguanylate, 2) splicing to remove non-coding intron regions, 3) 3'-terminal end cleavage and 4) polyadenylation [3]. Second, during transcription and prior to transport, a maturing mRNA also binds a number of protein factors to become a messenger ribonucleoprotein particle (mRNP). Recruitment of factors to the mRNA is dynamic, with multiple proteins binding, recruiting other proteins, and subsequently unbinding in an effort to produce an export-competent mRNP. In this regard, the transcription elongation-mRNA export (TREX) complex is crucial in producing export-competent mRNPs, as it recruits the NXF1/NXT1 (Mex67/Mtr2) heterodimeric export factor via a set of core proteins termed the transcription elongation (THO) complex and the associated protein Aly/REF (Yra1). These are subsequently removed from the mRNP prior to export, once NXF1/NXT1 are successfully bound [4]. Once the mRNA is processed and export receptors are recruited, the mRNP is considered export-competent and is recruited to the NPC, where it translocates to the cytoplasm via a series of binding and unbinding events with phenylalanine-glycine (FG) Nups.

The size of a typical mRNA (~2.2 kb) is larger than that of a typical protein that undergoes nucleocytoplasmic import, and much larger mRNAs such as the Balbiani ring mRNA (~37 kb) are known to translocate across the NPC. Subsequently, a model for NXF1/NXT1-mediated nucleocytoplasmic export of mRNA was proposed,

in which transport receptors are believed to bind multiple sites along the length of an mRNP. Unlike mRNA export pathways in which exportins of the karyopherin superfamily are recruited, bulk mRNA export via the NXF1/NXT1-mediated pathway relies on a different set of export factors to regulate transport directionality. The superfamily 2 DEAD-box ATPase DDX19 (Dbp5) is responsible for remodeling the mRNP as it translocates through the NPC and reaches the cytoplasm-facing Nups. DDX19 is localized to the cytoplasmic Nup124 (Nup159) and is activated by the export factor Gle1 along with its cofactor, inositol hexakisphosphate (IP6). Gle1/IP6 also remodels the mRNP by freeing transport receptors to return to the nucleus for a successive round of transport, and therefore prevents the return of the mRNP into the nucleus [5]. These mRNP/export factor interactions account for a virtual "Brownian ratchet" mechanism that ensures one-way transport of the mRNP through the NPC [6].

Although the mRNA ratcheting model appears feasible, to date, the number and distribution of transport receptors that are required to bind an mRNP for efficient transport remains unknown. Binding of the TREX complex, a factor involved in transport receptor recruitment, has been observed at the 5'-terminal end of mRNA [7]. Coincidentally, it has been reported that transport of very large Balbiani ring mRNA occurs with the 5'-terminal end leading through the pore as seen in electron microscopy (EM) experiments [8-10]. mRNA export orientation could be the result of transport receptors bound at the 5' end and a lack of receptors elsewhere. However, other groups have observed the binding of transport receptor recruiting factors such as REF along the length of the mRNA at the site of exon junction complexes (EJCs) [11-13], suggesting that transport receptors are distributed at multiple sites along an mRNA and lending support to the mRNA ratcheting model. Nevertheless, this leaves unanswered questions as to why there exists a preference for the 5' end leading orientation of export if multiple other transport receptors are present.

One could speculate that the reasons for this may be due to DNA being read from 3' to 5' end during transcription and complementary mRNA being produced in 5' to 3' end. It is possible that the 5' end simply recruits transport receptors before other sites have a chance to. Another possibility is that the lack of an EJC near the 3' and the presence of a poly-A tail leaves a segment of the mRNA lacking affinity for the pore. Although it would still be possible for the mRNA to initiate binding to the pore via a transport receptor located somewhere along the length of the mRNA, this would likely be a less favorable orientation for transport than threading through the pore via a leading 5' end orientation. Although the idea that transport receptors bind EJCs seems promising, one must consider that many lower eukaryotes lack the density of introns and subsequently EJCs as seen in higher eukaryotes [14]. This could suggest that even though the NXF1/NXT1-mediated export pathway is present in yeast via the homologs Mex67/Mtr2, the mechanism by which mRNPs mature and become export-competent may be slightly different. These questions remain to be answered and an understanding of the number and distribution of transport receptors that bind an export-competent mRNP will provide significant

insight into mRNA export dynamics with potentially far-reaching impacts in understanding human health and developing successful interventions for human disease.

In recent years, new molecular techniques have been developed for single molecule tracking of mRNA export, which provide further insight into export kinetics. In 2010, Grünwald and colleagues used the MS2-GFP system [15] to track the movement of β-actin mRNA (~3.3 kb) through the NPC, with temporal resolution of 20 milliseconds and spatial precision of 26 nm. They observed that, rather than transport through the central pore (5-20 ms in duration), docking to and release from the NPC are the rate limiting steps (each ~80 ms in duration) [16]. In 2012, the Kubitscheck group used light sheet fluorescence microscopy to track single export events of native mRNPs with hrp36 fluorescent tags with 20 millisecond temporal resolution and 10 nm spatial resolution [17]. In these experiments, an average transport time of 65 ms was observed for mRNA of all lengths that may be present within the nucleus. In this work, a fluorescent hrp36 protein that natively binds mRNA prior to export was introduced to the system. Since the protein binds mRNA at multiple sites, fluorescent hrp36 was added at sufficiently low concentration to ensure that each mRNA was tagged only once and to avoid tagging multiple proteins within the imaging plane. Subsequently, the group reports an average transport time of 65ms for the majority of mRNA, with transport times of up to several seconds for what are likely very large mRNAs. They further observe mRNA arrest strictly at the nuclear face of the pore that typically lasts ~50 ms.

The discrepancy between the transport times of these two methods (180 ms compared to 65 ms) could be due to the fact that in the MS2-GFP system, a specific mRNA of known length was tracked (β-actin mRNA, ~3.3 kb) whereas the hrp36 protein can bind mRNA of any length (average length of mRNA reported as ~2.2 kb). To further complicate matters, longer mRNA have a greater number of hrp36 binding sites, suggesting that larger mRNAs are more likely to be tagged. However, the typical length mRNAs tracked in the fluorescent hrp36 system also depends on the expression levels for each mRNA. Taken together, this suggests a level of uncertainty in the actual length of mRNAs tracked. Nevertheless, the fluorescent hrp36 system allows for tracking of native mRNA without introducing multimers of RNA stem loops that bind MS2 coat proteins, which likely alter the size and dynamics of the mRNA molecule [15]. It is likely that the addition of stem loops and the coat protein is a significant source of discrepancy between the observed kinetics of the two experiments, and likely explain why mRNA export was delayed at the cytoplasmic surface in the MS2-GFP system and not the fluorescent hrp36 system [18].

Interestingly, despite the discrepancy in transport kinetics between these works, both single molecule tracking studies as well as the previous EM studies report the observation of a rate-limiting step at the nuclear basket during transport. In the case of the EM experiments, this was thought to be a result of the 5'-end leading orientation of export and the time required for such a large polymer to be oriented

to satisfy this condition. It's not clear whether the observance of this rate-limiting step for smaller mRNAs is the result of the polymer searching for a favorable export orientation or whether there is a quality control mechanism or some unidentified processing step present at the nuclear basket that results in delayed entry and transport. Answering this question requires significant improvements in microscopy techniques.

Alternatively, the effective computational modeling of the system can provide further insight. Our results suggest different scenarios for mRNA export. The mRNP trajectory data suggest that there is a rate-limiting step at the nuclear basket, which could refer to finding a suitable mRNP configuration for entering the pore interior. The duration of this step could vary largely for mRNPs of different sizes. The actual translocation could be achieved by the action of Dbp5. An alternative hypothesis is suggested by comparison of the Dbp5 binding time (55 ms) with the export duration of the larger mRNPs (350 ms up to several seconds). Several Dbp5 molecules might act in sequence on a single mRNP to achieve the required export factor release and mRNP translocation. In this case the observed translocation pause is just valid for the specific incorporation site of the labeling hrp36 that remains in the basket until it is finally moving through the pore. Furthermore, several Dpb5 molecules could act at different locations on a single mRNP simultaneously.

Agent Based Modeling

Agent based modeling (ABM) is a robust computational technique used to simulate the spatiotemporal actions and interactions of real-world entities or "agents", in an effort to extract their combined effect on the system as a whole. Both space and time can be discretized in an ABM, giving these autonomous agents the ability to move and interact with other agents and their environment at each timestep over a given duration. Simple behavioral rules govern the movement and interaction of each individual entity in an effort to reproduce or predict more complex behaviors of multiple entities. Such a model attempts to simulate the emergence of complex phenomena that may not be apparent when simply considering individual entities. Agent based modeling has seen applications in a broad range of fields ranging from artificial intelligence and gaming to modeling emergent social behavior such as the spread of disease and outcomes of financial markets [22-25]. In their simplest form, on-lattice agent based models consist of a mesh of "cells" that make up the discretized space that agents occupy. The agents occupy these cells and are typically only aware of other agents within their "neighborhood"; in the simplest form a neighborhood consists of adjacent cells. Agents are given the ability to move into adjacent cells and to interact with other agents with some probability in conjunction with governing rules that define what movement and interactions are possible (Fig. 1). On-lattice agent based models have previously been applied to biological systems involving diffusion, binding and unbinding [26-28]; establishing methods for event probability selection – relating diffusion and rate constants to event

probability – will improve model accuracy and enable quantitative analysis of results from these models [29,30].

In the present work we adopt a method for modeling a collection of agents as dynamic polymers with properties representative of a freely diffusing mRNA molecule. We then build upon our ABM framework [29] to understand how binding, unbinding and diffusive events enable the efficient transport of long polymeric molecules across the nuclear envelope. We model the system using diffusion, binding and unbinding rates derived from *in vitro* experiments where possible. The agent based simulations were designed to address the following questions: (*i*) How does the number and distribution of transport receptors across an mRNA molecule affect export efficiency? (*ii*) Is tracking a single point of a polymeric chain transporting through a narrow pore sufficient for quantifying transport times?

To answer these questions, simulations were carried out using a computationally efficient, spatiotemporally detailed, three-dimensional agent-based model developed specifically for modeling molecular diffusion, binding and unbinding events with consideration for physical factors such as molecular crowding and steric repulsion. In addition to movement and interaction rules, event probabilities govern system dynamics in the agent-based model. Methods for accurate selection of movement, binding and unbinding probabilities to best represent actual diffusion coefficients and kinetic rate constants can build confidence in the output of agent based models and deductions from these models. The procedure for relating real world rate constants to molecular binding and unbinding probabilities is detailed in the Materials and Methods.

Figure 1. Simplified representation of the agent based model. Abstract cartoon representation of the nuclear pore structure environment (not to scale) projected onto a simplified, 2-dimensional, on-lattice ABM with agents representing proteins that move within the system and interact with other agents within their von-

Neumann neighborhood. The actual model consists of a three-dimensional representation of the NPC structure and physiologically relevant concentrations of biochemical factors and channel dimensions. In our model, the purple region representing the cytoplasmic periphery is treated as a compartmentalized volume containing non-interacting Nup and transport-receptor-interacting FG-Nup agents. Similarly, central channel (blue) and nuclear basket (green) regions are represented by compartmentalized volumes, containing both non-interacting and interacting Nup agents at physiologically meaningful concentrations. Grey regions of the diagram represent the scaffold and nuclear envelope regions of the model that are impermeable to diffusing species.

Materials & Methods

Probability Selection for Molecular Movement on an ABM Lattice

In our previous work we proposed a method for movement probability selection based on molecular diffusion coefficient (Eq. 1) along with algorithms for realistic consideration of crowding and steric repulsion [29] that was also used in the current model:

$$P_{move} = \frac{D \cdot \Delta t}{\Delta L^2} \qquad (1)$$

Here, movement probability of an agent is determined by its diffusion coefficient (D), simulation timestep (Δt) and lattice discretization length (ΔL). We implemented the *reduced probability (RP)* method to account for the steric effects of multiple agents occupying individual lattice sites [29].

Probability Selection for Molecular Binding and Unbinding Events on an ABM Lattice

In our more recent work, we proposed and validated a method for probability selection of binding and unbinding events in the ABM that was also used in the current model. Probability selection of unbinding (Eq. 2) and binding (Eq. 3) can be determined from kinetic rate constants as follows:

$$P_{off} = k_{off} \Delta t \qquad (2)$$

$$P_{on} = \frac{k_{on} \Delta t}{V \big/ N_{cells} \cdot N_{neighbors} \cdot N_{Avogadro}} \qquad (3)$$

Here, event likelihoods are determined from (k_{off}) and (k_{on}) which represent the real-world kinetic rate constants, simulation timestep (Δt), system volume (V), number of lattice cells (N_{cells}), number of lattice neighbors each cell has ($N_{neighbors}$) and the number of particles in a mole ($N_{Avogadro}$).

mRNA Dynamics

In the ABM, the mRNA is represented as a collection of bound agents, that are limited in their movements through the constraint of maintaining connection with their nearest neighbor to each side. Further, movements are only permitted in the diagonal direction into the nearest neighbor(s) von Neumann neighborhood. This ensures that all movements are of fixed length, and assures that a single movement probability can accommodate all movement events of a specific agent type. The probability of movement for agents belonging to the mRNA polymer was determined through successive simulations in which movement probability of a representative ABM polymer of length 2.2kb, composed of 130 agents in a cubic lattice of length 5nm was varied to determine the effective diffusion coefficient of the polymer. Effective diffusion coefficient was determined by calculating the mean squared displacement over 5 millions timesteps with a minimum sampling of 1000 timesteps. A movement probability of 0.5 was determined to yield and effective diffusion coefficient of ~0.01 μm^2/s, which is in agreement with previous *in vivo* measurements of mRNA diffusion [31]. Analysis of mean square displacement as a function of time also confirmed that diffusion of the ABM polymer followed a normal diffusion regime, which is expected from previous reports and our own comparison with Brownian dynamics simulations. Finally, to ensure that the agent based model correctly recapitulated the behavior of a freely jointed chain, we simulated ABM polymers of varying lengths to confirm that their average end-to-end length was in agreement with the length-dependent prediction postulated by the freely jointed chain model (Eq. 4).

$$\sqrt{\left\langle \vec{R}^{2} \right\rangle} = \sqrt{N}l \qquad\qquad (4)$$

Where $\vec{R}$ represents the total end-to-end vector of an ideal chain, N represents the number of monomers in the chain, which is equivalent to the number of agents that comprise a polymer chain in the ABM and l which represents the length of one of the monomers, which is equivalent to the discretization of the lattice in the ABM. We observed that for very short length polymer chains of 5 agents, there was a ~10% discrepancy between the predicted average end-to-end length (Eq. 4) and the average end-to-end length from 10,000 randomly sampled ABM configurations. This error was reduced to <1% when the number of monomers/agents was increased to 50. The error was determined to be negligible for the configuration of 130 monomers/agents that was used to simulate a typical mRNA.

ABM System and Simulation Details

The model environment consists of a 42,108 element, three-dimensional lattice comprised of elements with dimensions of 5nm x 5nm x 5nm. The lattice size was selected to accommodate the volume associated with the Stokes radius of the largest single-agent species in the system, in this case a collection of nucleotides representing the persistence length of the mRNP. Additionally, the model allowed for multiple agents of the same or different species type to occupy the same lattice element at any given time, so long as the available volume of a lattice element was not exceeded by agents diffusing into it. Discrete lattice elements belong to one of six region types, cytoplasmic, nuclear membrane, nucleoplasm, cytoplasmic filament periphery, central channel or nuclear basket. The cytoplasmic region contains a high concentration of Dbp5 in complex with Gle1 and IP6 while the nucleoplasm in each simulation contained a single 2.2kb mRNP, discretized into 130 agents. The 35nm thick nuclear membrane which partitions the two compartments is impermeable to all agent types and contains a single nuclear pore with diameter of 30nm at the center and 50nm at the peripheries. The cytoplasmic filament periphery consists of a 50nm diameter region that extends 30nm into the cytoplasm while the nuclear basket is composed of a basket shaped region that extends 55nm into the nucleoplasm [32,33]. The cytoplasmic periphery, central channel and nuclear basket each contain 24, 80 and 32 agents respectively, which represent the distribution of FG Nups [34]. In addition to these FG agents, non-FG agents are added to the channel to represent regions of the Nups that lack affinity for transport receptors but play a role in sterically repelling molecules, with the sum of the volume of these Nups corresponding to experimentally reported volumes [34]. The collection of agents representing the mRNP are free to diffuse throughout the system while FG agents and non-FG agents are restricted to movement within their respective pore regions in order to maintain the permeability barrier.

Distribution of transport receptors on the mRNA and their affinity for FG-Nups was varied across simulations. For each configuration investigated, 100 replicate simulations were generated and analyzed. Each simulation contained a single mRNA with a random initial configuration in the nucleoplasm. The 5' and 3' termini of the mRNA were tracked in each simulation. Additionally, a random site along the length of the mRNA was also tracked to emulate the use of a fluorescently labeled hrp36 molecule for comparison with experiments carried out by Siebrasse et al. [17]. This random site was varied across each simulation in the set of 100 replicates. Each simulation was carried out for duration of 20 seconds using a timestep of 2.5 microseconds.

The location of the 5' and 3' termini along with the location of the randomly placed hrp36 tag were tracked over the course of the simulation. To analyze these trajectories, kymographs were generated illustrating the location of the 5' and 3' termini of the mRNA over time (see Fig. 2). The trajectories were further analyzed to determine fraction of partial and successful transports per configuration, along with

125

mRNA residence times in the nuclear basket and central channel for comparison with previous *in vivo* observations.
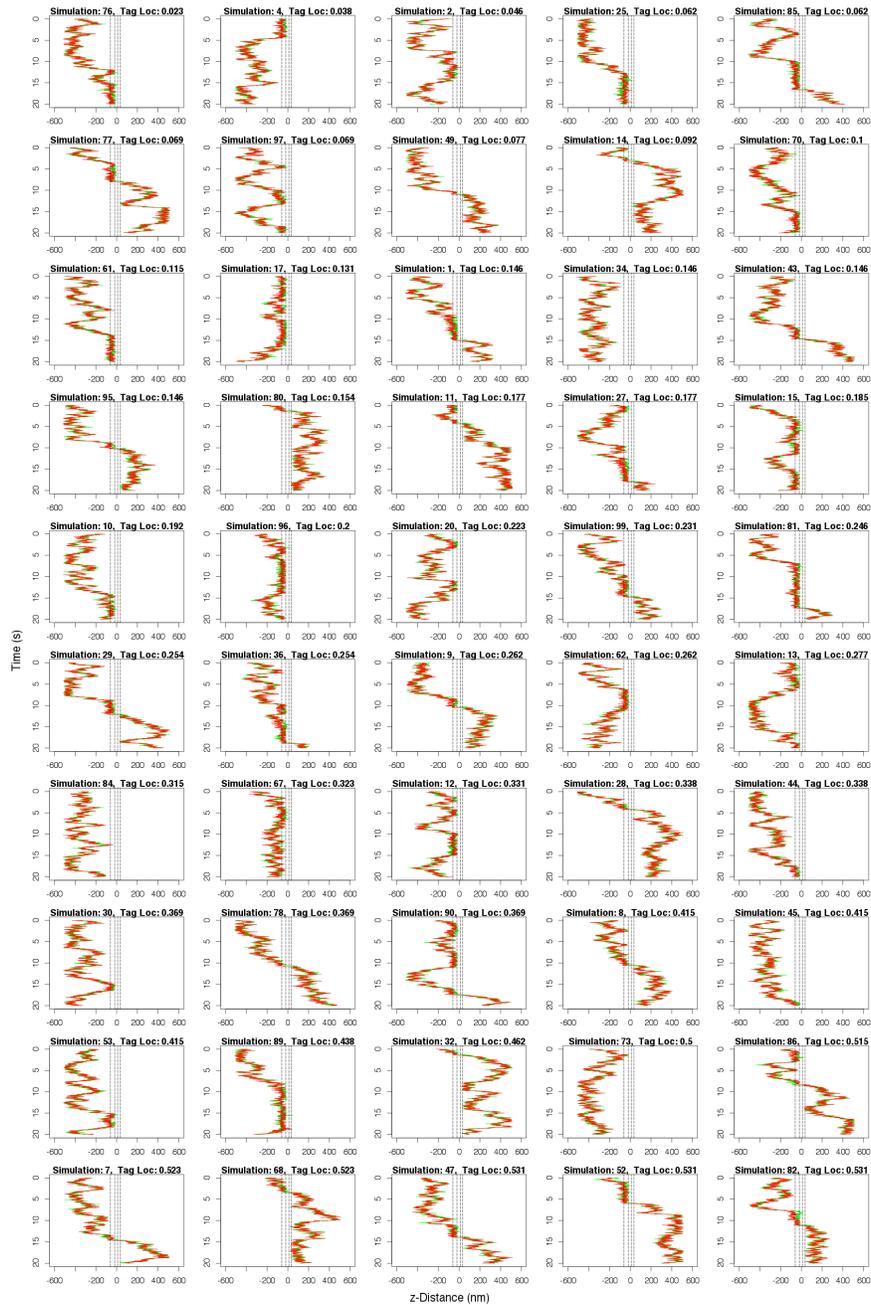
Figure 2. A collection of kymographs of 50 replicate simulations of mRNA export using the default configuration of transport receptor to FG-Nup affinity of K_D=100 μM and nine transport receptors along the length of the mRNA. The green lines represent the position of the 5' end while the red lines represent the position of the 3' end. The NPC central channel is centered at *z=0 nm* in each plot.

126

Model Validation

In silico experiments were performed to determine the model's ability to recapitulate experimentally-determined, size-dependent permeabilities for passive cargos as well as for Impβ [35,36]. These simulations served as a control to validate the model and associated algorithms' ability to simulate selective transport. Following a simulated microinjection of non-interacting species in the cytoplasm, the *in silico* pore is observed to inhibit the influx of larger species while allowing smaller species to diffuse through the pore (shown in Fig. 3). Influx rates of non-interacting species with Stokes radii of ~1nm are on the order of $0.1s^{-1}$ while larger species with Stokes radii of >2.5nm have influx rates of less than $0.001s^{-1}$. As expected, a reduced influx rate was not observed for larger species that had affinity for the FG-Nups. To test this behavior, experiments similar to those performed for non-interacting species were repeated, replacing the non-interacting species with 2.5μM labeled Impβ – in addition to the steady-state concentration of unlabeled Impβ. The influx rate of Impβ into the nucleus was observed to be $0.367s^{-1}$. This value comparable with an experimentally measured influx rate of $0.4s^{-1}$ for Impβ [36].
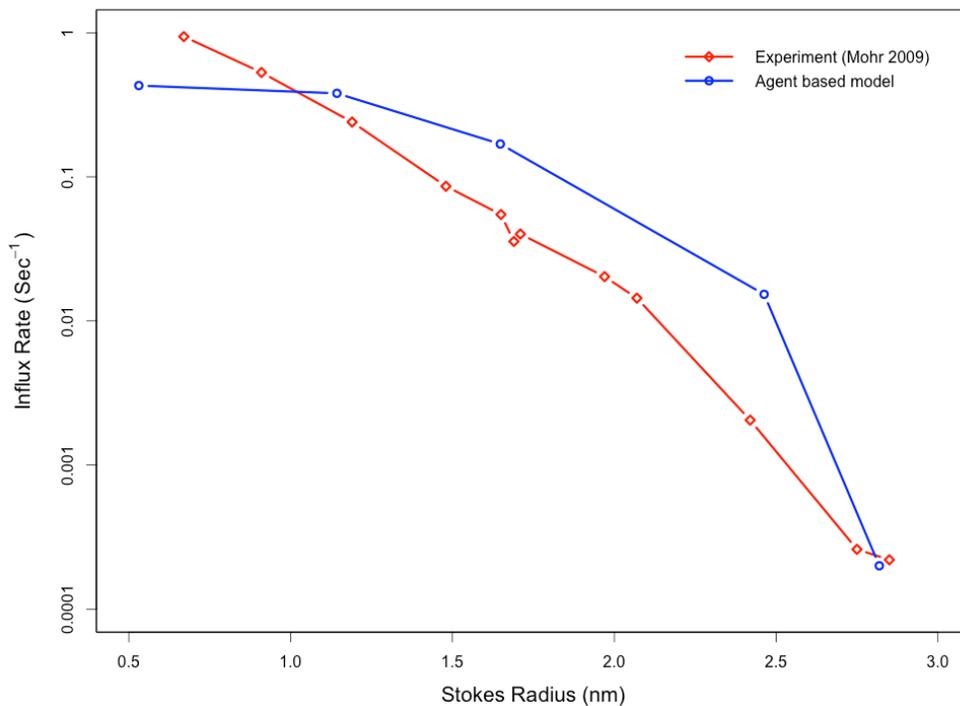
Figure 3. The agent based model recapitulates the experimentally observed size-dependent permeabilities of passive cargos through the nuclear pore. Following a simulated microinjection of non-interacting species in the cytoplasm, the *in silico* pore is observed to inhibit the influx of larger species while allowing smaller species to diffuse through the pore. This is in agreement with previous experimental observations.

127

Results

Effect of Nup-NXF1 Affinity on mRNA Export Time

Recent studies have reported an in vitro avidity for the NXF1 yeast homolog, Mex67 to the nucleoporin Nsp1 to be on the order of 100 nM [37]. It has been demonstrated experimentally that the affinities between Nups and transport receptors usually reported in vitro are much higher (~1000-fold) than actual in vivo affinities [38]. The use of in vitro affinities in the ABM was shown to have detrimental effects on transport rates in Section V, and it was demonstrated that affinities ~1000x weaker were required to reproduce transport characteristics of the NPC. To understand the export pathway's sensitivity to affinity, the affinity between NXF1 and FG Nups was varied from 20 μM to 200 μM (Fig. 4). The default number of NXF1 transport receptors bound to the mRNA was set to nine to represent the number of exon junction complexes present in a typical mRNA of this length. When mRNA-bound NXF1 was modeled to have a dissociation constant of 200 μM, no nucleocytoplasmic transport was observed. As affinity was increased incrementally towards 20 μM, the percent of simulations in which successful transport was observed increased. Average residence times in the basket and central channel for successful transport events slightly increased as well (Fig. 5a). On the other hand, the fraction of unsuccessful transport events (Fig. 4, red) did not appear to significantly change with affinity. Increasing NXF1's affinity to FG Nups appears to have caused more partial transport events (orange) to become successful ones (green). In this context, partial transport was defined as the initiation of transport from either the 5' or 3' end of the mRNA but failure in the export of the other end. (This includes transports that were not completed in the observation time, as well as rejected transports.)

The affinity between NXF1 and FG Nups in the ABM was then fixed at 100 μM. Mean mRNA residence time at the basket was measured as 462.9±265.4 ms, with a mean central channel residence time of 269.5±78.6 ms. Successful and partial transport across the pore was observed in 21% and 68% of simulations, respectively. The fraction of mRNA observed successfully trafficking out of the nucleus is in agreement with experimental observations[17]; however, the residence times observed in our in silico model appear to be significantly longer than those reported in in vivo studies. A potential explanation for this difference is provided in the next section.
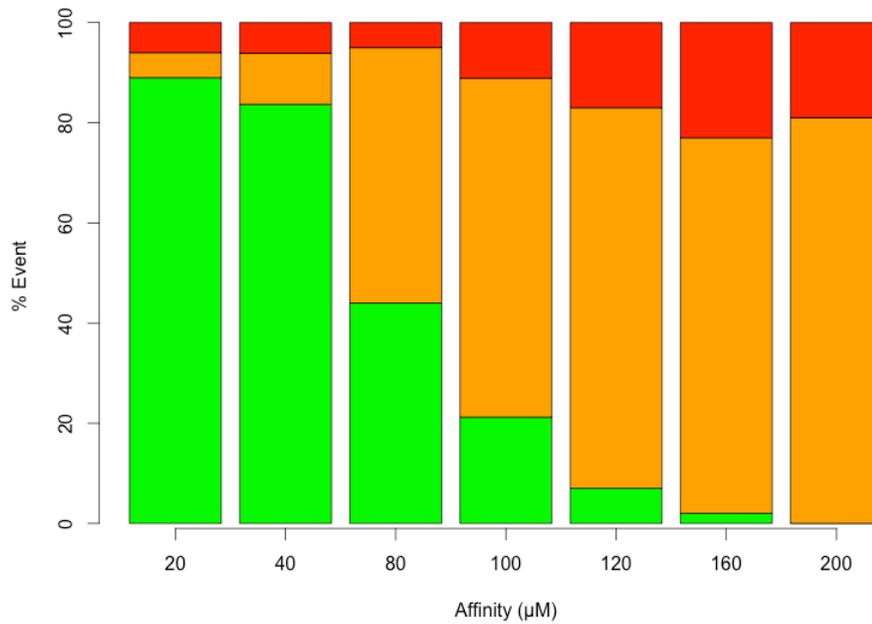
Figure 4. Chart showing the relative percent of successful (green), partial (orange) and unsuccessful (red) transport events for different affinities between NXF1 and FG-Nups.
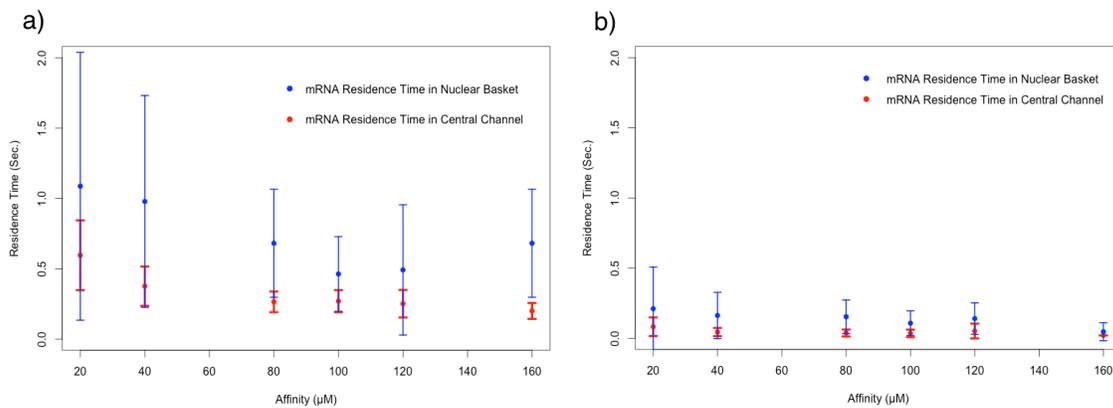
Figure 5. The effect of NXF1/FG-Nup affinity on nuclear basket and central channel residence times for successful export events using either a) probes placed at 5' and 3' ends or b) a single probe located randomly along the length of the mRNA (error bars represent one standard deviation).

129

Comparison with Single Molecule Tracking Transport Times for mRNA Export

While single probes placed at either the 3' end or a random site along the length of the mRNA have previously been used in experimental studies of mRNA export, they may inaccurately measure residence time in the pore as well as the fraction of successful transport. A single probe imprecisely indicates the location of the mRNA, leading to either over- or underestimation of true pore residence time. In addition, the fraction of successful transports may be overestimated if a probe on a partially transported mRNA reaches far enough ahead of the rest of the molecule within a particular time cutoff.

We determined residence times in the nuclear basket and central channel by tracking either a random site along the length of the mRNA or two probes placed at the 5' and 3' ends. This method of tracking provides a more meaningful measure of residence times for the whole polymer as compared to a single probe placed at any location on the length of the mRNA. For the same simulations depicted in Fig. 5a above, we tagged a single site along the length of the mRNA (randomized across 100 simulations) in order to measure residence times in a manner comparable to those reported by previous experiments. For a 100 μM affinity, nine transport receptors on the mRNA, and tracking using the single probe approach, the mean nuclear basket and central channel residence times were 107.3±87.8 ms and 34.5±27.0 ms, respectively. These are comparable to residence times measured *in vivo* (see Fig. 5b) [17]. The corresponding residence times using the labeled termini approach yielded 462.9±265.4 ms and 269.5±78.6 ms - both significantly higher. Wilcoxon rank-sum test of residence times using a single tag compared to two terminal tags confirmed that residence times were significantly different (p<<0.001) for all affinities except for 120 μM and 160 μM (likely, due to a small number of transport events).
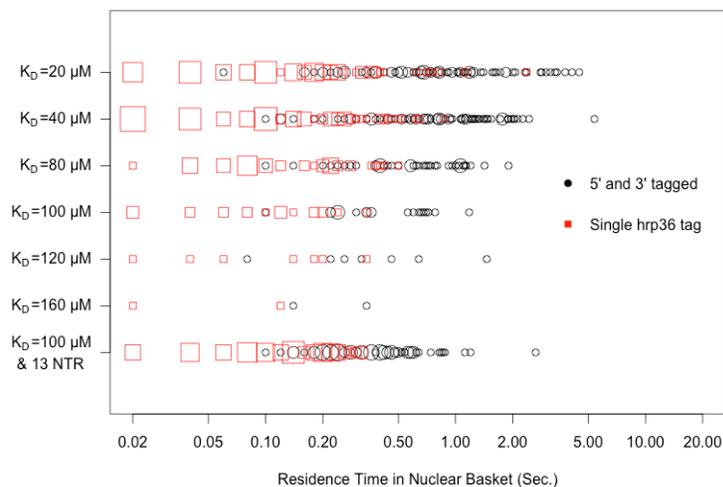
Figure 7. Distribution of mRNA residence times in the central channel for successful export events as calculated using a single probe (ie. hrp36 tag) directly compared to placement of probes at both 5' and 3' ends of the mRNA. Larger points indicate higher frequency of specified measurement.

Effect of Number and Distribution of mRNA-Bound NXF1 Transport Receptors

As previously mentioned, the default number of NXF1 transport receptors bound to the mRNA was set to nine to represent the number of exon junction complexes present in a typical mRNA of this length. Under this configuration, a Nup-NXF1 affinity of 100 μM yielded a frequency of successful transport events that was agreeable with experimentally reported values (Fig. 4). We assessed the sensitivity of mRNA export to the variation in the number of transport receptors for an mRNA of fixed length (i.e. receptor density). Increasing the number of transport receptors from the default nine to thirteen resulted in a significant increase in the number of successful transport events (see Fig. 8) at the cost of increased basket and central channel residence times (see Fig. 9). A decrease in number of transport receptors from nine to seven resulted in no successful transport events observed.

We also explored the effect of transport receptor localization on transport kinetics. The placement of transport receptors along the entire length of the mRNA resulted

131

in more robust transport than under partial coverage. As seen in Fig. 8, keeping transport receptor spacing fixed between trials, configurations of full-length coverage, ¾ length coverage and ½ length coverage yielded different rates of successful transport of 21%, 11% and 3% respectively for configurations in which partial coverage began at the 5' end and extended to either ½ or ¾ the length of the mRNA polymer. The number of transport receptors also differed with nine, seven and five transport receptors for the full length, ¾ length and ½ length configurations, respectively. Interestingly, comparing transport likelihood between the configuration of seven transport receptors uniformly distributed along the length of the mRNA to that where the seven transport receptors are distributed along the ¾ of the length (starting at the 5' end), we see that 11% of mRNA were exported when the transport receptors were placed within the ¾ length – suggesting that transport receptor spacing (or density) is likely a greater factor in determining export likelihood. Interestingly, when we positioned the five and seven transport receptors in the center ½ and ¾ length of the mRNA polymer, we did not observe any transport and the number of partial transports was also reduced.
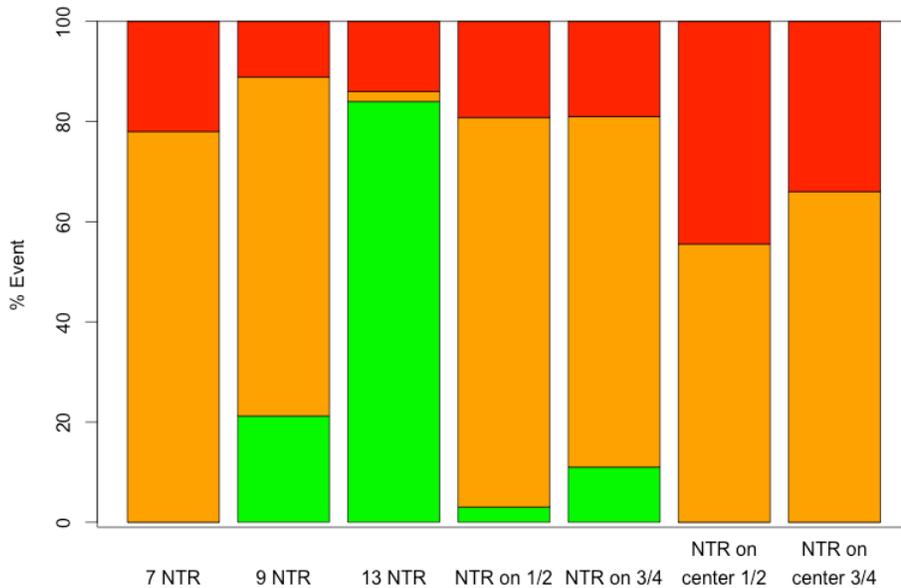
Figure 8. Chart showing the relative percent of successful (green), partial (orange) and unsuccessful (red) transport events for different distributions of nuclear transport receptors (NTR) on an export competent mRNA. Note: all configurations used the baseline NXF1 to FG-Nup affinity of 100 µM. *NTR on ½* and *NTR on ¾* represent configurations where transport receptors were placed on the terminal ½ and ¾ length of the mRNA respectively with the same spacing as used in the baseline configuration for a total of 5 transport receptors in the ½ configuration and 7 transport receptors in the ¾ configuration. *NTR on center ½* and *NTR on center ¾* represent configurations where transport receptors were placed in the center ½

and ¾ length of the mRNA respectively with the same spacing as used in the baseline configuration for a total of 5 transport receptors in the ½ configuration and 7 transport receptors in the ¾ configuration – these configurations lacked transport receptors near the 5' and 3' termini.
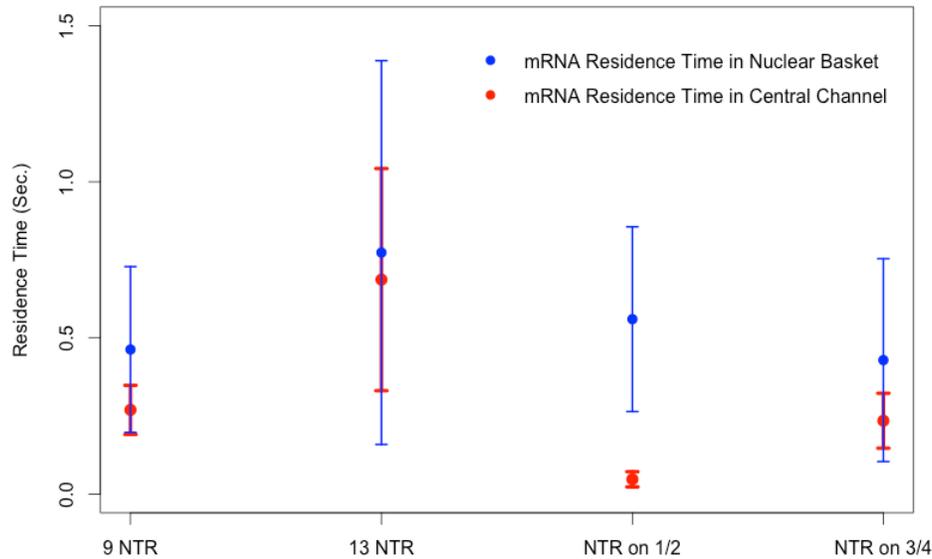
Figure 9. Effect of varying nuclear transport receptor (NTR) distribution along the length of an mRNA on nuclear basket and central channel residence times for successful export events (error bars represent one standard deviation). Note: all configurations used the baseline NXF1 to FG-Nup affinity of 100 μM. *NTR on ½* and *NTR on ¾* represent configurations where transport receptors were placed on the terminal ½ and ¾ length of the mRNA respectively with the same spacing as used in the baseline configuration for a total of 5 transport receptors in the ½ configuration and 7 transport receptors in the ¾ configuration.

Dynamics of an mRNA Polymer Undergoing Export

Previous studies have reported the lengthening of mRNA as it approaches the central channel [40]. This observation was for the Balbiani ring mRNA, which is much larger than the mRNA modeled in our simulations (~37 kb compared to ~2.2 kb). Nevertheless, we wanted to explore if such lengthening would be observed in our simulations. As shown in Fig. 10, the average end-to-end length of mRNA approaching the basket dropped initially from that of the globular conformation taken on when diffusing freely throughout the nucleoplasm. This reduced end-to-end length is the result of the mRNA taking on a more compact conformation as

NXF1 proteins bind multiple Nups in the basket. Average end-to-end length is observed to increase as the polymer translocates through the central channel. This behavior is consistent with what was observed in EM studies of mRNA export. Notably, increased affinity and number of transport receptors result in increased compaction of the mRNA polymer in the basket (see Fig. 10 – increased variability in average end-to-end lengths of low Nup-NXF1 affinity transports can be observed on the cytoplasmic side due to low sampling of successful transport events in those configurations).
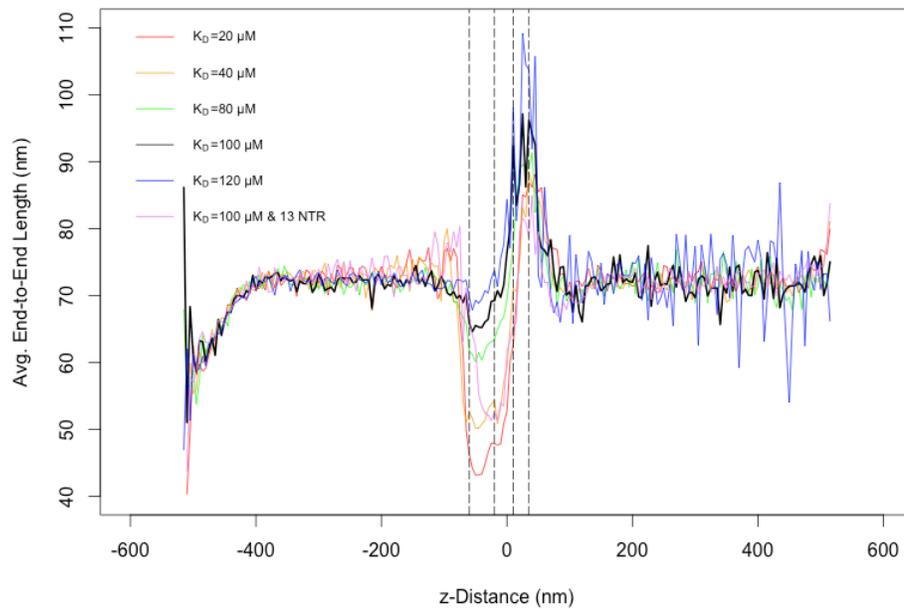
Figure 10. Plot of end-to-end length averaged over 100 simulations for multiple affinity and nuclear transport receptor (NTR) configurations in the ABM model. Note: where unspecified, the number of transport receptors was set to the baseline value of nine, distributed uniformly along the length of the mRNA. The x-axis represents the position along the axis perpendicular to the nuclear envelope (z-Distance) with x=0 set at the center of the central channel of the NPC. The first dashed line from the left represents the distal edge of the nuclear basket, the second dashed line from the left represents the nuclear edge of the central channel, the third dashed line from the left represents the cytoplasmic edge of the central channel and the fourth dashed line from the left represents the distal edge of the cytoplasmic filaments.

Discussion

Although accurate experimental techniques are ultimately essential for enhancing our understanding of mRNA export dynamics, effective computational models can aid in constructing hypotheses and experimental design while also revealing potential systematic limitations. Using the agent-based modeling modality, we have constructed a rudimentary model of mRNA export through the NPC that allows for testing how several kinds of parameters affect this process. We first assessed the impact of affinity between mRNA-binding nuclear transport receptor (NTR) NXF1 and the NPC's FG Nups. As shown in Fig. 4, an affinity of 100 μM yielded mRNA export efficacy that is most consistent with previous experimental observations. Furthermore, residence time within both the nuclear basket and the central channel was minimized under this affinity, Fig. 5. A balance between export likelihood and pore residence time of an mRNA is suggested by these results. Taken in the context of bulk export, this balance likely improves transport efficiency for other cargos. This is done by creating a pore that can remain occluded rather than one that can successfully export every mRNA that it encounters, which would result in very slow-transporting mRNAs that virtually "clog" the pore.

Additionally, the effect of density of NTR binding to the mRNA was also examined. Raising the density of NTRs bound to the mRNA lead to increased transport efficiency, while lowering it dramatically ablated transport. These observations hint at the degree of stringency behind the NTR-mediated method of export that cells would have to evolve to leverage the biophysics that are relevant to this process. In further support of this idea, while full coverage of the mRNA by NTRs appears to favor optimal transport conditions, coverage of at least one mRNA terminus may be necessary to ensure any successful export. To speculate on a potential explanation for this requirement, having an NTR-covered terminus may ensure that the mRNA can begin threading through the pore after sweeping the conformational space. Because similar trends in NTR distribution-dependent export efficacy were observed for case of 20 μM NXF1-Nup affinity (data not shown), they appear to be independent of affinity.

Importantly, our model of mRNA export sheds considerable insight on potential sources of systematic error that may be present in previously-utilized experimental approaches aimed at studying this process. Our results demonstrate that, regardless of affinity or NTR distribution, using a single mRNA tracking probe may lead to an overestimation of transport rate. A single probe only provides a local snapshot for the region along the mRNA that contains the tag. In addition, we observed that the use of a single tracking probe consistently overestimated the number of successful export events (data not shown). However, this over-estimation was in the range of 0-5% depending on affinity and transport receptor distribution. Subsequently, we can conclude that the use of a single tracking probe is sufficient for determining the fraction of successful mRNA export events. This is likely due to the relative duration

of a typical mRNA export event being much shorter than the duration of observation – in this case, ~1.0 second of pore residence compared to 20 seconds of observation. As the difference between these durations is reduced, the confidence in single probe measurements of transport likelihood becomes questionable.

With our ABM, we were able to observe the coarse elongation of the mRNA as it threaded through the NPC (Fig. 10), which is consistent with previous experimental findings. In addition, a collapse in configuration as the mRNA contacted the nuclear basket was observed. Both our simulations and previous experimental findings suggest that successful threading of the mRNA terminus is the rate-limiting step in mRNA transport. As such, it could be speculated that by forcing the mRNA into a more compact conformation, the basket reduces the number of entropic states that the mRNA must search in order to find an optimal conformation for export through the narrow central channel, thus accelerating the rate of transport. Crucially, such behavior of the mRNA cannot be observed in experimental studies that utilize one probe per mRNA molecule for labeling purposes.

As with other studies, ours has its own limitations that we wish to acknowledge. The ABM that we have produced recapitulates only coarse-grained behavior of an mRNA-like molecule as it transports through the NPC. Certainly, the detailed numbers behind our findings should be taken with a grain of salt. In addition, while it has been strongly suggested that different FG Nups within the NPC exhibit varying affinities NTR-cargo complexes, our model utilizes homogeneous affinities and thus potentially fails to capture intricacies in the export process. Despite these and other limitations, however, the model does faithfully recapitulate some of the coarse experimentally-observed processes, and therefore has potential for further refinement in the future.

Finally, it should be noted that the above discussion only addresses nuclear export of mRNA via CRM1 and NXF1/NXT1-mediated pathways via the NPC. In 2012, Speese and colleagues identified an alternate pathway for nuclear export of DFz2C mRNPs via budding of the nuclear envelope analogous to export pathways used by viruses such as herpes [19,20]. There remain many questions as to why these molecules are exported via nuclear budding as opposed to via the NPC (the authors don't exclude the possibility that some DFz2C mRNP could be exported via NPCs). Possible explanations for an alternative transport pathway such as size limits, co-export of multiple mRNPs, preserving mRNP structure for transcriptional repression during export and pathways presenting alternate regulation have been suggested [21]. This alternate pathway remains a topic of interest as more details become available.

Conclusion

These simulations highlight characteristics of mRNA export and may have further implications for targeted delivery and export of polymeric molecules into and out the nucleus – potentially in applications relating to gene therapy and/or disruption of viral entry or exit into and out of the nucleus. The key observations we made were that mRNA export appears very sensitive to the number of transport receptors lining the mRNA. A single transport receptor at the 5' end appeared insufficient for facilitating export. Increasing transport receptor coverage along the length of the mRNA improved the chances of successful export. Additionally, it was observed that the presence of a transport receptor near either the 5' or 3' terminus is required for successful export as it likely promotes the emergence of a favorable threading conformation and facilitates transport. The observations made for transport receptor number and distribution were observed to occur at other affinities, higher and lower, than the standard configuration affinity. Subsequently, we concluded that these effects were not an artifact of the affinity used in the standard configuration. Finally, we observed that the use of a single probe is typically sufficient for differentiating successful transports and failed transports, given that the observation time is sufficiently longer than the time of transport. However, it appears unlikely that a single probe can be used to correctly capture the transport time of an mRNA threading through a pore.

References

1. Kohler A, Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol 8: 761-773.
2. Rodriguez MS, Dargemont C, Stutz F (2004) Nuclear export of RNA. Biol Cell 96: 639-655.
3. Carmody SR, Wente SR (2009) mRNA nuclear export at a glance. J Cell Sci 122: 1933-1937.
4. Natalizio BJ, Wente SR (2013) Postage for the messenger: designating routes for nuclear mRNA export. Trends Cell Biol.
5. Montpetit B, Thomsen ND, Helmke KJ, Seeliger MA, Berger JM, et al. (2011) A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. Nature 472: 238-242.
6. Stewart M (2007) Ratcheting mRNA out of the nucleus. Mol Cell 25: 327-330.
7. Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, et al. (2006) Human mRNA export machinery recruited to the 5' end of mRNA. Cell 127: 1389-1400.
8. Mehlin H, Daneholt B, Skoglund U (1992) Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. Cell 69: 605-613.
9. Mehlin H, Daneholt B, Skoglund U (1995) Structural interaction between the nuclear pore complex and a specific translocating RNP particle. J Cell Biol 129: 1205-1216.
10. Visa N, Izaurralde E, Ferreira J, Daneholt B, Mattaj IW (1996) A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. J Cell Biol 133: 5-14.
11. Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, et al. (2000) Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. Mol Cell 6: 673-682.
12. Le Hir H, Izaurralde E, Maquat LE, Moore MJ (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J 19: 6860-6869.
13. Le Hir H, Gatfield D, Izaurralde E, Moore MJ (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J 20: 4987-4997.
14. Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet 7: 211-221.
15. Querido E, Chartrand P (2008) Using fluorescent proteins to study mRNA trafficking in living cells. Methods Cell Biol 85: 273-292.
16. Grunwald D, Singer RH (2010) In vivo imaging of labelled endogenous beta-actin mRNA during nucleocytoplasmic transport. Nature 467: 604-607.

17. Siebrasse JP, Kaminski T, Kubitscheck U (2012) Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy. Proc Natl Acad Sci U S A 109: 9426-9431.
18. Kaminski T, Siebrasse JP, Kubitscheck U (2013) A single molecule view on Dbp5 and mRNA at the nuclear pore. Nucleus 4: 8-13.
19. Speese SD, Ashley J, Jokhi V, Nunnari J, Barria R, et al. (2012) Nuclear envelope budding enables large ribonucleoprotein particle export during synaptic Wnt signaling. Cell 149: 832-846.
20. Hatch EM, Hetzer MW (2012) RNP export by nuclear envelope budding. Cell 149: 733-735.
21. Montpetit B, Weis K (2012) Cell biology. An alternative route for nuclear mRNP export by membrane budding. Science 336: 809-810.
22. Berry BJL, Kiel LD, Elliott E (2002) Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling. Proceedings of the National Academy of Sciences of the United States of America 99: 7187-7188.
23. Bonabeau E (2002) Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America 99: 7280-7287.
24. Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, et al. (2005) Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. Science 310: 987-991.
25. Helbing D, Farkas I, Vicsek T (2000) Simulating dynamical features of escape panic. Nature 407: 487-490.
26. Dong X, Foteinou PT, Calvano SE, Lowry SF, Androulakis IP (2010) Agent-based modeling of endotoxin-induced acute inflammatory response in human blood leukocytes. PLoS One 5: e9249.
27. Bonchev D, Thomas S, Apte A, Kier LB (2010) Cellular automata modelling of biomolecular networks dynamics. SAR QSAR Environ Res 21: 77-102.
28. Devillers J, Devillers H, Decourtye A, Aupinel P (2010) Internet resources for agent-based modelling. SAR QSAR Environ Res 21: 337-350.
29. Azimi M, Jamali Y, Mofrad MR (2011) Accounting for diffusion in agent based models of reaction-diffusion systems with application to cytoskeletal diffusion. PLoS One 6: e25306.
30. Jamali Y, Jamali T, Mofrad MRK (In Press) An Agent Based Model of Integrin Clustering: Exploring the Role of Ligand Clustering, Integrin Homo-Oligomerization, Integrin-Ligand Affinity, Membrane Crowdedness and Ligand Mobility. Journal of Computational Physics.
31. Mor A, Suliman S, Ben-Yishay R, Yunger S, Brody Y, et al. (2010) Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol 12: 543-552.
32. Adam SA (2001) The nuclear pore complex. Genome Biol 2: REVIEWS0007.
33. Loschberger A, van de Linde S, Dabauvalle MC, Rieger B, Heilemann M, et al. (2012) Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution. J Cell Sci 125: 570-575.

34. Yamada J, Phillips JL, Patel S, Goldfien G, Calestagne-Morelli A, et al. (2010) A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. Mol Cell Proteomics 9: 2205-2224.
35. Mohr D, Frey S, Fischer T, Guttler T, Gorlich D (2009) Characterisation of the passive permeability barrier of nuclear pore complexes. EMBO J 28: 2541-2553.
36. Riddick G, Macara IG (2005) A systems analysis of importin-{alpha}-{beta} mediated nuclear protein import. J Cell Biol 168: 1027-1038.
37. Hobeika M, Brockmann C, Gruessing F, Neuhaus D, Divita G, et al. (2009) Structural requirements for the ubiquitin-associated domain of the mRNA export factor Mex67 to bind its specific targets, the transcription elongation THO complex component Hpr1 and nucleoporin FXFG repeats. J Biol Chem 284: 17575-17583.
38. Tetenbaum-Novatt J, Hough LE, Mironska R, McKenney AS, Rout MP (2012) Nucleocytoplasmic transport: a role for nonspecific competition in karyopherin-nucleoporin interactions. Mol Cell Proteomics 11: 31-46.
39. Grunwald D, Singer RH, Rout M (2011) Nuclear export dynamics of RNA-protein complexes. Nature 475: 333-341.
40. Daneholt B (1997) A look at messenger RNP moving through the nuclear pore. Cell 88: 585-588.

Appendix A:

FORTRAN 90 Implementation of Agent Movement

```fortran
subroutine MOVE(noNeighbors,geometry,N,world,numTypes,maxParticles,pTypes)

integer N,geometry
TYPE (element) world(0:N-1)
integer numTypes,maxParticles
TYPE (particleTypes) pTypes(numTypes)
integer i,j
integer k,dk
integer t,ttemp,told
double precision RND, grnd
integer oldEl,newEl, partType(2)
integer noNeighbors
integer leftMono, currMono, rightMono
integer leftMonoID(2), rightMonoID(2)
integer polyMoveTrack(0:maxParticles-1),numPolyMoved

do i=1,numTypes
  if (pTypes(i).isPolymer.eq.0) then
    do j=0,maxParticles-1
      if(particles(i,j,1).ge.0) then
        oldEl=particles(i,j,1)
        RND=grnd()
        k=INT(RND*noNeighbors)
        if(k.lt.noNeighbors) then
          newEl=world(oldEl).neighbor(k)
          if( & (world(newEl).sumVolume+pTypes(i).partVolume).le.1.AND. &
            world(newEl).numParticles.lt.maxPerCell.AND. &
            oldEl.ne.newEl) then
            RND=grnd()
```

```
if(RND.lt.pTypes(i).diffC(world(newEl).elementType)) then
    RND=grnd()
    if(RND.lt.(1-world(newEl).sumVolume)) then
        world(newEl).numParticles=world(newEl).numParticles+1
        world(oldEl).numParticles=world(oldEl).numParticles-1
        told=particles(i,j,2)
        partType=world(oldEl).typeParticles(particles(i,j,2),:)
        t=0
        ttemp=1
        do while (t.le.maxPerCell.AND.ttemp.gt.0)
            t=t+1
            ttemp=world(newEl).typeParticles(t,1)
        enddo
        world(newEl).typeParticles(t,:)=partType
        world(oldEl).typeParticles(told,:)=(/0,0/)
        world(newEl).sumVolume= world(newEl).sumVolume+pTypes(i).partVolume
        world(oldEl).sumVolume= world(oldEl).sumVolume-pTypes(i).partVolume
        particles(partType(1),partType(2),1)=newEl
        particles(partType(1),partType(2),2)=t
    endif
    endif
    elseif(world(newEl).numParticles.ge.maxPerCell) then
        write(*,*) "ERROR, MAX PARTICLE PER CELL EXCEEDED"
    endif
    endif
    endif
    endif
    endif
    enddo
else
    polyMoveTrack = 0
    numPolyMoved = 0
```

143

```
do while (numPolyMoved.lt.maxParticles)
j=INT(grnd()*maxParticles)
if (polyMoveTrack(j).eq.0) then
polyMoveTrack(j) = 1
if(particles(i,j,1).ge.0) then
currMono = particles(i,j,1)
if (polyNeighbor(i,j,1).eq.-1) then
rightMono = world(currMono).neighbor(polyNeighbor(i,j,2))
rightMonoID = (/ polyNeighbor(i,j,5), polyNeighbor(i,j,6) /)
RND=grnd()
k=INT(RND*noNeighbors)
newEl=world(rightMono).neighbor(k)
if(world(newEl).numParticles.lt.maxPerCell.AND. &
rightMono.ne.newEl.AND. &
currMono.ne.newEl.AND. &
k.ne.dirOpposing(polyNeighbor(rightMonoID(1),rightMonoID(2),1))) then
RND=grnd()
if(RND.lt.pTypes(i).diffC(world(newEl).elementType)) then
RND=grnd()
if(RND.lt.(1-world(newEl).sumVolume)) then
world(newEl).numParticles=world(newEl).numParticles+1
world(currMono).numParticles=world(currMono).numParticles-1
told=particles(i,j,2)
partType=world(currMono).typeParticles(particles(i,j,2),:)
world(currMono).typeParticles(particles(i,j,2),:)
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
t=t+1
ttemp=world(newEl).typeParticles(t,1)
enddo
```

144

```fortran
world(newEl).typeParticles(t,:)=partType
world(currMono).typeParticles(told,:)=(/0,0/)
world(newEl).sumVolume= world(newEl).sumVolume+pTypes(i).partVolume
world(currMono).sumVolume= world(currMono).sumVolume-pTypes(i).partVolume
particles(partType(1),partType(2),1)=newEl
particles(partType(1),partType(2),2)=t
polyNeighbor(i,j,2) = dirOpposing(k)
polyNeighbor(rightMonoID(1),rightMonoID(2),1) = k
        endif
      endif
    endif
elseif (polyNeighbor(i,j,2).eq.-1) then
leftMono = world(currMono).neighbor(polyNeighbor(i,j,1))
leftMonoID = (/ polyNeighbor(i,j,3), polyNeighbor(i,j,4) /)
RND=grnd()
k=INT(RND*noNeighbors)
newEl=world(leftMono).neighbor(k)
if(world(newEl).numParticles.lt.maxPerCell.AND. &
leftMono.ne.newEl.AND. &
currMono.ne.newEl.AND. &
k.ne.dirOpposing(polyNeighbor(leftMonoID(1),leftMonoID(2),2))) then
RND=grnd()
if(RND.lt.pTypes(i).diffC(world(newEl).elementType)) then
RND=grnd()
if(RND.lt.(1-world(newEl).sumVolume)) then
world(newEl).numParticles=world(newEl).numParticles+1
world(currMono).numParticles=world(currMono).numParticles-1
told=particles(i,j,2)
partType=world(currMono).typeParticles(particles(i,j,2),:)
t=0
```

```
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
    t=t+1
    ttemp=world(newEl).typeParticles(t,1)
enddo
world(newEl).typeParticles(t,:)=partType
world(currMono).typeParticles(told,:)=(/0,0/)
world(newEl).sumVolume= world(newEl).sumVolume+pTypes(i).partVolume
world(currMono).sumVolume= world(currMono).sumVolume-pTypes(i).partVolume
particles(partType(1),partType(2),1)=newEl
particles(partType(1),partType(2),2)=t
polyNeighbor(i,j,1) = dirOpposing(k)
polyNeighbor(leftMonoID(1),leftMonoID(2),2) = k
            endif
        endif
    endif
else
rightMono = world(currMono).neighbor(polyNeighbor(i,j,2))
rightMonoID = (/ polyNeighbor(i,j,5), polyNeighbor(i,j,6) /)
leftMono = world(currMono).neighbor(polyNeighbor(i,j,1))
leftMonoID = (/ polyNeighbor(i,j,3), polyNeighbor(i,j,4) /)
newEl=world(leftMono).neighbor(polyNeighbor(i,j,2))
if (polyNeighbor(i,j,1) .ne. polyNeighbor(i,j,2) .AND. &
polyNeighbor(i,j,1) .ne. dirOpposing(polyNeighbor(i,j,2)) .AND. &
world(newEl).numParticles.lt.maxPerCell.AND. &
rightMono.ne.newEl.AND. &
leftMono.ne.newEl.AND. &
currMono.ne.newEl) then
RND=grnd()
if(RND.lt.pTypes(i).diffC(world(newEl).elementType)) then
```

```
RND=grnd()
if(RND.lt.(1-world(newEl).sumVolume)) then
world(newEl).numParticles=world(newEl).numParticles+1
world(currMono).numParticles=world(currMono).numParticles-1
told=particles(i,j,2)
partType=world(currMono).typeParticles(particles(i,j,2),:)
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
t=t+1
ttemp=world(newEl).typeParticles(t,1)
enddo
world(newEl).typeParticles(t,:)=partType
world(currMono).typeParticles(told,:)=(/0,0/)
world(newEl).sumVolume= world(newEl).sumVolume+pTypes(i).partVolume
world(currMono).sumVolume= world(currMono).sumVolume-pTypes(i).partVolume
particles(partType(1),partType(2),1)=newEl
particles(partType(1),partType(2),2)=t
polyNeighbor(leftMonoID(1),leftMonoID(2),2) = polyNeighbor(i,j,2)
polyNeighbor(rightMonoID(1),rightMonoID(2),1) = polyNeighbor(i,j,1)
polyNeighbor(i,j,1) = dirOpposing(polyNeighbor(leftMonoID(1),leftMonoID(2),2))
polyNeighbor(i,j,2) = dirOpposing(polyNeighbor(rightMonoID(1),rightMonoID(2),1))
endif
endif
elseif (polyNeighbor(i,j,1).eq.polyNeighbor(i,j,2)) then
RND=grnd()
k=INT(RND*noNeighbors)
newEl=world(leftMono).neighbor(k)
if((world(newEl).numParticles.lt.maxPerCell.AND. &
rightMono.ne.newEl.AND. &
```

```
leftMono.ne.newEl.AND. &
currMono.ne.newEl.AND. &
k.ne.dirOpposing(polyNeighbor(leftMonoID(1),leftMonoID(2),2))) then
RND=grnd()
if(RND.lt.pTypes(i).diffC(world(newEl).elementType)) then
RND=grnd()
if(RND.lt.(1-world(newEl).sumVolume)) then
world(newEl).numParticles=world(newEl).numParticles+1
world(currMono).numParticles=world(currMono).numParticles-1
told=particles(i,j,2)
partType=world(currMono).typeParticles(particles(i,j,2),:)
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
   t=t+1
   ttemp=world(newEl).typeParticles(t,1)
enddo
world(newEl).typeParticles(t,:)=partType
world(currMono).typeParticles(told,:)=(/0,0/)
world(newEl).sumVolume= world(newEl).sumVolume+pTypes(i).partVolume
world(currMono).sumVolume= world(currMono).sumVolume-pTypes(i).partVolume
particles(partType(1),partType(2),1)=newEl
particles(partType(1),partType(2),2)=t
polyNeighbor(i,j,1) = dirOpposing(k)
polyNeighbor(i,j,2) = dirOpposing(k)
polyNeighbor(rightMonoID(1),rightMonoID(2),1) = k
polyNeighbor(leftMonoID(1),leftMonoID(2),2) = k
endif
endif
endif
```

```
                endif
            endif
          endif
          numPolyMoved = SUM(polyMoveTrack)
        endif
      enddo
    endif
  enddo
end subroutine MOVE
```

Appendix B:

FORTRAN 90 Implementation of Agent Binding and Unbinding

```fortran
subroutine BIOCHEM(noNeighbors,N,world,numP,numTypes,maxParticles,pTypes,numrTypes,rTypes)

integer N
TYPE (element) world(0:N-1)
integer numP
integer numTypes,maxParticles,numrTypes
TYPE (particleTypes) pTypes(numTypes)
TYPE (rxnTypes) rTypes(numrTypes)
integer noNeighbors
integer i,j,r,k,dk,p
integer s,stemp,t,ttemp
double precision RND, grnd
integer rxnEl1,rxnEl2,pTyp1(2),pTyp2(2)
integer norxn
double precision RNDNeighbor
integer rxnsTrack(numrTypes),numRxnsComplete
integer leftMonoID(2), rightMonoID(2)

rxnsTrack = 0
numRxnsComplete = 0
do while (numRxnsComplete.lt.numrTypes)
r=INT(grnd()*numrTypes)+1
  if (r.le.numrTypes) then
    if (rxnsTrack(r).eq.0) then
      rxnsTrack(r) = 1
      i = rTypes(r).rxnInput(1)
      do j=0,maxParticles-1
        pTyp1(1)=i
        pTyp1(2)=j
```

151

```
rxnEl1=particles(i,j,1)
if (rxnEl1.ge.0) then
  if(rTypes(r).numInput.eq.1) then
    if (rTypes(r).numOutputs.eq.1) then
      RND=grnd()
      if(RND.lt.rTypes(r).kO) then
        world(rxnEl1).numParticles=world(rxnEl1).numParticles-1
        world(rxnEl1).typeParticles(particles(pTyp1(1),pTyp1(2),2),:)=(/0,0/)
        particles(pTyp1(1),pTyp1(2),:)=(/-1,-1/)
        world(rxnEl1).sumVolume= world(rxnEl1).sumVolume-pTypes(rTypes(r).rxnInput(1)).partVolume
        t=0
        ttemp=1
        do while (t.le.maxPerCell.AND.ttemp.gt.0)
          t=t+1
          ttemp=world(rxnEl1).typeParticles(t,1)
        enddo
        s=-1
        stemp=0
        do while (s.le.maxParticles-1.AND.stemp.ge.0)
          s=s+1
          stemp=particles(rTypes(r).rxnOutput(1),s,1)
        enddo
        world(rxnEl1).numParticles=world(rxnEl1).numParticles+1
        world(rxnEl1).typeParticles(t,1)=rTypes(r).rxnOutput(1)
        world(rxnEl1).typeParticles(t,2)=s
        world(rxnEl1).sumVolume= world(rxnEl1).sumVolume+pTypes(rTypes(r).rxnOutput(1)).partVolume
        particles(rTypes(r).rxnOutput(1),s,1)=rxnEl1
        particles(rTypes(r).rxnOutput(1),s,2)=t
        if (pTypes(i).isPolymer.eq.1) then
          leftMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),3), polyNeighbor(pTyp1(1),pTyp1(2),4) /)
```

```
rightMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),5), polyNeighbor(pTyp1(1),pTyp1(2),6) /)
polyNeighbor(rTypes(r).rxnOutput(1),s,1:6) = polyNeighbor(pTyp1(1),pTyp1(2),1:6)
if (polyNeighbor(pTyp1(1),pTyp1(2),1).ge.0) then
   polyNeighbor(leftMonoID(1),leftMonoID(2),5) = rTypes(r).rxnOutput(1)
   polyNeighbor(leftMonoID(1),leftMonoID(2),6) = s
endif
if (polyNeighbor(pTyp1(1),pTyp1(2),2).ge.0) then
   polyNeighbor(rightMonoID(1),rightMonoID(2),3) = rTypes(r).rxnOutput(1)
   polyNeighbor(rightMonoID(1),rightMonoID(2),4) = s
endif
polyNeighbor(pTyp1(1),pTyp1(2),1:6) = -1
endif
endif
elseif (rTypes(r).numOutputs.eq.2) then
RND=grnd()
if(RND.lt.rTypes(r).kO) then
numP=numP+1
world(rxnEl1).numParticles=world(rxnEl1).numParticles-1
world(rxnEl1).typeParticles(particles(pTyp1(1),pTyp1(2),2),:)=(/0,0/)
particles(pTyp1(1),pTyp1(2),:)=(/ -1,-1/)
world(rxnEl1).sumVolume= world(rxnEl1).sumVolume-pTypes(rTypes(r).rxnInput(1)).partVolume
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
   t=t+1
   ttemp=world(rxnEl1).typeParticles(t,1)
enddo
s=-1
stemp=0
do while (s.le.maxParticles-1.AND.stemp.ge.0)
```

```
s=s+1
      stemp=particles(rTypes(r).rxnOutput(1),s,1)
enddo
world(rxnEl1).numParticles=world(rxnEl1).numParticles+1
world(rxnEl1).typeParticles(t,1)=rTypes(r).rxnOutput(1)
world(rxnEl1).typeParticles(t,2)=s
world(rxnEl1).sumVolume= world(rxnEl1).sumVolume+pTypes(rTypes(r).rxnOutput(1)).partVolume
particles(rTypes(r).rxnOutput(1),s,1)=rxnEl1
particles(rTypes(r).rxnOutput(1),s,2)=t
if (pTypes(i).isPolymer.eq.1) then
leftMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),3), polyNeighbor(pTyp1(1),pTyp1(2),4) /)
rightMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),5), polyNeighbor(pTyp1(1),pTyp1(2),6) /)
polyNeighbor(rTypes(r).rxnOutput(1),s,1:6) = polyNeighbor(pTyp1(1),pTyp1(2),1:6)
if (polyNeighbor(pTyp1(1),pTyp1(2),1) .ge. 0) then
polyNeighbor(leftMonoID(1),leftMonoID(2),5) = rTypes(r).rxnOutput(1)
polyNeighbor(leftMonoID(1),leftMonoID(2),6) = s
endif
if (polyNeighbor(pTyp1(1),pTyp1(2),2) .ge. 0) then
polyNeighbor(rightMonoID(1),rightMonoID(2),3) = rTypes(r).rxnOutput(1)
polyNeighbor(rightMonoID(1),rightMonoID(2),4) = s
endif
polyNeighbor(pTyp1(1),pTyp1(2),1:6) = -1
endif
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
t=t+1
ttemp=world(rxnEl1).typeParticles(t,1)
enddo
s=-1
```

```fortran
        stemp=0
        do while (s.le.maxParticles-1.AND.stemp.ge.0)
          s=s+1
          stemp=particles(rTypes(r).rxnOutput(2),s,1)
        enddo
        world(rxnEl1).numParticles=world(rxnEl1).numParticles+1
        world(rxnEl1).typeParticles(t,1)=rTypes(r).rxnOutput(2)
        world(rxnEl1).typeParticles(t,2)=s
        world(rxnEl1).sumVolume= world(rxnEl1).sumVolume+pTypes(rTypes(r).rxnOutput(2)).partVolume
        particles(rTypes(r).rxnOutput(2),s,1)=rxnEl1
        particles(rTypes(r).rxnOutput(2),s,2)=t
      endif
    endif
  elseif(rTypes(r).numInput.eq.2) then
    RNDNeighbor=grnd()
    do dk=0,noNeighbors-1+1
      k=MOD(dk+INT(RNDNeighbor*(noNeighbors+1)),noNeighbors+1)
      if(k.lt.noNeighbors)then
        rxnEl2=world(rxnEl1).neighbor(k)
      else
        rxnEl2=rxnEl1
      endif
      do p=1,maxPerCell
        if(rxnEl1.eq.rxnEl2.AND.particles(pTyp1(1),pTyp1(2),2).eq.p)then
          goto 10
        endif
        pTyp2(1)=world(rxnEl2).typeParticles(p,1)
        pTyp2(2)=world(rxnEl2).typeParticles(p,2)
        if(pTyp2(1).eq.rTypes(r).rxnInput(2)) then
          if(pTypes(pTyp2(1)).diffC(world(rxnEl2).elementType).eq.0)then
```

```fortran
      goto 10
      endif
      RND=grnd()
      if(RND.lt.rTypes(r).kO) then
        world(rxnEl1).numParticles=world(rxnEl1).numParticles-1
        world(rxnEl2).numParticles=world(rxnEl2).numParticles-1
        world(rxnEl1).typeParticles(particles(pTyp1(1),pTyp1(1)),:)=(/0,0/)
        world(rxnEl2).typeParticles(particles(pTyp2(1),pTyp2(1)),:)=(/0,0/)
        particles(pTyp1(1),pTyp1(2),:)=(/-1,-1/)
        particles(pTyp2(1),pTyp2(2),:)=(/-1,-1/)
        world(rxnEl1).sumVolume= world(rxnEl1).sumVolume-pTypes(rTypes(r).rxnInput(1)).partVolume
        world(rxnEl2).sumVolume= world(rxnEl2).sumVolume-pTypes(rTypes(r).rxnInput(2)).partVolume
      if(rTypes(r).numOutputs.eq.1) then
        numP=numP-1
        s=-1
        stemp=0
      do while (s.le.maxParticles-1.AND.stemp.ge.0)
        s=s+1
        stemp=particles(rTypes(r).rxnOutput(1),s,1)
      enddo
      RND=grnd()
      if(RND.le.1.0) then
        t=0
        ttemp=1
      do while (t.le.maxPerCell.AND.ttemp.gt.0)
        t=t+1
        ttemp=world(rxnEl1).typeParticles(t,1)
      enddo
        world(rxnEl1).numParticles=world(rxnEl1).numParticles+1
        world(rxnEl1).typeParticles(t,1)=rTypes(r).rxnOutput(1)
```

156

```
world(rxnEl1).typeParticles(t,2)=s
world(rxnEl1).sumVolume=world(rxnEl1).sumVolume+pTypes(rTypes(r).rxnOutput(1)).partVolume
particles(rTypes(r).rxnOutput(1),s,1)=rxnEl1
particles(rTypes(r).rxnOutput(1),s,2)=t
if (pTypes(i).isPolymer.eq.1) then
leftMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),3), polyNeighbor(pTyp1(1),pTyp1(2),4) /)
rightMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),5), polyNeighbor(pTyp1(1),pTyp1(2),6) /)
polyNeighbor(rTypes(r).rxnOutput(1),s,1:6) = polyNeighbor(pTyp1(1),pTyp1(2),1:6)
if (polyNeighbor(pTyp1(1),pTyp1(2),1).ge.0) then
polyNeighbor(leftMonoID(1),leftMonoID(2),5) = rTypes(r).rxnOutput(1)
polyNeighbor(leftMonoID(1),leftMonoID(2),6) = s
endif
if (polyNeighbor(pTyp1(1),pTyp1(2),2).ge.0) then
polyNeighbor(rightMonoID(1),rightMonoID(2),3) = rTypes(r).rxnOutput(1)
polyNeighbor(rightMonoID(1),rightMonoID(2),4) = s
endif
polyNeighbor(pTyp1(1),pTyp1(2),1:6) = -1
endif
else
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
t=t+1
ttemp=world(rxnEl2).typeParticles(t,1)
enddo
world(rxnEl2).numParticles=world(rxnEl2).numParticles+1
world(rxnEl2).typeParticles(t,1)=rTypes(r).rxnOutput(1)
world(rxnEl2).typeParticles(t,2)=s
world(rxnEl2).sumVolume=world(rxnEl2).sumVolume+pTypes(rTypes(r).rxnOutput(1)).partVolume
particles(rTypes(r).rxnOutput(1),s,1)=rxnEl2
```

```
particles(rTypes(r).rxnOutput(1),s,2)=t
if (pTypes(i).isPolymer.eq.1) then
    leftMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),3), polyNeighbor(pTyp1(1),pTyp1(2),4) /)
    rightMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),5), polyNeighbor(pTyp1(1),pTyp1(2),6) /)
    polyNeighbor(rTypes(r).rxnOutput(1),s,1:6) = polyNeighbor(pTyp1(1),pTyp1(2),1:6)
    if (polyNeighbor(pTyp1(1),pTyp1(2),1).ge.0) then
        polyNeighbor(leftMonoID(1),leftMonoID(2),5) = rTypes(r).rxnOutput(1)
        polyNeighbor(leftMonoID(1),leftMonoID(2),6) = s
    endif
    if (polyNeighbor(pTyp1(1),pTyp1(2),2).ge.0) then
        polyNeighbor(rightMonoID(1),rightMonoID(2),3) = rTypes(r).rxnOutput(1)
        polyNeighbor(rightMonoID(1),rightMonoID(2),4) = s
    endif
    polyNeighbor(pTyp1(1),pTyp1(2),1:6) = -1
endif
endif
elseif(rTypes(r).numOutputs.eq.2) then
s=-1
stemp=0
do while (s.le.maxParticles-1.AND.stemp.ge.0)
    s=s+1
    stemp=particles(rTypes(r).rxnOutput(1),s,1)
enddo
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
    t=t+1
    ttemp=world(rxnEl1).typeParticles(t,1)
enddo
world(rxnEl1).numParticles=world(rxnEl1).numParticles+1
```

158

```
world(rxnEl1).typeParticles(t,1)=rTypes(r).rxnOutput(1)
world(rxnEl1).typeParticles(t,2)=s
world(rxnEl1).sumVolume= world(rxnEl1).sumVolume+pTypes(rTypes(r).rxnOutput(1)).partVolume
particles(rTypes(r).rxnOutput(1),s,1)=rxnEl1
particles(rTypes(r).rxnOutput(1),s,2)=t
if (pTypes(i).isPolymer.eq.1) then
leftMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),3), polyNeighbor(pTyp1(1),pTyp1(2),4) /)
rightMonoID = (/ polyNeighbor(pTyp1(1),pTyp1(2),5), polyNeighbor(pTyp1(1),pTyp1(2),6) /)
polyNeighbor(rTypes(r).rxnOutput(1),s,1:6) = polyNeighbor(pTyp1(1),pTyp1(2),1:6)
if (polyNeighbor(pTyp1(1),pTyp1(2),1).ge.0) then
polyNeighbor(leftMonoID(1),leftMonoID(2),5) = rTypes(r).rxnOutput(1)
polyNeighbor(leftMonoID(1),leftMonoID(2),6) = s
endif
if (polyNeighbor(pTyp1(1),pTyp1(2),2).ge.0) then
polyNeighbor(rightMonoID(1),rightMonoID(2),3) = rTypes(r).rxnOutput(1)
polyNeighbor(rightMonoID(1),rightMonoID(2),4) = s
endif
polyNeighbor(pTyp1(1),pTyp1(2),1:6) = -1
endif
s=-1
stemp=0
do while (s.le.maxParticles-1.AND.stemp.ge.0)
s=s+1
stemp=particles(rTypes(r).rxnOutput(2),s,1)
enddo
t=0
ttemp=1
do while (t.le.maxPerCell.AND.ttemp.gt.0)
t=t+1
ttemp=world(rxnEl2).typeParticles(t,1)
```

```
        enddo
        world(rxnEl2).numParticles=world(rxnEl2).numParticles+1
        world(rxnEl2).typeParticles(t,1)=rTypes(r).rxnOutput(2)
        world(rxnEl2).typeParticles(t,2)=s
        world(rxnEl2).sumVolume= world(rxnEl2).sumVolume+pTypes(rTypes(r).rxnOutput(2)).partVolume
        particles(rTypes(r).rxnOutput(2),s,1)=rxnEl2
        particles(rTypes(r).rxnOutput(2),s,2)=t
        endif
        goto 20
        endif
        endif
        endif
        endif
20      enddo
        enddo
10      enddo
        endif
        endif
        endif
        numRxnsComplete = SUM(rxnsTrack)
        endif
        enddo
        enddo
end subroutine BIOCHEM
```