MODELING MICROTUBULE DYNAMIC INSTABILITY

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Ivan Gregoretti, Licenciado en Ciencias Químicas

_____________________________________________________________________

Holly V. Goodson, Director

_____________________________________________________________________

Mark S. Alber, Director

Graduate Program in Chemistry and Biochemistry
Notre Dame, Indiana
December 2007
Abstract

by

Ivan Gregoretti

Microtubules are non-covalent polymers important for many processes such as vesicle trafficking and establishment of cell polarity and essential for cell division. A key property of microtubules is that they are highly dynamic. Microtubules constantly switch between phases of growth and shortening. This behavior is called dynamic instability. Although the importance of dynamic instability is well established from a cell-biological point of view, its regulation and mechanistic details are poorly understood. This doctoral thesis summarizes an effort to better understand dynamic instability by means of an interdisciplinary approach. The opening chapter presents an overview of microtubule dynamics and its pending questions. Then, an introduction to stochastic modeling in biology is presented for readers of diverse background. The dynamics of the microtubule network is then studied with a model at a mesoscopic scale (coarse grain) and a model at a microscopic scale (fine details). The mesoscopic modeling results indicate that many behaviors thought to require regulatory proteins are instead unavoidable outcomes of the physical constraints on a system of nucleated polymers competing in a confined space. This suggests that regulatory proteins tune microtubule dynamics rather than govern it. This conclusion has important evolutionary implications as microtubules are present in all eukaryotes and therefore their underlying mechanistic principles must be robust. The microscopic model is the first one built
at a the dimer scale to recapitulate dynamic instability. With this model, the mechanistic details of the once paradoxical microtubule dilution experiments are shown. As opposed to the canonical view, the microscopic model indicates that interprotofilament cracks are always present, even when the microtubule is growing, and it also indicates that there is GTP-tubulin binding to the shortening microtubule. Quantitative analysis concludes that it is the relationship between the lengths of cracks and the GTP cap what dictates microtubule dynamics, not the GTP cap alone. With its simulation speed and level of detail, the microscopic scale model finally opens the door to testing hypotheses of the mechanisms used by microtubule regulatory proteins.
To the memory and legacy of Haydeé Claudia Geary.
CONTENTS

FIGURES ................................................. vii
TABLES ................................................... ix

CHAPTER 1: MICROTUBULE DYNAMICS ......................... 1
  1.1 Introduction .................................... 1
  1.2 Structure of the microtubule ................. 3
  1.3 The origin of dynamic instability .............. 5
    1.3.1 A model to explain dynamic instability .. 5
    1.3.2 Experimental evidence behind the model of dynamic instability ......................... 6
    1.3.3 Structural states of the tubulin heterodimer .... 8
  1.4 References .................................... 9

CHAPTER 2: MONTE CARLO FUNDAMENTALS FOR COMPLEX BIOLOGICAL MODELING AND SIMULATION ............ 14
  2.1 Abstract ...................................... 14
  2.2 Introduction .................................. 14
  2.3 Markov process ................................ 16
  2.4 Poisson process ................................ 17
    2.4.1 Arrival times ............................. 18
    2.4.2 Memorylessness ............................ 18
    2.4.3 The cumulative probability function .... 20
  2.5 Simulating complex biological phenomena ....... 21
    2.5.1 Picking the single event that will take place ........ 22

CHAPTER 3: INSIGHTS INTO CYTOSKELETAL BEHAVIOR FROM COMPUTATIONAL MODELING OF DYNAMIC MICROTUBULES IN A CELL-LIKE ENVIRONMENT ........... 25
  3.1 Abstract ..................................... 25
  3.2 Introduction .................................. 26
CHAPTER 3: RESULTS AND DISCUSSION

3.3 Results and Discussion

3.3.1 Recapitulation of dynamic instability by our Monte Carlo model

3.3.2 Persistent growth of MTs in vivo and in silico

3.3.3 Relationship between total tubulin and soluble tubulin: expected behavior according to the classic model

3.3.4 Deviation from this expected behavior

3.3.5 Effect of small changes in soluble tubulin concentration on MT behavior

3.3.6 Effect of MT nucleation on MT dynamics

3.3.6.1 Implications for the transition to mitosis

3.3.6.2 Implications for the interpretation of nucleation mutants

3.3.6.3 Implications for the study of MAPs

3.4 Concluding remarks

3.5 Materials and Methods

3.5.1 Essential features of the simulation

3.5.2 Parameter determination

3.5.3 Other aspects of the model

3.5.4 Computation and analysis of the simulations

3.5.5 Classical relationship between total and soluble tubulin concentration

3.5.6 Limitations of the analysis

3.6 References

CHAPTER 4: ANALYSIS OF A MESOSCOPIC STOCHASTIC MODEL OF MICROTUBULE DYNAMIC INSTABILITY

4.1 Abstract

4.2 Introduction

4.3 Model Description, Parameters, and Notations

4.3.1 Observables

4.4 Cap Model

4.5 Ensemble dynamics of microtubules

4.5.1 Critical concentration of free tubulin

4.6 Competition for tubulin

4.6.1 Unbounded domain

4.6.2 Bounded domain

4.7 Discussion of the results

4.7.1 Comparison with existing cap models

4.7.2 Competition for tubulin and the edge effect

4.7.3 Non-steady-state phenomena

4.8 Conclusions

4.9 Additional technical details
6.5.1 Database searching and alignment ........................................ 170
6.5.2 Tree building ............................................................. 171
6.6 References ................................................................. 172

APPENDIX A: PROGRAMS ...................................................... 180
  A.1 Simulation code for the modeling of microtubules as simple linear
      polymers ................................................................. 180
  A.2 Visualization ............................................................ 188
  A.3 Analysis of the length of microtubules ............................... 192
  A.4 Running long simulations ............................................. 196
## FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The microtubule cytoskeleton.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Microtubule structure.</td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>GTP and the slowly hydrolyzable analogue GMPCPP.</td>
<td>8</td>
</tr>
<tr>
<td>3.1</td>
<td>Recapitulating dynamic instability.</td>
<td>31</td>
</tr>
<tr>
<td>3.2</td>
<td>Relationship between the concentration of total tubulin, the number of MTs and the behavior of MTs in a spatially constrained environment.</td>
<td>34</td>
</tr>
<tr>
<td>3.3</td>
<td>Relationships between total tubulin and soluble tubulin at steady-state.</td>
<td>37</td>
</tr>
<tr>
<td>3.4</td>
<td>Relationship between soluble tubulin and total tubulin under various conditions.</td>
<td>39</td>
</tr>
<tr>
<td>3.5</td>
<td>Dependence of MT behavior on the concentration of soluble tubulin available at steady-state, part I.</td>
<td>40</td>
</tr>
<tr>
<td>3.6</td>
<td>Dependence of MT behavior on the concentration of soluble tubulin available at steady-state, part II.</td>
<td>41</td>
</tr>
<tr>
<td>3.7</td>
<td>Dependence of MT behavior on the concentration of soluble tubulin available at steady-state, part III.</td>
<td>42</td>
</tr>
<tr>
<td>3.8</td>
<td>Impact of changes in nucleating activity.</td>
<td>46</td>
</tr>
<tr>
<td>4.1</td>
<td>Schematic representation of the model.</td>
<td>65</td>
</tr>
<tr>
<td>4.2</td>
<td>Frequency of catastrophe as a function of MT growth velocity.</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>Steady state concentration of tubulin in unbounded and bounded domains.</td>
<td>85</td>
</tr>
<tr>
<td>4.4</td>
<td>Histograms of MT lengths after 80 and 160 seconds from the beginning of polymerization. The concentration of free Tu in both cases is close to the steady state value: 7.63 and 7.51 µM, respectively.</td>
<td>90</td>
</tr>
<tr>
<td>4.5</td>
<td>Histogram of MT lengths after 150 seconds from the beginning of polymerization. The concentration of free Tu is 1.382 µM and has reached its steady state.</td>
<td>91</td>
</tr>
<tr>
<td>4.6</td>
<td>Oscillations in the amount of polymerized tubulin.</td>
<td>96</td>
</tr>
<tr>
<td>5.1</td>
<td>Microtubule structure and its modeled processes.</td>
<td>104</td>
</tr>
</tbody>
</table>
TABLES

2.1 Stochastic time per component per event: $t_{m,q}$ .......................... 23
3.1 Dynamic instability parameters derived from experiments and simulations. ............................................. 32
4.1 Notation highlights. ................................................................. 67
4.2 Comparison of simulated and theoretical results. .................................................. 87
5.1 Sudden dilution experiments performed with our model ................... 125
6.1 Eukaryotic set of analyzed proteins with corresponding accession numbers. .................................. 149
6.2 Interpreting phylogenetic trees. ................................................. 150
6.3 Pairwise sequence comparison between selected sequences of all three classes. ................................ 151
6.4 Eubacterial set of analyzed proteins with corresponding accession numbers. .................................... 165
6.5 Archaebacterial set of analyzed proteins with corresponding accession numbers. .......................... 166
1.1 Introduction

Microtubules are proteinaceous polymers present in all eukaryotic cells. Although their appearance under the microscope varies with cell type and cell cycle stage, the microtubules are commonly found as an array of fibers emanating from the center of the cell towards the periphery (Figure 1.1). In higher eukaryotes the fibers are the individual microtubules while in yeast these fibers are usually bundles of microtubules [1, 2].

The function of the microtubule network is threefold. First, when the cell is in interphase, the microtubules provide the tracks for organelle localization and establishment and maintenance of cell polarity [3, 4]. Second, microtubules are responsible for vesicle trafficking with the help of kinesin and dynein motor proteins. This function is particularly important in neurons, where axons can be more that one meter long and diffusion is insufficient to transport proteins from their site of synthesis to the distal portion of the dendrites [5]. Third, in dividing cells, the microtubules make up the mitotic spindle that segregates the chromosomes [6].

The microtubule network is often referred to as the microtubule cytoskeleton. However, this term is particularly misleading because it gives the impression of
Figure 1.1: The microtubule cytoskeleton. The filaments extend from the microtubule organizing center near the nucleus towards the periphery of the cell. Often the arrangement of the fibers is roughly radial as shown here but it varies with cell type and cell cycle stage. The figure shows a COS 7 cell fixed with methanol and stained with an anti-tubulin antibody. Courtesy of Gergana Ugrinova.
a static network. On the contrary, the microtubules are highly dynamic and this aspect is key for their function. The most illustrative example is that of dividing cells exposed to taxol. This drug makes the microtubules unable to depolymerize and therefore they fail to pull the chromosomes apart during mitosis. For this reason, taxol has been used as a potent anticancer drug [7] although with clinical limitations.

The signature behavior of microtubules is their switching between phases of prolonged polymerization and depolymerization. Mitchison and Kirschner termed this process dynamic instability and they attributed this behavior to the existence of a GTP cap whose experimental evidence was first discovered by Carlier and Pantaloni [8, 9]. The interconversion between phases is stochastic and is observed both in vitro and in vivo [8, 10]. Dynamic instability is the process by which microtubules explore their own dynamic cellular space. This is needed to bring microtubules into contact with vesicles and chromosomes that require active transport and to rapidly adapt to changes in cell shape [8, 11, 13].

1.2 Structure of the microtubule

Microtubules are non-covalent polymers of tubulin. The building block of a microtubule is the alpha-beta tubulin heterodimer. Both monomers are very similar at the sequence level and both have the ability to bind GTP. The binding site in the alpha monomer is known as the N-site. The binding site in the beta monomer is called the E-site. These sites are respectively non-exchangeable and exchangeable [14]. Beta tubulin has an intrinsic GTPase activity. Once the GTP has been hydrolyzed to GDP, it can be exchanged for a new GTP molecule once the heterodimer has been detached and released to the cytosolic pool. When
the E-site contains GTP, the heterodimer is thought to be straight and therefore prone to polymerize into the microtubule. Once the dimer is polymerized, there is a delay after which hydrolysis occurs and this produces a bending that makes the dimer prone to detach from the microtubule tip [15]. The cycle of tubulin polymerization, GTP hydrolysis, depolymerization and GDP exchange constitutes the life cycle of tubulin and the microtubule [16].

Cryoelectron microscopy and 3D reconstruction shows that the microtubule body comprises a tubular arrangement of parallel protofilaments [17]. The diameter of the microtubule is about 25nm. Protofilament is the name given to the head-to-tail linear arrangement of the constituting heterodimers. Specifically, the heterodimers position in an \(-(alpha-beta)\_n\) chain configuration. The microtubule end exposing the beta monomer is called the plus end. In in vitro spontaneously assembled microtubules, the end exposing the alpha monomer (called minus end) displays some dynamics but less significantly than the plus end. In in vivo observations, the microtubule minus ends are nucleated by stabilizing gamma-Tu complexes that make this end dynamically inactive [18, 19].

By far, the most common in vivo microtubule has 13 protofilaments although in vitro self-assembled microtubules range from 11 to 15 protofilaments. One of the distinguishing features of the 13 protofilament structure is that is has a seam. Across this seam, the alpha monomer is in lateral contact with the beta monomer of its neighbor while the beta is in contact with alpha neighbor. This singularity does not exist in the rest of the microtubule where alpha monomers are only in contact with another alpha monomer and analogously, a beta monomer only contacts a lateral beta monomer. An important characteristic of the seam is that there is a shift of 1.5 heterodimers that results from the non-orthogonal arrangement
Figure 1.2: Microtubule structure. Tubulin heterodimers are the building blocks of microtubules, which are made of 13 protofilaments arranged into a tube with a seam. Each heterodimer consists of one alpha tubulin and one beta tubulin monomer. The beta subunit binds GTP and eventually hydrolyze it to GDP. Note that microtubule ends are not as blunt and neatly organized as depicted here.

of the subunits in the lattice (Figure 1.2). The common microtubule is therefore not a helical structure but a so called B lattice [20]. Structural experimental and computational studies indicate that longitudinal contacts (intra-protofilament) are stronger than lateral contacts (inter-protofilament) [20, 21].

1.3 The origin of dynamic instability

1.3.1 A model to explain dynamic instability

Dynamic instability originates in the delay of GTP hydrolysis after the heterodimer is polymerized [9]. When the microtubule is growing, this results in a GTP-rich region of heterodimers that follow the polymerizing tip. This region is called the GTP cap [8] and its existence and control of microtubule dynamics are
the commonly accepted model of dynamic instability. According to this model, if the hydrolysis of GTP catches up with the growing microtubule, the tip loses its stability and the microtubule starts to shorten. This change in polymerization phase from growth to shortening is called catastrophe. The depolymerizing microtubule eventually restores its GTP cap by polymerization of fresh subunits and this triggers a change in phase back to growth. This process is called rescue.

This model offers an attractive explanation but it is conceptual rather than mechanical. The following chapters of this thesis present novel theoretical (both computational and non-computational) models of dynamic instability and provide insights into the underlying principles governing dynamic instability.

1.3.2 Experimental evidence behind the model of dynamic instability

There are two lines of evidence indicating that growing microtubules are stabilized at or near the plus end. First, in vitro studies showed that microtubules depolymerize rapidly immediately after they have been severed [8, 22]. The second line of evidence comes from the so-called rapid dilution experiments. In these assays, a polymerizing microtubule is subjected to a sudden decrease in soluble tubulin concentration. Walker et al. [23] showed that after the sharp dilution, there is a delay of a few seconds before the depolymerization starts. This supported the idea of a transient end-stabilizing structure. Furthermore, they found that the depolymerization delay is independent of the pre-dilution microtubule growth speed. This was surprising because intuitively it would be expected that microtubules growing at higher speeds would have larger stabilizing caps and therefore they should display longer delay times. This apparent paradox was later explained theoretically [24, 25]. The sum of these plus similar experiments es-
tablished the microtubule end as the site of dynamics control although did not resolve the characteristics of its structure.

Early efforts to detect the lagging GTP cap using a glass fiberfiltration [26] method were controversial due to poor time resolution. Finally, a novel assay continuously monitoring the release of $P_i$ with an enzyme-linked reaction [27] succeeded detecting the hydrolysis lag [28] indicating that the cap is 170–180 units in size. Yet, this finding is not fully conclusive because the experiment was performed using taxol-stabilized polymerizing microtubules. Later work then concluded that the GTP cap must be small, perhaps the size of a single ring of dimers [29–31] but this may not be true for species other than pig and cow from which the tubulin was extracted to perform the experiments. Experimental evidence indicates that chick and yeast tubulins have slower GTP hydrolysis rate [32, 33].

Studies of tubulin polymerization in the presence of nucleotide analogues helped to establish the significance of GTP hydrolysis. Of particular use was GMPCPP. In this nucleotide analogue, the linkage between the beta and gamma phosphates is largely resistant to the intrinsic GTPase activity of tubulin under standard conditions (Figure 1.3). Thus, although GMPCPP is a slowly-hydrolyzable for tubulin, for practical purposes it can be considered a non-hydrolyzable analogue.

In the presence of GMPCPP-Tu, microtubules polymerize but they are unable to depolymerize [34]. GTP hydrolysis therefore provides the energy for depolymerization. In terms of its mechanism, it is not known yet whether it is the hydrolysis or the release of $P_i$ that triggers heterodimer bending and depolymerization although conformational change is commonly correlated with phosphate release [35, 36].
Overall, it is not yet clear whether the stabilizing cap is chemical or conformational \[37\]. It is clear that the alpha-beta tubulin heterodimer exists in two conformational states, one straight and one bent, but it is not known whether a chemical reaction is instantaneously correlated with the change in conformation. The model that best explains the existing contradictory experimental results is one in which the change in dimer conformation is slow and comes after phosphate release so that a presumably straight conformation lasts a little longer after phosphate departure.

1.3.3 Structural states of the tubulin heterodimer

There are two reasons indicating that the heterodimer exists in two structural states. The first one is that GDP bound tubulin has more tendency to form rings than GTP bound tubulin \[38\] and the second reason is that microtubules depolymerize by outward curling of protofilaments \[20, 39, 40\].

While there is no high resolution structural information, GTP bound tubulin (GTP in both the alpha and beta subunits) is considered a dimer in straight conformation \[41\]. Perhaps a more realistic view would be to consider GTP-Tu as
a dimer in *straighter* rather than a straight conformation because GMPCPP-Tu dimers are curved although much less significantly than GDP-Tu oligomers \[42\].

The GDP-Tu dimer (GTP and GDP bound to the $\alpha$ and $\beta$ subunits respectively) is unquestionably bent. High resolution studies of tubulin structure show an approximate bending angle of 11 degrees \[13,43\] as compared to the tubulin structure inside of a lattice with straight protofilaments \[17,46,48\].

In summary, we know that the events occurring at the microtubule end control the dynamics but we do not know what these events are. What we want to know is what is the architecture and distribution of heterodimeric states at the tip. More importantly, we want to be able to relate the microscopic structure of the dynamic tip to its macroscopic behavior. What rescues and catastrophes look like at high resolution? Can these transitions be predicted? How variations in the microscopic structure affect the frequencies of rescue and catastrophe? How cellular environments alter dynamic instability? This dissertation thesis addresses these questions in detail and opens the door for the theoretical study of microtubule regulatory proteins.

1.4 References


2.1 Abstract

Due to the complex, discrete, and random nature of many biological processes, stochastic numerical simulations have become one of the most common modeling tools. In general, this approach involves the combination of biological knowledge with information and tools from different disciplines including mathematics, physics, chemistry, and computer science. Most life scientists are unfamiliar with the key probabilistic tools, their assumptions, and their implementation. The apparent obscurity of stochastic numerical simulations undermines their credibility, utility, and adoption. This work sheds light on some probabilistic aspects of this type of modeling for readers of diverse background. Specifically, we focus on the Poisson process, one of the most useful probabilistic concepts. We introduce the Poisson process with examples, mathematically show its key property, and finally explain how to use it for biological simulations.

2.2 Introduction

Biology has been traditionally an experimental discipline. In vivo work has shown that living organisms are remarkably complex. In an attempt to gain mech-
anistic understanding, biologists reduce the number of uncertainties into simpler and tightly controlled settings: in vitro experiments. Today, good research articles support their claims with both in vivo and in vitro experiments. Unfortunately, sometimes experiments are very difficult or impossible to perform\(^1\). To circumvent experimental limitations and to gain quantitative predictive understanding about a biological process, a growing number of these problems are studied via theoretical modeling. Often the biochemical or biophysical rules governing simple isolated processes are known. A common task for the modeler is then to build a larger system comprising many otherwise isolated smaller system so that they can interact and emergent properties can be studied. Because this approach usually involves the use of computers, this type of theoretical modeling is commonly called computational simulation or in silico modeling. This modeling offers an excellent alternative that lies between the complexity of in vivo and the often oversimplified in vitro approach. Its good modeling practice involves the creation of a computer model just as complex as the process being studied. For instance, protein catalysis is modeled in finer detail than protein folding and this, in turn, is modeled at higher detail than protein complex formation. Unlike that, in vitro experiments are usually far less versatile and its settings often removed from the composition and behavior of living systems. Finally, in silico modeling can not only fill the gap between in vivo and in vitro results but also integrate them.

Due to the complex, discrete, and random nature of many biological processes, stochastic numerical simulations have become one of the most common in silico modeling tools. The word discrete is meant to emphasize the non-continuous treatment of some variables. For instance, temperature as reported by a mercury thermometer represents a continuous variable while number of atoms is a discrete variable.

\(^1\)Sometimes the limitations are ethical, financial or temporal rather than experimental.
variable. Some variables are continuous but may be treated as discrete. *Time* for example, can be treated as discrete and assumed to advance by small well-defined amounts such as years or microseconds.

In general, *in silico* modeling involves combination of biological knowledge with information and tools from different disciplines including mathematics, physics, chemistry and computer science. Most life scientists are completely unfamiliar with the key probabilistic tools, their assumptions and their implementation. The apparent obscurity of stochastic numerical simulations undermines its credibility, utility, and adoption. This concise work intends to shed light on some probabilistic aspects of this type of modeling. We intend to make accessible to a wide audience the mathematical basis of sound discrete stochastic biological simulation. Specifically, we focus on the Poisson process, one of the most useful probabilistic concepts. We introduce the Poisson process with examples, mathematically show its key property, and finally explain how to use it for biological simulations. To the best knowledge of these authors, this is the first documentation of its kind.

### 2.3 Markov process

Many natural phenomena can be modeled as Markov processes. A Markov process is a discrete time model in which the state of the system at time $t_{i+1}$ depends only on the state of the model at time $t_i$, where $i$ is the number of a time step in a series. All prior history is irrelevant in a Markov process. Using this tool usually involves modeling long times as a succession of many small discrete time steps. These series of steps are referred to as the Markov Chain. The coupling of Markov chain approaches to stochastic modeling is a powerful simulation tool often referred to as *Markov Chain Monte Carlo* or simply MCMC.
In MCMC as in all stochastic models, the transition from the state at \( t_i \) to the state at \( t_{i+1} \) is decided probabilistically. These probabilistic decisions must be sound, which is always a compromise between mathematical simplicity and biological accuracy (usually complex). To achieve this, a simple and frequently used approach is to treat the simulated phenomenon as a special case: a Poisson process. What follows is not intended to be a formal mathematical treatise on Poisson processes but instead to be a practical description highlighting its advantages as a tool. Consistent with that, the concept is presented by examples, and mathematical demonstrations are shown when needed.

### 2.4 Poisson process

A random variable governed by a Poisson process is characterized by a Poisson distribution. This is, the probability function is

\[
P(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \quad (2.1)
\]

where \( N_t \) is the number of events observed during the duration of a time \( t \) and the integer \( k \) is the number of events. The parameter \( \lambda \) is the rate at which these events are observed. As an example, let’s assume that jams in a particular printer behave like a Poisson process and let’s say that the rate of jams occurrence \( \lambda \) is 4 times per month. The probability of experiencing 3 jams in 2 months is

\[
P(N_t = 3) = \frac{e^{-4 \times 2}(4 \times 2)^3}{3!} = 0.029 \quad (2.2)
\]
and the probability of not having jams for the next month is

\[ P(N_t = 0) = \frac{e^{-4 \times 1}(4 \times 1)^0}{0!} = 0.018. \tag{2.3} \]

2.4.1 Arrival times

It is easy to see from equation 2.3 (where \( k = 0 \)) that in general the waiting time until the next event – often called the arrival time – is

\[ P(N_t = 0) = e^{-\lambda t}. \tag{2.4} \]

This means that the waiting times in between events follow an exponential distribution.

2.4.2 Memorylessness

A key property of a Poisson process is that it is memoryless. To see that, imagine that \( T_1, T_2, T_3, \) etc. are times at which the events of a Poisson process occur. For instance \( T_3 \) is the time at which the third event occurs. Let’s call \( t \) the time since the beginning of the process. Clearly,

\[ N_t < k \iff T_k > t. \tag{2.5} \]

This equation is more obvious if analyzed with an example. For instance, let’s take \( k = 5. \) If at time \( t \) the number of observed events \( N_t \) is less than 5, then clearly \( T_5 \) must be greater than \( t, \)

\[ N_t < 5 \iff T_5 > t. \tag{2.6} \]
Because both sides of equation 2.6 imply each other, their probabilities must be equal:

\[ P(N_t < 5) = P(T_5 > t). \] (2.7)

If this time we take \( k = 1 \) and repeat the example above, we get

\[ P(N_t < 1) = P(T_1 > t) \] (2.8)

\[ P(N_t = 0) = P(T_1 > t) \]

and since \( N_0 = 0 \), then

\[ P(N_t - N_0 = 0) = P(T_1 > t) \] (2.9)

where \( T_1 \) is the time of the first event after \( t \). Applying the definition of the Poisson distribution (equation 2.1) into equation 2.9 we obtain

\[ \frac{e^{-\lambda t} (\lambda t)^0}{0!} = P(T_1 > t) \] (2.10)

\[ e^{-\lambda t} = P(T_1 > t) \]

and analogously

\[ P(N_{T_1} - N(t=0)) = 0) = \frac{e^{-\lambda T_1} (\lambda T_1)^0}{0!} = e^{-\lambda T_1}. \] (2.11)

Although it is not obvious yet, what we are really interested in is \( P(N_{T_1} - N_t) \), which is the probability of not having arrivals in the interval between time \( t \) and \( T_1 \). To calculate that, note that the probability \( P(N_{T_1} - N_0 = 0) \) must be equal
to the product of the probabilities of not having an arrival between time 0 and \( t \) and between \( t \) and \( T_1 \). Thus

\[
P(N_{T_1} - N_t = 0) \ P(N_t - N_0 = 0) = P(N_{T_1} - N_0 = 0) \tag{2.12}
\]

and upon rearrangement

\[
P(N_{T_1} - N_t = 0) = \frac{P(N_{T_1} - N_0 = 0)}{P(N_t - N_0 = 0)} \tag{2.13}
\]

which, by introduction of equations 2.4 and 2.11, leads to

\[
P(N_{T_1} - N_t = 0) = e^{-\lambda(T_1 - t)}. \tag{2.14}
\]

In conclusion, the distribution of these waiting times is also exponential. Importantly, equation 2.14 indicates that the waiting times for the next arrival \((T_1 - t)\) does not depend on when the past event occurred. The process is memoryless. For example, this implies that although the last event may have happened long ago, this does not make the next event more likely to occur sooner. Although this result still seems of little importance, its paramount value will be fully appreciated when we get to the simulation of complex biological phenomena.

2.4.3 The cumulative probability function

Equation 2.4 represents the probability of \textit{not} having an event as a function of time \( t \). Therefore, the probability of having one event is

\[
F(t) = 1 - e^{-\lambda t} \tag{2.15}
\]
which is the exponential cumulative probability function\(^2\).

2.5 Simulating complex biological phenomena

The inexorable outcome of the genomic era was the accumulation of large amounts of information from which predictable knowledge is waiting to be extracted. Now that we are much closer to knowing all the individual actors in the biological play, the challenge is to understand how these non-static individual components influence each other. We are now in the era of the system approaches. The scale and complexity of these approaches often demands the use of numerical simulations at least as a first step towards understanding. For instance, we may know all the components and reactions in a metabolic pathway and yet have scarce knowledge of its chemical kinetics, especially at the system level. Numerical calculations require high computer power but fortunately its cost have decreased in recent years with improvement on both processor design and parallelization based on consumer desktop computers. The strategy of modeling and simulation of a whole system involves modeling its individual dynamical components. The key question is how to soundly model such a system when all its components are changing with time. The purpose of this section is to explain how this type of complex modeling is done in praxis based on the solid mathematical concepts explained above.

Lets assume that we have a system composed of \(M\) components. Lets assume now that every individual component is susceptible to change upon occurrence of an event that affects the component. We are trying to simulate the system by letting it run over time and observing what happens. The time does not run

\(^2\)\(F(t)\) is not to be confused with its well known derivative, the exponential probability density function \(f(t) = \lambda e^{-\lambda t}\).
continuously, it runs based on very short single time steps. Although any particle may experience an event, we do not arbitrarily grant an event to every component in the system. Instead, at every time step, one event occurs in the whole system. How to decide what will be the next single event that will take place in the whole simulated system?

2.5.1 Picking the single event that will take place

The way this is done involves determining randomly the time that will take every possible event for every possible component. For every component $m$ and event type $q$, a waiting time $t_{m,q}$ is assigned by drawing stochastically from its appropriate probability distribution. Table 2.1 helps to visualize the randomly chosen times in a general case. Notice that the table shows the scenario in general assuming that not all $m$ components will experience exactly the same types of events $q$. Therefore, $q$ becomes a function of $m$. For example, let us imagine simple chemistry of amino acids in aqueous solution. The different amino acids are the components of the system. Some amino acids may lose H$^+$, some may get phosphorylated. Some amino acids may experience both, some may experience none. Furthermore, the same type of residue may not experience the same chemical reaction depending on the context as in a folded protein with a buried Serine residue unreachable by kinases.

From a purely computational standpoint, drawing times from the probability distribution function is very easy for a Poisson process. Let $r$ be a randomly generated number from a uniform distribution such that $0 \leq r < 1$, we can apply equation 2.15 and obtain

$$r = 1 - e^{-\lambda_{m,q} t_{m,q}}$$

(2.16)
and then

\[ e^{-\lambda_{m,q} t_{m,q}} = 1 - r \]  

(2.17)

\[ t_{m,q} = -\frac{\ln(1 - r)}{\lambda_{m,q}} \]  

(2.18)

where \( \lambda_{m,q} \) represents the Poisson parameter for the event of type \( q \) on component \( m \).

In simulations with multiple components and event types, the event that takes place is the one with the shortest waiting time \( t_{m,q} \), the one that will occur first. Then, after the event has been implemented and the simulation time advanced by \( t_{m,q} \) comes the question: How to calculate the waiting times next? All but one component have been waiting already for the next event. Should their waiting times be shorter then? It is here where the beauty of the Poisson process properties
comes into play. As shown in subsection 2.4.2, Poisson processes are memoryless so it does not matter when was the last time that a component experienced an event. At any point in time, the probability of the next event follows equation 2.15 and all previous history is irrelevant. So, every time an event is implemented and the time advanced $t_{m,q}$, all waiting times are recalculated and the whole procedure repeated.

As a biological example, we can consider simulating microtubule dynamics. A microtubule is a tubular polymer made up of $\alpha\beta$-tubulin monomers. The system would be the microtubule itself and every one of its 13 protofilaments would be a component. Every component could experience two types of events: growth or shortening, in both cases, by one monomer at a time. We would assume that these events can be modeled as a Poisson process. The simulation of the microtubule dynamics would involve randomly determining the time for every one of the 26 events and implementing the event that would have the shortest waiting time. Because a Poisson process is memoryless, next step of the simulation we would recalculate the waiting times for all 26 possible events and again implement the one with the shortest arrival time.
CHAPTER 3

INSIGHTS INTO CYTOSKELETAL BEHAVIOR FROM COMPUTATIONAL MODELING OF DYNAMIC MICROTUBULES IN A CELL-LIKE ENVIRONMENT

3.1 Abstract

Microtubule\(^1\) dynamic instability plays a fundamental role in cell biology, enabling microtubules to find and interact with randomly distributed cargo and spatially localized signals. In vitro, microtubules transition between growth and shrinkage symmetrically, consistent with the theoretical understanding of the mechanism of dynamic instability. In vivo, however, microtubules commonly exhibit asymmetric dynamic instability, growing persistently in the cell interior and experiencing catastrophe near the cell edge. What is the origin of this behavior difference? One answer is that the cell edge causes the asymmetry by inducing catastrophe in persistently growing microtubules. However, the origin of the persistent growth itself is unclear. Using a simplified coarse-grained stochastic simulation of a system of dynamic microtubules, we provide evidence that persistent growth is a predictable property of a system of nucleated, dynamic, microtubules containing sufficient tubulin in a confined space - MAP activity is not required. Persistent growth occurs because cell-edge-induced catastrophe increases the concentration

\(^{1}\)This thesis chapter has been published as a research article. IV Gregoretti, G Margolin, MS Alber and HV Goodson; 2006 Journal of Cell Science 119(22):4781–4788.
of free tubulin at steady-state. Our simulations indicate that other aspects of MT dynamics thought to require temporal or spatial changes in MAP activity are also predictable, perhaps unavoidable, outcomes of the ‘systems nature’ of the cellular microtubule cytoskeleton. These include the mitotic increase in microtubule dynamics and the observation that defects in nucleation cause changes in the behavior of microtubule plus ends. These predictions are directly relevant to understanding of the microtubule cytoskeleton, but they are also attractive from an evolutionary standpoint because they provide evidence that apparently complex cellular behaviors can originate from simple interactions without a requirement for intricate regulatory machinery.

3.2 Introduction

Microtubules (MTs) are components of the cytoskeleton, the network of proteinaceous fibers that endows the cell with structural integrity, motile properties, and internal organization. MTs play a particularly important role in cell organization: they pull the chromosomes apart at mitosis, act as a ‘railroad system’ for intracellular transport, and define the localization and structure of internal membrane systems [1–3]. Two characteristics of MTs are particularly significant for these functions. First, MT nucleation is regulated, and the limitation of nucleation to an organelle near the nucleus (the centrosome) endows most cell types with a radial organization. Second, and perhaps more importantly, the MT cytoskeleton is dynamic: individual MTs in the same cell (or same test tube) constantly change in length, either growing or shrinking with random transitions between these phases [4].

This counterintuitive behavior is termed dynamic instability and is fundamen-
tal to MT function. First, dynamic instability is a mechanism for exploring cellular space, bringing MT railroads into contact with poorly diffusible cargo, such as chromosomes for subsequent transport [5–8]. Second, this turnover ensures rapid response of the cytoskeleton to internal and external signals. Selective stabilization of dynamic MTs probably plays a key role in morphogenesis and appears to play a central role in the self-organizing properties of the mitotic spindle [8–10].

Understanding the role of MTs in cell organization requires a detailed understanding of dynamic instability and its regulation. MTs are non-covalent polymers of the protein tubulin. Dynamic instability originates in conformational changes that occur in tubulin subunits after polymerization. Briefly, tubulin subunits (which are obligate dimers of the polypeptides α- and β-tubulin) bind GTP. Upon polymerization, this GTP is hydrolyzed to GDP, but only after a short delay. This delay is thought to result in a ‘GTP cap’, which predisposes the MT to continued growth. The idea is that if this cap is lost (via hydrolysis or other mechanism), the exposed GDP-tubulin subunits rapidly depolymerize in an event termed catastrophe (reviewed in [4]). This elegant explanation has recently been modified to include the possibility that the conformational state of the tubulin subunits in the cap may be more important than the state of the bound nucleotide [11–13].

Dynamic instability occurs in solutions of pure tubulin and GTP, but it can be modulated by a variety of MT-binding proteins (also called microtubule-associated proteins or MAPs), including MAP4, stathmin, and CLIP-170 [4, 14, 15].

Although this molecular-level explanation has been extremely useful, many aspects of MT behavior remain poorly understood. In particular, what determines dynamic-instability parameters? One expects them to be a function of fundamental chemical parameters (association constants, dissociation constants...
and hydrolysis rates) and environmental parameters (the concentration of tubulin, the number of nucleation sites and the presence of spatial constraints), but no complete mathematical description of the relationship between these quantities has been derived. A major reason for this incomplete understanding is that dynamic instability is an emergent phenomenon—a behavior that arises from the independent interaction of many individual components, resulting in system-level properties that are not obviously predictable from the characteristics of the components.

One approach that has proven useful for studying such complex systems is computational modeling, in which the behavior of a system is simulated by allowing components of the system to interact according to defined rules. Insight into the system is obtained by seeing how the system changes when the rules are altered, and comparing these observations to expectation or experiment. Significant efforts have been made to predict and explain MT behavior using both deterministic models (systems of interacting equations) and stochastic simulations (Markov chain/Monte Carlo approaches) (e.g. [16–24]).

These studies have provided insight into numerous aspects of MT behavior, including the origin of dynamic instability, the nature of the stabilizing cap, and the effect of physical boundaries on MT length distributions. However, many of these efforts examine the behavior of single MTs. Those with multiple MTs occur in semi-infinite space and/or have variables, such as tubulin concentration or transition frequencies, defined to be constants. None of these studies examines how the behavior of MTs is influenced by the constraints of a cell-like environment, in which multiple MTs (but not an infinite number) compete for a limited pool of tubulin subunits and microtubule growth is spatially confined.
To begin to address these issues, we have performed a series of Monte-Carlo simulations of a system of dynamic MT in such a cell-like environment. In these simulations, MTs (nucleated by a defined number of seeds) compete with each other for free GTP tubulin subunits in a ‘cell’ of defined size and shape. MTs can be followed visually or statistically at the level of an individual or the population. Dynamic instability parameters (transition frequencies, growth rates, and shrinkage rates) are not set by the user but instead evolve from the interactions of the different parts of the system as the simulation proceeds. A recent article by Janulevicius et al. [25] describes a model that has similarities to the one we used here, but which was used to address different questions.

Using our model, we find that several cellular phenomena that were thought to require complex regulatory machinery are instead predictable outcomes of interactions between a system of dynamic MTs and its physical environment. More specifically, we find that the surprisingly persistent growth of MTs observed in vivo is a predictable property of a nucleated system of dynamic MTs containing sufficient tubulin and polymerizing in a constrained space. This perturbation of MT dynamics is an outcome of the increase in the steady-state concentration of free tubulin, resulting from interactions of MTs with the cell boundary. Similarly, changes in nucleation activity are expected to have major effects on MT length and transition frequencies, dictating changes such as those seen at the interphase-mitosis transition [26, 27]. These observations do not exclude MAP involvement (obviously, MAPs play central roles in these processes), but they imply that MAPs modulate these behaviors instead of creating them. These studies indicate that the classic concept of critical concentration requires revision when applied to cellular systems, and they provide a foundation for quantitative understanding of
MT dynamics *in vitro* and *in vivo*.

3.3 Results and Discussion

3.3.1 Recapitulation of dynamic instability by our Monte Carlo model

Using the rules outlined in the Materials and Methods, we have built a simplified coarse-grained stochastic simulation of MT dynamics that incorporates the major elements of the MT polymerization process *in vivo*, including spatially constrained nucleation, competition between MTs for tubulin subunits, and the imposition of physical limitations to MT polymerization by the cell edge. Adjustable parameters are the total tubulin concentration, cell size, rate of GTP hydrolysis, and association and dissociation rate constants for GTP and GDP tubulin. MT dynamic instability parameters (rate of growth, rate of depolymerization, catastrophe frequency, and rescue frequency) are not set by the modeler but instead change with time and conditions, emerging from the dynamic interactions of the system. This model recapitulates the obvious qualitative features of dynamic instability (Figure 3.1, Movie 1 in supplementary material of Gregoretti et al. [28]). When appropriate adjustable parameters are chosen (see Materials and Methods), it can approximate the quantitative features, including growth rate and transition frequencies (Table 3.1). Unless otherwise indicated, all data presented in this chapter are obtained from simulations run under a set of reference parameters chosen to approximate the behavior of tubulin *in vivo* in interphase [27], and are from steady-state (see Materials and Methods for details).

This model is similar to the stochastic simulations used by Hill and Chen [23], but there are two key differences. First, in our model, a system of dynamic
Figure 3.1: Recapitulating dynamic instability. (A) Snapshots of the simulation at different time steps (T). The MTs grow from seeds at left towards the ‘cell’ edge (right). The colors describe the state (red, GTP; green, GDP) of each subunit. At early times, when free tubulin is near the initial value, MTs grow persistently. As the polymer fraction increases and the concentration of free tubulin drops, catastrophe becomes more frequent. Eventually the steady-state is reached, and the system behavior exhibits behavior very similar to experimentally observed dynamic instability (see Movie 1, supplementary material of Gregoretti et al. [28]). (B,C) Comparison between life history plots obtained experimentally in vitro (B) and with our model (C). Experimental data were adapted from Fygenson et al. [29]. (C) shows three adjacent steady-state MTs from the simulation shown in (A). In this simulation, parameters were chosen arbitrarily; all other simulations reported in this manuscript are correlated to physiological concentrations and dimensions as described in Materials and Methods.
<table>
<thead>
<tr>
<th></th>
<th>in vivo†</th>
<th></th>
<th></th>
<th>in silico</th>
<th>Low [Tu]_{total}</th>
<th>High [Tu]_{total}</th>
<th>High [Tu]_{total}</th>
<th>High [Tu]_{total}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interphase</td>
<td>Mitosis</td>
<td></td>
<td></td>
<td>1 x no. of MTs</td>
<td>1 x no. of MTs</td>
<td>4 x no. of MTs</td>
<td>0.25 x no. of MTs</td>
</tr>
<tr>
<td>Catastrophe frequency</td>
<td>0.026±0.024</td>
<td>0.058±0.045</td>
<td>0.051±0.003</td>
<td>0.033±0.003</td>
<td>0.043±0.003</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>(seconds⁻¹)</td>
<td></td>
<td></td>
<td>(interphase-like)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rescue frequency</td>
<td>0.175±0.104</td>
<td>0.045±0.111</td>
<td>0.070±0.007</td>
<td>0.096±0.006</td>
<td>0.086±0.007</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>(seconds⁻¹)</td>
<td></td>
<td></td>
<td>(mitosis-like)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth rate</td>
<td>0.191±0.123</td>
<td>0.212±0.094</td>
<td>0.095±0.003</td>
<td>0.167±0.004</td>
<td>0.122±0.003</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>(µm/seconds)</td>
<td></td>
<td></td>
<td>(interphase-like)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortening rate</td>
<td>-0.218±0.140</td>
<td>-0.236±0.131</td>
<td>-0.18±0.01</td>
<td>-0.34±0.02</td>
<td>-0.28±0.02</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>(µm/seconds)</td>
<td></td>
<td></td>
<td>(mitosis-like)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean MT length</td>
<td>~85%‡</td>
<td>shorter than</td>
<td>0.09±0.01</td>
<td>0.76±0.01</td>
<td>0.23±0.01</td>
<td>0.95±0.03</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>(cell radius fraction)</td>
<td></td>
<td>in interphase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial [Tu] (µM)</td>
<td>–</td>
<td>–</td>
<td>5.0±0.0</td>
<td>14.0±0.0</td>
<td>14.0±0.0</td>
<td>14.0±0.0</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Steady-state [Tu] (µM)</td>
<td>–</td>
<td>–</td>
<td>4.10±0.06</td>
<td>6.4±0.1</td>
<td>4.7±0.2</td>
<td>11.64±0.07</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

The values of the simulations are the mean and standard deviation of 50 repetitions. †In vivo data are from Rusan et al. [27], except ‡, which is from Komarova et al. [30]. No numbers are given for the first four cells of the last column because microtubules in these simulations were too close to the boundary to allow accurate measurements. Note that though the precise values of the experiments and simulations do not match (and are not meant to), the trends are similar.
MTs is simulated instead of a single MT, and the concentration of soluble tubulin ([Tu]_{soluble}) in this system changes as MTs grow and compete for subunits (in most previous simulations, including those of Hill, [Tu]_{soluble} is a set parameter). Second, we have imposed MT-length limits (analogous to the limitations imposed by the cell boundary). The sum of these characteristics suggests that this model is useful for investigating the constraints on MT dynamics imposed by cell-like systems.

3.3.2 Persistent growth of MTs *in vivo* and *in silico*

Perturbation of the simulation by changing the concentration of total tubulin ([Tu]_{total}), leaving all other parameters unaltered, reveals an important relationship between tubulin concentration and MT behavior. At low [Tu]_{total}, MTs in the simulation mimic MTs observed *in vitro* at steady-state: catastrophe is frequent, rescue is relatively rare and MT lengths decay exponentially (short MTs significantly outnumber long MTs) (Figure 3.2A,E, Table 3.1, Movie 2 in supplementary material of Gregoretti *et al.* [28]). However, at relatively high [Tu]_{total}, dramatically different behavior is observed: MTs begin to grow persistently, meaning that many MTs reach the cell edge without ever undergoing catastrophe (Figure 3.2B and Table 3.1, Movie 3 in supplementary material of Gregoretti *et al.* [28]). Most catastrophes occur near the cell edge and rescue becomes more frequent, resulting in an accumulation of MT ends near the cell edge (Figure 3.2F). This behavior mimics that observed *in vivo* [30].

What is the origin of this change from an in-vitro-like behavior to an in-vivo-like behavior? The observation that catastrophe frequency displays position dependence *in vivo* has been addressed previously: Komarova *et al.* and Maly
Figure 3.2: Relationship between the concentration of total tubulin ([Tu]_{total}), the number of MTs and the behavior of MTs in a spatially constrained environment. (A-D) Life history plots of representative MTs in simulations run under the indicated conditions. (E,F) Distribution of MT lengths taken from a series of simulations conducted under the indicated conditions.
have suggested that the cell edge induces catastrophe in persistently growing MTs, causing the catastrophe asymmetry and resulting in the MT length distribution observed in vivo \[30, 31\]. Induction of catastrophe by physical boundaries has been observed experimentally \[32\]. We have incorporated these ideas into our model by stipulating that the cell edge prevents the addition of new tubulin subunits, resulting in loss of the GTP cap, which then leads to catastrophe.

However, although the ‘edge-induced depolymerization’ explanation is attractive, it is not complete. It explains the catastrophe asymmetry, but fails to address the key issue of the cause of the persistent growth. Komarova and colleagues proposed that MAP activities are involved, and provided evidence that the MT plus end tracking protein CLIP-170 plays a role in persistent growth \[33\]. However, the only difference between the simulations in Figures \[3.2A,B\] (one exhibiting persistent growth and one not) is the amount of total tubulin in the system.

Consideration of the behavior of our computational model leads to an alternative hypothesis for the origin of the persistent growth: (1) the cell boundary induces catastrophe prematurely \[30, 31\]; (2) this catastrophe induction causes \([Tu]_{\text{soluble}}\) at steady-state to be higher than it would be in the absence of the barrier; (3) this higher availability of \([Tu]_{\text{soluble}}\) is what allows the MTs to grow persistently, experiencing few catastrophes and undergoing more frequent rescues. This hypothesis predicts that a physical boundary can cause the \([Tu]_{\text{soluble}}\) steady-state to rise above its natural steady-state level, and this increase in \([Tu]_{\text{soluble}}\) can be sufficient to cause persistent MT growth.
3.3.3 Relationship between total tubulin and soluble tubulin: expected behavior according to the classic model

As a first test of this hypothesis, we investigated the relationship between [Tu]_{total} and [Tu]_{soluble} at steady-state. Before examining our system, we wanted to first understand what was expected from the existing literature. The classic understanding of monomer/polymer partitioning as defined by Oosawa and colleagues, and refined by Johnson and Borisy is shown in Figure 3.3A [34–36], see also Howard for a more recent discussion [37]. Examination of this figure shows that tubulin put into a system is expected to remain unpolymerized until the concentration of unpolymerized tubulin reaches a critical concentration, after which point all additional tubulin is incorporated into polymer. Under this model, [Tu]_{soluble} remains at the critical concentration no matter how much additional tubulin is added (assuming that all of it is active).

3.3.4 Deviation from this expected behavior

This Oosawa model has been the dominant framework for understanding bulk MT polymerization. Our model reproduces this expected behavior when MT length is unconstrained (compare red solid curves in Figure 3.3A,B). However, when MT length is physically constrained as it would be in a cell the concentration of soluble tubulin deviates significantly from the behavior according to the classic model (Figure 3.3B, dashed curve): there is no clearly observed critical concentration and, instead of plateauing after the initial rise, [Tu]_{soluble} continues to rise with [Tu]_{total}; at first slowly and then more quickly. Because the two curves of Figure 3.3B are produced under conditions that are identical except for the presence or absence of catastrophe induction by the cell edge, these observations
Figure 3.3: Relationships between total tubulin and soluble tubulin ([Tu]_soluble) at steady-state. (A) Classically expected behavior. Note that little polymer is seen until the critical concentration (C_c) is achieved. At total tubulin concentrations above C_c, all additional tubulin is incorporated into polymer, and the concentration of unpolymerized tubulin remains at C_c. See Materials and Methods for the equations used to plot these curves. (B) Relationship observed in our simulations. Solid red line: system without spatial confinement. Dashed red line: system with spatial confinement. The dotted grey line gives C_c (i.e. [Tu]_soluble) that is asymptotically approached as [Tu]_total increases. Notice that, in confined systems there is no easily observed C_c. Instead, the concentration of free tubulin continues to rise as total tubulin rises, at first slowly, and then more steeply. The curves in B are not fits to an equation but are provided to guide the eye as it follows the progression of the data.
suggest that the ‘edge effect’ predicted above can cause [Tu]_soluble to rise above that expected from standard considerations of critical concentration. Moreover, the effect increases as more tubulin is added to the system.

Note that the sharper transition to the plateau of the curve in panel A (compared with the curve in panel B) results from the fact that the model used to yield curve A incorporates a very unfavorable nucleation step, which normally exists when tubulin polymerizes in vitro. Our model, which is meant to simulate microtubule polymerization from stable nucleation sites in vivo, produces a similarly sharp transition if the difficulty in initial growth from the stable seeds is increased (Figure 3.4).

3.3.5 Effect of small changes in soluble tubulin concentration on MT behavior

We next examined the effect of such changes in [Tu]_soluble on MT behavior. The differences seen between the solid and dashed lines of the curves in Figure 3.3B are relatively small: when [Tu]_total rises from 5 µM to 14 µM, [Tu]_soluble rises only from ~4 to ~6 µM, respectively. Although this increase may seem insignificant, Figures 3.5 through 3.7 show that it is predicted to have a major effect on the behavior of the simulated MTs: Changing [Tu]_soluble from 4.1 µM to 6.4 µM shifts the behavior of the simulated MTs from classic dynamic instability to, apparently, persistent growth (Figure 3.5). Note that, to test the dependence of MT behavior on [Tu]_soluble as in Figure 3.5, it is necessary to hold [Tu]_soluble constant (i.e. we

2In 1987, Mitchison and Kirschner performed a theoretical analysis suggesting that [Tu]_soluble increases with [Tu]_total in mass- and number-limited systems of dynamic microtubules, such as those in a cell, even in the absence of spatial constraint [38]. Some of their conclusions are similar to those discussed here. However, their conclusions depend on the assumption that rescue does not occur. Because rescue is common in cellular systems, and because our simulations behave according to their predictions only when both rescue and spatial constraint are ‘turned off’ (Figure 3.4), we believe that the conclusions of their analysis were prescient but the equations leading to these conclusions are of limited use.
Figure 3.4: Relationship between soluble tubulin and total tubulin under various conditions. (A) Effect of changing the stringency of nucleation. Note that when nucleation is stringent (red solid line), the curve looks very much like the classic curve shown in Figure 3.3A. The dotted lines give the behavior of simulations performed when rescue is turned off (occurs by setting to zero the probability of GTP tubulin adding to a GDP end). (B) Effect of rescue on the relationship between soluble tubulin and total tubulin in the presence (edge on) and absence (edge off) of spatial confinement. Note that the shape of the blue line (no rescue, no spatial confinement) is very similar to that predicted by Mitchison and Kirschner [38].
Figure 3.5: Dependence of MT behavior on the concentration of soluble tubulin available at steady-state, part I. Life-history plots of individual MTs at different values of \([\text{Tu}]_{\text{sol}}\). * and **, average length of MTs at steady-state in 4.1 \(\mu\)M and 4.7 \(\mu\)M free tubulin, respectively. At 6.4 \(\mu\)M free tubulin, there is no steady-state length - the mass of polymer increases with time (MT growth is unbound; see also Figure 3.4). Notice that, in all three Figures 3.5, 3.6 and 3.7, the MTs do not compete with each other for free tubulin (because free tubulin is held constant at the indicated value), and the cell size is made so large so that no MTs interact with the edge during the course of the simulation. This is similar to an experimental situation \textit{in vitro} in which the pool of free tubulin is not depleted during the time course of the experiment.

altered the simulation so that in these trials the MTs do not compete for tubulin. In addition, there is no spatial constraint).

To better understand this transition to persistent growth, we examined the relationship between \([\text{Tu}]_{\text{sol}}\) and average MT length in more detail. When \([\text{Tu}]_{\text{sol}}\) is held constant at low to moderate levels, the length of MTs at steady-state increases as soluble tubulin increases, but it is finite across this concentration range (Figure 3.6). However, as \([\text{Tu}]_{\text{sol}}\) increases more, there is a narrow range in which the average steady-state length rises steeply, appearing to approach in-
Figure 3.6: Dependence of MT behavior on the concentration of soluble tubulin available at steady-state, part II. Relationship between [Tu]_{soluble} and the mean MT length at steady-state. The Error bars give the standard deviation of values observed from 50 different simulations at the indicated [Tu]_{soluble}. 


Figure 3.7: Dependence of MT behavior on the concentration of soluble tubulin available at steady-state, part III. Relationship between persistent growth and $[Tu]_{\text{soluble}}$. The data plotted on the y-axis give average rates of increase in polymer length as a function of $[Tu]_{\text{soluble}}$. The concentration of free tubulin required for the transition to persistent growth is indicated by the x-axis intercept of the dashed line. The Error bars give the standard deviation of values observed from 50 different simulations at the indicated $[Tu]_{\text{soluble}}$. 
finity at some threshold (between 5 and 6 µM soluble tubulin under these conditions) (Figure 3.6). This threshold is the transition to persistent growth - as can be seen in a plot of net MT growth rate as a function of [Tu]_{soluble} (Figure 3.7). When this threshold is passed, MTs still experience catastrophes but the balance between catastrophe and rescue is such that, the amount of polymer in such a system increases constantly with time (Figures 3.5 and 3.7). On the basis of this analysis, small increases in the steady-state concentration of soluble tubulin (like those caused by interaction of MTs with the cell edge) should induce persistent growth if the initial [Tu]_{soluble} is close enough to the persistent growth threshold. Note that the transition to persistent growth occurs at a concentration empirically similar to the critical concentration observed in the absence of spatial constraint (compare the dotted line in Figure 3.3B with the x-axis intercept of the dashed line in Figure 3.7), but we have not yet developed a formal demonstration of this relationship.

These observations are consistent with previous analyses indicating that average MT length should increase with [Tu]_{soluble}, and that MTs should show a transition from ‘bounded’ to ‘unbounded’ growth when [Tu]_{soluble} passes a threshold [18, 39]. In a physiological system, the total mass of tubulin is constant, and competition between MTs for this tubulin would be expected to keep [Tu]_{soluble} below the transition to persistent growth. However, if a physical boundary limits polymerization and the edge effect (Figure 3.3B) causes [Tu]_{soluble} to rise above the threshold, persistent growth should occur.

One conclusion of this work is that the classic concept of critical concentration must be used carefully in cellular systems - [Tu]_{soluble} at steady-state is not simply equal to the critical concentration. Instead, the steady-state [Tu]_{soluble} is a dynamic
parameter that depends on a host of variables including the number of nucleation sites, the total amount of tubulin and the cell size. Transition frequencies and MT length distribution are coupled to these same variables through \([\text{Tu}]_{\text{soluble}}\). As originally suggested by Mitchison and Kirschner [38], the increase of \([\text{Tu}]_{\text{soluble}}\) with \([\text{Tu}]_{\text{total}}\) is probably a major part of the mechanism, allowing total tubulin levels to be tightly regulated \emph{in vivo}.

On the basis of these simulations and the reasoning above, we propose that the persistent growth of MTs observed \emph{in vivo} is an expected outcome of placing a sufficient amount of tubulin in a spatially confined system under conditions where nucleation is limited. This reasoning does not exclude the involvement of MT-binding proteins in persistent growth \emph{in vivo}, but argues that they tune the behavior instead of generating it. Although physiological systems differ from our simulations (and from each other) in their quantitative details, this analysis suggests that persistent growth in the presence of an appropriate spatial constraint is an intrinsic property of any system of MTs or, indeed, of any nucleated two-state polymer system.

3.3.6 Effect of MT nucleation on MT dynamics

A key aspect of MT growth, both \emph{in vivo} and in our model, is that it normally occurs from stable nuclei or \emph{seeds}. We were interested to see whether our simulations could provide insight into two experimental observations related to nucleation: (1) the changes in MT length and dynamics seen together with increases in nucleation during mitosis; (2) the ability of mutations in \emph{S. pombe} nucleation factors to alter MT length and dynamics. In both cases, the effects on MT length and dynamics have been proposed to result from the direct action of MAPs.
3.3.6.1 Implications for the transition to mitosis

When cells transition into mitosis, the MT cytoskeleton changes from a relatively stable interphase array that has long, persistently growing MTs into an early mitotic array that has shorter, more numerous MTs that are also more dynamic [26, 27]. Using our computational model, we investigated the relationship between these characteristics. We found that, simply increasing the number of MT seeds in an interphase-like simulation causes the system to adopt mitosis-like dynamic instability characteristics: MTs become significantly shorter and more dynamic, undergoing more spontaneous (not-edge-induced) catastrophes and fewer rescues (Figure 3.2C, compare with 3.2B, Table 3.1).

These changes in length and dynamics result from the reduction in $[\text{Tu}]_{\text{soluble}}$ that occurs as the number of nucleation sites is increased (Figure 3.8). This reduction in $[\text{Tu}]_{\text{soluble}}$ results from the decreased likelihood of edge-induced catastrophe: when a given mass of tubulin is split between a greater number of microtubules, the MTs get shorter, eventually reaching a natural length that is too short to interact significantly with the cell boundary (compare Figure 3.2B with C). These considerations suggest that, the increase in MT dynamics observed \textit{in vivo} during mitosis is an obligatory consequence of the observed increase in MT nucleation. Note that, this assumes that nucleation is controlled directly. An alternative hypothesis is that MAPs could increase the frequency of catastrophes, causing $[\text{Tu}]_{\text{soluble}}$ to rise. Given that MT growth from centrosomes is reported to increase with $[\text{Tu}]_{\text{soluble}}$, an increase in $[\text{Tu}]_{\text{soluble}}$ could cause the observed increase in the number of MTs. The main point of this analysis is to stress the obligatory connection between $[\text{Tu}]_{\text{total}}$, $[\text{Tu}]_{\text{soluble}}$, microtubule number and transition frequencies.
Figure 3.8: Impact of changes in nucleating activity. Relationship in a spatially confined system between the number of nucleation sites and either the average MT length (A) or free tubulin concentration (B). All other parameters are identical for all simulations represented in this figure (the cell size and all other parameters are the same as used for the simulations analyzed in Table 3.1).
3.3.6.2 Implications for the interpretation of nucleation mutants

Similarly, reducing the number of seeds can be expected to have profound effects on MT length and dynamics (Table 3.1, Figure 3.2, Figure 3.8). Mutations in γ-tubulin-complex proteins, which cause defects in MT nucleation, also cause abnormalities in MT dynamics, including inappropriately continuous growth, abnormally long MTs, increased growth rate, and catastrophe defects [40]. A number of explanations have been proposed for these effects, including inappropriate regulation of MT-binding proteins, such as Tip1p. Our analysis suggests that, these effects are expected outcomes of the increase in the steady-state concentration of soluble tubulin that is predicted to result from reduction in the number of MTs (Figure 3.2D, compare with 3.2B; Table 3.1, Figure 3.8).

3.3.6.3 Implications for the study of MAPs

Although we did not directly study MAPs in these simulations, our analysis has implications for the understanding of MAP function. Experimental characterization of MAPs has often produced conflicting results (e.g. [41, 42]). We propose that some of the discrepancies stem from failure to take into account the fact that a solution of dynamic MTs in vitro is an evolving system. For example, a protein that suppresses catastrophe without altering other parameters under pre-steady-state conditions (when [Tu]_{soluble} is near the initial concentration) could have have little or no effect on catastrophe at steady-state because the protein would reduce [Tu]_{soluble} at steady-state, and catastrophe frequency increases as [Tu]_{soluble} decreases. Indeed, such a protein could have the paradoxical effect of slowing MT growth under steady-state conditions. These issues highlight the systems nature of the MT cytoskeleton and suggest that comprehensive understanding of MAP
function is likely to require integration between experimental and computational approaches.

3.4 Concluding remarks

We have provided evidence that several aspects of MT dynamics that are often thought to be imposed by MAP activity (or changes in MAP activity) are instead simple, perhaps unavoidable, outcomes of the systems nature of the cellular MT cytoskeleton. More specifically, we find that the surprising persistent growth of MTs observed in vivo is a predictable property of a nucleated system of dynamic MTs polymerizing in a constrained space. Moreover, our simulations indicate that the increase in MT dynamics seen in mitosis is an expected outcome of the mitotic increase in nucleation activity. Similarly, our work provides an explanation for the observation that defects in nucleation cause abnormally stable MT plus ends, without invoking action of nucleation factors at MT plus ends. Our simulations indicate that these effects are mediated by changes in the concentration of free tubulin in the system at steady-state.

We do not mean to imply that regulatory proteins are not involved in behaviors such as persistent growth and the transition to mitosis it is clear that MAPs play central roles in these processes. However, we suggest that, instead of causing these behaviors, this regulatory machinery tunes them, for example by shifting the amount of free tubulin available or the concentration of free tubulin at which persistent growth occurs.

A pressing question in the history of eukaryotic life is how complex cellular processes arise. Our work indicates that certain behaviors are intrinsic to a system of dynamic microtubules in a confined space. It seems plausible that cells with a
rudimentary microtubule array could then, over time, develop machinery to modulate these intrinsic behaviors. This hypothesis is attractive from an evolutionary perspective because it provides evidence that complex cellular behaviors can have simple origins.

3.5 Materials and Methods

3.5.1 Essential features of the simulation

MTs are modeled as simple linear polymers of tubulin subunits, with addition and loss of subunits from the tip occurring according to probabilities defined by the following rules: (1) Tubulin subunits have two states that are assumed to be GTP- and GDP-bound, but these states could represent other conformations. (2) Subunits polymerize (add to the end of a MT or seed) in the GTP-bound state, then undergo GTP hydrolysis after polymerization. (3) Hydrolysis of a given subunit occurs according to a set probability that is independent of the nucleotide state of the surrounding subunits, i.e. hydrolysis is a stochastic first order process, producing a shifting ‘cap’ of GTP subunits of variable length. (4) The probability of subunit-addition or -loss on a MT depends on the nucleotide state of the tubulin subunit exposed at the tip. (5) Subunit addition at the tip depends on the concentration of $[\text{Tu}]_{\text{soluble}}$, whereas subunit loss is independent of $[\text{Tu}]_{\text{soluble}}$. (6) The total number of tubulin subunits in the system is fixed (unless otherwise indicated), resulting in competition between microtubules for tubulin subunits and evolution of dynamic-instability parameters as the simulation proceeds. (7) The maximum length of microtubules is limited by a boundary (analogous to the cell edge). This is summarized in the following mathematical
terms:

$$P_{\text{hydrolysis}} = K_{\text{hydrolysis}} \Delta t, \quad (3.1)$$

$$P_{\text{growth}} = \begin{cases} K_{\text{growth}}^{\text{GTP}} [\text{Tu}]_{\text{soluble}} \Delta t & \text{if the tip of the MT is in GTP state} \\ K_{\text{growth}}^{\text{GDP}} [\text{Tu}]_{\text{soluble}} \Delta t & \text{if the tip of the MT is in GDP state}, \end{cases} \quad (3.2)$$

and

$$P_{\text{shortening}} = \begin{cases} 0 & \text{if the tip of the MT is in GTP state} \\ K_{\text{shortening}}^{\text{GDP}} \Delta t & \text{if the tip of the MT is in GDP state}. \end{cases} \quad (3.3)$$

Note that, because this model is stochastic instead of deterministic, probability transition constants ($K$) are used in place of the kinetic rate constants. These kinetic rate constants are often symbolized in deterministic literature using the letter $k$ as in

$$-\frac{d[A]}{dt} = k[A] \quad (3.4)$$

from a simple chemical kinetic equation

$$A \xrightarrow{k} B. \quad (3.5)$$

3.5.2 Parameter determination

The number of MT-nucleation sites, the size of the pool of tubulin subunits and the size of the cell are set by the user. Unless otherwise indicated, the simulations were run at $[\text{Tu}]_{\text{total}} = 14 \mu\text{M}$, 128 stable seeds (nucleation sites) and cell radius =10 $\mu\text{m}$. These values were arbitrarily chosen within the range that is biologically
plausible (while mammalian cells have many MTs, yeast cells have few) and computationally rapid. Probability transition constants can be varied, but for the purpose of these simulations they were set to constant values chosen to qualitatively approximate the behavior observed in mammalian cells during interphase \cite{27}. The specific value of the constants were $K_{growth}^{GTP} = 2\mu M^{-1} \text{seconds}^{-1}$, $K_{growth}^{GDP} = 0.1\mu M^{-1} \text{seconds}^{-1}$, $K_{shortening}^{GDP} = 48 \text{seconds}^{-1}$, $K_{shortening}^{GTP} = 0 \text{seconds}^{-1}$, and $K_{hydrolysis} = 0.1 \text{seconds}^{-1}$. Note that, because the microtubules are modeled as simple linear polymers with subunits that are larger than tubulin (20 nm vs 8 nm), these rate constants do not correspond to experimentally determined values.

We emphasize that, the purpose of these simulations is to explore the relationships common to microtubule systems in general, not to quantitatively replicate specific experimental observations. All other parameters of the system (growth rates, depolymerization rates, transition frequencies, concentrations of polymeric and soluble tubulin) are emergent properties. Varying the user-defined constants over a large range changed the specific values of the emergent parameters but did not alter the basic relationships discussed in the text (data not shown).

3.5.3 Other aspects of the model

Induction of catastrophe by the simulation boundary (cell edge) is an indirect result of inhibition of new subunit addition at the boundary, which leads to loss of the GTP cap and eventual depolymerization. In some cases (indicated in the text), simulations were run without spatial constraint, allowing unlimited MT length. Some simulations (as indicated) were also run without constraints on the mass of total tubulin (i.e. constant soluble tubulin). Both tubulin diffusion and regeneration of GTP-tubulin from detached GDP-tubulin were assumed to
be fast and, therefore, modeled as instantaneous [43, 44]. Therefore, although we present the shape of the cell as rectangular in Figure 3.1 and as radial in the movies (see supplementary material of Gregoretti et al. [28], Movies 1-3), the actual cell geometry is undefined. Investigations into the effects of finite diffusion rates or alternative geometries are important avenues for future study. We refer to recent work by Janulevicius et al. for analysis of the effects of limited compartment volume, such as those that might be found in cell processes, on microtubule dynamics [25].

3.5.4 Computation and analysis of the simulations

The algorithm was coded in C language and run on PCs using Linux operating systems. The output was analyzed and visualized with Matlab 7.0 SP2 (MathWorks, Natick, MA); movies of Gregoretti et al. [28] were made from Matlab output with QuickTime Pro (Apple Computer, Inc., Cupertino, CA). Calculation of transition-frequencies and -rates was performed with a Matlab script that filtered out the noise, and detected persistent growth and shortening phases. Phase transitions occurring within 1% of cell radius from MT seeds or cell edge were excluded from the calculations. Time conversion is: 1 simulation step equals 0.02 seconds. Error bars depicted in figures represent the ± standard deviation of 50 repetitions of each simulation.

3.5.5 Classical relationship between total and soluble tubulin concentration

The prediction depicted in Figure 3.3A was plotted as described by Johnson and Borisy [15]. Considering fiber formation as a simple linear polymerization at equilibrium, the whole process can be described as multiple competing single step
reaction processes. If $u_1$ represents one monomer and $u_i$ represents a polymer $i$ units long, the polymerization can be described in a general fashion as

<table>
<thead>
<tr>
<th>Single Reaction</th>
<th>Thermodynamic Equilibrium Constant</th>
<th>Concentration of Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1 + u_1 \rightleftharpoons u_2$</td>
<td>$K_1$</td>
<td>$C_2 = K_1 C_1^2$</td>
</tr>
<tr>
<td>$u_2 + u_1 \rightleftharpoons u_3$</td>
<td>$K_2$</td>
<td>$C_3 = K_1 K_2 C_1^3$</td>
</tr>
<tr>
<td>$u_3 + u_1 \rightleftharpoons u_4$</td>
<td>$K_3$</td>
<td>$C_4 = K_1 K_2 K_3 C_1^4$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$u_{i-1} + u_1 \rightleftharpoons u_i$</td>
<td>$K_{i-1}$</td>
<td>$C_i = K_1 K_2 K_3 K_{i-1} C_1^i$</td>
</tr>
</tbody>
</table>

Therefore, if $C_0$ represents the total concentration of $u_1$ and $C_1$ represents the concentration of polymer of length $i$, and if all equilibrium constants are equal

$$C_0 = C_1 + 2C_2 + 3C_3 + \ldots + iC_i = \sum_{i=1}^{\infty} iC_i$$

Equation (3.6) establishes the relationship between $C_0$ and $C_1$ in the simplest scenario for a linear polymer in equilibrium conditions. Note that in this subsection we name $C_0$ and $C_1$ as such because we want to be consistent with existing literature nomenclature. In the rest of this work we use $[Tu]_{\text{total}}$ and $[Tu]_{\text{soluble}}$ as they are meant to be applied in particular to microtubules.

Similarly, it can be described a polymerization process in which the very first steps are less favorable than the rest. Assuming for instance a polymerization in which the first 3 steps are less favorable than the rest, we can write

$$K_1 = K_2 = K_3 = K'$$

(3.7)
and

\[ K_4 = K_5 = K_6 = \ldots = K. \] (3.8)

Replacing equation 3.8 into the left hand side of equation 3.6 we can write

\[ C_0 = C_1 + 2K_1C_1^2 + 3K_1K_2C_1^3 + \sum_{i=4}^{\infty} i(K_1K_2K_3K^{-4})(KC_1)^i \] (3.9)

and introducing now equation 3.7 followed by rearrangement we get the general equation

\[ C_0 = \sum_{i=1}^{s-1} iK'(i-1)C_1^i + \sum_{i=s}^{\infty} iK'(s-1)K^{-s}(KC_1)^i \] (3.10)

Figure 3.3A was drawn using equation 3.10 with a seed size of 6 monomers (then \( s = 5 \)) in agreement with previous analysis [45] and chose \( K = 2.86 \times 10^{-1} \) and \( K' = 1.43 \times 10^{-2} \) to produce an appropriate shape and plateau position. \( K \) and \( K' \) have units of inverse concentration.

3.5.6 Limitations of the analysis

This model is an approximation chosen to be general (i.e. not dependent on unknown details of the polymerization process) - consistent with MT behavior observed microscopically - and computationally fast, while still incorporating key aspects of tubulin biochemistry. Necessarily, the model has simplifications, the most obvious of which is that MTs were modeled as simple linear polymers instead of tubes composed of protofilaments. Incorporation of structural detail is an important goal for future work, but we do not expect this simplification to alter our basic conclusions because the rules governing the simulation are largely independent of this structural detail (i.e. the rate of growth is expected to be a linear function of subunit concentration [46], regardless of the size of the subunits.
or the number of protofilaments). A second simplification is that, our simulation assumes that tubulin has two states, with a stochastic transformation between them. It is important to notice that the identity of the two states (GTP and GDP, GDP-Pi and GDP, or open sheet and closed tube) is irrelevant - what is important is that there are two states with different characteristics. There may in fact be several conformational states at the end of the MTs but we assume that one of the transitions is rate-limiting, and it is this rate-limiting transition that we are simulating. We chose stochastic transitions instead of the commonly assumed vectorial hydrolysis because previous modeling efforts have indicated that vectorially catalyzed transitions do not fit the data well [19]. Given the similarity of the simulation to the behavior of MT systems observed experimentally (Table 3.1, Figures 1 and 2), we believe that this model is an informative approximation.

3.6 References


CHAPTER 4

ANALYSIS OF A MESOSCOPIC STOCHASTIC MODEL OF MICROTUBULE DYNAMIC INSTABILITY

4.1 Abstract

A theoretical model of dynamic instability of a system of linear one-dimensional microtubules (MTs) in a bounded domain is introduced for studying the role of a cell edge in vivo and analyzing the effect of competition for a limited amount of tubulin. The model differs from earlier models in that the evolution of MTs is based on the rates of single-mesoscopic-unit (e.g., a heterodimer per protofilament) transformations, in contrast to postulating effective rates and frequencies of larger-scale macroscopic changes, extracted, e.g., from the length history plots of MTs. Spontaneous GTP hydrolysis with finite rate after polymerization is assumed, and theoretical estimates of an effective catastrophe frequency as well as other parameters characterizing MT length distributions and cap size are derived. We implement a simple cap model which does not include vectorial hydrolysis. We demonstrate that our theoretical predictions, such as steady-state concentration of free tubulin and parameters of MT length distributions, are in agreement with the numerical simulations. The present model establishes a quantitative link between mesoscopic parameters governing the dynamics of MTs and macroscopic

\footnote{This thesis chapter has been published as a research article. G Margolin, IV Gregoretti, HV Goodson and MS Alber: 2006 Physical Review E 74:041920.}
characteristics of MTs in a closed system. Last, we provide an explanation for non-exponential MT length distributions observed in experiments. In particular, we show that the appearance of such non-exponential distributions in the experiments can occur because a true steady state has not been reached and/or due to the presence of a cell edge.

4.2 Introduction

Microtubules (MTs) are intracellular polymers which constitute part of the cytoskeleton and are responsible for many cell functions including division, organelle movement, and intracellular transport.

For this purpose, MTs possess a property called dynamic instability, which enables them to promptly switch between two modes, growth and shortening [1–3]. This is achieved through MT having a stabilizing cap which keeps the MT from disassembling. The MT tends to depolymerize when the cap is lost [2–5]. The cap gradually hydrolyzes and becomes unstable as well, and so for the MT to survive it has to grow to renew its cap.

The existence of a guanosine triphosphate (GTP) cap at the end of MTs [4] and the phenomenon of dynamic instability [1] were discovered in the early 1980s. Hill and Chen used a Monte Carlo approach to simulate this behavior [2, 6], employing a representation of a MT in which its cap could consist of many units (heterodimers). One of the main outcomes of their work was a suggestion that a two-phase (cap, no cap) model of dynamic instability, based only on observable macroscopic rates of phase and length changes, was sufficient to understand the behavior of the ensemble of MTs (cf. [5], Figs. 4–6). This phenomenological approach has been prevalent since then in modeling the behavior of an ensemble of
MTs at the cellular level and in vitro [5, 7–15].

In order to advance the understanding of the assembly of individual MTs, in 1990 Bayley et al. [16] developed the computational molecular-level lateral cap model, in which the cap consisted of a single layer of GTP-Tubulin. Quite recently many computational molecular-level models of a single MT began to emerge [17–21] which try to incorporate biological details observed due to advances in the experimental techniques. In particular, it is now known from the experiments that the tips of MTs can have geometrical configurations typical to growing and shortening MTs, which differ from one another (e.g., [3]). This is closely related to the idea of the structural, and not necessarily a GTP cap [17], when due to tensile stresses inside the elastic body of a MT, its shape deforms from a cylinder near the tip.

Flyvbjerg et al. [22, 23] introduced an elegant analytical model of the GTP cap dynamics based on a one-dimensional (linear) representation of a MT. It incorporated constitutive processes of spontaneous and vectorial hydrolyses inside the MT and fluctuating growth of the cap size. The mesoscopic scale of these processes is larger than the molecular scale, but is smaller than the scale of the phenomenological macroscopic changes. It roughly coincides with the scale of resolution of a microscope, and it allows us to define a unit with the length of the order of a tubulin heterodimer.

In both in vitro and in vivo experiments, the dynamics of MTs has been observed under a large variety of physical conditions and in various chemical environments. A lot of data at macroscopic scale have been accumulated including parameter values describing the MT dynamics and length distributions. Can these values be predicted based on the conditions of the experiments? How would
a change in ambient conditions or the presence of spatial constraints affect observables? These questions are difficult, if not impossible, to answer using the models with postulated observable (macroscopic) rates.

In this paper we analyze a model of MT dynamics in a finite domain bounded by the cell edge, which involves competition between individual MTs for tubulin. The model is based on a linear one-dimensional (1D) approximation of a MT structure. We consider the role of the boundary and extend the model to incorporate finite hydrolysis. Our model is different from earlier works \[11, 15\] addressing the role of the edge in that we explicitly consider the concentration dependence of the dynamic instability parameters, as well as a competition for a limited tubulin pool. Although the cap model described in the first part of the paper deals with a single MT, the second part of the paper focuses on studying cellular level behavior of many MTs using the link between the mesoscale and the macroscale provided by the cap model.

Namely, we use a generalization of a mesoscopic model of MTs introduced in \[24\]. Instead of postulating macroscopic rates \[2, 5\] or deducing them from numerical simulations \[6\], as was done by Hill and Chen, we estimate them analytically from basic mesoscopic rates of (de)polymerization and hydrolysis of a single unit. This results in a higher resolution analytical model which may be more suitable for today’s higher-resolution experiments and can partially address the questions posed two paragraphs above.

Recent papers \[25, 26\] deal with modeling similar biological problems. The method in \[25\] is based on using macroscopic rates, while \[26\] does not consider finite domain size and competition for tubulin. The analyses differ as well.

There are two reasons that justify our choice of a mesoscale approach to model-
ing of dynamic microtubules. (i) The intermediate scale of our model is appropriate for studying the behavior of MTs at the cellular level, but without postulating the macroscopic/observable rates. (ii) A mesoscale approach is appropriate given currently available experimental data. In an ideal case, experimental results would provide a modeler with biochemical properties (such as rate constants) that determine the interactions between molecules such as tubulin. The first experimental results on microtubule dynamics at molecular resolution have been published only very recently [27]. Although these results are important, they leave several central structural questions unanswered.

This chapter is organized as follows. The conceptual model and its computational implementation are presented in Sec. II. Next, in Sec. III we develop a cap model where the cap can have any number of units –see Figure 4.1(b). This cap model differs from previous models in that it does not involve vectorial or induced hydrolysis. Using this model we derive approximate expressions for observable rates. We then describe in Secs. IV and V a quantitative theoretical analysis of a lower-resolution model with the cap being treated as a single unit –see Figure 4.1(a). The influence of the cell edge is also studied there. Section V describes balance between polymerized and free tubulin in a bounded domain with a fixed total amount of tubulin present. Finally, we discuss and summarize our findings in Secs. VI and VII.

4.3 Model Description, Parameters, and Notations

In this section we describe a basic model of dynamics of MTs. We consider a domain of size \( L_x \times L_y \times L_z \) with \( N_n \) available nucleation sites for MTs in its center. \( N_n \) is the maximal number of MTs –see Table 4.1 For simplicity, in our
Figure 4.1. Schematic representation of the model.
study of the role of the boundary (e.g., cell edge) we assume that all MTs have an identical maximal allowed length. All MTs grow from nucleation sites (there is no spontaneous nucleation), and MTs grow at one end (usually the so-called plus end) only. There is a fixed amount of total tubulin in the domain. This tubulin is present in two forms: free tubulin in the solution and polymerized tubulin constituting the MTs. Free tubulin is taken up by growing (polymerizing) MTs and is released back into the solution by shortening MTs. In general, free tubulin (Tu) diffuses inside the domain. In this paper we assume that the diffusion of free Tu is fast and does not lead to a diffusion-limited reaction rates. This is in agreement with [28]. Moreover, we assume uniform concentration of free tubulin throughout the domain which implies instantaneous diffusion. (For studies of the effects of tubulin diffusion the reader is referred to [29, 30].)

In this paper a MT is represented at each moment in time in the form of a 1D straight line consisting of a certain number of units of a predefined length. Each unit belonging to a MT can be in either a growth-prone state or a shortening-prone state. We will refer to them as GTP (or T) state or GDP (or D, from guanosine diphosphate) state, respectively. All free tubulin is assumed to be in a T state. When a unit joins the MT it is initially in a T state. The probability that the internal units have hydrolyzed (transformed to a D state) increases with time. When MT disassembles (shortens) these D units, upon becoming terminal, have higher probability to disassemble and return to the solution than the terminal T units. Upon return to the solution they immediately switch to the T state. The terminal T unit does not hydrolyze but can with a certain probability depolymerize (drop from the MT end). Incorporation of a new unit at the MT tip triggers the hydrolysis process of the previously terminal unit. This description seems
## Table 4.1

### Notation Highlights.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c)</td>
<td>concentration of free tubulin</td>
<td>(\mu M)</td>
</tr>
<tr>
<td>(c_{eq})</td>
<td>critical concentration of free tubulin</td>
<td>(\mu M)</td>
</tr>
<tr>
<td>(c_{tot})</td>
<td>total concentration of tubulin</td>
<td>(\mu M)</td>
</tr>
<tr>
<td>(k_{gT})</td>
<td>second order rate of adding a unit on top of a terminal T unit</td>
<td>(\mu M^{-1}s^{-1})</td>
</tr>
<tr>
<td>(k_{gD})</td>
<td>second order rate of adding a unit on top of a terminal D unit</td>
<td>(\mu M^{-1}s^{-1})</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>parameter of exponential distribution (number of units)</td>
<td>–</td>
</tr>
<tr>
<td>(\ell)</td>
<td>characteristic cap size (number of units)</td>
<td>–</td>
</tr>
<tr>
<td>(m)</td>
<td>mean length (number of units) of MTs</td>
<td>–</td>
</tr>
<tr>
<td>(n)</td>
<td>coarsened step size in the cap model (number of units)</td>
<td>–</td>
</tr>
<tr>
<td>(\rho)</td>
<td>((n - 1)/(\ell - 1))</td>
<td>–</td>
</tr>
<tr>
<td>(K_c)</td>
<td>rate of the edge-induced catastrophe</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_h)</td>
<td>rate of hydrolysis of internal T units</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{gT})</td>
<td>rate of adding a unit on top of a terminal T unit</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{gT}^{eff})</td>
<td>effective rate of growing by one unit in growth phase</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{obs}^{gT})</td>
<td>rate of growing by (n) units in growth phase, equal to (K_{gT}^{eff}/n)</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{gD})</td>
<td>rate of adding a unit on top of a terminal D unit</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_n)</td>
<td>nucleation rate of a MT</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{sT})</td>
<td>rate of depolymerization of terminal T unit</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{sT}^{eff})</td>
<td>(nK_{sT}^{obs})</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{obs}^{sT})</td>
<td>rate of shortening by (n) units in growth phase, equal to catastrophe frequency</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(K_{sD})</td>
<td>rate of depolymerization of terminal D unit</td>
<td>(s^{-1})</td>
</tr>
<tr>
<td>(L)</td>
<td>maximal length (number of units) of MTs in domain with upper bound</td>
<td>–</td>
</tr>
<tr>
<td>(L_x, L_y, L_z)</td>
<td>domain sizes in the numerical simulations</td>
<td>(m)</td>
</tr>
<tr>
<td>(M_g(l))</td>
<td>number of MTs of length (l) number of units in growth phase</td>
<td>–</td>
</tr>
<tr>
<td>(M_g(l))</td>
<td>number of MTs of length (l) number of units in shortening phase</td>
<td>–</td>
</tr>
<tr>
<td>(N_0)</td>
<td>number of free nucleation seeds</td>
<td>–</td>
</tr>
<tr>
<td>(N_{MT})</td>
<td>number of MTs, (\leq N_n)</td>
<td>–</td>
</tr>
<tr>
<td>(N_n)</td>
<td>total number of nucleation seeds</td>
<td>–</td>
</tr>
</tbody>
</table>

The dash (–) means dimensionless.
appropriate in view of [3].

The dynamics of the MTs is determined by five mesoscopic rates, $K_{gT}$ ($K_{gD}$) and $K_{sT}$ ($K_{sD}$), which are the rates of MT growth and shortening (i.e. adding one more unit from the solution on top of the current terminal unit, and losing this current terminal unit to the solution) when the terminal unit is in state T (D), and the hydrolysis rate $K_h$ of the internal units which are in state T. If the terminal unit has to hydrolyze in order to depolymerize (and its hydrolysis rate is not faster than that for the internal units) then $K_{sT} < K_h$.

For numerical simulations, shortening rates are taken to be independent of $c$ while growth rates are assumed proportional to $c$ at the location of MT tip:

$$K_{gT,gD} = k_{gT,gD}c. \quad (4.1)$$

Such specific dependence of the growth and shortening rates on $c$, though, is not required for many of the theoretical results we report.

When the MT reaches boundary of the domain it is not allowed to grow anymore and will eventually lose its terminal unit initiating with certain probability a shortening phase. There are two more rates at the domain boundary of importance in the model: a rate $K_n$ of nucleation from existing seeds and a rate $K_e$ of edge-induced catastrophe, which can also depend on $c$. Section 4.9.1 contains a brief description of a numerical algorithm we used in our simulations.

In what follows, we will impose restriction on a maximal length of a MT (upper bound, e.g., due to a cell edge). We will call zero a lower bound.
4.3.1 Observables

The standard experimental observables describing the dynamic behavior of a single MT are derived from the MT length versus time plot. Typical length history plots of MTs are shown in Fig. 2. Due to the two-state nature of the tubulin units inside the MT, the fluctuations in this length may be large and even in the (macroscopically) steady state each MT can repeatedly change its length all the way from zero to some characteristic length or to the boundary. If no boundary is present and if free tubulin concentration stays high enough (if $K_{gT}$ and $K_{gD}$ are high enough), MTs can grow unbounded [7, 29, 31]. From a sawtooth-like evolution of a MT length four parameters can be extracted: the velocity/rate of growth, the velocity/rate of shortening, the average time of growth before switching to shortening, and the average time of shortening before switching to growth. The inverses of these times define the so-called catastrophe frequency and the rescue frequency, respectively. Note that brief growth or shortening intervals may pass unnoticed in the analysis of experimental data. Some models of dynamics of MTs use these four parameters as given constants for constructing analytical solutions [7, 11] ignoring their microscopic origin. In Sec. III we use mesoscopic rates to derive the observable growth velocity and catastrophe frequency instead of setting them from the beginning.

4.4 Cap Model

Carlier and colleagues have provided experimental evidence that the GTP cap of the MTs is not restricted to units at the very tip [4, 31, 32]. This suggests that the hydrolysis is not instantaneous, a conclusion also supported by work on yeast tubulin [33, 34]. To incorporate this feature into our approach we develop a model
for the cap. In this section we show that using this model, based on the underlying mesoscopic laws, one can predict the observables: the catastrophe frequency and the velocity (rate) in the growth phase of MTs. In other words, the mesoscopic laws of the MT dynamics, governing single-unit polymerization/depolymerization and hydrolysis, can be related to the macroscale (observable) dynamics of a MT.

When $K_h < \infty$, the MT cap in our model consists mainly of T units [cf. Fig. 1(b)] and possibly of a few D units and has some characteristic length (number of units) $\ell > 1$. When $K_h \to \infty$, then $\ell = 1$ because only the terminal unit is not allowed to hydrolyze and it is in a T state. Our approach is based on coarsening the resolution in the growth phase so that only blocks of the order of cap size, $n \sim \ell$, are resolved. By catastrophe we understand the loss of the cap. In what follows we will establish a connection between coarsened observed rate constants $K_{\text{obs}}^{\text{gT},sT}$ of growing or shortening by one block of $n$ units (in the growth phase) and the original rates $K_{gT,sT}$ and $K_h$. As will become apparent from the following derivations, $K_{sT}^{\text{obs}}$ approximates the catastrophe frequency, while $nK_{gT}^{\text{obs}}$ is the effective rate of addition of a single unit on top of the MT, in the growth phase. There is no need to rescale $K_{gD,sD}$ if the free tubulin concentration is not too low. Therefore, we only consider a model for the MT cap and do not alter the rest.

Let us consider the cap model in detail. In what follows we neglect the fluctuations in $\ell$ due to randomness in hydrolysis, and we assume that each T unit is hydrolyzed after staying an internal unit for a time $\Delta t_h = 1/K_h$. After the rescue or nucleation event occurs the cap begins to grow. It has time $\Delta t_g + \Delta t_h$ to elongate, where $\Delta t_g = 1/(K_{gT} + K_{sT})$ is the time of a single-unit step in the growth phase. After that its average length remains constant (under assumption that no
catastrophe occurs during this time). For the catastrophe to occur the cap should be lost, due to fluctuations in cap size and in the growth velocity \[2, 5, 35\]. In our analysis we consider two scenarios for cap loss: (i) roughly half of the cap is lost due to random nature of MTs growth, and the second half gets hydrolyzed during this time, or (ii) the whole cap is lost due to random fluctuations in MT growth. Keeping in mind that the terminal unit cannot be lost as a result of hydrolysis, in description (i) we require that \((\ell + 1)/2\) and \((\ell - 1)/2\) units be lost due to fluctuations in MT growth and propagation of hydrolysis front, respectively. We define

\[
\rho = \frac{n - 1}{\ell - 1}. \quad (4.2)
\]

In case (i) \(n = (\ell + 1)/2\) and \(\rho = 1/2\), while in case (ii) \(n = \ell\) and \(\rho = 1\). It is important to stress that \(\rho\) is introduced as a fixed parameter set \textit{a priori}, based on the scenarios of cap loss similar to (i) and (ii). The two descriptions (i) and (ii) determine the duration of the coarsened step:

\[
\Delta t_{\text{obs}}^g \equiv \frac{1}{K_{\text{gT}}^{\text{obs}} + K_{sT}^{\text{obs}}} = \rho \Delta t_h + \Delta t_g \equiv \rho \cdot \frac{1}{K_h} + \frac{1}{K_{gT} + K_{sT}}. \quad (4.3)
\]

For a given \(n\), in order to rescale/coarsen the dynamics in the growth phase we require that both the average velocity and the diffusion coefficient of the MT tip remain unchanged, in the hypothetical case of no hydrolysis. For a random walk on a line, with probability \(p\) to jump to the right and \(q = 1 - p\) to jump to the left, the average velocity is \(v = (p - q)\Delta x/\Delta t\) and the diffusion coefficient is \(D = 2pq\Delta x^2/\Delta t\). Here \(\Delta x\) is the step length and \(\Delta t\) is the time per step. In the case of the original walk \(p = p_g = K_{gT}\Delta t_g\), \(\Delta x = 1\) and \(\Delta t = \Delta t_g\), while in the case of the rescaled walk \(p = K_{gT}^{\text{obs}}\Delta t_{\text{obs}}^g\), \(\Delta x = n\) and \(\Delta t = \Delta t_{\text{obs}}^g\) is given by
Equation (4.3). After introducing the effective rates of adding or losing one unit (as opposed to one block),

\[ K_{gT,sT}^{\text{eff}} \equiv nK_{gT,sT}^{\text{obs}}, \quad (4.4) \]

and using conservation of \( v \) and \( D \) we obtain that

\[ K_{gT}^{\text{eff}} - K_{sT}^{\text{eff}} = K_{gT} - K_{sT} \quad (4.5) \]

\[ \frac{K_{gT}^{\text{eff}} K_{sT}^{\text{eff}} n}{K_{gT}^{\text{eff}} + K_{sT}^{\text{eff}}} = \frac{K_{gT}K_{sT}}{K_{gT} + K_{sT}}. \quad (4.6) \]

From eqs. (4.3) and (4.4), \( n/(K_{gT}^{\text{eff}} + K_{sT}^{\text{eff}}) \) can be found and substituted into Equation (4.6), resulting in

\[ K_{gT}^{\text{eff}} K_{sT}^{\text{eff}} = \beta, \quad (4.7) \]

where

\[ \beta \equiv \frac{K_{gT}K_{sT}}{1 + \rho \cdot \frac{K_{gT} + K_{sT}}{K_h}}. \quad (4.8) \]

After solving Equations. (4.5) and (4.7) and choosing only positive solutions we obtain that

\[ K_{gT}^{\text{eff}} = \frac{K_{gT} - K_{sT} + \sqrt{(K_{gT} - K_{sT})^2 + 4\beta}}{2} \quad (4.9) \]

\[ K_{sT}^{\text{eff}} = -K_{gT} + K_{sT} + \frac{\sqrt{(K_{gT} - K_{sT})^2 + 4\beta}}{2}. \quad (4.10) \]

Expression (4.9) displays some expected features: namely, when \( K_h \to \infty \Rightarrow \)
\( K_{gT}^{\text{eff}} \to K_{gT} \) and when

\[
K_h \to 0 \Rightarrow K_{gT}^{\text{eff}} \to (K_{gT} - K_{sT} + |K_{gT} - K_{sT}|)/2 = \begin{cases} 
K_{gT} - K_{sT}, & K_{gT} \geq K_{sT} \\
0, & K_{gT} \leq K_{sT}
\end{cases}
\]

(4.11)

In general, \( \max(0, K_{gT} - K_{sT}) \leq K_{gT}^{\text{eff}} \leq K_{gT} \).

Equation (4.6) combined with Equations (4.9) and (4.10) yields

\[
n = \frac{K_{gT}K_{sT}}{K_{gT} + K_{sT}} \sqrt{(K_{gT} - K_{sT})^2 + 4\beta/\beta}
\]

(4.12)

and Equation (4.2) can now be used to determine \( \ell \). On the other hand, \( \ell \) can be approximated as follows:

\[
\ell \approx K_{gT}^{\text{eff}} \Delta t_h + 1 = \frac{K_{gT}^{\text{eff}}}{K_h} + 1,
\]

(4.13)

where the term \( K_{gT}^{\text{eff}} \Delta t_h \) approximates the number of added units after the beginning of a growth phase, before the hydrolysis front starts moving. In fact, Equation (4.13) can be used as a definition of \( \ell \) and \( n \) can be found from Equation (4.2). Then there is no need for Equation (4.3) as Equations (4.5), (4.6) and (4.13) form a closed set of equations. This approach, however, leads to a cubic equation for \( K_{gT}^{\text{eff}} \), while in the above approach we need to solve a quadratic equation, which is much simpler. Nevertheless, the definition of \( \ell \) through Equation (4.13) seems to work better in the limit of \( K_{gT} \to 0 \). Namely, substituting eq. (4.13) into Equation (4.2) yields \( n = 1 + \rho K_{gT}^{\text{eff}}/K_h \) and the only non-negative solution of this equation together with Equations (4.5) and (4.6), in the limit \( K_{gT} \to 0 \), is \( K_{gT}^{\text{eff}} = 0, K_{sT}^{\text{eff}} = K_{sT} \) and \( n = 1 \). Indeed, it seems reasonable to postulate that \( n \to 1 \), i.e., there is no rescaling, when \( K_{gT} \to 0 \). This does not follow from
Equations (4.8) and (4.12), which lead to \( n \to 1 + \rho K_s T / K_h > 1 \) instead.

Using the above developments, it is possible to derive scaling behaviors of various quantities as functions of, e.g., \( c \) and \( K_h \). For example, substituting Equations (4.2) and (4.13) into equation (4.8) and using Equation (4.7) together with Equation (4.4) leads to

\[
K_{sT}^{\text{obs}} \approx \frac{1}{n} \cdot \frac{K_g T K_s T}{K_g^{\text{eff}} + (n - 1)(K_g T + K_s T)}
\]

\[
\approx \begin{cases} 
\frac{K_s T}{n^2}, & K_g T \gg K_s T \\
\frac{K_s T}{2n(n - 1)}, & K_g T \gg K_g^{\text{eff}}
\end{cases} \quad (4.14)
\]

so that the catastrophe rate (frequency) \( K_{sT}^{\text{obs}} \) scales as \( n^{-2} \propto \ell^{-2} \). This is characteristic of diffusive scaling because the time to catastrophe is determined by the diffusive movement of the hydrolysis front relative to the MT tip. If free tubulin concentration \( c \) is not too small, then Equations (4.12), (4.8) and (4.1) yield \( n \propto c \) and hence \( K_{sT}^{\text{obs}} \propto c^{-2} \), which is in at least qualitative agreement with previous predictions [13, 36]. The scaling \( K_{sT}^{\text{obs}} \propto n^{-2} \) might have been postulated, as well, which would have led us to an additional version of the solution of the cap model.

To provide another scaling example, let us now assume that \( K_h \) is small and \( K_g T \approx K_s T \). Then from Equation (4.8) it follows that \( \beta \approx K_h K_s T / (2\rho) \propto K_h \). If \( \beta \gg (K_g T - K_s T)^2 \), i.e., if \( K_h \) is not too small, then Equation (4.12) yields \( n \propto K_h^{-1/2} \) and hence \( \ell \propto K_h^{-1/2} \) as well. Notice, that it is the same scaling as derived for actin polymers (Equation 3 in [37]).
4.5 Ensemble dynamics of microtubules

In this section we treat cap as a single effective unit - cf. Fig. 4.1. Thus the model essentially reduces to the two-phase model proposed in [5]. First, we re-derive length distribution of MTs, which are known in the literature. In particular, we study the role of upper bound (e.g., cell edge). We use these results for analyzing in the next section competition for a finite tubulin pool. We also consider the steady state critical concentration of free Tu.

In what follows, we use either the discrete or the continuous description of MT dynamics, whichever is convenient. We assume that the continuous model provides a good approximation of the discrete model. The continuous approach was discussed in [7, 38, 39] while an analogous discrete approach was developed in [5, 11]. Following [7] we write down the equations for length distributions of MTs in growth and shortening phases in the form

\[ \partial_t M_g = -K_{sT}^{obs} M_g + K_{gD} M_s - K_{gT}^{eff} \partial_l M_g \]  
(4.15)

\[ \partial_t M_s = K_{sT}^{obs} M_g - K_{gD} M_s + K_{sD} \partial_l M_s, \]  
(4.16)

where \( M_{g,s}(l, t) \) are densities of MTs of length \( l \) at time \( t \), in the growing (g) and shortening (s) phases.

Equations (4.15) and (4.16) can be used to describe regular diffusion with drift, if we do not distinguish between the phases (see also [15]). However, it is important to stress that these equations don’t have diffusion terms for \( M_{g,s}(l, t) \) and hence switching phases back and forth is the only mechanism of spreading of these distributions present. This is in agreement with our simulations in the case
of instantaneous hydrolysis of internal units. Notice that diffusion terms are used in [5].

First, consider a semi-infinite domain. Equations (4.15) and (4.16) with the boundary condition $M_{g,s}(l = \infty) = 0$ have the following steady state solution

$$M_g = Ae^{-l/\lambda}, \quad (4.17)$$

$$M_s = \frac{K^{eff}_g T}{K^{eff}_s} Ae^{-l/\lambda}, \quad (4.18)$$

where

$$\frac{1}{\lambda} \equiv \frac{K^{obs}_T}{K^{eff}_g} - \frac{K^{obs}_D}{K^{obs}_s}. \quad (4.19)$$

The necessary condition for the existence of a steady state in the case without a boundary is given by $\lambda > 0$. The prefactor $A$ is a normalization coefficient which depends on the total number of MT nucleation seeds present and on the nucleation probability.

We now add a constraint limiting the maximal length of MTs to be $L$. MTs cannot become longer due to a barrier, for example a cell edge, as is often the case in vivo, especially when the cell is in interphase and the MTs are relatively long. For simplicity we assume that $L$ is identical for all MTs. We still can use Equations (4.15) and (4.16) inside the domain, for $0 < l < L$, and consider a steady state. Adding up these two equations then leads to $\partial_t (-K^{eff}_g T M_g + K^{obs}_s D M_s) = 0$, which means that the spatial derivative of the flux (of the MT tips, considered as random walkers) is zero meaning that the flux is uniform. However, in the closed system
this flux must be zero and hence
\[ M_s = \frac{K_{\text{eff}} gT}{K_{sD}} M_g. \] (4.20)

It follows that Equations 4.17, 4.18, and 4.19 still hold inside the domain, except for \( \lambda \) not being necessarily positive, which is in agreement with previous work [11]. This means qualitatively that there might be a steady state distribution of MTs in which most of them are close to the upper boundary (e.g., cell edge), while only a few are short.

4.5.1 Critical concentration of free tubulin

Let us consider the limiting case \( 1/\lambda = 0 \) – cf. Equation (4.19). This defines the upper limit of the concentration of free tubulin \( c_{eq}^\infty \) at which the steady state in the semi-infinite domain still exists (cf. [7]). Let us use Equation (4.1) and define
\[ a = \frac{K_{sT}}{k_{gT}}, \quad b = \sqrt{\frac{K_{sT} K_{sD}}{k_{gT} k_{gD}}}. \] (4.21)

Because the MTs in the growth phase are less likely to shorten than are MTs in the shortening phase, \( a < b \). In general, \( c_{eq}^\infty \in [a, b] \). Notice that slowdown of hydrolysis reduces \( c_{eq}^\infty (K_h) \). When hydrolysis is instantaneous then \( c \) reaches its maximal value \( c_{eq}^\infty (\infty) = b \). When there is no hydrolysis at all then \( c_{eq}^\infty (K_h) \) reaches its minimal value \( c_{eq}^\infty (0) = a \) meaning that the average growth rate in this case, \( k_{gT} c_{eq}^\infty (0) - K_{sT} \), is zero.

One can get a scaling estimate of \( c_{eq}^\infty \) if it is far enough from both \( a \) and \( b \). Assume that \( K_{gT} \gg K_{sT} \) and \( K_{gT} \gg K_h \). Using Equation (4.1), then \( K_{gT} \approx K_{gT} \approx K_{gT} c \). Now, \( \rho \) is of order of 1, so that from Equation (4.8) and Equation (4.12)
it follows that $\beta \sim K_{sT} K_h \ll K_{sT}^2$ and $n \sim K_{gT}/K_h \propto c/K_h$, respectively. Hence $K^\text{obs}_{sT} \approx K_{sT}/n^2 \propto (K_h/c)^2$. Substituting these scaling relations into $1/\lambda = 0$ and using Equation (4.19) yields $c^\infty_{eq} \propto \sqrt{K_h}$.

Notice that if there are no rescues, $k_{gD} = 0$, then $c^\infty_{eq}$ is infinite (see also [36]) and unbounded growth can not happen. This is so because without rescues the MT depolymerizes completely after the catastrophe, no matter how long it was before.

4.6 Competition for tubulin

In Section 4.5 we have shown the existence of a steady state distribution of MTs inside a domain. It is conceivable that by sensing and controlling free tubulin concentration and the number of MTs the cell regulates MT dynamics, as suggested in [24, 36]. In what follows we show in detail how to determine a steady state concentration of free tubulin $c$, which is the key to finding steady state characteristics of MTs in a closed system. The main goal of this section is to derive expressions for the average number of units per MT $m$, and a number of MTs $N_{MT}$ as functions of $c$ and the other parameters. These functions are needed to determine $c$ from the conservation of total tubulin. Because $m$ can not extend beyond the domains’ boundary, and because $N_{MT}$ is less or equal than a number of nucleation sites $N_n$, the number of polymerized units in a bounded domain stays restricted as the amount of total Tu grows.

In what follows we assume that the total amount of tubulin is constant and we consider bounded domain of volume $V$. We also assume instantaneous diffusion so
that the concentration of free tubulin is uniform throughout the domain. Hence,

$$N_{tot} = N_{free} + mN_{MT},$$  \hspace{1cm} (4.22)$$

where $N_{tot}$ is a total number of tubulin units in the domain and $N_{free}$ is a number of free tubulin units. By dividing this formula by $V$ we obtain an expression for concentrations measured in micromolars ($\mu M$):

$$c_{tot} = c + \frac{mN_{MT}}{10^{-3}N_AV},$$  \hspace{1cm} (4.23)$$

where $V$ is given in $m^3$ and $1 \mu M$ is equal to $10^{-6}N_A$ units per liter or $10^{-3}N_A$ units per $m^3$; $N_A \approx 6.022 \times 10^{23} \text{mol}^{-1}$ is the Avogadro’s constant. We will use this expression for studying MTs in cases of unbounded and bounded domains.

4.6.1 Unbounded domain

It has been shown in Section 4.5 that the steady state distribution of MT lengths in this case is exponential as described by Equations (4.17) and (4.18) with $\lambda$ representing mean length of a MT:

$$m = \lambda.$$

(4.24)

To find $N_{MT}$ for a given $N_n$, we use a balance equation for the number of available nucleation sites $N_0 \equiv N_n - N_{MT}$:

$$K_{sT}^{obs}M_g(l = 1) + K_{sD}M_s(l = 1) = K_nN_0,$$

(4.25)
where $K_n$ is a nucleation rate, which in general depends on $c$. The left-hand side of eq. (4.25) describes the rate of production of available nucleation sites by completely depolymerizing MTs. The first term represents those MTs which experience a catastrophe, while the second term stands for those MTs which are already in the shortening phase. Using (4.17), (4.18), and assuming that $\lambda \gg 1$, yields

$$A(K_{sT}^{obs} + K_{gT}^{eff}) \approx K_n N_0.$$  

(4.26)

In addition, approximating summation by integration,

$$N_{MT} = \sum_{l=1}^{\infty} (M_g(l) + M_s(l)) \approx \left(1 + \frac{K_{gT}^{eff}}{K_{sD}}\right) A \lambda$$  

(4.27)

results in

$$N_{MT} \approx \frac{N_n}{K_{sT}^{obs} + K_{gT}^{eff}} \cdot \frac{1}{1 + \frac{\lambda K_n (1 + K_{gT}^{eff}/K_{sD})}{K_{sT}^{obs}}}.$$  

(4.28)

Substituting Equations (4.24), (4.28) and (4.19) into Equation (4.23), and using dependence of the rates on concentration $c$, eq. (4.1), yields an equation $c_{tot} = F(c)$, at steady state. This equation relates free tubulin concentration $c$ and total concentration $c_{tot}$, as a function of all given parameters.

### 4.6.2 Bounded domain

Here again we impose limitation on the maximal possible length of MTs not to exceed $L$. In the case of a bounded domain a steady state solution of Equations (4.15) and (4.16) can be calculated, and then $m$ and $N_{MT}$ are determined in a way similar to the previous case. Notice that the steady state does exist even if $c > c_{eq}^{\infty}$. Since $0 \leq m \leq L$ and $0 \leq N_{MT} \leq N_n$ always hold, the second term in
Equation (4.23) is non-negative and bounded. Therefore, when $c$ goes from 0 to $\infty$ so does $c_{\text{tot}}$. If the right-hand side of Equation (4.23) monotonically increases with $c$, then there is a unique physically meaningful $c$ for each $c_{\text{tot}}$ at steady state.

It is shown in Section 4.9.2 that now

$$N_{MT} \approx \frac{N_n}{K_{sT}^{\text{obs}} + K_{qT}^{\text{eff}}}$$

$$1 + \frac{K_n[(1 + K_{gT}^{\text{eff}}/K_{sD})(1 - e^{-L/\lambda})\lambda + (K_{gT}^{\text{eff}}/K_e)(1 + K_{gD}/K_{sD})e^{-L/\lambda}]}{(1 + K_{gT}^{\text{eff}}/K_{sD})\lambda(1 - e^{-L/\lambda}) + (K_{gT}^{\text{eff}}/K_e)(1 + K_{gD}/K_{sD})e^{-L/\lambda}}$$

and

$$m \approx \frac{(1 + K_{gT}^{\text{eff}}/K_{sD})\lambda(\lambda - e^{-L/\lambda}(L + \lambda)) + L(K_{gT}^{\text{eff}}/K_e)(1 + K_{gD}/K_{sD})e^{-L/\lambda}}{(1 + K_{gT}^{\text{eff}}/K_{sD})\lambda(1 - e^{-L/\lambda}) + (K_{gT}^{\text{eff}}/K_e)(1 + K_{gD}/K_{sD})e^{-L/\lambda}}.$$  

(4.29)

(4.30)

When $\lambda > 0$ and $L \to \infty$ eqs. (4.29) and (4.30) reduce to eqs. (4.28) and (4.24), respectively. When $K_{gT}^{\text{eff}} \to 0$ and $K_{gD} \to 0$, then $\lambda \downarrow 0$ (so that $\lambda > 0$), $N_{MT} \to 0$ and $m \to 0$. When $K_{gT}^{\text{eff}} \to \infty$ and $K_{gD} \to \infty$, then $\lambda \uparrow 0$ (so that $\lambda < 0$), $N_{MT} \to N_n$ and $m \to L$, as expected.

4.7 Discussion of the results

4.7.1 Comparison with existing cap models

The cap model presented in this paper differs from the approaches used in [2, 5] [22] [23]. First, we don’t postulate catastrophe frequency and growth velocity, or derive them from numerical simulations, as was done in [2, 5]. Instead, we analytically derive these macroscopic rates from small-scale rates (such as chemical rate constants). Second, we employ only spontaneous hydrolysis and don’t use induced, or vectorial, hydrolysis; both types of hydrolysis were used in [22, 23].
Figure 4.2: Frequency of catastrophe $K_{sT}^{\text{obs}}$ as a function of MT growth velocity $v_g$. Dots represent experimental data \cite{22,40}. Solid and dashed lines correspond to $\rho = 1/2$ and $\rho = 1$, respectively, and were obtained using the model (Section 4.4). Inset: Two upper curves on the right are $K_{gT}^{\text{eff}}/K_{gT}^{\text{eff}_C}$ and two lower curves on the right are $K_{sT}^{\text{eff}}/K_{gT}^{\text{eff}}$.

Our model agrees with the experimental data analyzed in \cite{22,23} as can be seen from the main panel in Fig. 4.2. Specifically, the predicted dependence of the catastrophe frequency on MT growth velocity is in agreement with experimental data.

The cap dynamics in the model proposed by Flyvbjerg et al. \cite{22,23} is modeled by addition of tubulin from the solution to the MT tip. This addition (polymerization) is faster than the propagation of the induced hydrolysis front (low end of the cap). Therefore, the cap length would grow infinite were it not for the spontaneous hydrolysis at some point inside the cap. When it occurs, the cap is redefined as an interval between this spontaneous hydrolysis point and the MT tip. In this way,
the average cap size can be kept constant at steady state. According to Flyvbjerg et al. [22, 23], catastrophe occurs when this cap is lost, and it is postulated that the remaining GTP-Tu units below the cap are not capable of rescuing the MT. This assumption is made in order to allow for catastrophe to occur. Otherwise, in many cases the rescue would immediately follow the cap loss. While this picture is widely accepted, in our alternative model the picture is even simpler. We use only one type of hydrolysis and we don’t need to make any additional assumptions. In our model, there is a hydrolysis front due to spontaneous hydrolysis of old enough units (see Section 4.4). The velocity of this front is governed by the age of the units inside the MT and hence it is always approximately equal (with fluctuations) to the growth velocity. Faster growth velocity leads to a longer cap, reducing the catastrophe frequency.

Dilution experiments have shown that sharp reduction in the concentration of free Tu to low or zero values results in collapse of the MTs after a certain delay. Importantly, this delay is practically independent of the initial free Tu concentration [41]. Flyvbjerg et al. explain this phenomenon by arguing that the dilution results in domination of spontaneous hydrolysis which regulates the waiting time before the collapse. Therefore, this time is almost independent of the initial cap size. Our model yields the following simple explanation. When concentration of free Tu becomes very low, two events must occur for the cap to disappear. The terminal unit should be lost (rate $K_{sT}$) and the next unit should hydrolyze (rate $K_h$). If this next to last unit is old enough to hydrolyze then, with high probability, the rest of the cap has already hydrolyzed.

Dilution experiments reported in [41] and cited in [22] determine the average waiting time before the catastrophe as roughly 5-10s. Therefore, in Fig. 4.2 we
set \( K_h = K_{sT} = 0.15 \text{s}^{-1} \). If the shortening velocity \( (K_{sD}) \) is much larger than \( K_{sT} \), and if the loss of the terminal unit is conceptualized as a two-stage process of hydrolysis and then falling, the equality \( K_h = K_{sT} \) would indicate that the hydrolysis rate of the terminal unit equals the hydrolysis rate of the internal unit. Notice that we were not able to fit the catastrophe frequency data if \( K_{sT} \) and \( K_h \) were significantly different. In the dilution experiments the spatial resolution was about 0.25\( \mu \text{m} \) \([11]\), which is about 30 heterodimers (per protofilament), so that actual waiting time before losing the terminal unit (heterodimer) might be faster than the reported waiting time before the collapse. We used Equation (4.1) with \( k_{gT} = 0.3 \mu \text{M}^{-1} \text{s}^{-1} \), however any value of \( k_{gT} \) can be used, because it enters the formulas only through \( K_{gT} = k_{gT} c \) and there is no explicit \( c \)-dependence in the figure. The unit length is taken to be the length of one heterodimer of Tu, 8nm and hence 1 \( \text{unit/s} \approx 0.5 \mu \text{m/min} \). Therefore \( v_g (\mu \text{m/min}) \approx K_{gT}^{\text{eff}} / 2 \text{(units/s)} \).

In the inset of Fig. [4.2] we show the values of \( K_{sT}^{\text{eff}} / K_{gT}^{\text{eff}} \) and \( K_{gT}^{\text{eff}} / K_{gT} \). It is seen that for \( v_g > 0.2 \) one has \( K_{sT}^{\text{obs}} < K_{sT}^{\text{eff}} \ll K_{gT}^{\text{eff}} / K_{gT} \) and hence \( v_g \propto K_{gT}^{\text{eff}} \propto c \).

### 4.7.2 Competition for tubulin and the edge effect

Here we study competition for a limited pool of free tubulin and combine it together with the cap model. At steady state, the dependence of the free tubulin concentration \( c \) on the total tubulin concentration, \( c_{\text{tot}} \), is governed by tubulin mass conservation Equation (4.23). The resulting value of \( c \), in turn, defines the dynamics of MTs and their ensemble distributions.

In the case of an unbounded spatial domain, we have reproduced the prediction of the Oosawa-Kasai model \([42]\) including the existence of a critical concentration of free tubulin \( c_{eq}^{\infty} \) (see thin lines in Fig. [4.3]). The steady state cannot exist above
Figure 4.3: Steady state concentration of free vs. total Tu in the unbounded (thin lines) and bounded (thick lines) domains. For a bounded domain maximal MT length \(L\) is roughly 1400 units. For a bounded domain \(c\) reaches its asymptotic value \(c^\infty_{eq}(K_h)\). Hydrolysis rate is \(K_h = 10 \text{s}^{-1}\). Full and dashed lines correspond to \(\rho = 1/2\) and \(\rho = 1\), respectively. Other parameters are \(N_n = 200, V = 10^{-17} \text{m}^3 = 10^{-8} \mu\text{L}, k_{gT} = 5 \mu\text{M}^{-1} \text{s}^{-1}, K_{sT} = 5 \text{s}^{-1}, k_{gD} = 0.5 \mu\text{M}^{-1} \text{s}^{-1}, K_{sD} = 500 \text{s}^{-1}, K_n = K_{gD}, K_e = K_{sT}\), and we use Equation (4.1).

this value. Depending on the choice of parameters, the transition from almost linear growth of \(c\), \(c \approx c_{tot}\) for low \(c_{tot}\), to the asymptotic value \(c = c^\infty_{eq}\) can be made sharp or smooth. (In the Oosawa-Kasai model, this transition is assumed to be sharp meaning that when \(c_{tot} < c^\infty_{eq}\) there are no MTs and when \(c_{tot} > c^\infty_{eq}\) all the excess tubulin above \(c^\infty_{eq}\) goes into polymerized state.) When the probability of rescue is zero then \(c^\infty_{eq} \rightarrow \infty\), our model describes the situation considered in [36].

If there is an upper bound on MT lengths, then at sufficiently high concen-
trations of total tubulin, this bound inhibits further polymerization of MTs (edge effect). Hence, the steady-state free tubulin concentration can rise above its critical value. This edge effect is demonstrated by the thick lines in Fig. 4.3 and was discussed in Gregoretti et al. [24]. Under our reference conditions, it is seen that for $c_{\text{tot}} > 20 \mu M$ the edge starts to play an important role. At sufficiently high $c_{\text{tot}}$, the edge reestablishes linear growth of $c$ with respect to $c_{\text{tot}}$. The implications of this effect are as follows. If $c_{\text{tot}}$ is high enough, the MTs grow persistently up until hitting the edge which triggers their catastrophe. This is consistent with recent experimental observations of persistent growth of MTs \textit{in vivo} [43].

From our model it follows that by changing the number of nucleation sites while keeping the total amount of Tu constant the cell could regulate the transition between mitotic (short) and interphase (long) MTs. As an example, assume $\rho = 1/2$, $c_{\text{tot}} = 20$ and the rest of the parameters, except $N_n$, as given in Fig. 4.3. Then $L \approx 1400$. For $N_n = 50$ there are a few long MTs with $m = 1100$ and $\lambda = -334$, while for $N_n = 500$ there are many short MTs with $m \approx \lambda = 149$. The physiological relevance of such a behavior is addressed in detail in our work [24].

Next we compare in Table 4.2 Monte Carlo simulations (see Section 4.9.1) with the results obtained by using our continuous model. We choose large domain size so that MTs never reach the boundary and $m = \lambda$ - cf. Equation (4.24). We run simulations for different parameter sets until steady state is reached and then determine free tubulin concentration, number of MTs and their mean length, and estimate the cap size. Recall that only MTs in growth phase have caps. Therefore, we estimate the cap size $\ell$ of MTs in our simulations by $\ell \sim (1 + K_{\text{eff}}^{\text{gT}}/K_{\text{sD}}) \cdot \#T/N_{MT}$, where $\#T$ is the number of polymerized Tu units in T state. We also use Equation (4.20). For the parameter values used in Table 4.2 simplified formula
### TABLE 4.2

**COMPARISON OF SIMULATED AND THEORETICAL RESULTS.**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Simulated results</th>
<th>Theoretical estimates: ( n = (\ell + 1)/2, \quad n = \ell ) (see text)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_h )</td>
<td>( c_{tot} )</td>
<td>( N_n )</td>
</tr>
<tr>
<td>( \infty )</td>
<td>36</td>
<td>2000</td>
</tr>
<tr>
<td>( \infty )</td>
<td>28</td>
<td>3000</td>
</tr>
<tr>
<td>( \infty )</td>
<td>28</td>
<td>( 10^5 )</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>30</td>
<td>2000</td>
</tr>
<tr>
<td>( 10 )</td>
<td>10</td>
<td>2000</td>
</tr>
<tr>
<td>( 10 )</td>
<td>5</td>
<td>2000</td>
</tr>
<tr>
<td>( 0.5 )</td>
<td>5</td>
<td>2000</td>
</tr>
<tr>
<td>( 0.03 )</td>
<td>2</td>
<td>2000</td>
</tr>
<tr>
<td>( 0.01 )</td>
<td>2</td>
<td>2000</td>
</tr>
</tbody>
</table>

All values reported here are at or very close to steady state. Cells containing two numbers show theoretical predictions for \( \rho = 1/2 \) and 1, respectively. Domain size is \( L_x = L_y = 10^{-4} \) and \( L_z = 10^{-7} \) (m). It is ensured that the domain is long enough so that the MTs don’t reach the boundary for the given parameters. Here \( k_gT = 5, k_gD = 0.5, K_sT = 500, K_{gT,gD} = k_{gT,gD}c \). For nucleation rate we use \( K_n = k_gD \), i.e., a D seed, except for the third line, where it is a T seed, \( K_n = k_gT \). \( \#T/N_n \) is an estimate for the cap size - see Section 4.7.2. Initially all Tu is free and its concentration is \( c_{tot} \). Notice that \( c < c_{eq}^\infty \).

Table 4.2 also contains theoretical estimates corresponding to the simulated values discussed above. In addition, the table includes theoretical estimates of \( c_{eq}^\infty \) and of the cap size based on Equation (4.13). In most cases the simulated results lie in between our two theoretical approximations, for \( \rho = 1/2 \) and \( \rho = 1 \) respectively, in agreement with the model description of cap evolution. These approximations are given as two adjacent numbers in the cells of the table displaying.
our theoretical estimates. When, however, $K_h$ becomes small, and $c$ approaches
Equation (4.21), our approximations seem to consistently overestimate the
number of MTs, $N_{MT}$. This should be improved by rescaling the rest of the rates,
$K_{gD}$, $K_{sD}$, $K_n$ and $K_e$, which is outside of the scope of this paper.

4.7.3 Non-steady-state phenomena

It is often mentioned in the literature that the steady state length distri-
bution of MTs observed in the experiments is either exponential or bell-shaped
8–10, 44–46. Exponential distribution agrees with our model. Inability to obtain
a bell-shaped distribution seems to indicate a limitation to our model. While we
do not exclude the possibility that some rates might depend on the MT length,
as proposed by 8, 9, or on time spent in a given phase 17, 18, we suggest two
alternative ways of obtaining bell-shaped distributions under certain conditions.
First, one should be careful in determining when the system reaches the steady
state in the experiment or simulation. As our simulations demonstrate, the sys-
tem reaches the constant free tubulin concentration and MTs reach constant mean
length relatively quickly. (The number of MTs does not change much henceforth.)
By that time the MT length histogram is often bell-shaped as illustrated in Fig-
ures 4.4 and 4.5. This can be explained as follows. When MTs start growing from
nucleation sites there is an excess of free tubulin. Therefore, the growth is origi-
nally unbounded leading to a Gaussian shape. If the cell edge (upper boundary)
is far away, in the course of this growth the free tubulin concentration drops and
reaches its steady state value. At this time the shape can still be close to a Gauss-
ian (cf. Figure 4 in 10). This is followed by a process of a shape change of the
MT length histogram with free tubulin being constant. Eventually, this results in
an exponential shape and in system reaching true steady state. The shape relaxation is relatively fast in Figure 4.4, while it is very slow in Figure 4.5. The shape relaxation occurs through diffusive exchange of polymer mass among the MTs, which might be orders of magnitude slower than initial rate of polymerization due to excess of free tubulin. This behavior is well described in [49, 50].

A bell-shaped distribution can be also obtained in a bounded domain when the steady state concentration of free tubulin is high enough for unbounded growth of MTs if it were not for a cell edge (i.e., if \( c > c_{eq}^\infty \)). In this situation we predict positive exponential distribution of MT lengths, in the case when all MTs can reach identical maximal length restricted by the edge, consistent with [11, 15]. As can be seen in experiments, however, MTs are curved and cell shape is not ideally spherical or circular, so that different MTs experience different restrictions (e.g., [43, 45]). This can lead to an MT length histogram of a bell-shaped form, in true steady state. Notice that in some simulations the free Tu concentration approaches its steady state value in a non-monotonous way—see Section 4.9.3.

4.8 Conclusions

In this paper we analyze a model of MT dynamics in a domain bounded by the cell edge which involves competition of individual MTs for tubulin. The model is based on a mesoscopic linear 1D approximation of a MT structure and includes finite hydrolysis of polymerized GTP-Tu units.

We start by deriving analytical formulas linking mesoscopic parameters (those describing the addition, loss, and hydrolysis of individual units) to the macroscopic characteristics of a system of dynamic microtubules. Specifically, we are able to predict all four dynamic instability parameters: velocities of growth and shorten-
Figure 4.4: Histograms of MT lengths after 80 (top) and 160 (bottom) seconds from the beginning of polymerization. The concentration of free Tu in both cases is close to the steady state value: 7.63 and 7.51 µM, respectively. Although the concentration has reached steady state after 80 sec, the length distribution of MTs is still changing. Here $K_h = 10s^{-1}$, $c_{tot} = 10\mu M$ and the rest of the parameters are specified in Table 4.2.
Figure 4.5: Histogram of MT lengths after 150 seconds from the beginning of polymerization. The concentration of free Tu is 1.382 \( \mu M \) and has reached its steady state value (to within random fluctuations). \( K_h = 0.1 \text{s}^{-1} \), \( c_{tot} = 3 \mu M \) and rest of the parameters are specified in Table 4.2. Note the bimodal character of the distribution.
ing and frequencies of catastrophe and rescue (see Section 4.4). We demonstrate how to recapitulate the macroscopic steady-state behavior of MTs using mesoscopic rates and, vice versa, extract mesoscopic rates from macroscopic behavior. Hence, it is possible to analytically and quantitatively predict the effect of changes in mesoscopic parameters on observable features, as well as to deduce mesoscopic changes from observed changes in macroscopic behavior, when relevant geometry and chemistry are taken into account.

The key ingredient in establishing a link between mesoparameters and macroparameters is the cap model, which allows one to replace the actual cap consisting of many units with an effective single unit. We demonstrate that the cap model behavior agrees with experiments measuring catastrophe frequency as a function of free tubulin concentration as well as with dilution experiments.

The model yields the following additional new results. For a cell of a given size, number of MT nuclei, and amount of total tubulin, we analyze conservation of tubulin mass equations (4.22, 4.23) governing the balance between polymerized and free tubulin. This results in description of partitioning of tubulin, distribution of MT lengths, and macroscopic rates of the dynamic instability. We also demonstrate that, by restricting the growth of MTs, the edge can raise the free Tu concentration above its critical value for an unbounded domain, leading to a persistent MT growth inside the cell (in agreement with numerical simulations [24]). Also, an increase in nucleation activity results in an increase of the number of MTs and a decrease in their average length, reducing the edge effect. Thus, by regulating the nucleation activity, the cell can transition between interphase and mitotic arrays of MTs. These predictions can be experimentally verified.

Last, we show a very good agreement between our Monte Carlo simulations and
analytical results. We also use the Monte Carlo model to provide an explanation for the non-exponential MT length distributions observed in experiments. This might be happening because of these distributions not having enough time for relaxing to an exponential shape characteristic of a steady state (see also [49, 50]).

We are currently working on incorporating a molecular level model describing individual MT protofilament structure into a unified multiscale model for investigating effects of various microtubule associated proteins (MAPs) on dynamic instability. It might be feasible to go first from the molecular to the mesoscopic level, either numerically or, perhaps, even analytically. Then, using the results of the present paper, one could go from mesoscopic to macroscopic scale and compare multiscale model predictions with experimental data at the macroscopic scale.

4.9 Additional technical details

4.9.1 Computational model

In what follows we provide a short description of the numerical algorithm used for our simulations, which differs slightly from that used in [24]. At time zero, the MTs begin to grow from the nucleation seeds. At each simulation step, the time of this step is calculated by defining the rates of change in length (either growth or shortening) for each MT. If the MT tip is in T (D) state, then this rate of change is $K = K_{gT(gD)} + K_{sT(sD)}$. We demand that the maximal average number of changes for each MT will be 1, which means that we find a maximal value among all $K$ and then set $\Delta t = 1 / \max_{MTs} \{K\}$ (a technical note – we set $\Delta t$ slightly below this value because the rate $K$ for a given MT can change slightly as $c$ is affected after each MT changes its length; these changes in $c$ are usually very small). Then, in general, each MT will have a chance to grow, shorten, or retain...
its length during $\Delta t$. We do not allow a distribution of the possible number of length changes for a MT during $\Delta t$ (only zero or one change is possible). After the length of each MT is updated, we update $c$ accordingly. After updating the lengths of all MTs, the hydrolysis cycle runs through all internal units of all MTs. The probability that a unit will hydrolyze during time $\Delta t$ is taken as $1 - e^{-K_h \Delta t}$, assuming Poisson statistics.

All MTs have their first unit in the D state and this unit cannot be lost –this constitutes a simple nucleation seed with lower growth probability than when the MT tip is in the T state (these units are not counted when calculating the lengths of the MTs). Similarly, when the edge is relevant we can assume $K_e = K_{sT}$. Such choices are made purely to reduce the number of parameters in the system and are not essential for our purposes.

4.9.2 Explicit solution in the bounded domain

In what follows we obtain expressions for $N_{MT}$ and $m$ in the bounded domain at the steady state. Notice that in our model all the MTs of the maximum length $L$ are technically in the growing phase, because their terminal unit can never become internal and therefore does not hydrolyze. (These MTs cannot grow because of the edge.) The edge-induced catastrophe rate $K_e$ will be governed by the smallest of the rates $K_{sT}$, $K_h$. The discrete version of Equations (4.15) and (4.16) determining a steady state at the boundary

\begin{align*}
0 &= -K_e M_g(L) + K_{gD} M_s(L) + K^eff gT M_g(L) - 1, \quad (4.31) \\
0 &= K_e M_g(L) - K_{gD} M_s(L) - K_{sD} M_s(L).
\end{align*}

94
After summing up these two equations we recover

\[ M_s(L - 1) = (K^{eff}_{gf}/K_{sD})M_g(L - 1), \]  

(4.33)

which is already known –cf. Equation (4.20)– and so one of these equations is superfluous. Another way to find \( M_{g,s} \) is to write the general solution of Equations (4.15) and (4.16), 

\[ M_g = Ae^{-z/\lambda} + B, \quad M_s = (K^{eff}_{gf}/K_{sD})Ae^{-z/\lambda} + (K^{obs}_{gf}/K_{gD})B. \]

and plug it into Equation (4.33), yielding \( B = 0 \), unless \((K^{obs}_{gf}K_{sD})/(K^{eff}_{gf}K_{gD}) = 1\) in which case \( B \) is arbitrary. But this last condition implies \( \lambda \to \infty \) and hence \( M_{g,s} \) are constant inside the domain. At the lower domain boundary, Equation (4.25) still holds.

The number of MTs is given now by

\[ N_{MT} = \sum_{l=1}^{L-1} (M_g(l) + M_s(l)) + M_g(L) = N_n - N_0. \]  

(4.34)

Using Equations (4.17), (4.18), (4.25) and (4.31) and replacing the summation by the integration from 0 to \( L \) we can determine \( A \) and hence \( N_{MT} \), which is given in Equation (4.29).

Similarly, mean MT length is

\[ m = \frac{\sum_{l=1}^{L-1} l(M_g(l) + M_s(l)) + LM_g(L)}{\sum_{l=1}^{L-1} (M_g(l) + M_s(l)) + M_g(L)} \]  

(4.35)

leading to Equation (4.30).
Small oscillations in the amount of polymerized Tu observed in numerical simulations for $K_h = 0.1 s^{-1}$, $c_{tot} = 50 \mu M$ and $N_n = 10^4$. Other parameters are as in Table 4.2. Inset is a blowup of the region of oscillations.

4.9.3 Oscillations

In some of the simulations we observed an overshoot in free Tu concentration before the steady state was reached. Moreover, in some cases there were slight oscillations of free Tu concentration, as shown in Figure 4.6. Overshoots and large oscillations have been reported and modeled in the literature [9, 12, 30, 51–55]. It is believed that slow conversion of D- into T-tubulin in solution, after the depolymerization, is the key to understanding of such oscillations. These oscillations occur if the initial free Tu concentration is sufficiently large.

Free D-tubulin cannot polymerize. In our model we assume its conversion into T-tubulin to be instantaneous. We also use a linear (and not higher order) dependence of the nucleation rate of the free Tu concentration and a fixed number of nucleation sites. It is remarkable that under these restrictive assumptions the
model produced some oscillations. We suggest the following explanation for their existence, which is in agreement with [53]. If the hydrolysis is slow enough, the MTs grow quickly in the beginning, resulting in a large cap. When the free Tu concentration changes quickly, the cap needs a relatively long time to adjust. This leads to a delayed response and possibly to oscillations. Hence it might be that the ability to produce oscillations is inherent to MT structure and that it can be magnified under certain experimental conditions.

4.10 References


CHAPTER 5

DEPENDENCE OF RESCUES AND CATASTROPHES ON MICROSCALE EVENTS NEAR THE MICROTUBULE TIP

5.1 Abstract

Dynamic\(^1\) instability is fundamental to microtubule function, but little is known about the microscopic events that drive the transitions between growth and depolymerization. Technical challenges have restricted experimental approaches, and the impact of computational models has been limited by structurally inaccurate assumptions or computational intensiveness. To address these problems, we have constructed a computational model of microtubule assembly that takes into account the detailed structure of microtubules, is based on a minimal number of assumptions, and exhibits dynamic instability similar to that observed experimentally. In particular, the model explicitly considers the formation and breakage of lateral bonds between protofilaments, and operates over time scales similar to those used in experiments. We validate the model by showing that it reproduces the experimentally observed behavior in sudden dilution experiments, and use it to investigate the processes of catastrophe and rescue and their relation to the structure of the microtubule tip. Our analysis suggests that the GTP cap has

\(^1\)This thesis chapter has been submitted for publication as a research article. IV Gregoretti, G Margolin, TM Cickovski, HV Goodson and MS Alber; Dependence of rescues and catastrophes on microscale events near the microtubule tip, submitted 2007. IVG and GM contributed equally to this work. HVG and MSA co-directed this work.
no easily-defined structure under conditions of dynamic instability and that dy-
namic interplay between the length of the stabilizing cap and the depth of the
inter-protofilament cracks governs both catastrophe and rescue.

5.2 Introduction

Microtubules (MTs) are long proteinacious tubular polymers found in all eu-
karyotes. MTs act as tracks for vesicle transport, segregate the chromosomes
during cell division and help to establish cell polarity. A key property of MTs
necessary for these activities is that they are highly dynamic: individual MTs
transition frequently between phases of elongation and shortening. This behavior
is termed dynamic instability, and it is observed both in vivo and in vitro [1]. The
resulting length fluctuations allow the MTs to explore space and respond rapidly
to both local and global signals [2]. The transitions from growth to shortening
and vise versa are known as catastrophe and rescue respectively. Elongation is
achieved by incorporation of new subunits while shortening occurs by subunit
detachment. Both processes occur exclusively at the MT tip.

5.2.1 MT structure and the origin of dynamic instability

MTs are noncovalent polymers of the protein tubulin and typically consist
of 13 parallel protofilaments arranged in a hollow tube. Each protofilament is
composed of a linear chain of alpha-beta tubulin heterodimers, resulting in an
\(-(\text{alpha-beta})_n-\) chain configuration with the so called plus end exposing the
beta monomer. The minus end is usually bound to the nucleation site and is less
dynamic. The subunits in the protofilaments are arranged in a B lattice (alpha
monomers laterally bind alpha monomers and beta bind beta), except at the
Figure 5.1: Microtubule structure and its modeled processes. A: tubulin heterodimers are the building blocks of MTs, which are made of 13 protofilaments arranged into a hollow tube with a seam. Each heterodimer consists of one alpha tubulin and one beta tubulin monomer. The beta subunits bind GTP and eventually hydrolyze it to GDP. We symbolize those states as GTP-Tu and GDP-Tu respectively. Note that MT ends are not as blunt and neatly organized as depicted here. B: individual processes explicitly treated in our model.

seam, where there is a helical shift of three monomers between the first and last protofilaments, resulting in an A lattice (alpha monomers bind laterally to beta) (Figure 5.1A) [3]. Structural studies suggest that the longitudinal bonds between the heterodimers are significantly stronger than the lateral contacts [3, 4].

Dynamic instability originates in conformational changes that occur in the tubulin heterodimers after polymerization. Tubulin subunits bind the nucleotide GTP. Upon polymerization, this GTP is hydrolyzed to GDP after a short delay. This delay is thought to result in formation of a GTP cap, which predisposes the MT to growth. The idea is that if this cap is lost (via GTP hydrolysis or other mechanism), the exposed GDP tubulin subunits rapidly depolymerize (undergo catastrophe) [5, 6] (note that GDP cannot be exchanged for GTP until a sub-
More recently this explanation has been modified to include the idea that the conformation of the tubulin subunits in the cap may be more important than the state of the bound nucleotide [7–10]. While there is general agreement about the existence of some kind of stabilizing cap, the characteristics of the cap and the mechanisms of the catastrophe and rescue transitions remain unclear [11–15].

Due to its involvement in many cellular processes, the MT cytoskeleton is a subject of intense study. The complexities of MT structure and behavior pose a challenge to experimentalists trying to understand the origins of catastrophe and rescue using traditional cell biological and biochemical techniques. There has been remarkable progress in the field of MT structural biology (for example, see review [16]) but its contribution to understanding MT dynamics has been limited by the static nature of the techniques. The sum of these difficulties has resulted in a situation where little is known about the mechanistic differences between growing and shortening MTs or the structural changes driving catastrophe and rescue. To address these and other issues related to MT dynamics we have constructed a dimer-scale kinetic Monte Carlo model of a single MT which is based on the structural and biochemical attributes described above and has a minimal number of assumptions.

The most important distinguishing feature of our model in comparison with previous dimer-scale models [4, 17–19], is its explicit treatment of lateral bonds. Specifically, chains of bonds between two neighboring protofilaments grow and shorten (due to the formation and breakage of bonds between individual subunits) just as the protofilaments themselves can grow and shorten. This allows for the formation of frayed ends and curled protofilaments observed experimentally (the so
called “ram’s horns” [20]). Note that though our modeled MTs are drawn as planar representations, each one is a lattice bent on itself in 3D, forming a tube with a seam; unbonded regions of GDP-Tubulin protofilaments should be interpreted as curving out. This feature is important because these curled, laterally unbonded protofilaments have significant influence on microscopic processes at MT tips, particularly rescue. In addition, we incorporate the left-right asymmetry of lateral interactions observed in analysis of the MT structure [21].

VanBuren et al. [4] recognized the importance of modeling ram’s horns and developed a mechanochemical model [22] based on elastic energy considerations between neighboring dimers. In our model, the impact of mechanical influences (such as the energetic barriers to forming lateral bonds between curling protofilaments) is incorporated into the kinetic rate constants governing the different dimer-scale events. The main advantage of using the kinetic approach relative to the mechanochemical one is that the kinetic approach is orders of magnitude faster, allowing simulation of experimentally relevant time spans (tens of minutes), which in turn allows observation of both catastrophies and rescues and generation of experiment-like “life history” plots (needed for analysis of dynamic instability parameters).

As shown below, the model recapitulates dynamic instability. We further validate it by showing that it reproduces the surprising relationship between growth rate and time before catastrophe observed in sudden dilution experiments. This suggests that a simple GTP cap model (one in which GTP hydrolysis is a stochastic first order process) can account for experimentally observed behavior. We use the model to investigate the microscopic characteristics of the tips of growing and shrinking MTs and the mechanisms of catastrophe and rescue. Analysis of
these data suggests that the GTP has no easily-defined structure, and that dy-
namic interplay between the length of the stabilizing cap and the depth of the
inter-protofilament cracks governs both catastrophe and rescue.

5.3 Results and Discussion

5.3.1 Overview of the model

Briefly, we treat the MT as a lattice bent on itself forming a tube with a
seam (Figure 5.1A); normally the MT has 13 protofilaments, but this can be
adjusted. The MT subunits (tubulin heterodimers) have two states, one prone
to polymerization and the other prone to disassembly; these states are denoted
as GTP-Tu and GDP-Tu, but they could represent other conformational states.
Finally, the model is based on five events: protofilament growth, protofilament
shortening, inter-protofilament bond growth, inter-protofilament bond shortening,
and the transition of GTP-Tu to GDP-Tu, i.e. GTPase activity. The GTPase
is modeled as a simple first-order event that occurs only on internal subunits,
consistent with structural evidence that alpha monomers act as GAPs (GTPase
activating proteins) for beta monomers [23]. These events occur at the level of
individual dimers and are governed by kinetic rate constants that are input by the
user. Unless otherwise indicated, the simulations presented here use a common set
of reference parameters chosen to be consistent with existing understanding and
to give rise to transition frequencies that are in agreement with those observed
in experiment. We emphasize that this parameter set is not unique in its ability
to display dynamic instability, as would be expected from the observation that
tubulins from different organisms exhibit significant biochemical differences. A
sensitivity analysis is also provided to test the model’s design robustness.
5.3.2 Recapitulation of dynamic instability

A typical life history plot, obtained for the reference set of parameter values (see Methods, subsection 5.5.4), is shown in Figure 5.2. The plot shows clear growth and shortening phases, as well as catastrophes and rescues (specific examples are indicated by points A-D respectively). A decrease in growth rate around the time of 400 a.u. (arbitrary unit of time, see below) may represent a pause. To calculate the dynamic instability parameters, 10 simulations were run under identical conditions (the reference parameters) for $10^7$ steps each. From the resulting life-history plots the following values were extracted: growth velocity $v_g = 5.27 \pm 0.09$, shortening velocity $v_s = 11.8 \pm 0.3$, catastrophe frequency $f_{cat} = 0.0069 \pm 0.0008$ and rescue frequency $f_{res} = 0.0016 \pm 0.0007$. These velocities are given in dimer lengths/a.u. and the frequencies in inverse a.u.. If one a.u. is set to 1 second and the dimer length is set at the experimentally measured value of 8nm, these values are consistent with \textit{in vitro} observations \cite{24, 26}.

5.3.3 Sensitivity analysis

With the intention of helping the reader relate the results in different sections to each other, we performed our simulations with the parameter values fixed at our reference set unless stated otherwise. In this section we show that the ability of our model to recapitulate dynamic instability depends on its design rather than a particular set of parameter values. In our model the adjustable parameters are the rate constants for all processes: growth, shortening, bonding, breaking and hydrolysis, for all combinations of nucleotide states and positions (seam or regular bond). The number of protofilaments per MT and the concentration of tubulin can also be set by the user although these are not parameters \textit{per se}.
Figure 5.2: Macroscopic view of dynamic instability recapitulated with our model. The length of the MT is expressed in number of dimers and the time in arbitrary units. The figure shows 26 -mostly overlapping- curves. The length of the 13 protofilaments are represented by the blue curves. The length of the interprotofilament bonds are in black. The black curves are largely invisible because they are drawn under the overlapping blue curves. Microscopic snapshots of the simulation at points A-D are shown in Figures 5.4, 5.6 and 5.7. This figure shows the first $10^6$ steps of a simulation $10^7$ steps long. For comparison to experimental data assume that 1 a.u. = 1 second.
We took the reference parameters set and varied all rate constants randomly by ±10% generating 73 new sets. Then we used these sets for running simulations for $10^6$ steps and graphed the outcome to observe dynamic instability. We classified the outcomes into three categories: (i) persistent growth without catastrophe, (ii) conventional dynamic instability, and (iii) inability to grow persistently. We found that among those 73 randomly generated sets, 12 fell into category (i), 56 into (ii) and 5 into (iii). Because the probability of dimer attachment to the protofilament tip depends directly on the concentration of free tubulin dimers, we realized that the sets that grew permanently did so mostly because their random value of protofilament growth kinetic constant was too high for our standard 14µM free tubulin concentration. We then randomly picked 2 of the sets in category (i) and re-ran them at 9µM recovering dynamic instability (Figure 5.3 A, B shows one of those 2 sets). Note that although the reader unfamiliar with MT dynamics would expect little qualitative change upon small tubulin concentration changes, we have shown with a mesoscopic model that it has a profound effect [26, 37]. The explanation for the 5 sets that did not grow persistently lies in the design of the model itself. Basically, it is a two-states model in which lateral bonds between two GTP-Tu dimers must be more stable than bonds between two GDP-Tu dimers. We observed that all sets in (iii) had parameter values that made the bonding between two GTP-Tu less stable than between GDP-Tu. We then randomly picked 2 of these sets and decreased only the probability of lateral breakage between GTP-Tu, resulting in the recovery of dynamic instability. Figure 5.3 C, D shows one of the sets recovering dynamic instability upon decreasing only its breakage kinetic constant from 31.1 to 24.8. Although not in systematic fashion, the model was run with other parameters sets also showing dynamic instability. In particular, when
the parameters were matched as closely as possible\(^2\) to those of the kinetic model of VanBuren et al. \([4]\), we obtained length history plots similar to that showed in their article. We conclude that the model is robust.

In what follows, we use the simulated MTs to gain insight into the structure of the tip during growth, shortening, and the catastrophe and rescue transitions. In interpreting the images of the tips, it is important to note that although the figures display the MT as a flat object, the model’s MT representation itself is tubular. The last and first protofilament are connected via the seam bond, which is indicated to the right of the rightmost protofilament, with a shift of 1.5 dimers.

5.3.4 Structure of the MT tip during growth and shortening phases

As noted above, growing MTs are thought to have a stabilizing cap consisting of a region near the tip that has high concentration of GTP-containing heterodimers. In addition, one would expect that a large fraction of these GTP heterodimers will be laterally bonded \([5]\). Our simulations are consistent with these ideas, as seen in Figure 5.4. This figure further suggests that the primary difference between growth and shortening is the following: in depolymerizing MTs, the cracks between the laterally unbonded protofilaments extend from the tip down into the GDP rich region (Figure 5.4B), while in growing MTs, these cracks terminate in the GTP-rich region and contact few GDP heterodimers.

The conclusions derived from examination of single simulation snapshots are supported by quantitative analysis of the MT tip, shown in Figure 5.5. For instance, during the first growth phase (times from 0 through 135a.u.) the average length of the GTP cap (as approximated by the distance of the center of mass of the GTP-Tu from the tip) is approximately 25 dimers. In contrast, the laterally

---

\(^2\)Not all the parameters can be matched because the models are different.
Figure 5.3: Sensitivity analysis for testing model robustness. Based on the reference set of parameter values (see subsection 5.5.4) used throughout this paper, 73 new parameters sets were generated by assigning values randomly ±10% around the reference. The parameters are the rate constants for all processes (see text for more details). Tubulin concentration is not a parameter. We use 14µM as our standard. A, example of a parameters set that instead of showing dynamic instability it grows permanently at 14µM. B, the same set as in A but displaying dynamic instability now that the concentration has been decreased to 9µM. C, example of parameters set unable to grow persistently. D, same set as in C but with a less unstable GTP-Tu/GTP-Tu bond now displays dynamic instability. 56 out of the 73 sets showed dynamic instability without any alteration.
Figure 5.4: Sample close-ups of the structure of the MT tip during growth and shortening (panels A and B respectively; which corresponds to points A and C in Figure 5.2). Each column represents a protofilament. Red and green symbolize respectively GTP-Tu and GDP-Tu heterodimers. Small white rectangles between protofilaments symbolize the lateral bonds between two heterodimers. The image is a planar representation of the tubular polymer; the first and last columns appear distant in this representation but are actually in contact. The white rectangles at the right of the rightmost column represents the seam bond. In addition, the model features a 1.5 dimer shift between the first and last columns but for clarity this is not shown in the figure. Only the top portion of the MT is displayed. The dimers are stretched in horizontal direction, in comparison to their actual proportions.
bonded GTP cap length (see Methods for means of estimation) is \( \sim 18 \) dimers. This means that the cracks are \( \sim 7 \) dimers deep and that the GTP-rich region of the MT extends roughly 18 dimers below these destabilizing cracks. This in turn means that the protofilaments are likely to keep their cohesiveness.

During shortening, the situation is the opposite: the number of GTPs at the tip is small (a few dimers) (Figure 5.5, panel C), and the cracks extend through these GTP subunits into the GDP-rich region (indicated by the negative values for the laterally bonded GTP cap length, panel D). The crack at the seam is even bigger. The numerical values for the crack depths are parameter dependent, but the greater average depth of the seam crack is expected from the weaker nature of lateral bonds at the seam [3]. These characteristics result in formation of a frayed end with a GDP rich tip and outwardly curled protofilaments which favors continuation of the shortening phase. An additional characteristic to note is that in the model, depolymerizing MTs typically have a small number of newly polymerized GTP-Tu heterodimers at the tip. However, these GTP-Tu are unable to laterally bind to each other because cracks between the protofilaments (similar to the experimentally observed rams horns) keep them apart [20].

5.3.5 Structure of the MT tip during phase transitions

Points B and D in Figure 5.2 indicate examples of catastrophe and rescue. Microscopic views of the MT at these points are shown in Figures 5.6 and 5.7.

*Mechanism of catastrophe (Figure 5.6).* The number of GTP-Tu subunits at the tip fluctuates randomly during the growth phase. Occasionally, the number of GTP-Tu subunits becomes very low. Our analysis suggests that catastrophe becomes likely when the cracks (regions of laterally unbonded protofilaments)
Figure 5.5: Statistical analysis of the dynamic MT end as a function of time. A: average GTP-Tu position (green) and average MT length (red). As expected, the GTP-Tu dimers are localized around the tip and then their positions follow closely the position of the MT end (compare to Figure 5.2 which shows the same stochastic realization). B: number of GTP-Tu dimers in the whole MT. C: GTP-Tu cap size, estimated as the distance of the center of mass of the GTP-Tu from the average MT length. The blue bold curve is the average of the 13 protofilaments, the black curves are the average ± standard deviation. D: Length of the laterally bonded GTP cap for regular bonds (magenta) and for the seam (cyan) –see Methods for details. All lengths are in dimers.
extend into regions rich in GDP-Tu. This shifts the architecture of the tip from a growth prone structurally cohesive tubule to a shortening prone frayed end. First, the laterally unbonded parts of the adjacent protofilaments are more likely to depolymerize (see Figure 5.6 D for a planar representation). Subsequently, their neighbors loose more lateral bonds and the process repeats (Figure 5.6 E-H).

Mechanism of rescue. Figure 5.7 shows the progression of a rescue from the cessation of depolymerization to the beginning of growth. As noted above, an important aspect of depolymerizing MTs is that cracks between protofilaments typically extend from the tip down into the GDP rich region (see Figure 5.7 B for a planar representation). However, on occasion a crack disappears, resulting in the lateral bond extending to the end of a protofilament (e.g., laterally bonded GDP-Tu could be exposed when pieces of protofilament are lost). This sets the stage for a rescue because when new GTP-Tu dimers bind on top of this protofilament, they can potentially establish a lateral bond. Our analysis suggests that if this sequence of events occurs approximately simultaneously on several protofilaments, a new GTP-Tu cap is likely to form, resulting in a rescue. Note that in our model we use asymmetric GTP-Tu|GDP-Tu vs GDP-Tu|GTP-Tu lateral bond strengths. This stipulation is based on experimental evidence showing that the changes in lateral contacts induced by GTP hydrolysis are different [21] (see Methods for further details).

Examination of Figure 5.5 A (or Figure 5.2) reveals an event during which the GTP cap was partially lost and the MT was close to a catastrophe: around 400 a.u. of time both the size of the GTP cap and the number of GTP subunits dropped to about half of the normal growth phase values (Figure 5.5 panels B and C). This event may correspond to a pause. It has been suggested that pauses are
Figure 5.6: Sample close-ups showing progression of a catastrophe from cessation of growth (A) to the beginning of shortening (H). The microscopic snapshots in this series are approximately equidistant in time (1000 steps apart) and correspond to point B of Figure 5.2, from around 135 to 140 a.u. of time. Only the top portion of the MT is shown. For color codes see Figure 5.4.
Figure 5.7: Sample close-ups showing progression of a rescue event from its onset (A) to the beginning of a growth phase (H). The microscopic snapshots in this series are approximately equidistant in time (1000 steps apart) and correspond to point D of Figure 5.2, from around 163 to 168 a.u. of time. Only the top portion of the MT is shown. For color codes see Figure 5.4.
the outcome of the activity of MT associated proteins or defects in microtubule structure. While we have never seen extended (i.e., order of a minute) pauses in our simulations, the observation of short pause-like events in multiple simulations (not shown) suggests that at least some of the occasional pauses observed in vitro are expected outcomes of normal microtubule polymerization processes. The idea that pauses may be due to partial losses of GTP caps was also suggested by VanBuren et al. [22].

One might expect that characteristics such as tip raggedness or crack depth at the seam would be predictive of catastrophe and/or rescue. Examination of tip behavior at high spatial and temporal resolution (Figure 5.8) shows that there is no obvious correlation between these aspects of the MT tip and the likelihood of phase transition. This lack of correlation is initially surprising, but it is justified by the observation that the tip structure fluctuates at a rate that far exceeds the frequencies of the phase transitions. However, these issues may warrant further investigation.

The GTP cap. These considerations raise the issue of the nature of the GTP cap. Statistical arguments based on MT behavior in the presence of GMPCPP (a non-hyrolizable GTP analog) have suggested that a single layer of GTP tubulin is sufficient to cap MT ends [12, 27]. These observations together with difficulties in experimental detection of the cap [12, 28, 29] have led to the idea that the cap consists of single layer of GTP-Tu (e.g., [18, 19, 29]). While a single layer of laterally bonded GTP-Tu may be sufficient to promote growth, the behavior of the model presented here indicates that the MT tip rarely if ever corresponds to a single ring. Instead, the model suggests that the cap is a constantly changing structure that typically contains many GTP-Tu. However, the precise extent of
Figure 5.8: Detailed analysis of the architecture of the MT tip during catastrophe and rescue. The arrival of the phase transitions seems to occur without an advance warning. Panel A presents in its upper sub-panel a life history plot (as in Figure 5.2) matched to properties (see below) that characterize the MT tip architecture (lower sub-panels). Panels B, C, and D show respectively in higher detail the first catastrophe (as is Figure 5.6), the first rescue (as in Figure 5.7) and the second catastrophe. MT length is represented by the length of each of the 13 protofilaments (blue) and their average (red). $\sigma_{\text{MT length}}$ is the standard deviation of the length of the 13 protofilaments. $\sigma_{\text{MT length}}$ is a measure of MT tip raggedness. $\Delta_{\text{reg}}$ and $\Delta_{\text{seam}}$ are the MT length minus the averaged length of 12 regular bonds and minus the length of the seam, respectively. They show how much the bonded parts of the protofilaments lag behind the average MT length. The lowest sub-panels show the true interprotofilament cracks defined as lengths of laterally unbonded parts of the shorter one among each pair of adjacent protofilaments. This is shown in cyan for the seam and in magenta for the 12 regular cracks (average and ± standard deviation in solid and dotted lines respectively). For all vertical axes, the units are dimer lengths. Time is expressed in arbitrary units (see text).
the functionally significant cap (i.e., the cap involved in suppressing catastrophe) is difficult to define because some GTP-Tu in the outer cap are likely not yet laterally bonded and those of the inner cap are intermixed with GDP-Tu.

With our reference parameter set, the cap commonly contains \(~400\) GTP subunits. This value varies with tubulin concentration and other parameters (see \([5.9]\), but is similar in magnitude to the value measured by Melki et al. \([11]\) (\(~170\) GTP-Tu subunits). Our observation of a large GTP cap seems at odds with experiments that detect little/no GTP in dynamic MT, implying that the cap is small if it exists at all. We concur with \([11]\) in suggesting that experimental difficulties in detecting a GTP cap result from two problems: first, even an extended cap (e.g., a depth of 10-20 dimers) is a tiny fraction of the length of most experimentally generated MT, creating signal/noise problems; second, the GTP in the cap has such a short half life that all GTPs except those in the terminal subunits are lost in the dead time of most experiments (if we assume that 1 a.u. = 1 sec, the GTP-Tu half life is 5 seconds, within the range measured experimentally \([11]\) and used in modeling \([4, 22]\)). However, it is also important to note that in our model GTP-Tu symbolizes a growth-prone conformation. It is possible that this conformation persists beyond the point of phosphate release.

5.3.6 Sudden dilution experiments

To further test the physiological significance of our model, we subjected the simulated MTs to sudden dilution experiments. In the dilution simulations the strength of the lateral bond between two GTP-Tu dimers was increased in comparison with the value used in the reference parameter set, to obtain larger variations in \(v_g\). One might expect that faster growing MTs would have longer GTP caps,
Figure 5.9: Statistical analysis of the dynamic MT end as a function of time for $[\text{Tu}] = 3.5\mu\text{M}$. This Figure is analogue to Figure 5.5 that was obtained at $[\text{Tu}] = 14\mu\text{M}$.
and that MTs with longer GTP caps would survive longer after being suddenly deprived of new subunits by sudden dilution. However, experiments show that in real MTs, there is little correlation between initial growth rate and the time to catastrophe [34]. These observations have been used to argue against the existence of a GTP cap [34]. Remarkably, the simulated MTs also show little correlation between initial growth rate and the time to catastrophe (Figure 5.10), even though faster growing MTs do have larger GTP caps—compare Figure 5.5 ($v_g = 5.27$) to Figure 5.9 ($v_g = 2.25$). We conclude that the processes ongoing in our simulated MTs are useful reflections of processes occurring in real MTs. We also conclude that, contrary to some suggestions, the results of sudden dilution experiments are consistent with idea that GTP hydrolysis occurs by a simple stochastic first order process. There is no need to include vectorial hydrolysis as has been previously suggested from mathematical analysis [35] or postulated on the basis of a tube-closure mechanism [36]. Similar conclusions about the sufficiency of a simple stochastic hydrolysis mechanism were obtained by VanBuren et al. [4] and in our analytical study of MT dynamics [37], which also suggested that that delay time is proportional to the inverse of $K_{hydrolysis}$.

5.4 Conclusions

We have developed a computational model of MT dynamics that is structurally detailed, based on experimentally-derived characteristics, and runs fast enough to simulate typical in vitro dynamic instability experiments (tens of minutes). This model reproduces the expected microtubule behavior as observed in life history plots and approximates experimental measurements of growth rates, depolymerization rates, transition frequencies, and time to catastrophe observed in sudden
Figure 5.10: Life history plot close-ups of the MT tip under sudden dimer dilution. Time (a.u.) and length (dimers) are relative to the moment when the dimer concentration is dropped to zero and to the pre-dilution MT length. The red bold curve represents the MT length as an average of all protofilaments (blue). Bond lengths are in black. A and B are examples of results at $k_{\text{hydrolysis}}=0.2$ with $v_g=2.9$ and $v_g=17$ respectively. C and D show results at $k_{\text{hydrolysis}}=0.1$ with $v_g=1.4$ and $v_g=15$ respectively. A statistical assessment of these results is provided in Table 5.1.
TABLE 5.1

SUDDEN DILUTION EXPERIMENTS PERFORMED WITH OUR MODEL.

<table>
<thead>
<tr>
<th>$k_{\text{hydrolysis}}$ (inverse time a.u.)</th>
<th>Pre-dilution [Tu] (μM)</th>
<th>Pre-dilution $v_g$ (dimers/time a.u.)</th>
<th>Delay time (a.u.)</th>
<th>Δlength 10 dimers</th>
<th>Δlength 30 dimers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>2</td>
<td>2.9</td>
<td></td>
<td>3.6 ± 0.9</td>
<td>5.2 ± 0.9</td>
</tr>
<tr>
<td>0.2</td>
<td>16</td>
<td>17</td>
<td></td>
<td>4.2 ± 0.7</td>
<td>7.4 ± 0.8</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>1.4</td>
<td></td>
<td>5.1 ± 1.3</td>
<td>7.0 ± 1.2</td>
</tr>
<tr>
<td>0.1</td>
<td>1.5</td>
<td>2.3</td>
<td></td>
<td>6.8 ± 1.0</td>
<td>9.2 ± 1.0</td>
</tr>
<tr>
<td>0.1</td>
<td>10</td>
<td>15</td>
<td></td>
<td>7.7 ± 0.9</td>
<td>12.2 ± 0.9</td>
</tr>
</tbody>
</table>

The results are expressed as means and standard deviations of 10 simulations. Time is in arbitrary units (a.u.) and lengths in dimers (∼8 nm). The delay times are shown in two columns for two arbitrary decreases in MT length (10 and 30 dimers). Thirty dimers is the approximate resolution of the light microscope. Close-ups of individual MT ends are shown in Figure 5.10. The table indicates that the delay times in our model are mostly independent of pre-dilution tubulin concentration and roughly equal to the inverse of the hydrolysis rate.
dilution experiments. In agreement with recent experimental work, our model indicates that MT growth at the nanoscale is highly variable and that the growth phase contains brief shortening excursions [30].

This model is firmly rooted in previous modeling efforts from a variety of research groups, but it is unique in its combined structural foundation, explicit treatment of lateral bonds between protofilaments, and rapid simulation speed. For example, the lateral cap model of [18] produces dynamic instability, but requires both non-physiological diagonal dimer-dimer interactions and instantaneous hydrolysis. Our model has several similarities to the pseudo-mechanical model of VanBuren et al. [4], but unlike this work we explicitly model lateral bonds, producing transition frequencies similar to those observed experimentally without the need for arbitrary rules to approximate the effects of lateral bonds. The later mechanochemical model of VanBuren et al. [22] incorporates explicit treatment of lateral bonds and provides significant insight into both the energetics of tubulin assembly and the structure of the MT tip, but it is so computationally intensive that it can simulate only short spans of time, prohibiting generation of life-history plots or the observation of transitions. These combined attributes make our model unique in its ability to investigate the mechanisms of the dynamic instability transitions.

Our analysis leads to several predictions about the structure of MT tips. First, it suggests that under conditions that promote dynamic instability, depolymerizing protofilaments frequently have GTP-Tu at the tips. Addition of GTP-Tu to GDP protofilaments has been incorporated into previous models (see, e.g., [22]), and it appears to be an essential aspect of the mechanism of rescue (see more below).

Second, our analysis suggests that growing MTs have a short region near the
tip in which GTP-Tu-containing protofilaments are laterally unbonded (i.e., the tip of the GTP cap has cracks in it). We make no predictions about the depth of these cracks, since this would be extremely parameter-dependent, but we think that the existence of these cracks is likely for several reasons. In particular, entropic considerations argue that these cracks (which are different from the cracks between GDP protofilaments) are expected because the activation energy barrier for forming lateral and longitudinal bonds simultaneously should be higher than that for binding to one face and then the other.

Third, while it is clear that growing MTs have more GTP-Tu than shrinking MTs, our simulations suggest that the more significant difference between growing and shrinking MTs is the number of lateral bonds between tip-localized GTP-Tu subunits: in depolymerizing MTs, cracks between protofilaments (the laterally unbonded regions/ram’s horns) extend from the tip, through the GTP-rich regions, down into the GDP rich region (Figure 5.4B), while in growing MTs, these inter-protofilament cracks terminate in the GTP-rich region. This leads to the suggestion that the GTP cap is not defined by the presence of GTP-Tu (since GTP-Tu is also at the ends of depolymerizing MTs), but by the presence of laterally bonded GTP-Tu.

These discussions raise the question of the minimum number of laterally bonded GTP-Tu necessary for an effective cap. Although this question bears further investigation, we propose that there is no single answer. First, the number needed for an effective cap will be very parameter dependent. Tubulins from different organisms differ considerably in their biochemical characteristics [31–33], so caps should differ from organism to organism. Indeed, yeast MTs contain much more GTP than do mammalian MTs [31, 32]. The existence of an extended GTP cap
even in mammalian MT is supported by recent nanoscale measurements of MT dynamics [30]. Second, even with a particular parameter set, it is unlikely that it will be possible to define a minimum effective cap (i.e., the minimum cap capable of preventing near-term catastrophe) because of the stochastic nature of the polymerization process and the many possible futures for a MT tip in a given configuration.

Finally, the model indicates that both catastrophe and rescue are governed by the lengthening/shortening of cracks (laterally unbonded regions) between the protofilaments. We suggest that catastrophe becomes likely when the cracks extend into regions rich in GDP-Tu. In contrast, rescue becomes possible when cracks between GDP protofilaments disappear approximately simultaneously on several protofilaments (perhaps due to loss of extended protofilament pieces). Once this occurs, the new GTP-Tu subunits that are constantly adding (but usually falling off) may be able to establish lateral bonds (Figure 5.7). We emphasize that both catastrophe and rescue are the outcome of a series of contingent stochastic microscopic changes and are difficult, if not impossible, to anticipate in advance.

The common significance of cracks to both catastrophe and rescue that is suggested by our model may explain how MT associated proteins such as MAP2 and tau are able to alter both catastrophe and rescue. We predict that any regulatory protein with the ability to alter the balance between hydrolysis rate and crack formation should have a strong effect on MT dynamics.
5.5 Methods

5.5.1 Processes taking place in the model

As noted above, the model simulates at the dimer scale the events occurring on a 13-protofilament microtubule. It includes five processes: protofilament growth, protofilament shortening, inter-protofilament bond growth, inter-protofilament bond shortening, and the transition of each GTP-Tu subunit to GDP-Tu (i.e. GTPase). At each step the algorithm checks all possible events which can occur in any of the protofilaments and bonds and determines the fastest event. This event is then implemented. After that the hydrolysis cycle is run, randomly converting GTP-Tu dimers into GDP-Tu dimers. Then the next step begins.

Growth is an addition of a single subunit (single tubulin heterodimer) from the solution to the end of a protofilament. The probability of growth depends on soluble tubulin concentration and the identity (i.e., GTP/GDP) of the dimer at the tip. Shortening is a depolymerization event which is independent of soluble tubulin concentration. It depends instead on the presence/absence of lateral bonds between subunits, which are themselves influenced by the nucleotide states of the subunits in question. Any part of a laterally unbonded protofilament end can detach, meaning that multiple subunits can fall off simultaneously, which is consistent with experimental observation [16]. Bonding is a formation of a new lateral bond between two dimers of neighboring protofilaments. A new lateral bond can form only if the subunits below it are already laterally bonded (similar to a zipper). This is reasonable given the likely separation between unbonded protofilaments. Breaking is a loss of such a bond. An existing bond can break only if it is the highest (last) bond between the two protofilaments. (This rule is based on the structural constraints inside the MT lattice.) The bonding and breaking
rules result in formation of continuous bonded segments between protofilaments, from the seed up to the point at which a crack between protofilaments starts. Hydrolysis is modeled as a stochastic first order process. Hydrolysis can occur only in interior subunits (i.e., it does not occur on the terminal subunit of a protofilament) which is consistent with structural data indicating that binding of alpha tubulin is necessary to induce the beta tubulin GTPase [16]. In the present version of the model the rate of hydrolysis is not influenced by the state of surrounding subunits.

5.5.2 Asymmetric stability of the dimer’s lateral bonds

The rates of bonding/breaking events depend separately on the identities of the left and right neighbors. This is physiologically relevant because analysis of the structure of subunits comparing straight and bent conformations implies that such asymmetries exist [21]. Specifically, the M-loop, H3 helix, and H10 helix (Figure 5.11) display heterogeneous changes in position rather than undergoing a block movement as part of the general $11^\circ$ bending of the dimer axis. We think that the heterogeneous shifts in these lateral structures could make one side of the dimer much more prone to breaking or bonding than the other side. We use the lateral interaction stability to model GDP-Tu|GDP-Tu and GDP-Tu|GTP-Tu bonds as being weak and GTP-Tu|GTP-Tu and GTP-Tu|GDP-Tu bonds as being strong. For a more detailed quantitative description see subsection 5.5.4.

5.5.3 Additional details of the computational model

The rate $k_i$ of each possible event is calculated based on the identities of the neighboring dimers except in the case of growth, where it also depends on
Figure 5.11: Structure of the tubulin heterodimer highlighting the lateral contacts undergoing important changes upon switching from a straight to a bent conformation. The M-loop, H3, and H10 helices undergo heterogeneous changes rather than simply accompanying the general $11^\circ$ bending of the dimer axis (Krebs et al. 2005 *EMBO Rep* 6, 227-32). These changes could make one lateral side of the dimer much more prone to breaking—or bonding—than the other side.
soluble tubulin concentration. All rates are provided in the configuration file. We model the events as Poisson processes with waiting times following an exponential distribution. Then the time for each event is drawn from its respective rate:

\[ t_i = -\ln r / k_i \]

where \( r \) is a random number uniformly distributed between 0 and 1. If this time is shorter than the previous shortest time, the present event becomes the candidate for implementation. Figure 5.12 shows the structure of the program.

The rate of bond breaking depends on whether there are neighboring bonds at the same height. Effectively, this keeps MT shortening somewhat coordinated after catastrophe, so that it shortens roughly simultaneously along all the protofilaments. This dependence can be justified by arguing that the steric constraints of being in the middle of a tight lattice might make it harder for a dimer to break free.

Processes occurring on the protofilaments next to the seam are governed by different rates because the MT structure near the seam differs from that in the rest of the MT. Also, for the presently considered half-integer seam shift of 1.5 dimers, the bonds at the seam grow and shorten by half a dimer, in contrast to other bonds that evolve by a full dimer length. This is done to avoid introducing artificial asymmetry at the seam.

5.5.4 The reference parameter set

The parameter values that we used throughout this article are as follows:

Free tubulin concentration \( c = 14 \mu M \).

Our units of time are arbitrary (a.u.) and the rates are given in the inverses of these units.
Figure 5.12: Structure of our stochastic algorithm for modeling MT dynamics. After setting the initial seed for the MT, we execute the subsequent loop for a desired duration. At each iteration of the loop, we keep track of the shortest time \( t_{\text{shortest}} \) (initially set to the largest representable floating point value on our machine, \( \text{MAX\_FLOAT} \)), associated event \( k_{\text{shortest}} \), and bond/protofilament (PF) number \( i_{\text{shortest}} \). Looping over all events (for bonds, formation and breaking and for PFs, growth and shortening) we stochastically compute the time to execute the event using an exponential distribution with scaling factor dependencies based on local interactions with adjacent dimers. Once we have computed the time to execute all possible events, we execute the shortest event, update the MT structure, and run hydrolysis.
Hydrolysis rate $K_h = 0.2$.

Polymerization rate of a dimer

$$K_{grow} = \kappa \frac{c}{c + c_{1/2}}$$

where $\kappa = 500$ is the maximal possible growth rate and $c_{1/2} = 200 \mu M$. For $c \ll c_{1/2}$ we thus have $K_{grow} \approx k_+ c$ with $k_+ = \kappa / c_{1/2} = 2.5 (\mu M^{-1} \text{a.u.time}^{-1})$. Note that only the longitudinal bond is formed during the growth (polymerization) event. Possible lateral bonds can be formed in separate steps.

Shortening rate $K_{shorten} = 1$ is the rate of breakage of a longitudinal bond, following which the whole part of the laterally unbonded protofilament above it de-polymerizes. Note that shortening of laterally bonded parts of the protofilaments cannot happen.

Lateral bond formation rate $K_{bond} = 30$ for regular bonds, while for the seam $K_{bond, seam} = 60$. Note that the seam bond grows/shortens by half a dimer, as opposed to all other bonds which evolve by a whole dimer length.

Lateral bond breakage rate depends on the left and right neighbors, and its reference values are given below for the specified left|right neighbors. For regular bonds

$$K_{break} = \begin{cases} 
30, & \text{GTP-Tu|GTP-Tu} \\
40, & \text{GTP-Tu|GDP-Tu} \\
80, & \text{GDP-Tu|GTP-Tu} \\
80, & \text{GDP-Tu|GDP-Tu}
\end{cases}$$
while for the seam

\[
K_{\text{break,seam}} = \begin{cases}
60, & \text{GTP-Tu|GTP-Tu} \\
150, & \text{GTP-Tu|GDP-Tu} \\
150, & \text{GDP-Tu|GTP-Tu} \\
150, & \text{GDP-Tu|GDP-Tu}
\end{cases}
\]

For the simulated dilution experiments, the only change was lowering the GTP-Tu|GTP-Tu breakage rates to 10 for regular bonds and to 20 for the seam.

Note that while we chose to express the influence of the neighbor dimer states through the lateral bond breakage rates alone, the program allows to make all the rates \(K_{\text{grow}}, K_{\text{shorten}}, K_{\text{bond}}\) and \(K_{\text{break}}\) neighbor-dependent.

In addition, there is another important parameter, \(\pi_{\text{break}} = 10\): the rate of breaking the lateral bond (\(K_{\text{break}}\) or \(K_{\text{break,seam}}\)) decreases by a factor of \(\pi_{\text{break}}\) if there are two presently formed lateral bonds to the left and to the right of the considered one. As is stated previously, this reflects the fact that it is harder to break out of the tight lattice.

The above reference rates can be related to the longitudinal and lateral bond energies, in the same fashion as it is done in \([1, 22]\). Using the standard conditions of \(c = 1M = 10^6 \mu M\) and linear approximation \(K_{\text{grow}} \approx k_+ c\) yields

\[
\Delta G_{\text{long}}^* = -k_B T \ln \frac{k_+ 10^6}{K_{\text{shorten}}} \approx -14.7 k_B T
\]

for the longitudinal bond energy (the asterisk indicates that the entropic contri-
The lateral bond energy is also included - cf. [4, 22]. Similarly, for the lateral bond energy we use

\[ \Delta G_{lat} = -k_B T \ln \left( \frac{K_{bond}}{K_{break}} \right) \]

for bonds lacking two existing neighboring lateral bonds. The numerical values of the lateral energies for GTP-Tu|GTP-Tu, GTP-Tu|GDP-Tu, GDP-Tu|GTP-Tu, and GDP-Tu|GDP-Tu bonds in units of \( k_B T \) are 0, 0.3, 1 and 1 respectively. This means that bond formation is energetically unfavorable. On the other hand, in the presence of the two neighboring lateral bonds we use

\[ \Delta G_{lat} = -k_B T \ln \left( \frac{K_{bond}}{K_{break}/\pi_{break}} \right) \]

with corresponding values of -2.3, -2, -1.3 and -1.3 for GTP-Tu|GTP-Tu, GTP-Tu|GDP-Tu, GDP-Tu|GTP-Tu, and GDP-Tu|GDP-Tu bonds. This means that bond formation is energetically favorable. The energies for the seam can be estimated in a similar fashion.

These bond energy values are qualitatively similar to those used in [4] and estimated in [3]. In [4], for the value of \( k_+ = 2 \mu M^{-1} s^{-1} \) the authors predict that \( \Delta G_{\text{long}} = -9.4 k_B T \) and \( \Delta G_{\text{GTP}} = -3.2 k_B T \), where now we add a superscript GTP to emphasize that this was the estimation of the lateral energy for the simulations of MT growth. For the simulations of MT shortening, it is shown in [4] that \( \Delta G_{\text{kink}} = 2.1 k_B T \) leading to \( \Delta G_{\text{GDP}} = \Delta G_{\text{GTP}} + \Delta G_{\text{kink}} = -1.1 k_B T \). These energies \( \Delta G_{\text{GTP}} \) and \( \Delta G_{\text{GDP}} \) can be interpreted in terms of our model as the lateral bond energies between two GTP-Tu dimers and between two GDP-Tu dimers and in the case when all the lateral neighbor bonds exist. Notice that the model in [4] cannot have broken lateral bonds. In addition, the model in [4] does
not have lateral asymmetry.

The structural computational studies [3] estimate that \( \Delta G_{\text{long}} - \Delta G_{\text{lat}} \approx -7 \text{kcal/mol} \approx -11.7 k_B T \) for room temperature, which is also in agreement with our choice of parameters.

Finally, we would like to mention that the lateral bonds are weak and it well might be the case that they continuously form and break on a fast time scale. From the numerical point of view, we don’t want to dedicate most of the steps of the code to simulate these fast variables, because we want to see the evolution in the MT length. To reduce the number of bonding/breaking events while keeping the same mesoscopic behavior, the principal parameters that must be preserved are the mean velocities of lateral bond breaking (for a shortening MT) and lateral bond formation (for a growing MT). These velocities affect the velocities of the MT shortening and growth. If the lateral bond energy between two dimers is \( \Delta G \) then the bonding and breaking rates are related through \( k_{\text{bond}} / k_{\text{break}} = e^{-\Delta G / kT} \). If the potential barrier between laterally bonded and unbonded states of two dimers is small then both \( k_{\text{bond}} \) and \( k_{\text{break}} \) are large. The mean velocity of lateral bond growth between the neighboring protofilaments is given by \( k_{\text{bond}} - k_{\text{break}} = k_{\text{bond}}(1 - e^{\Delta G / kT}) \). Our goal is to use some effective slow bonding and breaking rates, \( k_{\text{bond}}^{\text{eff}} \) and \( k_{\text{break}}^{\text{eff}} \), while preserving \( k_{\text{bond}} - k_{\text{break}} = k_{\text{bond}}^{\text{eff}} - k_{\text{break}}^{\text{eff}} = k_{\text{bond}}(1 - e^{\Delta G^{\text{eff}} / kT}) \). If \( k_{\text{bond}}^{\text{eff}} < k_{\text{bond}} \) then necessarily \( |\Delta G^{\text{eff}}| > |\Delta G| \). Hence, the more stable state (either bonded or unbonded) becomes even more stable. The purpose of this discussion is to argue that we do not necessarily have to use the parameters in our model which agree with the measured lateral bond energies. Of course, by reducing/upscaling the rates we reduce the fluctuations around the average behavior.
5.6 References


CHAPTER 6

MOLECULAR EVOLUTION OF THE HISTONE DEACETYLASE FAMILY:
FUNCTIONAL IMPLICATIONS OF PHYLOGENETIC ANALYSIS

6.1 Abstract

Histone\(^1\) deacetylases (HDACs) modify core histones and participate in large regulatory complexes that both suppress and enhance transcription. Recent studies indicate that some HDACs can act on non-histone proteins as well. Interest in these enzymes is growing because HDAC inhibitors appear to be promising therapeutic agents against cancer and a variety of other diseases. Thus far, 11 members of the HDAC family have been identified in humans, but few have been characterized in detail. To better define the biological function of these proteins, make maximal use of studies performed in other systems, and assist in drug development efforts, we have performed a phylogenetic analysis of all HDAC-related proteins in all fully sequenced free-living organisms. Previous analyses have divided non-sirtuin HDACs into two groups, classes 1 and 2. We find that HDACs can be divided into three equally distinct groups: class 1, class 2, and a third class consisting of proteins related to the recently identified human HDAC11 gene. We term this novel group class 4 to distinguish it from the unrelated class 3 sirtuin

\(^1\)This thesis chapter has been published as a research article. IV Gregoretti, YM Lee and HV Goodson; 2004 Journal of Molecular Biology 338(1):17–31.
deacetylases. Analysis of gene duplication events indicates that the common ancestor of metazoan organisms contained two class 1, two class 2, and a single class 4 HDAC. Examination of HDAC characteristics in light of these evolutionary relationships leads to functional predictions, among them that self-association is common among HDAC proteins. All three HDAC classes (including class 4) exist in eubacteria. Phylogenetic analysis of bacterial HDAC relatives suggests that all three HDAC classes precede the evolution of histone proteins and raises the possibility that the primary activity of some histone deacetylase enzymes is directed against non-histone substrates.

6.2 Introduction

Histone deacetylases (HDACs) reverse the regulatory acetylation of histone proteins, influencing nucleosome structure and altering gene transcription. Histone acetylation is often correlated with gene activation, suggesting that histone deacetylases act to silence genes, but genetic experiments in *Drosophila* and *Saccharomyces cerevisiae* have indicated that deacetylase activity can contribute to gene activation as well [1][4]. Acetylation of core histones has also been correlated with cellular processes, including chromatin assembly, DNA repair, and recombination (reviewed by Yang et al. [5]).

HDACs are members of an ancient enzyme family found in plants, animals, and fungi, as well as archaeabacteria and eubacteria [6]. Thus far, eukaryotic HDACs have been classified into two groups (class 1 and class 2) on the basis of sequence similarity. Recently, a pair of HDAC-related proteins from *Arabidopsis* and humans have been reported to be more similar to each other than to either class 1 or class 2 enzymes [7], possibly representing an additional class. In addition,
a group of unrelated NAD-dependent deacetylase enzymes related to the yeast protein Sir2 have sometimes been called class 3 HDACs [3, 4, 8, 9]. To reduce confusion, we refer to these NAD-dependent enzymes as sirtuins (see Grozinger & Schreiber [1]). The domain structures and class affiliation of human HDACs, as well as representative fungal and prokaryotic proteins, are shown in Fig. 6.1.

As is obvious from their name, it has generally been assumed that the activity of HDACs is directed at histones. However, many HDACs are at least partially cytoplasmic [11–14], and evidence is accumulating that some fraction of these proteins can act on non-histone substrates, including the cytoskeletal protein tubulin and transcription factors such as p53 [11] [15] [16]. Indeed, it has been suggested that regulatory acetylation/deacetylation is considerably more widespread than presently appreciated, acting in a manner similar to phosphorylation and dephosphorylation [17].

HDACs have recently enjoyed increased attention because HDAC inhibitors (e.g. hydroxamic acids, depsipeptide) act as efficient anti-proliferative agents in tissue culture and inhibit tumor progression in rodent models. Several HDAC inhibitors are now in phase I and II human trials [2, 18] [19]. Most of the known HDAC inhibitors have broad specificity for both class 1 and class 2 deacetylases, although some (e.g. trapoxin) exhibit some specificity [2, 20]. Large synthetic efforts are now focused on developing new inhibitors, with particular interest on development of isoform-specific inhibitors [4].

Although understanding of HDAC activity and function is growing rapidly, most members of this family have received only initial characterization. A classic way to enhance understanding of one HDAC is to use information gained from study of related HDACs. However, maximal use of this information requires a
Figure 6.1: Schematic representation of HDAC structure, drawn to scale. Class assignments are based on the analysis of Fig. 6.2 and Figure 6.4, names are described in the corresponding legends. The HDAC6 metal-binding motif was reported by Hook et al. [10], while coiled-coil assignments are derived from the analysis described in Figure 6.3.
detailed understanding of the evolutionary relationships within the HDAC family. More specifically, making useful functional predictions about human HDACs based on those from model systems depends on our ability to establish which HDACs are orthologs (diverged as a result of species divergence) or paralogs (diverged by gene duplication). Orthologs are generally expected to have similar functions and characteristics because of the constancy of basic cell biological processes between organisms (even divergent ones). Paralogs are expected to be functionally differentiated because the initial duplication releases evolutionary constraint, giving the duplicated genes the opportunity to acquire new functions, partition old functions, or gain tissue-specific distributions [21, 22]. Phylogenetic trees also provide other types of functional information: conserved proteins with ancient origins are expected to participate in basic processes conserved across organisms, while recent origins suggest organism or tissue-specific functional specialization. Observation that some proteins exhibit unusually rapid rates of evolution indicates that these proteins have undergone a change in selection pressure relative to other members of the family, likely correlated with loss or change of function.

Phylogenetic analyses have been included in other studies of HDAC structure and function [6, 7, 23], but limited sequence diversity and statistical analysis have restricted interpretation of these studies. To address these issues as they relate to the HDAC family, we undertook the following analysis of evolutionary relationships between all recognizable HDAC relatives in all fully sequenced free-living organisms, eukaryotic and prokaryotic.
6.3 Results and Discussion

6.3.1 Phylogenetic analysis: alignment and tree building

We have used two phylogenetic methods, neighbor joining (as implemented by ClustalX) [24, 25], and Bayesian analysis (as implemented by the program MrBayes) [26] to determine evolutionary relationships between recognizable HDAC-related proteins in fully sequenced free-living organisms. Fig. 6.2 shows the inferred relationships between eukaryotic sequences, with selected prokaryotic proteins included, and Figure 6.4 shows the relationships between prokaryotic sequences, with selected eukaryotic sequences included (see Table 6.2 for a discussion of how to interpret phylogenetic trees). HDAC sequences were identified by systematic PSI-BLAST [27] searches of the relevant genomes using human and yeast HDACs as probes. Alignment was performed by ClustalX as described in Methods. Only the core region conserved between all family members was included in the phylogenetic analysis.

6.3.2 Eukaryotic histone deacetylases fall into three, not two, related classes

Non-sirtuin HDACs have previously been divided into two groups (class 1 and class 2) on the basis of molecular mass and pairwise sequence similarity, with some sequences such as *S. cerevisiae* Hos3 often remaining ungrouped [6, 8, 30, 31]. However, Fig. 6.2 indicates that HDACs in the sequenced eukaryotic organisms fall cleanly into three (not two) equally distinct groups: the class 1 HDACs, class 2 HDACs, and a group of proteins similar to human HDAC11 [7]. This tripartite division is based on strong statistical support for the existence of three distinct groups, the absence of support for any larger groups, and on the observation that
Figure 6.2: Bootstrapped neighbor joining phylogenetic tree of eukaryotic HDAC sequences. See Table 6.2 for information on interpreting the tree. All clearly distinct and recognizable full-length HDAC sequences in all fully sequenced freeliving eukaryotic organisms are included in this analysis (see Methods); selected prokaryotic proteins (in italics) are included as well. The deepest well-supported nodes (large black dots) divide the sequences into classes; subgroups are designated as subclasses if they contain proteins from divergent organisms (minimally vertebrates and invertebrates) and have strong statistical support. Nomenclature: sequences are named by abbreviations of genus and species followed by the published name, the genetic locus, or (when the sequence is otherwise uncharacterized) the NCBI gi number. The set of analyzed proteins with corresponding accession numbers (if not already included in the name) is shown in Table 6.1 for eukaryotes and in Figure 6.4 for prokaryotes.
TABLE 6.1
EUKARYOTIC SET OF ANALYZED PROTEINS WITH CORRESPONDING ACCESSION NUMBERS.

<table>
<thead>
<tr>
<th>Species</th>
<th>NCBI database record</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>AtHDA2 gi21105771, AtHDA6 gi15242626, AtHDA7 gi10176806, AtHDA8 gi18390898, AtHDA9 gi15230483, AtHDA14 gi18418220, AtHDA15 gi18401915, AtHDA19 gi2318131, At9759454 (motifs a and b, correspond to HDAC5 and HDAC18 of Pandey et al. [28])</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>CeF41H10 (motifs a and b) gi17540334, CeRpd3like gi17561978, Ce17508561, Ce17534739, CeC10E2 gi17550310, Ce17537347, Ce17533993 C. elegans sequence C35A5.9 (gi17558426) is related to class 4 HDACs but was not included in this analysis because the reported sequence appears to be truncated.</td>
</tr>
<tr>
<td>Danio rerio (zebrafish, not fully sequenced)</td>
<td>Dr33416863 (included to allow analysis of the rate of HDAC8 evolution)</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>DmHDAC2 gi18860039 (motifs a and b), DmCG10899 gi7301219, DmRpd3 gi7292522, DmHDAC3 gi7296744, DmCG1770 gi18860041</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>HsHDAC1 gi3128860, HsHDAC2 gi21411359, HsHDAC3 gi3128862, HsHDAC4 gi5174481, HsHDAC5 gi4885531, HsHDAC6 gi3128864 (motifs a and b), HsHDAC7 gi3259524, HsHDAC8 gi8923769, HsHDAC9 gi5590680, HsHDAC10 gi15213867 or gi15213865, HsHDAC11 gi13376228</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>MmHDAC1 gi6680193, MmHDAC2 gi6680195, MmHDAC3 gi12643653, MmHDAC5 gi6911182, MmHDAC6 gi20983042, MmHDAC7 gi20903371, MmHDAC8 gi20984875, MmHDAC10 gi20903151, MmHDAC11 gi20828622</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>ScRpd3 gi6323999, ScHos2 gi6321244, ScHos1 gi6325325, ScHda1 gi6324307, ScHos3 gi6325141</td>
</tr>
<tr>
<td>Schizosaccharomyces pombe</td>
<td>Sp19112125, Sp19114991, Sp19111896</td>
</tr>
</tbody>
</table>

The gi numbers are shown when the name of the sequence does not include it.
### TABLE 6.2

<table>
<thead>
<tr>
<th>Interpreting phylogenetic trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>The tips of the branches of the phylogenetic tree (see Figure 6.2 and Figure 6.4) represent individual sequences, and branch lengths are proportional to sequence divergence (sequence non-identity).</td>
</tr>
<tr>
<td>The relationship between time and branch length is positive (ancient times are towards the center of the tree), but it is uncalibrated and varies over different parts of the tree. The trees are unrooted (there is no assumption as to which sequence is closest to the common ancestor), resulting in a star or tumbleweed topology.</td>
</tr>
<tr>
<td>Nodes (connection points between branches) indicate sequence divergence, which results from either gene duplication or species divergence. To distinguish between these possibilities, one compares the observed branching pattern with that expected from evolutionary relationships between organisms.</td>
</tr>
<tr>
<td>Because most trees have some branching patterns that are well-supported by the data and others that derive from sequence noise, one must estimate the validity of the groupings before attempting to interpret any phylogenetic tree.</td>
</tr>
<tr>
<td>We have used the statistical method of bootstrap resampling (1000 trials) to estimate confidence in particular groupings, and have indicated these results by colored dots at the relevant nodes in Figure 6.2 and Figure 6.4. Groups found in &gt;95% of bootstrap trials can be considered well-supported, those found in &gt;80% moderately supported, those &gt;50% are suggestive, and groups found in fewer than 50% of bootstrap trials should be regarded as completely unresolved.</td>
</tr>
</tbody>
</table>
The percentage identity/similarity (PI and PS respectively) shows that sequences from extremely divergent organisms are more closely related to their own class than to any other class. Both identity and similarity were computed using the program OldDistances from the GCG Package. The similarity was scored using the BLOSUM62 matrix.

All three groups contain proteins from extremely divergent organisms: animals, plants, and bacteria (Fig. 6.2). Pairwise sequence comparisons (Table 6.3) further strengthen the conclusion that HDAC11-related HDACs are as distinct from the class 1 and class 2 HDACs as either class is from the other. Pandey and colleagues recently reached a similar conclusion in their study of *Arabidopsis* HDACs [28].

On the basis of this analysis, we propose that HDAC11-related enzymes be given their own class designation, and we refer to this group as class 4 HDACs (the name class 3 is already commonly used to refer to the unrelated sirtuin deacetylases).

All HDACs except class 4 are found in all fully sequenced free-living eukaryotic
organisms, and class 4 proteins are found in all of these organisms except fungi\(^2\) (Figure 6.2). This observation suggests that all three HDACs existed in the ancestral eukaryote (bacterial HDAC origins are discussed below). The conservation of these proteins in such divergent organisms suggests that each HDAC class has a non-redundant role in basic cell biological processes: though some functional overlap between these proteins may exist, each class must have some unique function(s) resulting in selective pressure for maintenance of these genes over time \[21, 22\]. This is important to consider in light of observations that HDACs often appear to have similar or overlapping substrates.

The combination of the early origin of these proteins and their conservation across divergent organisms suggests that representatives of all three HDAC classes will be found in most eukaryotic organisms and most metazoan cell types. Although class 2 HDACs are sometimes described as being tissue-specific, recent SAGE analysis of human HDACs\[23\] is consistent with the prediction that most tissues will express at least one class 2 HDAC. Given that fungi are more closely related to animals than plants\[32, 33\], it seems most likely that class 4 proteins were lost somewhere on the line leading to fungi. Perhaps fungal-specific changes in physiology, lifestyle, or HDAC function led to fungal-specific loss of class 4 utility. Such changes may be significant in interpreting differences between fungal and vertebrate HDAC function.

6.3.3 Subclass analysis: definitions and interpretations

Given that humans have 11 different HDAC genes, dividing HDACs into three classes is useful but insufficient: the HDAC family must contain functionally dif-

\(^2\)Examination of the full NCBI database also failed to identify any fungal class 4 proteins. A sequence with clear similarity to class 4 HDACs does exist in the worm database (gi17558426) but it was not included in our analysis because of its apparently fragmentary nature.
ferentiated subgroups resulting from gene duplications that occurred at various points along the evolutionary pathway. Understanding when these duplications occurred relative to organismal evolution is essential for judging the likelihood of functional overlap between these proteins, and also for predicting the properties of uncharacterized sequences on the basis of known proteins.

For the purposes of our analysis, we define subgroups to be subclasses if they contain proteins from divergent organisms (minimally vertebrates and invertebrates) and have strong statistical support. This definition allows us to identify the set of HDACs common to most/all metazoan organisms. Names were assigned to these subclasses by the human proteins contained within them (i.e. proteins HsHDAC1 and HsHDAC2 are members of the HDAC1/2 subclass).

6.3.3.1 Class 1 HDACs

HDAC class I contains two subclasses common to vertebrates and Drosophila: HDAC1/2 and HDAC3. The HDAC1/2 subclass consists of HDAC1 and HDAC2 in vertebrates, the single RPD3 protein in Drosophila, and (most likely) a pair of proteins in Caenorhabditis elegans (Figure 6.2). Bootstrap support for inclusion of the C. elegans proteins is weaker. Biochemical evidence described below suggests that the subfamily may extend into single-celled organisms. The branching pattern indicates that the two mammalian and two C. elegans proteins are the result of independent gene duplications and that the common metazoan ancestor contained one HDAC1/2 protein.

The strong sequence constraint observed within the HDAC1/2 subclass and the relatively recent nature of the gene duplication leading to HDAC1 and HDAC2 suggests that HDAC1 and 2 have undergone little functional divergence either
from the ancestral HDAC1/2 protein or from each other (to estimate sequence constraint, compare branch lengths for the human-mouse and human-fly divergences in different HDAC subclasses). Consistent with this prediction, HDAC1 and HDAC2 exist in similar complexes including Sin3, NuRD/ NRD/Mi2, and CoREST [2, 34, 35]. They are also both widely expressed [23 36]. However, the observation that homozygous disruption of mouse HDAC1 causes early embryonic lethality [37] demonstrates that HDAC2 cannot fulfill all HDAC1 functions and that some significant difference in function or expression does exist. The C termini of the human and mouse HDAC2 proteins contain a previously unrecognized region predicted to have strong propensity for coiled-coil (Figure 6.3). The existence of this predicted coiled-coil outside of the HDAC oligomerization domain (discussed below) suggests that this region is involved in additional protein-protein associations and may account for some amount of functional differentiation. The coiled-coil propensities of the analogous regions in HDAC1 and the single Drosophila HDAC1/2 protein are much weaker but do not rule out coiled-coil formation (Figure 6.33).

The HDAC3 subclass:

The HDAC3 subclass exists as a single protein in sequenced organisms from humans to Drosophila (Figure 6.2). Biochemical evidence discussed below suggests that this subclass may extend to single-celled organisms. The C. elegans protein Ce17534739 may be a member of this group, but if so it is surprisingly divergent (tree reconstruction methods do not detect any relationship to the HDAC3 subclass, see Figure 6.2) and therefore it likely differs from other HDAC3 proteins in functionally significant ways. Consistent with being the only vertebrate
Figure 6.3: Propensity for formation of alpha-helical coiled-coils within HDAC proteins as estimated by the COILS program of Lupas [38]. The 28 amino acid window is considered to be the most stringent. All human and yeast HDAC proteins were analyzed, but only HDAC1/2, HDAC3, and HDAC4/5/9 proteins displayed significant coiled-coil propensity (other data not shown).
member of an ancient subclass, human HDAC3 is widely expressed \[36, 39, 40\]. HsHDAC3 (and likely HDAC3 members in other organisms) participates in complexes including SMRT/N-Cor, and it binds to members of the nuclear receptor corepressor family and the transducin B-like protein [5]. HDAC3 interacts with and regulates the nuclear export of the NF-kB subunit RelA [41]. The observation that human HDAC3 interacts with most class 2 proteins (HDAC4, 5, and 7 as well as HDAC10[42, 43]) suggests that interaction between HDAC3 proteins and class 2 HDACs is an ancestral phenomenon likely conserved in many organisms. HDAC3 is reported to have both nuclear and cytoplasmic localization [44]. Though this dual localization may be a means of regulating activity against nuclear substrates (reviewed by Grozinger et al. [4]) it is interesting to consider the possibility that part of HDAC3 function is directed against non-nuclear substrates.

**HDAC8:**

HDAC8 was not given subclass status in our analysis because it appears to be restricted to vertebrates (Figure 6.2). This restriction suggests that it does not play an important role in conserved processes, though it may have developmental or tissue-specific functions consistent with its apparently tissue-specific expression [23]. The deep divergence of the HDAC8 branch might be taken to suggest that HDAC8 is an ancient protein lost from invertebrate organisms. However, it seems more likely that this branch is mispositioned. It is well-established that unusually rapid divergence along a branch can cause misplacement of that branch [45]. Consistent with this explanation, HDAC8 does appear to be evolving faster than other class I HDACs: the human-mouse distance is greater on the HDAC8 branch than on other class I branches, and the HDAC8 human-zebrafish distance is similar to
the human-fly distance on other class I branches (Figure 6.2). This increased rate of HDAC8 evolution suggests that HDAC8 has undergone significant functional specialization relative to other class 1 HDACs.

**Non-metazoan class I proteins:**

Relationships between metazoan and non-metazoan class 1 HDACs are not resolved by our phylogenetic analysis. This is unfortunate, because maximal use of the genetic analyses performed in yeast requires precise understanding of relationships between human and yeast proteins. However, biochemical evidence suggests that the division of class I HDACs into two subclasses may extend all the way to yeast: *S. cerevisiae* RPD3 participates in a complex similar to Sin3 [46], implying that it may be part of the HDAC1/2 subclass. HOS2 is found in a complex related to SMRT [47], providing evidence that it is an HDAC3 protein. The remaining yeast class 1 protein (Hos1) appears to be unique to part of the fungal lineage (the *Schizosaccharomyces pombe* genome has no homologue, but a *Candida albicans* homologue has been reported [48]).

**Self-association between class I proteins:**

Human HDAC1 associates both with itself and with HDAC2 [49 50]. This interaction is mediated by an N-terminal region that includes part of the conserved HDAC domains [49]. This observation suggests that dimerization is an ancestral characteristic of HDAC1/2 proteins and that the single HDAC1/2 proteins found in simpler organisms will also dimerize. Moreover, human HDAC3 forms homooligomers [5]. The observation that both class I subclasses contain self-associating proteins suggests that self-association may be an ancestral property common to
most if not all class 1 HDACs. It is interesting to note that *C. elegans* and *Arabidopsis* class 1 HDACs have each undergone duplications independent of the one leading to HDAC1 and HDAC2 (Figure 6.2). This leads to the speculation that these organism-specific protein pairs might associate into heteromeric complexes analogous to those formed by HDAC1 and HDAC2, though they could be differentiated in other ways.

6.3.3.2 Class 2 HDACs

Metazoan organisms contain two class 2 subclasses. These are HDAC4/5/7/9 and HDAC6/10. Each group consists of the indicated human proteins, their mouse homologues, and single proteins in *Drosophila* and *C. elegans*. Recently, these subfamilies have also been called 2a and 2b, respectively [51].

**HDAC4/5/7/9 subclass:**

Consistent with their relatively recent divergence, members of this subfamily interact with similar proteins, including HDAC3, nuclear receptors N-Cor and B-Cor, and the complex SMRT11 [52], (reviewed by Verdin *et al.* [51]). They also associate with the muscle-specific transcription factor MEF2 and have well-demonstrated roles in muscle development [53–57]. These proteins have similar expression profiles, with particularly strong expression in muscle [8, 12, 23, 52, 53]. The common nature of these attributes suggests that they are ancestral and (for example) that the single *Drosophila* HDAC4/5/7/9 protein has similar characteristics.

While these proteins have many common features, their unusually rapid sequence divergence (compare the position of the vertebrate/invertebrate divergence
on this and other branches) suggests that functional divergence may also have occurred. Indeed, it is unlikely that these genes would be maintained across the period of time separating mice and humans unless they had at least partially non-overlapping functions [21]. Observation that movement of HDAC4 to the nucleus is correlated with terminal muscle differentiation has led to the suggestion that the specialized functions of these proteins may relate to particular stages of muscle differentiation [13, 58]. However, it is important to note that Drosophila only has one HDAC4/5/7/9 protein, demonstrating that functional specialization of these proteins cannot be essential to muscle development per se.

**HDAC6/10 subclass:**

HDAC6 proteins contain two tandem HDAC domains, while HDAC10 contains one full and one vestigial HDAC domain (the vestigial HDAC10 domain was not included in Figure 6.2) [14, 59]. Alignment of full-length human HDAC6 and HDAC10 proteins against the single Drosophila HDAC6/10 member indicates that HDAC10 arose from an HDAC6-like enzyme sometime after the vertebrate/invertebrate divergence (see also Guardiola & Yao [14]). The relatively rapid rates of evolution on the HDAC6 and HDAC10 branches (compare branch lengths between human and mouse orthologs in this and other subfamilies; note that only the N-terminal catalytic motifs are included in the analysis) suggests significant functional divergence between HDAC6 and HDAC10.

One clue to the nature of this divergence is the observation that HDAC6 is primarily cytosolic, while HDAC10 is predominantly nuclear, although there is some variability in reported localizations [11, 14, 59, 60]. This suggests that HDAC6 and HDAC10 may have partitioned between them the functions of the
single ancestral protein, becoming specialized to substrates in their different locations. Consistent with this idea, HDAC6 deacetylates the cytoplasmic protein tubulin [11, 56, 61]. In addition, HDAC10 is reported to associate with HDAC3 and SMRT [43, 62], while HDAC6 apparently lacks these associations and instead associates with the class 4 HDAC11 protein [7]. The observation that HDAC3 associates with proteins on both branches of the class 2 family suggests that association between HDAC3 and class 2 proteins may predate the divergence between the HDAC4/5/7/9 and HDAC6/10 subclasses. These considerations lead to the prediction that the single Drosophila HDAC6/10 protein (and possibly the class 2 proteins of organisms such as yeast) will act in complexes containing HDAC3 and will have activity against both nuclear and cytoplasmic substrates.

Other class 2 proteins:

Examination of fungal and plant class 2 proteins suggests that the common ancestor of plants, fungi, and animals contained one class 2 protein with a simple structure lacking the N and C-terminal extensions seen in the metazoan proteins. S. cerevisiae contains a second protein (Hos3) that is extremely divergent and appears to be specific to Saccharomyces and its relatives. Orthologs have been reported in Candida [48] and Aspergillus nidulans [63] but do not appear to exist in the S. pombe genome. Arabidopsis contains two divergent class 2 proteins (HDA8 and HDA14) that appear to be specific to the plant lineage. At9759454 groups with the other class 2 proteins and contains what appear to be two duplicated HDA motifs (Figure 6.2). However, according to some analyses these belong to two neighboring but separate proteins, HDA18 and HDA5 [28].
Domain duplication and self-association between class 2 proteins:

Extremely divergent class 2 proteins self-associate, suggesting that association with itself is an ancestral characteristic common to many or most class 2 proteins. These self-associating proteins include the yeast class 2 proteins HDA1 and Hos3 [64, 65], the acetylpolyamine amidohydrolase from the eubacterium *Mycoplasma ramose* [66, 67], and mammalian HDAC4.

HDAC4 self-associates via an N-terminal coiled-coil motif [68]. HDACs 5 and 9 contain a similar but previously unrecognized coiled-coil motif (Figure 6.3), suggesting that they will also self-associate through this domain. It will be interesting to determine whether these motifs can mediate heterodimerization. This same region is implicated in interactions with the transcription factor MEF2 [53, 54]. Although HDAC7 does interact both with other HDACs and with MEF2 [57], its reported sequence lacks a region with strong coiled-coil propensity (Figure 6.3). This suggests that it may interact with these proteins in a unique way, or that like HDAC9 it may be alternatively spliced [52, 69].

As yet, neither the significance of the HDAC6 domain duplication nor the vestigial nature of the second HDAC10 domain are clear. We speculate that class 2 proteins have an ancestral requirement for dimerization, and in HDAC6 this requirement is satisfied by fusion of two HDAC domains. This duplication may have happened more than once: as mentioned above, the predicted *Arabidopsis* protein gi9759454 contains two independently duplicated HDAC motifs (Figure 6.2), although this sequence needs to be confirmed (see Pandey *et al.* [28]).

Nucleocytoplasmic shuttling of class 2 proteins:

All characterized mammalian class 2 proteins (with the possible exception of
HDAC6) move between the cytoplasm and nucleus [13, 14, 59, 60, 70–72]. This observation suggests that nucleocytoplasmic shuttling is an ancestral characteristic common to eukaryotic class 2 proteins. It has been suggested that this shuttling is a mechanism for regulating the activity of these proteins against nuclear substrates [4, 13, 70]. However, the demonstration that HDAC6 deacetylates tubulin emphasizes the possibility that other class 2 proteins may also have cytoplasmic substrates.

6.3.4 Prokaryotic HDAC1-related histone deacetylases

Previous studies have noted that prokaryotic organisms contain proteins with similarities to both class 1 and class 2 HDACs. Little is known about the function of these enzymes, but they are variously known as acetoin utilization proteins (because they are necessary for efficient utilization of the carbon source acetoin) or acetylpolyamine amidohydrolases (because they deacetylate polyamines such as spermine) [6]. The vast majority exist only as uncharacterized sequences. To understand the relationship between these prokaryotic proteins and their eukaryotic relatives, we included prokaryotic HDAC-like proteins in the phylogenetic analysis.

Figure 6.2 shows how selected prokaryotic HDAC-related sequences fit into the eukaryotic tree. Figure 6.4 shows the tree resulting from analysis of all HDAC relatives in all fully sequenced prokaryotes, together with selected sequences from eukaryotes (all human and yeast HDACs) and additional bacteria. These trees show that all analyzed prokaryotic HDAC-like sequences group unambiguously with either class 1, class 2, or class 4, demonstrating that all three HDAC classes exist in bacteria (note that some apparently fragmentary sequences were not in-
cluded in the analysis, see the legend). This conclusion is further supported by pairwise distance matrix analysis (Table 6.3), which demonstrates that eukaryotic and prokaryotic proteins within a given class are more closely related to each other than proteins in different classes.

6.3.5 Distribution of HDACs in prokaryotes

The observation that class 1, class 2, and class 4 HDACs exist in bacteria suggested that all three classes of non-sirtuin HDAC-related proteins might pre-date the divergence of eukaryotes and prokaryotes. Alternatively, prokaryotic organisms may have obtained eukaryotic HDAC proteins by gene transfer. To address this question, it is necessary to examine the distribution and evolution of prokaryotic HDACs and compare the observations to those expected from the evolutionary relationships between organisms. Unfortunately, there is as yet little consensus about bacterial relationships because analysis of different bio-molecules gives different answers, in part because of the common nature of gene transfer and gene loss in these organisms [73, 74]. Therefore, we used major taxonomic divisions (as defined by the NCBI taxonomy database) as an approximate means of differentiating between groups of related bacteria (Figure 6.4, bottom).

Examination of prokaryotic HDAC distribution within the major taxonomic divisions (Figure 6.4) leads to three immediate conclusions. First, class 2 HDAC-related proteins are found in the widest range of prokaryotic organisms, being identified in all groups of archaeabacteria except Thermoplasma and in many groups of eubacteria. Second, both class 1 and class 4 enzymes appear to exist exclusively (class 4) or almost exclusively (class 1) in eubacteria, although this con-
Figure 6.4: Bootstrapped neighbor joining tree of prokaryotic HDAC-related proteins. HDAC family members from selected eukaryotes (H. sapiens and S. cerevisiae) and additional bacteria are included to better represent HDAC diversity. Major bacterial taxonomic groups (as defined by the NCBI taxonomy database) and the distribution of HDACs within these groups are listed at the bottom. A group is designated as containing an HDAC class if at least one organism in that group contains that HDAC class. Nomenclature: sequence names contain abbreviations of the taxonomic group, the genus and species, and the HDAC class as defined by the deepest well-supported nodes. When more than one sequence of a given class exists in an organism, the last two digits of the accession number are included. Organisms, protein abbreviations and accession numbers are described in Table 6.4, Table 6.5, and Figure 6.2. Note: HDAC-related sequences from Thermosynechococcus elongates (NP_0681355), Methanosarcina barkeri (gi23049645), and Archaeoglobus fulgidus (NP.071115, NP.070114) were removed from analysis because they appeared to be fragmentary and/or contain artificial data.
<table>
<thead>
<tr>
<th>Species</th>
<th>NCBI database record</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Agrobacterium tumefaciens</em></td>
<td>AgroTu2 NP_353772.1</td>
</tr>
<tr>
<td><em>Aquifex aeolicus</em></td>
<td>AquiAe2 NP_214446.1, AquiAe1 NP_213698.1</td>
</tr>
<tr>
<td><em>Bacillus anthracis</em></td>
<td>BacAn1 NP_058697.1</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>BacSu1 NP_390849.1</td>
</tr>
<tr>
<td><em>Bacillus halodurans</em></td>
<td>BacHa1 NP_244103.1</td>
</tr>
<tr>
<td><em>Bradyrhizobium japonicum</em></td>
<td>BradJa2,45 gi27377345, BradlyJa2,64 gi27377764</td>
</tr>
<tr>
<td><em>Brucella suis</em></td>
<td>BrucSu2 NP_007458.1 (this sequence is very similar to <em>Brucella melitensis</em> NP_540421.1, not included)</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em></td>
<td>CaenEl4 gi23137215</td>
</tr>
<tr>
<td><em>Cytophaga hutchinsonii</em></td>
<td>CytoHu4 gi23137215</td>
</tr>
<tr>
<td><em>Deinococcus radiodurans</em></td>
<td>DeinRa4 NP_294557.1</td>
</tr>
<tr>
<td><em>Desulfotobacterium hafniense</em></td>
<td>DsoHu2 gi23112899</td>
</tr>
<tr>
<td><em>Desulfovibrio desulfuricans</em></td>
<td>DsoDes2 gi23475839.</td>
</tr>
<tr>
<td><em>Geobacter metallireducens</em></td>
<td>GeobMe1 gi23053171</td>
</tr>
<tr>
<td><em>Leptospira interrogans</em></td>
<td>LeptIn2 NP_713103.1, LeptIn2 NP_711096.1</td>
</tr>
<tr>
<td><em>Magnetococcus sp. MC-1</em></td>
<td>MagnSp1 gi23001994</td>
</tr>
<tr>
<td><em>Mesorhizobium loti</em></td>
<td>MesoLo4 NP_106295, MesoLo2_59 gi13472559, MesoLo2_04 gi13470404, MesoLo2_97 gi13475797, MesoLo2_10 gi13476210</td>
</tr>
<tr>
<td><em>Neisseria meningitidis</em></td>
<td>NeisMc2 NP_273649.1</td>
</tr>
<tr>
<td><em>Nostoc sp. PCC 7120</em></td>
<td>NostSp4 NP_487390.1, NostSp2 NP_484882.1</td>
</tr>
<tr>
<td><em>Oceanoviboccus iheyensis</em></td>
<td>OceaIh1 NP_693414.1</td>
</tr>
<tr>
<td><em>Psecudomonas aeruginosa</em></td>
<td>PseuAe2_06 gi15596966, PseuAe2_09 gi15598969, PseuAe2-18 gi15595518</td>
</tr>
<tr>
<td><em>Pseudomonas putida</em></td>
<td>PseuPu2 NP_747444.1, PseuPu4 gi26991444</td>
</tr>
<tr>
<td><em>Ralstonia solanacearum</em></td>
<td>RalsSo2 NP_519040.1, RalsSo4 NP_519727</td>
</tr>
<tr>
<td><em>Rhodobacter sphaeroides</em></td>
<td>RhodSp1 gi22958568</td>
</tr>
<tr>
<td><em>Shewanella oneidensis</em></td>
<td>ShewOne NP_717423</td>
</tr>
<tr>
<td><em>Sinorhizobium meliloti</em></td>
<td>SinoMe2 NP_284989.1</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>StaphAu1_07 gi21283407 (NP_272259.1 is closely related)</td>
</tr>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>StreCo1 gi7479936</td>
</tr>
<tr>
<td><em>Synechocystis sp. PCC 6803</em></td>
<td>SyneSp1 gi440029.1</td>
</tr>
<tr>
<td><em>Synechococcus sp. WH 8102</em></td>
<td>SyneSp1 gi21133452</td>
</tr>
<tr>
<td><em>Thermus caldophilus</em></td>
<td>TherCa1 gi30025976</td>
</tr>
<tr>
<td><em>Thermostreptobacter tengcongensis</em></td>
<td>TherTe2 NP_622753</td>
</tr>
<tr>
<td><em>Thermotoga maritima</em></td>
<td><em>Thermotoga maritima</em></td>
</tr>
<tr>
<td><em>Thermoxanthus axonopus</em></td>
<td>XantAx2 NP_641221.1</td>
</tr>
<tr>
<td><em>Thermoxanthus campestris</em></td>
<td>XantCa2 NP_636186.1</td>
</tr>
</tbody>
</table>
### Table 6.5

<table>
<thead>
<tr>
<th>Species</th>
<th>NCBI database record</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Archaeoglobus fulgidus</em></td>
<td>ArchFu2 NP_068969</td>
</tr>
<tr>
<td><em>Aeropyrum pernix</em></td>
<td>AeroPv2_83 gi14602183, AeroPv2_12 gi14601312</td>
</tr>
<tr>
<td><em>Halobacterium sp.</em></td>
<td>HaloSp2 NP_279276.1</td>
</tr>
<tr>
<td><em>Methanothermobacter thermautotrophicus</em></td>
<td>MethTh2 NP_276322.1</td>
</tr>
<tr>
<td><em>Methanocaldococcus jannaschii</em></td>
<td>MethJa2 NP_247514.1</td>
</tr>
<tr>
<td><em>Methanopyrus kandleri</em></td>
<td>MethKa2 NP_614338.1</td>
</tr>
<tr>
<td><em>Methanosarcina acetivorans</em></td>
<td>MethAc2 NP_617785</td>
</tr>
<tr>
<td><em>Methanosarcina mazei</em></td>
<td>MethMa2 NP_632091.1</td>
</tr>
<tr>
<td><em>Pyrococcus abyssi</em></td>
<td>PyroAb2 NP_126833.1</td>
</tr>
<tr>
<td><em>Pyrococcus horikoshii</em></td>
<td>PyroHo2, NP_143159.1</td>
</tr>
<tr>
<td><em>Pyrococcus furiosus</em></td>
<td>PyroFu2 NP_578547.1</td>
</tr>
<tr>
<td><em>Pyrobaculum aerophilum</em></td>
<td>PyroAe, NP_560403.1</td>
</tr>
<tr>
<td><em>Sulfolobus solfataricus</em></td>
<td>SulfSo2_79 gi15896979, SulfSo1_99 gi15896999 SulfSo1_78 gi15897978</td>
</tr>
<tr>
<td><em>Sulfolobus tokodaii</em></td>
<td>SulfTo2 NP_378326.1, SulfTo1_38 gi15922638 SulfTo1_98 gi15922198</td>
</tr>
</tbody>
</table>
clusion is preliminary because many groups of archaeabacteria are represented by single genomes. Finally, although HDACs are widely distributed among divergent eubacterial groups, most individual eubacterial species (for example, *Escherichia coli*, *Mycobacterium tuberculosis*) lack recognizable HDACs (Figure 6.4 and data not shown).

6.3.6 HDAC origins

The relationships between class 2 enzymes (Figure 6.4, tree) are broadly consistent with the expected bacterial and archaeabacterial groupings (Figure 6.4, bottom). This observation, together with the existence of class 2 HDACs in all fully sequenced free-living eukaryotes (Figure 6.2), suggests that class 2 HDAC proteins existed in the common ancestor of extant life forms and have been lost by the particular organisms lacking them. One case of apparent gene transfer was observed: the class 2 HDACs of certain *Clostridia* (a type of eubacteria) group at high confidence with archaeabacterial enzymes, suggesting that they obtained these HDACs by horizontal transfer from archaeabacteria (see the placement of TherTe2 and DesuHa2 in Figure 6.4). Other cases of gene transfer may exist, but they are difficult to substantiate because of weak bootstrap support for some interior nodes (see the placement of MethTh2 and HaloSp2), uncertainty in taxonomy, and poor understanding of the order of prokaryotic evolution [74].

Explaining the sparse but diverse distribution of class 1 and class 4 enzymes is more challenging. Three hypotheses can account for this observation: (a) these proteins existed in the common ancestor of eubacteria and eukaryotes and were lost by the large majority of individual species; (b) early eukaryotes obtained these proteins by gene transfer from eubacteria; or (c) eubacteria obtained these pro-
teins by gene transfer from eukaryotes. Whichever scenario is correct, it is likely that additional gene transfers have occurred between bacterial species, providing additional complication. The observation that eukaryotic HDACs of a given class are monophyletic (more closely related to each other than to any prokaryotic proteins) is inconsistent with the third hypothesis. Moreover, the general consistency between the topology of the bacterial HDAC tree and established bacterial taxonomy (Figure 6.4) suggests that if gene transfer was responsible for introducing class 1 and 4 HDACs into prokaryotes, the primary event must have occurred very early in prokaryotic evolution.

Histone-related proteins exist in archaebacteria but appear to be absent from eubacteria: the characterized histone-like proteins in eubacteria are similar to histones in that they are small basic proteins that bind DNA, but differ at both the sequence and structural levels [75–77]. These observations, together with the observed phylogenetic relationships, suggest that all three classes of histone deacetylase enzymes evolved in the absence of histone proteins.

6.4 Concluding Discussion

Whatever their ultimate origin, the observation that all three types of histone deacetylase exist in eubacteria demonstrates that these proteins have functions in the absence of histone proteins. The analysis above suggests that they pre-date the evolution of histones. These conclusions are significant for the function of eukaryotic HDAC proteins because it is unlikely that all three HDAC classes would lose activity on their ancestral substrates (whatever they are) and become uniquely directed at histones in eukaryotes. The probability that HDACs have other physiologically important substrates is further emphasized by observations
(discussed above) that human HDAC6 acts on tubulin and HDAC1 on transcription factors including the tumor suppressor p53 [11, 15, 61, 78]. It is reasonable to speculate that HDACs have additional (perhaps many additional) non-histone protein targets. Possible non-protein substrates that deserve more attention include polyamines and metabolic intermediates. It is interesting to consider the possibility that the pharmacological activity of HDAC inhibitors is partially or even primarily due to effects on these non-histone substrates. This concern becomes even more valid when one considers that a number of the most potent extant anti-cancer drugs (taxol, vinblastine, vincristine) are directed against the cytosolic HDAC substrate tubulin [79].

One recurring theme that emerges from this analysis is the common nature of association between HDAC molecules (see also Grozinger et al. [8], and Verdin et al. [51]). First, it is well established that class 1 and class 2 HDACs are often found together as components of larger complexes. Whether or not this association is direct is in most cases unclear. However, there also appear to be many examples of direct self-association. Dimerization is seen in most human class 2 proteins, yeast HDA1 and HOS3, and even acetylpolyamine amidohydrolase from the eubacterium M. ramose (references given above). The broad phylogenetic distribution of this self-association suggests that it is an ancestral feature of class 2 proteins and leads to the prediction that many class 2 proteins will associate with identical or related proteins. Similarly, all three human class I proteins form dimers/oligomers, suggesting that self-association may be an ancestral feature common among class 1 proteins as well.

What is the functional significance of association between HDAC molecules? Is dimerization (homo or hetero) required for activity? Can a given protein dimer-
ize with more than one partner? If so, does such mixed heterodimerization play a role in regulation, perhaps akin to that seen in the differential activities of different complexes of helix-loop-helix transcription factors? Recognition of the self-associating properties of HDAC proteins may give insight into the function of HDAC gene and domain duplications. We speculate that some duplicated HDAC pairs (e.g. HsHDAC1/2) evolved to be functionally differentiated halves of what originated as a homomeric complex. Duplication of class 2 catalytic domains (seen in HDAC6 and perhaps At9759454) may have occurred as a way of ensuring self-association. As yet the answers to these questions are unclear, but it is interesting to consider them in light of drug development efforts. If, for example, only dimers were active, one could inhibit both members of a heterodimeric pair with a chemical directed against only one. Phenomena such as these may contribute to explaining the surprisingly broad spectrum of most HDAC inhibitors.

6.5 Methods

6.5.1 Database searching and alignment

NCBI nucleotide and protein sequence databases for fully sequenced genomes (November 2002) were scanned for proteins related to human HDAC1 using either PSI-BLAST (protein databases) or tBLASTn (nucleotide databases) \[27\]. After initial sequence collection, the full NCBI database was probed with individual human and yeast HDAC sequences to enhance the chances of finding sequences related to particular divergent HDACs and better define the range of organisms containing these proteins. Sequences were aligned using ClustalX \[24\] with default alignment parameters. Adjustments were made in the resulting initial alignment by asking ClustalX to realign specified sequences (across the entire length) or
regions (all sequences were realigned in the specified region), resulting in an otherwise good alignment that contained unnecessary gaps. The final alignment was obtained by realigning the adjusted alignment after resetting the gaps (in this procedure the guide tree is calculated before the gaps are removed). As described in the text, the eukaryotic alignment and tree contain sequences from the prokaryotic genomes that can be selected on the main PSI–BLAST page. To identify the full set of prokaryotic HDAC relatives, PSSM matrices defined in the eukaryotic PSI–BLAST searches were used to probe the database of fully sequenced prokaryotic organisms accessed at the microbial genomes BLAST page\(^3\). HDACs from some unsequenced bacteria were identified by similar methods and are included to increase the diversity of organisms represented on the tree. An alignment of these bacterial sequences and selected eukaryotic HDACs (yeast and human) was built as described for the eukaryotic sequences.

6.5.2 Tree building

Phylogenetic analysis was performed on the conserved core of these alignments (corresponding to amino acid residues 11319 of HsHDAC1) by the neighbor joining algorithm of ClustalX \(^{24}\) using default parameters (gapped regions were included). When sets of sequences >95% identical were identified in the same organism, only one sequence was included to avoid analysis of duplicate sequences. Sequences failing to align along the length of the core HDAC domain (and therefore containing potential sequencing/splicing artifacts) were also excluded. Bootstrap analysis (1000 trials) provided a measure of confidence for the detected relationships as described above. The resulting tree was graphed by the program unrooted provided with ClustalX, and was prepared for presentation by Adobe Illustrator

\(^3\)http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi
9.01. Phylogenetic analysis was also performed by MrBAYES 2.01 [26] using the same alignment. After the initial burn-in phase the program was allowed to run for $1 \times 10^6$ generations saving one tree every 100 generations. The resulting sample of $1 \times 10^4$ trees were loaded into PAUP 4.0 and a 50% majority rule consensus tree was generated. MrBayes followed by PAUP analysis was repeated five times to ensure that the consensus tree did not depend on the starting random trees. No significant changes in the topology of the trees were observed when comparing MrBayes to ClustalX results, although differences in statistically unsupported regions of the trees were common.

6.6 References


173


41. L Chen, W. Fischle, E. Verdin, and W. C. Greene. Duration of nuclear nf-

42. Wolfgang Fischle, Franck Dequiedt, Michael J Hendzel, Matthew G Guenther,
Mitchell A Lazar, Wolfgang Voelter, and Eric Verdin. Enzymatic activity as-
associated with class ii hdacs is dependent on a multiprotein complex containing

43. Jenny J Tong, Jianhong Liu, Nicholas R Bertos, and Xiang-Jiao Yang. Iden-
tification of hdac10, a novel class ii human histone deacetylase containing a

44. Y. Takami and T. Nakayama. N-terminal region, c-terminal region, nuclear
export signal, and deacetylation activity of histone deacetylase-3 are essential
for the viability of the dt40 chicken b cell line. *J Biol Chem*, 275(21):16191–

45. D. M. Hillis, J. P. Huelsenbeck, and C. W. Cunningham. Application and

46. D. Kadosh and K. Struhl. Repression by ume6 involves recruitment of a
complex containing sin3 corepressor and rpd3 histone deacetylase to target

47. W. W. Pijnappel, D. Schaft, A. Roguev, A. Shevchenko, H. Tekotte, M. Wilm,
G. Rigaut, B. Sraphin, R. Aasland, and A. F. Stewart. The s. cerevisiae set3
complex includes two histone deacetylases, hos2 and hst1, and is a meiotic-

48. T. Srikantha, L. Tsai, K. Daniels, A. J. Klar, and D. R. Soll. The hi-
stone deacetylase genes hda1 and rpd3 play distinct roles in regulation
of high-frequency phenotypic switching in candida albicans. *J Bacteriol*,

Homo-oligomerisation and nuclear localisation of mouse histone deacetylase

50. C. A. Hassig, J. K. Tong, T. C. Fleischer, T. Owa, P. G. Grable, D. E. Ayer,
and S. L. Schreiber. A role for histone deacetylase activity in hdac1-mediated
transcriptional repression. *Proc Natl Acad Sci U S A*, 95(7):3519–3524, Mar
1998.


APPENDIX A

PROGRAMS

A.1 Simulation code for the modeling of microtubules as simple linear polymers

The code below is written in C. It simulates microtubules as simple linear polymers competing for the pool of soluble tubulin heterodimers. It is designed to limit the length of the microtubules to a maximum value so that cell confinement is recapitulated.

/* sm03.c Simple Model (of microtubule dynamic instability)
   * by Ivan Gregoretti */

#include <stdio.h>
#include <stdlib.h>
#define ROW 128 /* number of rows, this is the number of microtubules */
#define COLM 512 /* number of columns, this is the maximum MT size */
#define TUCONC 10 * 1.0 /* initial number of tubulin units per node */
#define STEPS 20000 /* number of steps to run the simulation */
#define SEEDSIZE 5 /* number of units that make up a seed */

/* Parameters of growth, shrinkage and hydrolysis */
float Kgtpright = 0.04; /* GTP, growth */
float Kgdpright = 0.0020; /* GDP, growth */
float Pgdpleft = 0.96; /* GDP, shrinkage */
float Pgtptogdp = 0.01 * 2.0; /* hydrolysis */
double brick = 10; /* number of monomers used per step of growth */

long int n = ROW; /* number of rows, this is the number of microtubules*/
long int m = COLM; /* number of columns */

short int M[ROW][COLM];
unsigned short int E[ROW];

/* Building the main matrices M and E */
int buildM() {
    long int a, b; /* rows and columns counters respectively */
    for (a = 0; a < n; ++a)
        M[a][0] = 2;
    for (a = 0; a < n; ++a) {
        for (b = 1; b < m; ++b)
            M[a][b] = 0;
    }
    return(0);
}

int buildE() {
    int a;
    for (a = 0; a < n; ++a)
        E[a] = 0;
    return(0);
}

/* Printing M and E to the screen */
int printMtoscreen() {

long int a, b; /* rows and columns counters respectively */
for (a = 0; a < n; ++a) {
    for (b = 0; b < m; ++b)
        printf("%hd", M[a][b]);
    printf("\n");
}  
return(0);

int printEtoscreen() {
    int a;
    for (a = 0; a < n; ++a)
        printf("%hd\n", E[a]);
    return(0);
}

/* Setting up the initial concentration of tubulin */
double SumTu = ROW * COLM * TUCONC; /* total amount of units in the system */
double DeltaSumTu = 0;

/**********************
* End of initializations
* *****************************/

/* Tubulin concentration look up function */
float Tu(int x, int EndPlusOne) {
    /* the value of x is discarded at the moment because this code assumes
    * instant diffusion */
    float coef = 1.0; /* coefficient to modify the activity of soluble Tubulin.
    * It will reduce the the probability of growth when the
    * MT is small thus giving the model a cooperative behaviour.*/
    if (EndPlusOne < SEEDSIZE) coef = 0.5;
return(coef * (float)SumTu / (float)(n * m));

/* the "central" function is the one that makes the microtubule either grow,
 * shrink or pause. */
int central() {
    int a; /* just a counter for the "for" loop */
    for (a = 0; a < n; a++) {
        int dice2 = n * drand48(); /*picking a microtubule randomly */
        float dice3 = drand48();
        /* First: the evolution of the mt with a GTP tip */
        if (2 == M[dice2][E[dice2]]) {
            if (dice3 < (Kgtpright * Tu(dice2, E[dice2] + 1))) {
                /* Next "if" prevents growth beyond maximum */
                if ((m - 2) > E[dice2]) {
                    E[dice2] += 1;
                    M[dice2][E[dice2]] = 2;
                    DeltaSumTu -= brick;
                }
            }
        } else if (dice3 < Pgdpleft) {
            /* Next "if" prevents shrinkage below minimum */
            if (0 < E[dice2]) {
                E[dice2] -= 1;
                M[dice2][E[dice2] + 1] = 0;
                DeltaSumTu += brick;
            }
        } else if (dice3 < Pgdpleft + Kgdpright * Tu(dice2, E[dice2] + 1)) {
            /* Next "if" prevents growth beyond maximum */
            if ((m - 2) > E[dice2]) {
                /* code */
            }
        }
    }
}

return(coef * (float)SumTu / (float)(n * m));
E[dice2] += 1;
M[dice2][E[dice2]] = 2;
DeltaSumTu -= brick;
}
}
/* Third: updating the amount of tubulin in the system */
SumTu += DeltaSumTu;
DeltaSumTu = 0; /* just clearing this variable */
}
return(0);
}

/* Hydrolysis of GTP */
int hydrolysis() {
    int a, b; /* just two counters */
    float dice;
    for (a = 0; a < n; a++) {
        for (b = 1; b < m; b++) {
            if (M[a][b] == 2) {
                dice = drand48();
                if (dice < Pgtptogdp)
                    M[a][b] = 1;
            }
        }
    }
    return(0);
}

/* Printing to text file some input and output variables */
/* This function outputs:
   1)total [Tu] 2)hydrolysis probability 3)# MTs 4)Boxsize

184
5) final soluble [Tu] 6) number of tips with GTP 7) final mean MT length 8-and-up) final length of all MT. */

int printIOvariables() {

    /* Calculating soluble tubulin concentration: */
    double SolTuConc;
    SolTuConc = SumTu / (double)(n * m);
    /* Calculating the average MT length */
    int a;
    int NumberTipsGTP = 0;
    for (a = 0; a < n; ++a) if (M[a][E[a]] == 2) NumberTipsGTP += 1;
    double SumOfLengths = 0.0;
    double MeanMTLength;
    for (a = 0; a < n; ++a) SumOfLengths += E[a];
    MeanMTLength = SumOfLengths / (double)(n);
    /* Writing the file */
    FILE *iovars;
    iovars = fopen("iovars.txt", "a");
    fprintf(iovars, "%3.1f %4.3f %3ld %5ld %4.2f %d %6.1f", TUCONC,
            Pgtptogdp, n, m, SolTuConc, NumberTipsGTP, MeanMTLength);
    for (a = 0; a < n; ++a) fprintf(iovars, " %d", E[a]);
    fprintf(iovars, "\n");
    fclose(iovars);
    return(0);
}

int main() {

    int a; /* just a counter for the "for" loop */
    long seed = 0;
    srand48(seed);

    return(0);
}
/ * File output declarations */
FILE *file_M;
int write_M;
file_M = fopen("M.dat", "w");
FILE *file_E;
file_E = fopen("E.dat", "w");
int write_E;

buildM();
buildE();
printMtoscreen();
printEtoscreen();
for (a = 0; a < STEPS; ++a) {
    central();
    hydrolysis();
    write_M = fwrite((char *)&M, 1, sizeof(M), file_M);
    write_E = fwrite((char *)&E, 1, sizeof(E), file_E);
}
printMtoscreen();
printEtoscreen();
fclose(file_M);
fclose(file_E);
printIOvariables();

return(0);
}

The code above is compiled at the Linux command line by issuing the following instruction

gcc -D_GNU_SOURCE -D_FILE_OFFSET_BITS=64 -D_LARGEFILE_SOURCE64 -o sm03 sm03.c
and then the program is executed by entering

`.sm03`

at the command line without arguments. The program is not designed to take arguments but instead to be properly edited at the code level, followed by compilation. Execution of the program generates three files: `E.dat`, `M.dat`, and `iovars.txt`. The file `E.dat` contains a snapshot of the E vector at every simulation step. The E vector stores the length of all simulated microtubules. Analogously, the `M.dat` contains a snapshot of the array M every simulation step. M stores the state of all units along the body of all microtubules. The state can be either GTP-Tu or GDP-Tu. The `iovars.txt` file contains the most important input variables and some output variables. The identity of those variables is described as a comment in the code and it depends on the simulation conditions.

After the last step of the simulation, the program `sm03` is designed to write the data into `iovars.txt` as a single, usually long line. The program `sm03` opens `iovars.txt` in `append` mode, this means that if `sm03` is executed multiple times in the same directory, the data is appended to `iovars.txt` as a new line every time. The file `iovars.txt` is never deleted by `sm03`.

With the default parameters, the simulation will run for 20000 equally spaced steps and then stop. It should be noted that the outputted `M.dat` file will be 2500 MB. If back up of `M.dat` is desired, the user might prefer running a compression algorithm. For example, running

`bzip2 -v9f M.dat`

generates `M.dat.bz2`, a 6 MB losslessly compressed version of the 2500 MB `M.dat`. 
A.2 Visualization

Once a simulation’s execution has been finished, the results can be visualized using Matlab (Mathworks). A Matlab script with a graphical user interface is provided below.

```matlab
% readGUI.m
% by Ivan Gregoretti

clf
clear all

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% build the GUI
% define the plot button
plotbutton=uicontrol('style','pushbutton',...
    'string','Run', ...
    'fontsize',12, ...
    'position',[20,190,50,20], ...  
    'callback', 'run=1;');

% define the stop button
erasebutton=uicontrol('style','pushbutton',...
    'string','Stop', ...
    'fontsize',12, ...
    'position',[20,160,50,20], ...  
    'callback','freeze=1;');

% define the Quit button
quitbutton=uicontrol('style','pushbutton',...
    'string','Quit', ...
number = uicontrol('style','text', ...
    'string','1', ...
    'fontsize',12, ...
    'position',[20,100,50,20]);

medianlength = uicontrol('style','text', ...
    'string','1', ...
    'fontsize',12, ...
    'position',[20,70,50,20]);

meanlength = uicontrol('style','text', ...
    'string','1', ...
    'fontsize',12, ...
    'position',[20,40,50,20]);

SDlength = uicontrol('style','text', ...
    'string','0', ...
    'fontsize',12, ...
    'position',[20,10,50,20]);

ROW = 128;
COLM = 512;

file_M = fopen('M.dat', 'r');
Mtemp = fread(file_M, [COLM, ROW], 'ushort=>double');
fclose(file_M);
M = Mtemp;
file_E = fopen('E.dat', 'r');
Etemp = fread(file_E, [1, ROW], 'ushort=>double');
fclose(file_E);
E = Etemp';

% setting the counter of the "while if" loop
a = 0;

% build an image and display it
A = zeros(1000, 3);
s1 = subplot(2,1,1);
imh1 = image(30*M);
set(imh1, 'erasemode', 'none')
axis image off
s2 = subplot(2,1,2);
imh2 = plot(A);
set(imh2, 'erasemode', 'none')
axis([0 1000 0 COLM]);

%the loop of the program

% Main event loop
stop = 0; % wait for a quit button push
run = 0; % wait for a draw
freeze = 0; % wait for a freeze

while (stop==0)
    if (run==1)

        file_M = fopen('M.dat', 'r');
fseek(file_M, a * ROW * COLM * 2, 'bof');
Mtemp = fread(file_M, [COLM, ROW], 'short=>double');
fclose(file_M);
M = Mtemp';
file_E = fopen('E.dat', 'r');
fseek(file_E, a * ROW * 2, 'bof');
Etemp = fread(file_E, [1, ROW], 'short=>double');
fclose(file_E);
E = Etemp';
a = a + 1;

% draw the new image
% first subplot
set(imh1, 'cdata', (30*M))
% now the second subplot
B = A(2:1000,:);
B(1000,:) = [median(E) mean(E) std(E)];
A = B;
imh2 = plot(A);
axis([0 1000 0 COLM]);
% update the step number display
stepnumber = 1 + str2num(get(number,'string'));
set(number,'string',num2str(stepnumber))
stepmedianlength = median(E);
set(medianlength, 'string', sprintf('%4.1f', stepmedianlength))
stepmeanlength = mean(E);
set(meanlength, 'string', sprintf('%4.1f', stepmeanlength))
stepSDlength = std(E);
set(SDlength, 'string', sprintf('%4.1f', stepSDlength))

end;
if (freeze==1)
    run = 0;
    freeze = 0;
end;
drawnow
end;

To run this script, open Matlab and cd to the directory where E.dat and M.dat are located. Executing the script opens a figure window with three buttons to Run, Stop and Quit the visualization (see Figure [A.1]). There are also four parameters that are updated and displayed at each step: step number, microtubule length median, microtubule length mean and microtubule length standard deviation. The bottom sub-figure shows the history of these parameters for the last 1000 steps. The top sub-figure displays all microtubules with their states clearly marked (red for GTP-Tu and green for GDP-Tu).

A.3 Analysis of the length of microtubules

Most of the analyses of microtubule dynamics is done through measurements of their lengths as a function of time. These analyses range from very simple to very complex. A good numerical analysis tool such as Matlab is then recommended. The following script allows loading the output history of MT lengths into Matlab for all-purpose analysis.

% loadE.m
% by Ivan Gregoretti.

MaximumMT = 128; % number of microtubules
MaxLengthMT = 512; % maximum number of subunits in the polymer
Figure A.1: Graphical user interface with Matlab.

InitialStep = 0000; % if "100", it will start reading the snapshot number 101.
Steps = 20000;

SnapshotInterval = 1;
StepsInvitro = floor(Steps/SnapshotInterval);

% Warning the user that reading E.dat takes some time.
disp('Reading E.dat. This may take a while...')

% reading the data from the binary file
file_Es = fopen('E.dat', 'r');
for i = 1:StepsInvitro
    fseek(file_Es,
InitialStep * MaximumMT * 2 + (i - 1) * MaximumMT * SnapshotInterval * 2,
'bof');
tempEs = fread(file_Es, MaximumMT, 'ushort=>double');
Es(:, i) = tempEs + 1;
end
fclose(file_Es);
save('Es.mat', 'Es');

The script above reads the file \texttt{E.dat} and extracts all the \textit{Es} stored at each step. To run it, open Matlab and \texttt{cd} to the directory where \texttt{E.dat} is located. When the execution of this script is finished, Matlab has loaded in memory all the \textit{E} vectors as a matrix named \textit{Es} and also has written to disk \textit{Es} as a Matlab matrix format file called \texttt{Es.mat}. As an example, a classic life history plot can be drawn by issuing the following sequence of orders at the Matlab command line:

\begin{verbatim}
loadE
T = [1:1:20000];
plot(T, Es(1,:));
\end{verbatim}

The resulting plot is shown in Figure A.2. It should be noted that the curve is drawn for the \textit{first} microtubule, this is, the microtubule whose length is stored first in the \texttt{E.dat} file. For example, to plot the life history of the \textit{second} microtubule, use instead \texttt{plot(T, Es(2,:))}.

The script \texttt{loadE.m} runs rather slowly for two reasons. First, it contains a \texttt{for} loop at its core; Matlab is well known for its inability to process loops fast. Second, the loaded \textit{Es} matrix is generally big and each element must be explicitly cast from \texttt{unsigned short} to the default Matlab data type \texttt{double float}. 

194
Figure A.2: Drawing a life history plot with Matlab.
A.4 Running long simulations

Occasionally are needed simulations much longer than the default. On a Linux operating system, under the default conditions including the 20000 steps, the generated M.dat is 2500 MB and it grows linearly with the number of steps. Depending on the availability of disk space, this M.dat file may become a problem at larger number of steps. Because in many situations the length of the microtubules (E vectors) is what needs to be analyzed rather than the state of the microtubules (M matrix), there is a way to solve this problem. The solution is to modify the code sm03.c so that it does not write to disk M.dat but it still writes E.dat. To modify it, the following lines

```c
/* File output declarations */
FILE *file_M;
int write_M;
file_M = fopen("M.dat", "w");
FILE *file_E;
file_E = fopen("E.dat", "w");
int write_E;

buildM();
buildE();
printMtoscreen();
printEtoscreen();
for (a = 0; a < STEPS; ++a) {
    central();
    hydrolysis();
    write_M = fwrite((char *)&M, 1, sizeof(M), file_M);
    write_E = fwrite((char *)&E, 1, sizeof(E), file_E);
}
```
printMtoscreen();
printEtoscreen();
fclose(file_M);
fclose(file_E);

must be commented out like this

/* File output declarations */
///////////FILE *file_M;
///////////int write_M;
///////////file_M = fopen("M.dat", "w");
FILE *file_E;
file_E = fopen("E.dat", "w");
int write_E;

buildM();
buildE();
printMtoscreen();
printEtoscreen();

for (a = 0; a < STEPS; ++a) {
    central();
    hydrolysis();
    ///////////write_M = fwrite((char *)&M, 1, sizeof(M), file_M);
    write_E = fwrite((char *)&E, 1, sizeof(E), file_E);
}
printMtoscreen();
printEtoscreen();
///////////fclose(file_M);
fclose(file_E);

Although the absence of M.dat makes the visualization by readGUI.m unavailable, the history of microtubule lengths is available through E.dat and loadE.m.