

# Microtubule assembly and disassembly dynamics model: Exploring dynamic instability and identifying features of Microtubules' Growth, Catastrophe, Shortening, and Rescue



Evgenii Kliuchnikov <sup>a</sup>, Eugene Klyshko <sup>a,1</sup>, Maria S. Kelly <sup>b</sup>, Artem Zhmurov <sup>c</sup>, Ruxandra I. Dima <sup>b,\*</sup>, Kenneth A. Marx <sup>a</sup>, Valeri Barsegov <sup>a,\*</sup>

<sup>a</sup> Department of Chemistry, University of Massachusetts, Lowell, MA 01854, USA

<sup>b</sup> Department of Chemistry, University of Cincinnati, Cincinnati, OH 45221, USA

<sup>c</sup> KTH Royal Institute of Technology, Stockholm, Sweden

## ARTICLE INFO

### Article history:

Received 26 October 2021

Received in revised form 26 January 2022

Accepted 27 January 2022

Available online 31 January 2022

## ABSTRACT

Microtubules (MTs), a cellular structure element, exhibit dynamic instability and can switch stochastically from growth to shortening; but the factors that trigger these processes at the molecular level are not understood. We developed a 3D Microtubule Assembly and Disassembly DYnamics (MADDY) model, based upon a bead-per-monomer representation of the  $\alpha\beta$ -tubulin dimers forming an MT lattice, stabilized by the lateral and longitudinal interactions between tubulin subunits. The model was parameterized against the experimental rates of MT growth and shortening, and pushing forces on the Dam1 protein complex due to protofilaments splaying out. Using the MADDY model, we carried out GPU-accelerated Langevin simulations to access dynamic instability behavior. By applying Machine Learning techniques, we identified the MT characteristics that distinguish simultaneously all four kinetic states: growth, catastrophe, shortening, and rescue. At the cellular 25  $\mu\text{M}$  tubulin concentration, the most important quantities are the MT length  $L$ , average longitudinal curvature  $\kappa_{long}$ , MT tip width  $w$ , total energy of longitudinal interactions in MT lattice  $U_{long}$ , and the energies of longitudinal and lateral interactions required to complete MT to full cylinder  $U_{long}^{add}$  and  $U_{lat}^{add}$ . At high 250  $\mu\text{M}$  tubulin concentration, the most important characteristics are  $L$ ,  $\kappa_{long}$ , number of hydrolyzed  $\alpha\beta$ -tubulin dimers  $n_{hyd}$  and number of lateral interactions per helical pitch  $n_{lat}$  in MT lattice, energy of lateral interactions in MT lattice  $U_{lat}$ , and energy of longitudinal interactions in MT tip  $u_{long}$ . These results allow greater insights into what brings about kinetic state stability and the transitions between states involved in MT dynamic instability behavior.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Microtubules (MTs) are essential for many fundamental processes involved in the organization and dynamics of substructures necessary for the health and viability of different eukaryotic cell types [1]. MTs are long, micrometer sized, hollow cylinders consisting of a lattice that can be formed from 9 to 16 laterally aggregating protofilaments [2]. Each protofilament strand is comprised of linearly arrayed  $\alpha\beta$ -tubulin dimers, interacting noncovalently via so-called longitudinal interactions. Within non-dividing cells,

MTs help form the cytoskeleton, a multi-protein poly-filament interlinked superstructure comprised of MTs, intermediate filaments and actin fibers, that helps to organize and dynamically alter cells in response to their environment [3]. The MTs interact with many types of accessory proteins and microtubule associated proteins (MAPs), that can alter the structural and dynamic properties of the MT [4]. MTs possessing a different number of laterally aggregated protofilaments have been observed in different cell types and *in vitro* [5]. However, the 13 protofilament MTs appear to be dominant, representing the presumptive major canonical biological structure [2,6,7]. Another important MT function within most cell types involves the transport of molecular cargo along tracks on its surface by a variety of ATP dependent motor proteins [8]. For example, in neural origin cells, the prominent tau protein binds multiple  $\alpha\beta$ -tubulin dimers along a protofilament [9] that

\* Corresponding authors.

E-mail addresses: [dimari@ucmail.uc.edu](mailto:dimari@ucmail.uc.edu) (R.I. Dima), [Valeri\\_Barsegov@uml.edu](mailto:Valeri_Barsegov@uml.edu).

<sup>1</sup> Current address: Department of Physics, University of Toronto, Toronto, ON M5S 1A7, Canada.

helps: to greatly stabilize MTs, to the point of promoting and stabilizing lattice defects, to promote MTs' lateral aggregation behavior [10], and to provide long neural processes, traversed by such bundled MTs in association with specialized proteins ability to transport molecular cargo over considerable distances (tens of micrometers) [11].

The ability of MTs to undergo stochastic cycles of growth and shortening, a property called dynamic instability, is critically important in many cellular processes [12–16]. A prime example is cell mitosis, when dividing cells must accurately segregate their duplicated chromosomes into two identical daughter cells prior to the cell dividing. MTs are principal components of the spindle apparatus that organize this process. When the chromosomes are correctly organized and positioned, depolymerizing MTs exert pulling forces on chromosomes [17] to bring about accurate, error free, segregation of the identical duplicated sister chromatids to the daughter cells [18]. Depolymerizing MTs act as a biological motor creating poleward chromosome motion from the hydrolysis of tubulin-bound GTP to GDP that drives destabilization and depolymerization of the MT filaments [19,20]. GTP hydrolysis results in formation of the GDP-bound bent metastable state of the  $\alpha\beta$  dimer, as opposed to the GTP-bound straight stable conformation [18]. Not surprisingly then, MT polymerization is stimulated by the stretching of cells [21]. The  $\alpha\beta$ -tubulin dimer subunit structure has been the target of drugs (colchicine, vinblastine, taxanes, epothilones) designed to bind specific sites on these subunits, leading to arrest of rapid cell growth in cancer cells by abolishing the MTs' ability to segregate chromosomes [22].

During MT assembly, tubulin dimers possessing the GTP-bound  $\beta$ -tubulin subunits are added to protofilaments at the tip of a growing MT with a slight bend at the inter-dimer region. Due to GTP hydrolysis, protofilament bound dimers can transform into the more bent GDP bound dimer conformation. In the disassembly or shortening phase, weaker lateral tubulin bonds, responsible for protofilaments' lateral aggregation, dissociate prior to the stronger longitudinal bonds, thereby causing a splaying out of individual protofilaments at the MT end away from the cylindrical axis direction. This creates the so-called "rams' horns" structures, which have been observed in many high resolution electron microscopy and cryotomography studies [23–26]. A number of investigators have measured the kinetics of MT assembly and attempted to discover the rationale for the dynamic instability behavior of MTs (continual transitions from assembly to disassembly and back to assembly), one extreme aspect of which is the so-called catastrophe behavior of MTs, namely the sudden transition from assembly to a significant disassembly event [27–29]. It has been hypothesized that the presence of the GTP cap at the growing MT tip stabilizes the MT structure [13,30]. Due to the stochastic nature of the assembly process, occasionally the tip is depleted of its GTP cap via hydrolysis to GDP, and then the catastrophe event occurs via a mechanism that is unclear [31,32]. MTs' tip conformation and its role in the catastrophe phenomenon remains unclear; thus, they represent important targets for the present and future studies. One view is that the MT grows as an opened cylinder of PFs, which subsequently closes into a tube [33,34]. However, most EM structure studies of the growing MT tip suggest that it is only represented by a closed tapered structure comprised of varying length protofilament extensions.

MTs are critically important structures within cells, in which the interplay between the biomechanics of MT protofilaments and biochemistry of the  $\alpha\beta$ -tubulin heterodimers defines the MT properties; yet, the mechano-chemical aspects of MT dynamics are not well understood. While considerable experimental efforts have been expended to elucidate the mechanisms of MT catastrophe [29], almost nothing is known about the specific factors that trigger the onset of MT rescue, in part because a high-resolution

dynamic view of MTs undergoing growth and shortening is not available experimentally. This, in addition to the complex multi-protofilament structure of MTs, makes it difficult to establish a direct correspondence between the microscopic properties of MTs and MT properties observed experimentally. These experimental limitations necessitate the use of high-resolution computer-based modeling to provide mechanistic insights. Several mechano-chemical properties of the 13 protofilament MT have been explored in a recent *in silico* study of MT indentation [35]. This has provided a high resolution structural and energetics view of the deformation process and has determined the tubulin-tubulin lateral and longitudinal bond strengths between adjacent protofilaments and within a single MT protofilament, respectively.

Due to the complexity of MTs, their large system size (~1,000 tubulin monomers for the purpose of modeling) and the long time-scales involved (from seconds to minutes), *in silico* modeling of MT dynamics is challenging. Different kinetic and mechanical models of MT assembly and disassembly have been reported in the literature [25,28,36–39]. These models vary in the types of representation of molecular processes, and the extent to which they agree with experiments [14]. Here, we briefly describe several of the most elaborate models [28,40,41]. Michaels et al. have developed a model of MT dynamics combining the MT lattice deformation mechanics with the kinetic description of MT growth and tubulin hydrolysis. The authors designed a simulation pipeline based on 3D geometry of MTs, which combines the MT wall deformation mechanics, the kinetics of MT growth and tubulin hydrolysis, and quenched disorder in the longitudinal and lateral mechanical parameters. The results of these studies suggest a possible rescue mechanism via random GTP-islands, i.e. patches of remaining GTP-bound tubulins within the MT wall [40]. Gudimchuk et al. employed a Monte Carlo algorithm to model the kinetics of MT growth, GTP hydrolysis and Brownian dynamics of MT depolymerization, where the motion of each of 13 MT protofilament is constrained to a 2D plane. Gudimchuk et al. explored the effect of activation barriers on MT dynamics, the MT growth dependence on assisting force, and studied mechanisms of generation of pushing force by growing MTs and pulling by shortening MTs [41]. Despite these recent advances in theoretical modeling of MTs, a molecularly detailed, physically realistic, thermodynamically accurate, and complete 3D molecular representation of MT dynamics has not yet been explored. There is a critical need for physically rigorous computational models capable of describing both MT polymerization and depolymerization, as well as transitions from growth to shortening (catastrophe), and from shortening to growth (rescue).

Here, we developed the Microtubule Assembly and Disassembly DYnamics (MADDY) model for Langevin dynamics simulations of the canonical 13 PF MT's 3D assembly and disassembly in biologically relevant times (seconds). The simulated kinetics of these processes compare well with the experimental values for MT rates of assembly and disassembly [33,42,43]. Stalling forces simulated on the Dam 1 ring agree with the experimental force distributions [44,45]. The extent of curving of protofilaments at the + end of MTs to form "rams' horns" structures correlate with experimental data [46,47], and MT tip shape and length agree with experimental EM images [47,48]. To help understand the nature and origins of dynamic instability exhibited by MTs at the  $\alpha\beta$ -tubulin-dimer level of molecular detail, we used the MADDY model to carry out very long simulations (4–5 s) accelerated on Graphics Processing Units (GPU). We observed different kinetic states in the dynamic instability profiles of the MTs, i.e. MT growth, shortening, catastrophe and rescue. Next, we employed a range of Machine Learning techniques, including the Support Vector Machines, Logistic Regression, Random Forest and XGBoost, to identify the molecular parameters that can effectively discriminate between these different kinetic

states based upon 14 different MT characteristics (see Results) quantitated for structures within the individual kinetic states. For the experimentally relevant and intracellular concentration of soluble  $\alpha\beta$ -tubulin heterodimers, the most important features are: the MT length, MT tip width, the total energies of lateral and longitudinal interactions in MT lattice, and the energies of lateral and longitudinal interactions required to complete the MT to a full cylinder. Further refining the importance of these identified MT parameters to characterize the specific kinetic states occurring in dynamic instability of MTs can help guide new experiments to prove these parameters' importance in defining the molecular causes of state stability and of transitions between specific states.

## 2. Results

### 2.1. Microtubule assembly and disassembly DYnamics (MADDY) model

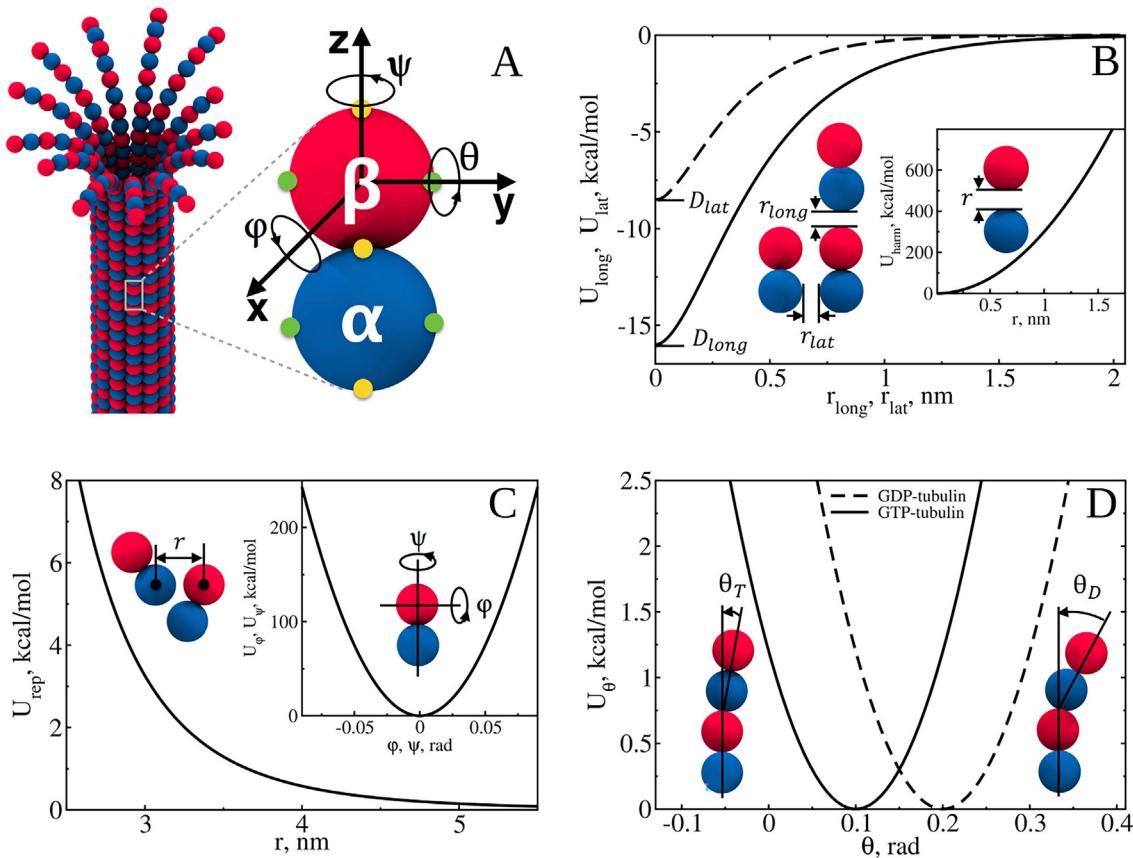
In the molecularly detailed MADDY model, an MT structure is represented by a cylindrical array of tubulin heterodimers arranged into a 13\_3 lattice. Each  $\alpha$ - and  $\beta$ -tubulin monomer is modeled by a spherical bead with radius  $R = 2$  nm. Position and orientation of each monomer in the MT lattice is fully defined by six degrees of freedom: three translational for the coordinates of the subunit's center of mass  $r = \{x, y, z\}$  and three rotational for the orientation angles  $\chi = \{\psi, \theta, \varphi\}$ . These angles correspond to rotations around the MT cylinder axis  $z$ , and the radial axes  $y$  and  $x$ , respectively (Fig. 1A). Each  $\alpha$ - and  $\beta$ -tubulin monomer has four non-covalent interaction sites, enabling them to form bonding interactions with their next-neighbor subunits, which are described by the bond potential  $U_{bond}$  (first term in Eq. (1) in Materials and methods). Each monomer can form two longitudinal (intra-protofilament) non-covalent bonds and two lateral (interprotofilament) non-covalent bonds (Fig. 1A). These are described by the longitudinal and lateral Morse potentials  $U_{long}$  and  $U_{lat}$ , respectively (Fig. 1B), expressed in terms of the binding strength  $D_{long}$  and  $D_{lat}$  and interaction range  $a_{long}$  and  $a_{lat}$  (Eq. (2)). In addition, there is a non-covalent bond between the  $\alpha$ - and  $\beta$ -tubulin monomers within the  $\alpha\beta$ -tubulin dimer described by the harmonic potential  $U_{harm}$  (the inset in Fig. 1B). The total potential energy for non-covalent bonding interactions between the  $\alpha$ - and  $\beta$ -tubulin monomers within the protofilament and between the next-neighbor protofilaments, and within the  $\alpha\beta$ -tubulin dimer  $U_{bond}$  is, therefore, the sum of three potentials  $U_{long}$ ,  $U_{lat}$  and  $U_{harm}$  (Eq. (2)). The MADDY model also takes into account protofilament bending described by the potential energy  $U_{bend}$  (second term in Eq. (1); see also Eq. (3)), which includes the bending potential  $U_\psi$  for monomer rotation around the longitudinal  $z$ -axis, and bending potentials  $U_\theta$  and  $U_\varphi$  for monomer rotation around the  $y$ - and  $x$ -axes, respectively (Fig. 1A).  $U_\theta$  describes protofilaments' splaying outward resulting in formation of "ram's horns" structures (Fig. 1D).  $U_\psi$  and  $U_\varphi$  account for monomers' rotation around the other orthogonal directions (the inset to Fig. 1C). All three potentials are expressed in terms of the trigonometric cosine function with the bending rigidities  $B_\psi$ ,  $B_\theta$  and  $B_\varphi$  and equilibrium bending angles  $\psi_0$ ,  $\theta_0$ , and  $\varphi_0$  (Eq. (3)), which for small angle differences,  $\psi - \psi_0$ ,  $\theta - \theta_0$ , and  $\varphi - \varphi_0$  reduce to the harmonic potentials  $U_\psi \sim (\psi - \psi_0)^2$ ,  $U_\theta \sim (\theta - \theta_0)^2$ , and  $U_\varphi \sim (\varphi - \varphi_0)^2$  (Fig. 1C, D). To reflect the difference in the equilibrium bending angles for the GTP-bound vs GDP-bound  $\alpha\beta$ -tubulin dimer, we set  $\theta_0 = 0.1$  rad and 0.2 rad for the respective GTP-bound and GDP-bound tubulins (Fig. 1D). The MADDY model also accounts for the excluded volume interactions between the tubulin monomers (Fig. 1C) using the repulsive Lennard-Jones potential,  $U_{rep}$ , which depends on the strength of repulsion  $\epsilon$  and the range of repulsion  $\sigma$  (third term in Eq. (1); see also Eq. (4)). The dynamics of MT assembly-

disassembly is generated using the Langevin dynamics in the over-damped (high-friction) limit. This amounts to solving numerically  $6N$  Langevin equations of motion for the coordinates  $r_i = \{x_i, y_i, z_i\}$  (see Eq. (5)) and angles  $\chi_i = \{\psi_i, \theta_i, \varphi_i\}$  (see Eq. (6)) for all particle  $i = 1, 2, \dots, N$  ( $\alpha$ - or  $\beta$ -tubulin monomers). To speed up the simulations, the Langevin dynamics are implemented on a Graphics Processor Unit (GPU) using the numerical algorithms developed in our previous studies [49–51].

### 2.2. MADDY force field parameterization:

The MADDY force-field has 11 parameters described in Materials and methods, which must first be determined before running Langevin simulations. To achieve this goal, we varied the parameters of the MADDY force-field, using the procedure described below, until we found the best match between the experimental and simulation results. We assumed that the equilibrium conformations of the GTP- and GDP-bound tubulin dimer is nearly straight and slightly bent with the equilibrium bending angle  $\theta_0 = 0.1$  rad and 0.2 rad, respectively (Table 1, Fig. 1D). The choice of  $\theta_0 = 0.2$  rad is consistent with the configuration of the "ram's horns" structures at the end of a depolymerizing MT [46,48]. The value of  $\theta_0 \approx 0.1$  rad for the GTP-bound tubulin comes from the analysis of experimental GMPCPP-containing tubulin structures [48,52]. Rotations of monomers around  $\varphi$ - and  $\psi$ -angles are of small amplitude, and so we set  $\varphi_0 = 0$  and  $\psi_0 = 0$  [28] which agrees with electron crystallography [53] and Cryo-EM data [54]. The initial estimate for bending rigidity range  $B_\theta = 1.2 \times 10^3$ – $1.7 \times 10^3$  pN·nm/rad<sup>2</sup> was based upon our previous work on MT indentation *in silico* [35]. Both for GTP- and GDP-bound tubulin dimers, we searched for the best value of  $B_\theta$  in the  $1.4 \times 10^3$ – $2.1 \times 10^3$  pN·nm/rad<sup>2</sup> range using a  $\sim 140$ -pN·nm/rad<sup>2</sup> step. Because the other bending rigidities  $B_\varphi$ ,  $B_\psi$  are not known, we varied the values of these parameters in the  $7.0 \times 10^2$ – $7.0 \times 10^5$  pN·nm/rad<sup>2</sup> range. In our *in silico* indentation studies [35], the longitudinal bonds were found to be stronger than the lateral bonds, and the binding energies come to  $14.9 \pm 1.5$  kcal/mol (longitudinal bond) versus  $6.9 \pm 0.4$  kcal/mol (lateral bond). On the other hand, the tubulin-tubulin bond energies are unlikely to exceed the 20-kcal/mol value known for the biotin-streptavidin bond, which is considered to be the strongest non-covalent bond occurring in biological systems [55]. Therefore, we varied  $D_{long}$  and  $D_{lat}$  in the 10–17-kcal/mol range with a 1 kcal/mole step and in the 5–9-kcal/mol range with a 0.5-kcal/mol step, respectively. These parameters define the depth of potential energy wells for the longitudinal and lateral binding energies. The ranges of longitudinal and lateral interactions are given by  $a_{long}$  and  $a_{lat}$ . The typical protein–protein interaction range, over which the strength of interaction becomes negligible, does not exceed 2 nm [26,56]. Therefore, we spanned the 0.5–2.0-nm interaction range using a 0.5-nm step. We set  $K = 2.1 \times 10^3$  pN/nm for interaction between the  $\alpha$ - and  $\beta$ -tubulin monomers in the same heterodimer [28].

The entire parameter space consists of  $6 \times 10 \times 10 \times 8 \times 9 \times 4 \times 4 = 691,200$  combinations of all possible values of all 11 parameters of the MADDY force field. However, after running test simulations of MT disassembly for the first few hundred sets, we gained intuition on how to correctly tune these parameters, which helped us reduce the number of test simulations from a few hundred thousand to a few thousand tests. For each set, we carried out 10 simulation runs of MT disassembly to attain an acceptable level of statistical significance (Materials and methods). Our first goal was to obtain a reasonable parameterization of the MADDY model that would result in agreement with the experimental 417 nm/s (25  $\mu$ m/min) rate of MT disassembly [14]. The length–time profiles



**Fig. 1. MADDY model:** **A)** The 13 protofilament helical microtubule with splayed protofilament ends forming “ram’s horns” structures is shown along with a representation of the  $\alpha$ - and  $\beta$ -tubulin monomers displayed as rigid spheres with 3 coordinates,  $x$ ,  $y$ ,  $z$ , and 3 angles,  $\psi$ ,  $\theta$ ,  $\varphi$ . The  $\alpha$ - and  $\beta$ -tubulin monomers possess 4 interaction sites: 2 lateral (shown in green) and 2 longitudinal (yellow). The potential energy terms are described in Materials and methods. **B)** Longitudinal and lateral energies  $U_{long}$  and  $U_{lat}$  vs distance between adjacent interaction sites:  $r_{ij} = r_{long}$  (intra-protofilament bond) and  $r_{ij} = r_{lat}$  (inter-protofilament bond), respectively (Eq. (2)).  $U_{long}$  and  $U_{lat}$  are described by the bond-energy scale  $D = D_{long}$  and  $D_{lat}$  and interaction-length scale  $a = a_{long}$  and  $a_{lat}$  to account for the interactions between monomers from different tubulin dimers (Eq. (2)). The inset shows the potential  $U_{har}$  vs distance between longitudinal interaction sites in the same dimer  $r_{ij}$  (Eq. (2)). **C)** Repulsive Lennard-Jones potential  $U_{rep}$  vs distance between monomers’ centers of mass  $r_{ij} = r$  (see Eq. (4)). The inset shows bending potentials  $U_\psi$  and  $U_\varphi$  vs angles  $\psi = \psi_{ij}$  and  $\varphi = \varphi_{ij}$  (Eq. (3)). **D)** Potential energy functions for protofilament outward bending  $U_\theta$  for the GTP- and GDP-bound tubulin dimers vs bending angle between tubulin dimers  $\theta = \theta_{ij}$  (Eq. (3)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 1**

**Optimized parameters of MADDY force field:** the strength of longitudinal and lateral non-covalent bonds between the tubulin monomers  $D_{long}$  and  $D_{lat}$ ; characteristic length scale for longitudinal and lateral interactions  $a_{long}$  and  $a_{lat}$ ; flexural rigidities of protofilament bending around the  $y$ -axis  $B_\theta$  (rotation angle  $\theta$ ),  $z$ -axis  $B_\psi$  (angle  $\psi$ ), and  $x$ -axis  $B_\varphi$  (angle  $\varphi$ ); equilibrium angles for rotation around the  $y$ -axis for GTP- and GDP-bound tubulin dimers  $\theta_0^{GTP}$  and  $\theta_0^{GDP}$ ; equilibrium angles for rotation around the  $z$ - and  $x$ -axes, respectively,  $\psi_0$  and  $\varphi_0$ ; and spring constant for non-covalent bond between the  $\alpha$ - and  $\beta$ -tubulins in the  $\alpha\beta$ -tubulin heterodimer  $K$  (see Materials and methods). The values of  $D_{long}$  and  $D_{lat}$  used in the simulations of MT shortening and MT growth are separated by a slash.

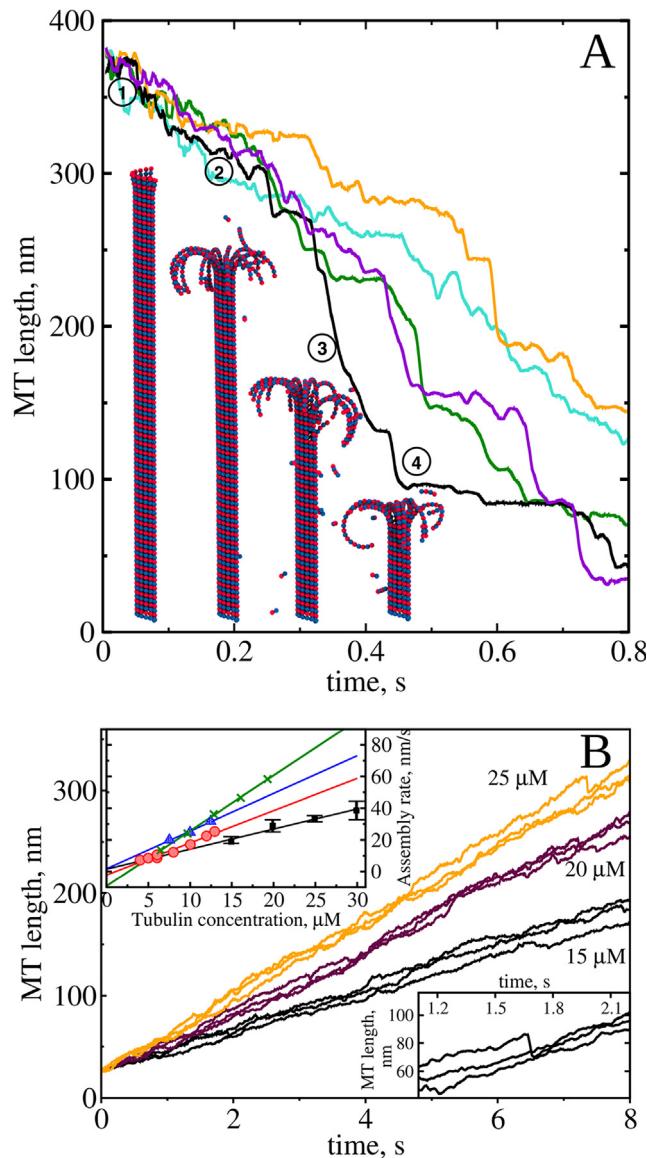
$D_{long}$ , kcal/mol	$D_{lat}$ , kcal/mol	$a_{long}$ , nm $^{-1}$	$a_{lat}$ , nm $^{-1}$	$B_\theta$ , pN·nm/rad $^2$	$B_\psi$ , pN·nm/rad $^2$	$B_\varphi$ , pN·nm/rad $^2$	$\theta_0^{GTP}$ , rad	$\theta_0^{GDP}$ , rad	$\psi_0$ , rad	$\varphi_0$ , rad	$K$ , pN/nm
12.0/16.5	6.7/8.5	3.0	4.0	$1.7 \times 10^3$	$5.6 \times 10^5$	$5.6 \times 10^5$	0.1	0.2	0	0	$2.1 \times 10^3$

for a shortening MT fragment are displayed in Fig. 2A (see also Movie S1). The best agreement with the experimental MT disassembly rate was obtained for the longitudinal and lateral bond energies  $D_{long} = 12.0$  kcal/mol and  $D_{lat} = 6.7$  kcal/mol, respectively (Table 1). Curling protofilaments almost always form singlets [33,46,57,58]. We used this fact to set the lower and upper bounds for  $B_\varphi$  and  $B_\psi$ :  $4.9 \times 10^5$ – $6.3 \times 10^5$  pN·nm/rad $^2$ . Our next goal was to further improve the MADDY model parameterization to generate the dynamics of MT assembly (Materials and methods) with the experimental 17–67 nm/s (1–4  $\mu$ m/min) rate for the 7–15.5  $\mu$ M tubulin concentration range (the inset in Fig. 2B) [14]. The length-time profiles for a growing MT are displayed in Fig. 2B (see also Movies S2 and S3). The best agreement with the experimental MT growth rate was obtained for the longitudinal and lateral bond energies,  $D_{long} = 16.5$  kcal/mol and  $D_{lat} = 8.5$  kcal/mol,

respectively (Table 1). The optimized parameters of MADDY force field are displayed in Table 1, which lists the final values of  $D_{long}$ ,  $D_{lat}$ ,  $a_{long}$ ,  $a_{lat}$ ,  $B_\theta$ ,  $B_\psi$ ,  $B_\varphi$ ,  $\theta_0$ ,  $\psi_0$ ,  $\varphi_0$ , and  $K$ .

### 2.3. MADDY model validation:

We explored the statistics of MT protofilaments’ length and curvature by analyzing the structure of an MT tip from 20 simulation runs (a total of 20 s of biological time) of disassembling MT fragments (Materials and methods). First, we analyzed the statistics of MT protofilaments’ length. The protofilaments were aligned along the MT axis (z-axis; Fig. 1A) such that they bend outward starting from the same point on the y-axis ( $\sim 30$  nm on the y-axis). The histogram of lengths of splayed MT protofilaments is directly compared with the experimental histogram of the same



**Fig. 2. Dynamics of MT assembly and disassembly:** A) MT length vs time plots for MT shortening (see Table 1 for force-field parameters). Each curve represents the results from a MADDY based simulation run. Snapshots of an MT fragment for the black curve are from four different frames, corresponding to the time points indicated by the circled numbers 1–4. The average disassembly rate is  $\sim 410$  nm/s ( $\sim 24.6$   $\mu\text{m}/\text{min}$ ;  $n = 10$  runs; see Movie S1). B) MT length vs time plots for MT growth for 15, 20 and 25  $\mu\text{M}$  concentration of soluble tubulin ( $n = 3$  simulation runs for each tubulin concentration). The top inset shows a comparison of the experimental and theoretical rates of MT growth (black squared) for the 15–30- $\mu\text{M}$  range of soluble tubulin concentration. The average growth rate for 15  $\mu\text{M}$  is 20 nm/s (1.2  $\mu\text{m}/\text{min}$ ); for 20  $\mu\text{M}$  is 30 nm/s (1.8  $\mu\text{m}/\text{min}$ ); for 25  $\mu\text{M}$  is 35 nm/s (2.1  $\mu\text{m}/\text{min}$ ); and for 30  $\mu\text{M}$  is 40 nm/s (2.4  $\mu\text{m}/\text{min}$ );  $n = 10$  runs for each concentration; see Movies S2 and S3. The experimental data points are taken from Refs. [33,42,43]. The bottom inset shows  $\sim 1$ -s long dynamics of MT self-assembly from 3 simulation runs for 25- $\mu\text{M}$  tubulin concentration with transient stoppage and acceleration of growth.

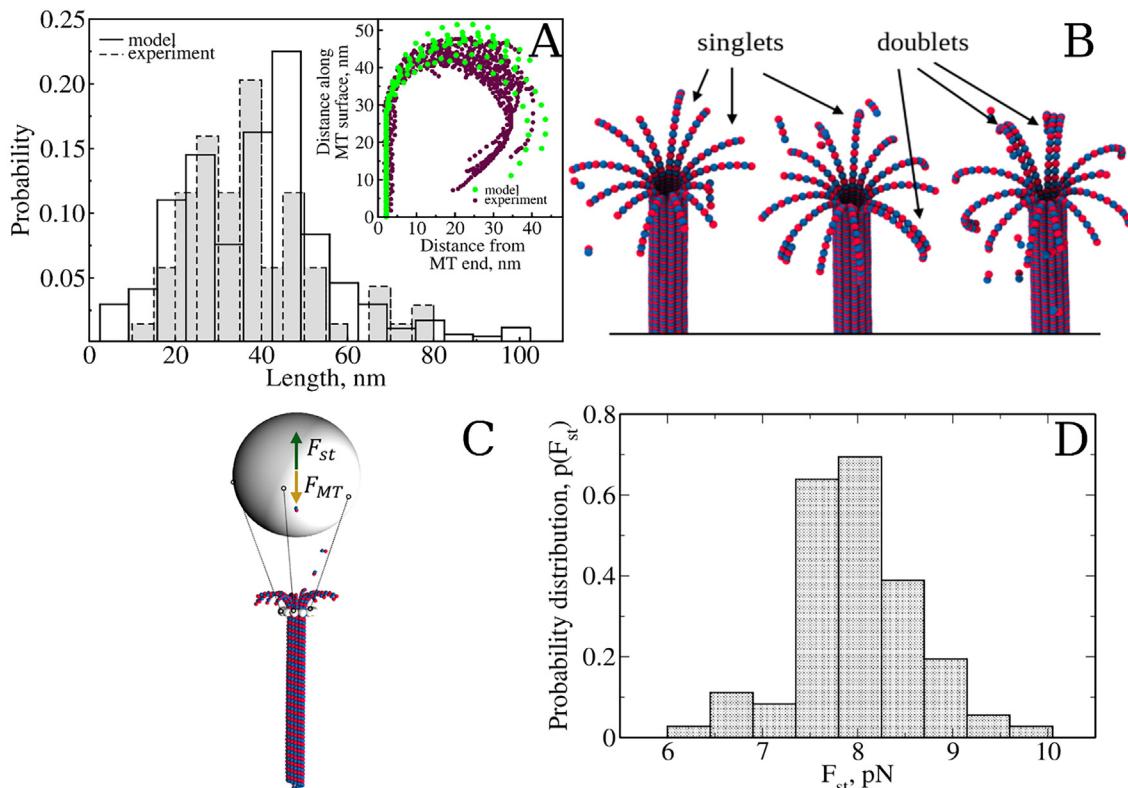
quantity from *in vitro* data reported in Refs. [46,47] and is presented in Fig. 3A. We see that the experimental data and theoretical results for MT protofilament lengths agree very well. The average protofilament length in a disassembling MT come to:  $39 \pm 17$  nm for the MADDY based simulations ( $n = 1280$  protofilaments used in analysis), versus  $40 \pm 15$  nm from experiments ( $n = 72$  protofilaments), which is a very good agreement. The two broad histograms also agree in that both are skewed to the

right. Next, we analyzed the statistics of MT protofilaments' curvature. The scatterplots of the protofilaments' profiles, i.e. the length values measured along the longitudinal  $z$ -axis as a function of the distance from the MT end, from the MADDY based simulations are compared to those from experimental measurements [43,44] in Fig. 3A, which shows very good agreement. The MT protofilaments are capable of bending outward as single structures (singlets), as dimeric structures (doublets), and occasionally as trimeric structure bundles (triplets), where the lateral interactions maintain the nearest neighbor associations along the doublet and triplet protofilaments (Fig. 3B). This agrees with experiment [33,46,57–59].

We also analyzed the force-generating properties of depolymerizing MTs. A disassembling MT must generate sufficiently strong forces to pull apart chromosomes via kinetochore-binding protein complexes such as the Dam1 in yeast. These forces are estimated *in vivo* to be as high as  $\sim 20$  pN with 50% of experiments showing forces  $< 10$  pN [19]. We carried out 80 simulation runs to explore the statistics of stalling forces of disassembling MTs upon the Dam1 ring (Materials and methods). The simulation setup is displayed in Fig. 3C and the histogram of equilibrium stalling forces from a single depolymerizing MT is shown in Fig. 3D. With the flexural rigidity  $B_0 = 1.7 \times 10^3$  pN·nm/rad $^2$  used, a disassembling MT polymer produces an average stalling force of  $8.0 \pm 0.6$  pN ( $n = 80$ ) with the forces reaching as high as 10 pN (Fig. 3D). Thus, the MADDY model describes the force-generating properties of MTs coupled to the Dam1 complex reasonably well. These multiple points of agreement between the results of MADDY based simulations and experimental data, namely the rates of MT growth and shortening, the structure of MT tip, statistics of MT protofilaments' length and curvature, and the force-generating properties of depolymerizing MTs, validate the MADDY model.

#### 2.4. Dynamics of MT disassembly and assembly *in silico*:

Next, we explored the dynamics of MT growth and shortening at the tubulin-monomer level of detail. We first focused on the dynamics of MT disassembly, carrying out a total of 10 simulation runs of MT shortening using the longitudinal bond with the strength  $D_{long} = 12.0$  kcal/mol and the lateral bond with the strength  $D_{lat} = 6.7$  kcal/mol (Table 1). Distance vs time curves for MT disassembly are displayed in Fig. 2A. Overall, MT disassembly is non-monotonic and highly variable with phases of shortening followed by episodes of transient arrest of MT shortening. The duration of stoppage times and transient shortening rates are positively correlated. The longer duration MT shortening-arrested episodes (regions with a near-zero slope in Fig. 2A) are followed by the more rapid disassembly periods (regions with a large negative slope), whereas the shorter length arrested fragments correspond to slower disassembly periods. Hence, these results imply that MT disassembly is a cooperative process. We also observed outward curled "rams' horns" in agreement with the prior studies [46,47] (Fig. 2A, see also Movie S1). The presence of "rams' horns" structures at the end of shortening MT polymers is evident in the structure snapshots 2–4 shown in Fig. 2A and Fig. 3B. This fraying behavior of mostly individual protofilaments results from the lateral bonds being weaker than the longitudinal bonds (Table 1), and also agrees with the previous studies [28,41]. When the MT protofilaments are detached from each other over longer lengths (long "rams' horns") – the most metastable configuration of an MT tip taking a long time to form, the disassembly rate and extent are the largest. From model parameterization, flexural rigidities  $B_\phi$  and  $B_\psi$  affect the frequency of observing doublets and triplets. The higher the values of these quantities, the higher the propensity for singlets' formation in the simulations.



**Fig. 3. Length statistics of curled protofilaments:** **A)** Experimental (dashed bars) vs theoretical (solid bars) histogram-based estimates of the distribution of lengths of the curled protofilaments as observed in 'ram's horn' structures. The inset shows a comparison of the colored dot tracings from the curled protofilaments observed *in vitro* and *in silico* (experimental data were taken from Ref. [47] based on the experiment reported in Ref. [46]). **B)** Structural snapshots showing transient conformations of shortening MT tip. Most of the protofilaments are singlets (single MT protofilaments), albeit doublets (i.e. two laterally associated protofilaments) were also detected in half of the simulation runs ( $n = 50$  runs). **C)** Schematic of the simulation setup used the simulations of depolymerizing MTs (colored cylinder) coupled to the Dam1 ring (white ring) for the calculation of stalling forces. In these simulations,  $F_{MT}$  is the force generated by depolymerizing MT when PFs push on the Dam1 ring and  $F_{st}$  is the opposing stalling force due to the cantilever tip displacement. The equilibrium is reached when the MT disassembly comes to a stall, i.e.,  $\vec{F}_{st} = -\vec{F}_{MT}$ . **D)** Histogram based estimate (bars) of the distribution of stalling forces  $F_{st}$ , generated by a depolymerizing MT on the Dam1 ring, from 80 simulation runs. Freedman-Diaconis rule was used to select the bin size in panels **A** and **D** [94].

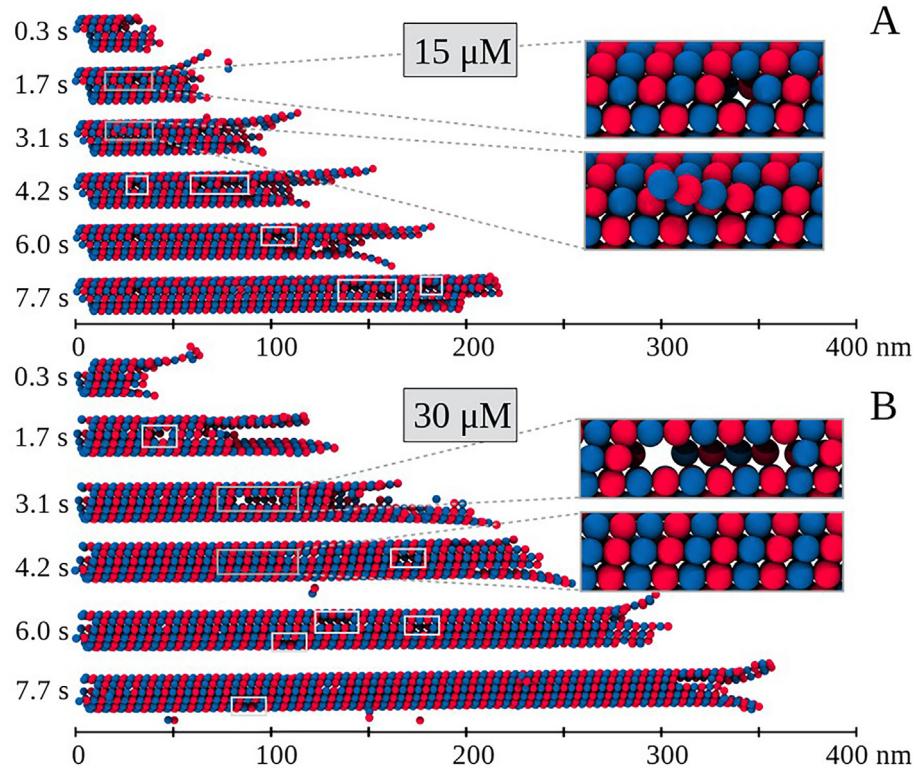
Next, we explored the dynamics of MT growth. We performed a total of 10 simulations of MT assembly in the 15–30  $\mu\text{M}$  range of soluble tubulin concentration, where the binding energies for longitudinal and lateral bonds were set to  $D_{long} = 16.5 \text{ kcal/mol}$  and  $D_{lat} = 8.5 \text{ kcal/mol}$ , respectively (Table 1). MT length vs time profiles are displayed in Fig. 2B. Although the *in silico* rate of MT growth as a function of tubulin concentration has a lower slope than the experimental dependence (top inset in Fig. 2B), given the bead-per-monomer nature of the MADDY model the agreement is very good, especially at lower tubulin concentrations. Similar to MT disassembly behavior, the MT growth is non-monotonic and it proceeds with a variable growth rate, with phases of growth followed by frequent stoppage times (bottom inset in Fig. 2B). To summarize, achieving good agreement with the experimental rates of MT assembly and disassembly (Fig. 2) led us to next explore additional aspects of MT structure and dynamics described below.

### 2.5. MT lattice defects, lattice vacancies, and MT tip structure

Next, we examined dynamic remodeling of the tip of a growing MT. Snapshots of an MT tip from transient MT structures are displayed in Fig. 4 (see also Movie S4). A tip of growing MT is formed by outwardly curved single protofilaments, which gradually straighten when the number of laterally coupled protofilaments increases. We observed formation of "lattice defects", i.e. erroneous attachments formed by tubulin dimers on the MT lattice, and "lattice vacancies", i.e. empty lattice sites lacking tubulin

dimers (Fig. 4 and Movie S4). These lattice defects and lattice vacancies occur dynamically; they form and disappear over time when locally some protofilaments elongate at a more rapid rate than others. These dynamic error features are more frequently observed at higher tubulin concentrations. The lattice defects and vacancies are transient features of MT lattice structure that accumulate within the MT wall and subsequently self-correct over time. A large number of accumulated improper attachments and empty lattice sites results in a transient slowdown or a complete stoppage of MT growth, but when these lattice defects and vacancies are corrected dynamically, through small magnitude, short term fluctuations between assembly and disassembly, normal MT growth resumes.

Next, we examined the MT tip structure, since ideas regarding MT dynamic instability behavior are often thought to originate in properties of the tip or tip-proximal region of MT filament. The uneven length of an MT tip, owing to disparity in the rates of growth of individual protofilaments, are reflected in the probability distribution of extensions of the tip of a growing MT. We defined a tip extension as a length-difference between the shortest and the longest protofilaments at the MT tip. The histogram-based estimates of the probability distributions of MT tip extensions are compared in Fig. 5A for 15  $\mu\text{M}$  and 30  $\mu\text{M}$  tubulin concentration. At higher 30  $\mu\text{M}$  tubulin concentration, the MT-tip extension is more variable because the distribution has a longer tail compared to that for the lower 15  $\mu\text{M}$  tubulin concentration. The average length of an MT tip extension is  $\sim 65 \text{ nm}$  ( $65.6 \pm 30.3 \text{ nm}$ ) for 15  $\mu\text{M}$  tubulin



**Fig. 4. MT assembly and growing MT tip structures:** Snapshots of growing MTs from 8-s simulation runs for 15  $\mu\text{M}$  (panel A) and 30  $\mu\text{M}$  (panel B) concentration of soluble tubulin, taken at the times points: 0.3, 1.7, 3.1, 4.2, 6.0, and 7.7 s. The insets show detailed views of transient structure defects: from the inter-protofilament closure error in the MT lattice and resulting in site vacancy and oligomer erroneously attaching to the MT lattice at a site of vacancy (panel A), and from the assembly error resulting in formation of large vacant sites lacking tubulin dimers before and the same site after dynamic error correction occurs (panel B).

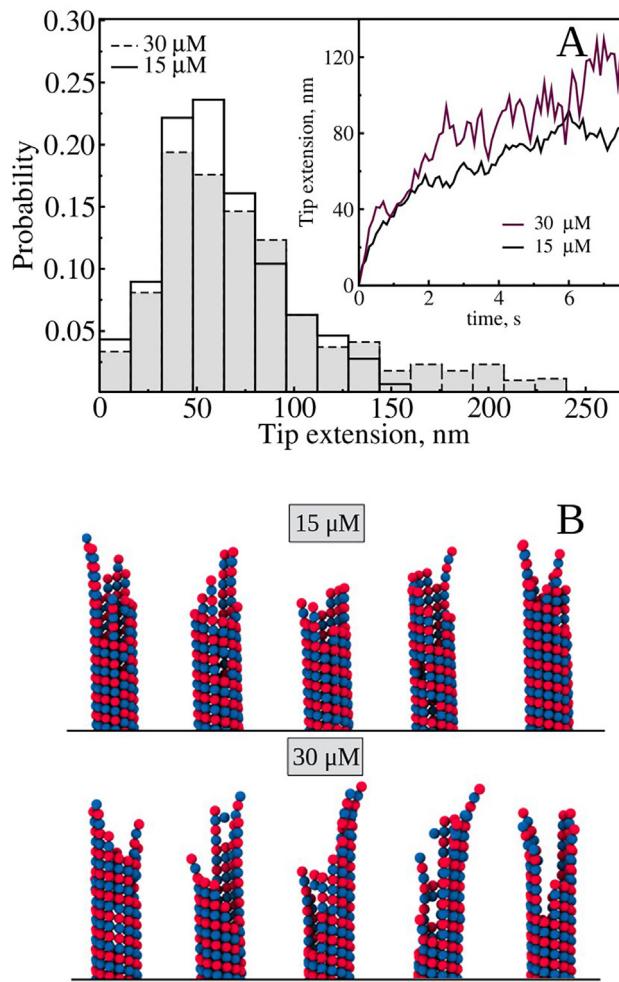
solution vs  $\sim 80$  nm ( $80.5 \pm 49.6$  nm) for 30  $\mu\text{M}$  solution (Fig. 5A). Representative trajectories of tip extension for 15  $\mu\text{M}$  and 30  $\mu\text{M}$  tubulin solution are compared in the inset to Fig. 5A. In the simulations, we observed the curved tips reaching  $\sim 160$ -nm ( $\sim 20$  tubulin dimers) for 15  $\mu\text{M}$  tubulin solution and to  $\sim 245$ -nm in length ( $\sim 30$  tubulin dimers) for 30  $\mu\text{M}$  tubulin solution. These maximal extensions agree with experimental findings [59], which show that the average tip extension is 180 nm long and that extensions shorter than 75 nm represent  $\sim 66\%$  of the whole population. We analyzed transient conformations of the MT tip. The most typical MT tip structures are compared for 15  $\mu\text{M}$  and 30  $\mu\text{M}$  tubulin solutions in Fig. 5B. At the lower 15  $\mu\text{M}$  concentration of soluble tubulin, the MT tip is more blunt, whereas at the higher 30  $\mu\text{M}$  tubulin concentration the MT is more sharp. Hence, the shape of the MT tip is tubulin concentration dependent.

## 2.6. Dynamics of MT assembly-disassembly in silico

Next, we explored the dynamics of MT polymerization-depolymerization (Materials and methods). We carried out 6 sets of simulations – Case Studies 1–6, in which we varied the concentration of soluble tubulin dimers and reaction rate constant for GTP hydrolysis (5 simulation runs for each Case Study; see Table 2). For 25  $\mu\text{M}$  tubulin concentration, we set the GTP hydrolysis rate constant  $k_{hyd}$  to 2  $\text{s}^{-1}$  (Case Study 1), 4  $\text{s}^{-1}$  (Case Study 2), and 6  $\text{s}^{-1}$  (Case Study 3). For 250  $\mu\text{M}$  tubulin concentration, we used  $k_{hyd} = 5 \text{ s}^{-1}$  (Case Study 4), 8  $\text{s}^{-1}$  (Case Study 5), and 10  $\text{s}^{-1}$  (Case Study 6). The experimental  $k_{hyd}$  value is  $\sim 0.5\text{--}1.0 \text{ s}^{-1}$  [28,60], but in our experiments *in silico* we increased  $k_{hyd}$   $\sim 4\text{--}40$  fold. Experimental concentration of soluble tubulin is 10–30  $\mu\text{M}$ , but in our simulations we increased tubulin concentration 1–10-fold (Table 2). Carrying out Case Studies 1–6 at higher tubulin concen-

tration and larger GTP hydrolysis rate constant enabled us to access many transitions between different kinetic states, i.e., from MT growth to MT shortening and from MT shortening back to MT growth, and to observe the MT catastrophe and rescue.

For Case Studies 1 and 6, the MT length vs time trajectories from 4 runs of the MADDY based simulations of MT assembly-disassembly are displayed in Fig. 6, which shows the results of long ( $\sim 1\text{--}4$  s) simulations of the MT dynamics in all four kinetics states – assembly, catastrophe, shortening, and rescue. We detected a total of 34 MT catastrophes and 21 MT rescues for 25  $\mu\text{M}$  tubulin solution, and 41 MT catastrophes and 35 MT rescues for 250  $\mu\text{M}$  tubulin solution (Table 2). We posed a question: what structural and energy properties of MT filament in general and MT tip in particular best describe the MT kinetic states: assembly, catastrophe, shortening, and rescue? We surveyed the following 14 quantitative features: MT length  $L$  (feature 1), number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice  $n_{hyd}$  (feature 2), total energy of lateral interactions in MT lattice  $U_{lat}$  (feature 3), total energy of longitudinal interactions in MT lattice  $U_{long}$  (feature 4), energy of lateral interactions in MT tip  $u_{lat}$  (feature 5), energy of longitudinal interactions in MT tip  $u_{long}$  (feature 6), average longitudinal curvature  $\kappa_{long}$  (feature 7), average lateral curvature  $\kappa_{lat}$  (feature 8), MT tip length  $l$  (feature 9), MT tip width  $w$  (feature 10), average number of longitudinal interactions per protofilament in MT lattice  $n_{long}$  (feature 11), average number of lateral interactions per helical pitch in MT lattice  $n_{lat}$  (feature 12), energy of lateral interactions required to complete MT to full cylinder  $U_{lat}^{add}$  (feature 13), and energy of longitudinal interactions required to complete MT to full cylinder  $U_{long}^{add}$  (feature 14). These 14 features were chosen based upon current ideas of dynamic instability concerning the control of MT growth and shortening. The time-dependent profiles for all 14 features listed above for Case Studies 1 and 6 are presented in Figs. S1 and S2.



**Fig. 5. Evolution of protofilament extensions at the growing MT tip:** A) Histogram-based estimates of the distributions of tip extensions for 15  $\mu\text{M}$  (solid bars) and 30  $\mu\text{M}$  (dashed bars) concentration of soluble tubulin defined as the length-difference between the shortest and the longest protofilaments at the end of the growing MT. The Freedman-Diaconis rule was used to select the bin size [94]. The average extension ( $\text{mean} \pm \text{standard deviations}$ ) comes to  $65.6 \pm 30.3 \text{ nm}$  for 15  $\mu\text{M}$  tubulin concentration ( $n = 970$  data points) and to  $80.5 \pm 49.6 \text{ nm}$  for 30  $\mu\text{M}$  tubulin concentration ( $n = 780$ ). The inset shows the temporal evolution of MT tip extension averaged over 10 simulation runs for 15  $\mu\text{M}$  and 30  $\mu\text{M}$  tubulin concentrations. B) Conformations of a growing MT tip with long sharp extensions curving outward. In the simulations, the curved tips were observed to grow up to 50 nm in length (~6–7 dimers).

### 2.7. Machine Learning: Identifying features that distinguish microtubules' growth, catastrophe, shortening, and rescue in MTs

Next, we employed statistical modeling to identify structural and energetics features of MT filaments described above capable

of distinguishing between the four kinetic states of MT dynamics (assembly, catastrophe, shortening, and rescue); see Materials and methods. Classifier methods such as Random Forests (RF), Gradient Boosting (XGB), and Support Vector Machines (SVMs) have been used in the following biological applications i) to learn the features that distinguish protein sub-cellular localization [61,62], ii) to characterize biological activities of small molecules [63], iii) to predict molecular properties and determine binding drug partners [63,64], and iv) to predict hot spots on protein interfaces that control the binding free energy [65].

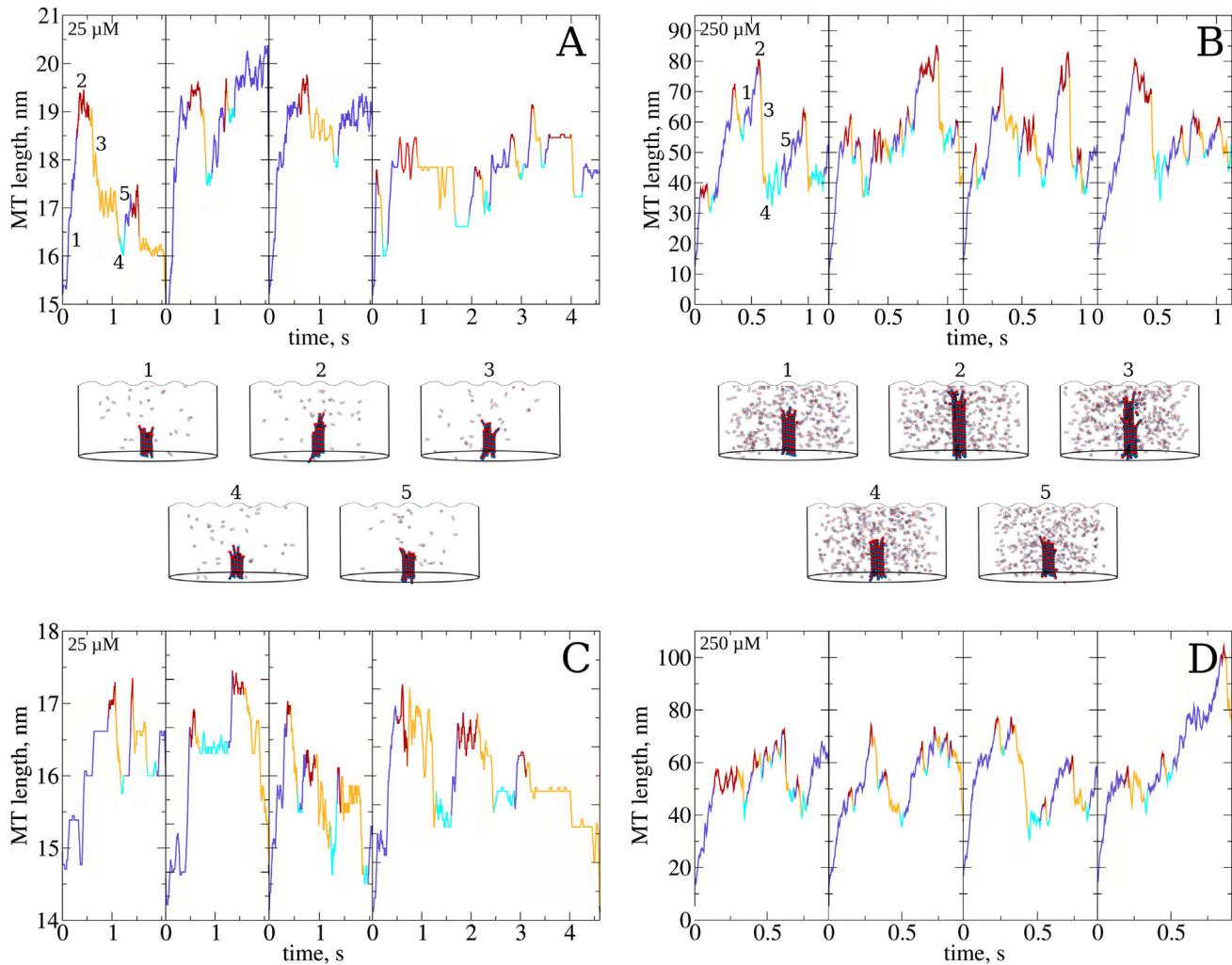
For each Case Study 1–6, we used the distributions of the 14 features from the output of MADDY based simulations (coordinate and energy files) and the information about the four kinetic states. The distributions of these features are displayed in Fig. 7 for Case Study 2. The first step was to determine the prediction accuracy of several classification and ensemble methods in order to find the best method for distinguishing the four kinetic states of MT dynamics. The baseline accuracy, which we evaluated using the Dummy Classifier method, ranged from 33 to 59% for all Case Studies 1–6 based on the size of the largest class, which was typically the state of growth or the state of catastrophe. Since these two classes represented a large percentage of the data compared to the other kinetic states (shortening and rescue), we applied the SMOTE technique to account for the uneven spread of data by replicating samples from the less populated classes [62]. The classifier methods (LDA, SVM, LR, KNN), with and without SMOTE, gave a range of prediction successes. Compared to the baseline accuracy for each Case Study, LR without SMOTE performed the worst (43.1–65.1% accuracy), while SVM and KNN with SMOTE gave the highest accuracy (64.2–84.1%). Next, we used ensemble methods (BAG, RF, XGB), which incorporate the grouping of multiple decision trees. Application of these methods resulted in the highest accuracies out of all methods. The results obtained for Case Study 2 are displayed in Fig. 8 for LDA, SVM, BAG, RF, ET, and XGB (for Case Studies 1, and 3–6 see Figs S3–S7). The accuracy level attained with RF, BAG, and XGB ranged between 74.0 and 87.3% (Table 2). The best models with their associated accuracy level for each Case Study 1–6 are accumulated in Table 2.

Because the application of RF and XGB methods resulted in the best or next to the best prediction accuracy for the majority of Case Studies, we next analyzed the quality of RF and XGB based predictions. To determine the Case Studies, for which the classification of the kinetic states is the most robust, we extracted the confusion matrices, which label the class chosen for each prediction against the true (known) class (Materials and methods). Fig. 9 shows the confusion matrices for XGB for all Case Studies 1–6 (confusion matrices for RF are displayed in Fig. S8). Each matrix compares the predictions to their correct classes, and, for each matrix entry, we report the number of predictions, the total number of samples per class, and the percentage of predictions per total number of samples. Overall, we found that the population of a state for any Case Study correlates with the goodness of prediction from the classification methods. For example, we find a high precision in

**Table 2**

**MT Feature analysis and best prediction models:** Feature analysis performed on the numerical output from  $N_{\text{traj}} = 5$  simulation runs for each Case Study 1–6. Case Studies 1–3 correspond to 25  $\mu\text{M}$  concentration of soluble tubulin dimers and the GTP hydrolysis rate constants of  $k_{\text{hyd}} = 2, 4$ , and  $6 \text{ s}^{-1}$ . Case Studies 4–6 correspond to 250  $\mu\text{M}$  concentration of soluble tubulin dimers and the hydrolysis rate constants of  $k_{\text{hyd}} = 5, 8$ , and  $10 \text{ s}^{-1}$ . Reported for each Case Study are the total number of data points  $N_{\text{dp}}$ , the number of catastrophes  $N_{\text{cat}}$  and rescues  $N_{\text{res}}$  observed (separated by a slash). Summarized are Machine Learning models that have resulted in the highest accuracy compared to all other classification and ensemble methods used.

Case study	$N_{\text{traj}}$	$N_{\text{dp}}$	$N_{\text{cat}}/N_{\text{res}}$	$\alpha\beta\text{-tubulin concentration, } \mu\text{M}$	GTP hydrolysis rate constant $k_{\text{hyd}}, \text{ s}^{-1}$	Best model	Accuracy, %
1	5	3249	16/16	25	2	RF	74.0
2	5	3277	9/7	25	4	RF/XGB	87.1
3	5	3246	9/6	25	6	XGB	76.1
4	5	1718	11/10	250	5	BAG	87.3
5	5	2319	11/8	250	8	XGB	77.2
6	5	2781	19/17	250	10	XGB	76.9

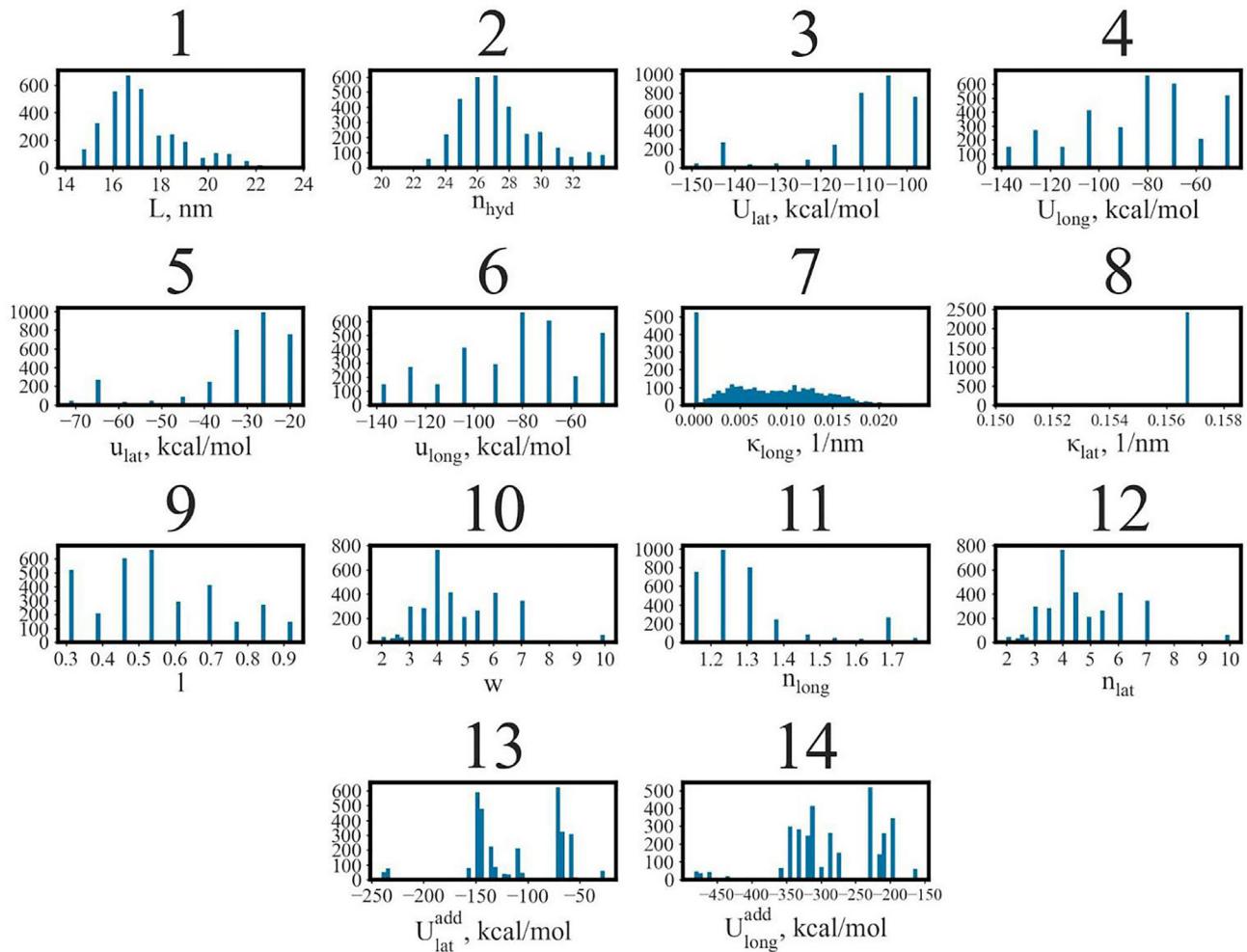


**Fig. 6. Dynamics of MT assembly-disassembly:** MT length vs time profiles for Case Study 1 for soluble tubulin concentration 25  $\mu\text{M}$  and GTP hydrolysis rate constant  $k_{hyd} = 2 \text{ s}^{-1}$  (panel A), Case Study 6 for tubulin concentration 250  $\mu\text{M}$  and  $k_{hyd} = 10 \text{ s}^{-1}$  (panel B), Case Study 3 for tubulin concentration 25  $\mu\text{M}$  and  $k_{hyd} = 6 \text{ s}^{-1}$  (panel C), and Case Study 5 for tubulin concentration 250  $\mu\text{M}$  and  $k_{hyd} = 8 \text{ s}^{-1}$  (panel D). The profiles are shown for 4 independent runs of MT assembly-disassembly for each Case Study. Each profile was divided into portions that correspond to the following kinetic states (classes): MT growth (shown in blue color), MT catastrophe (red), MT shortening (yellow), and MT rescue (cyan); see Materials and methods. In panels A and B, snapshots numbered 1–5 display the time moments of MT for growth (1), MT catastrophe (2), MT shortening (3), MT rescue (4) and MT growth after the rescue (5), as observed in reaction volume. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

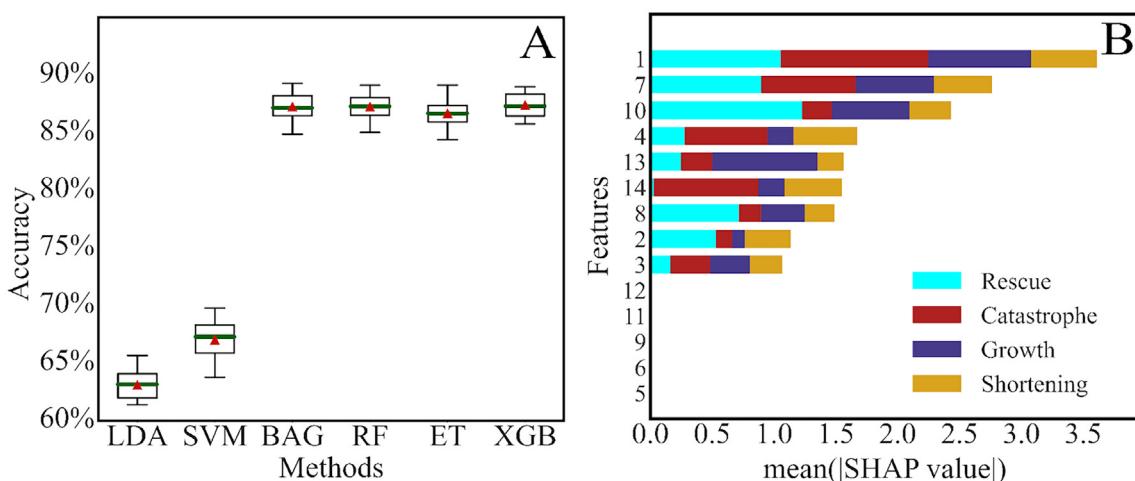
predicting the MT growth or MT catastrophe, depending on which class had the largest population in the respective case. The MT rescue and MT shortening were repeatedly found to be difficult to predict. Case Studies 2 and 4 received the best comprehensive classification when both XGB or RF were used. Next, we calculated the Receiver Operating Characteristic (ROC) curves and the Precision-Recall (PR) curves as performance measures (Materials and methods). Fig. 10 and Fig. S9 display the PR curves for Case Study 2 and Case Study 4, respectively. The ROC curves for Case Study 2 and Case Study 4 are shown in Figs. S10 and S11, respectively. The corresponding values of area under curve (AUC values) are accumulated in Table 3 for the PR curves and in Table S1 for the ROC curves. We constructed the PR and ROC curves for two datasets (Case Study 2 and Case Study 4) to make a comparison between RF, XGB, and SVM methods. Because of the imbalance among the populations of the four states, the results obtained from the PR curves were made a priority, since the PR curves focus on the smaller sets of true positives. The AUC values from the PR plots, which we used to quantify the overall prediction accuracy level, showed that RF and XGB methods classified the kinetic states

better than SVM (Table 3). For example, RF and XGB produced excellent classification for the MT growth and MT catastrophe ( $>0.95$ ) for several Case Studies, with XGB having slightly higher AUC values overall. Also, from the confusion matrices XGB predicted all classes (i.e. kinetic states) more consistently compared to RF. Therefore, we selected XGB for identification of the most important (critical) features for class (kinetic state) discrimination.

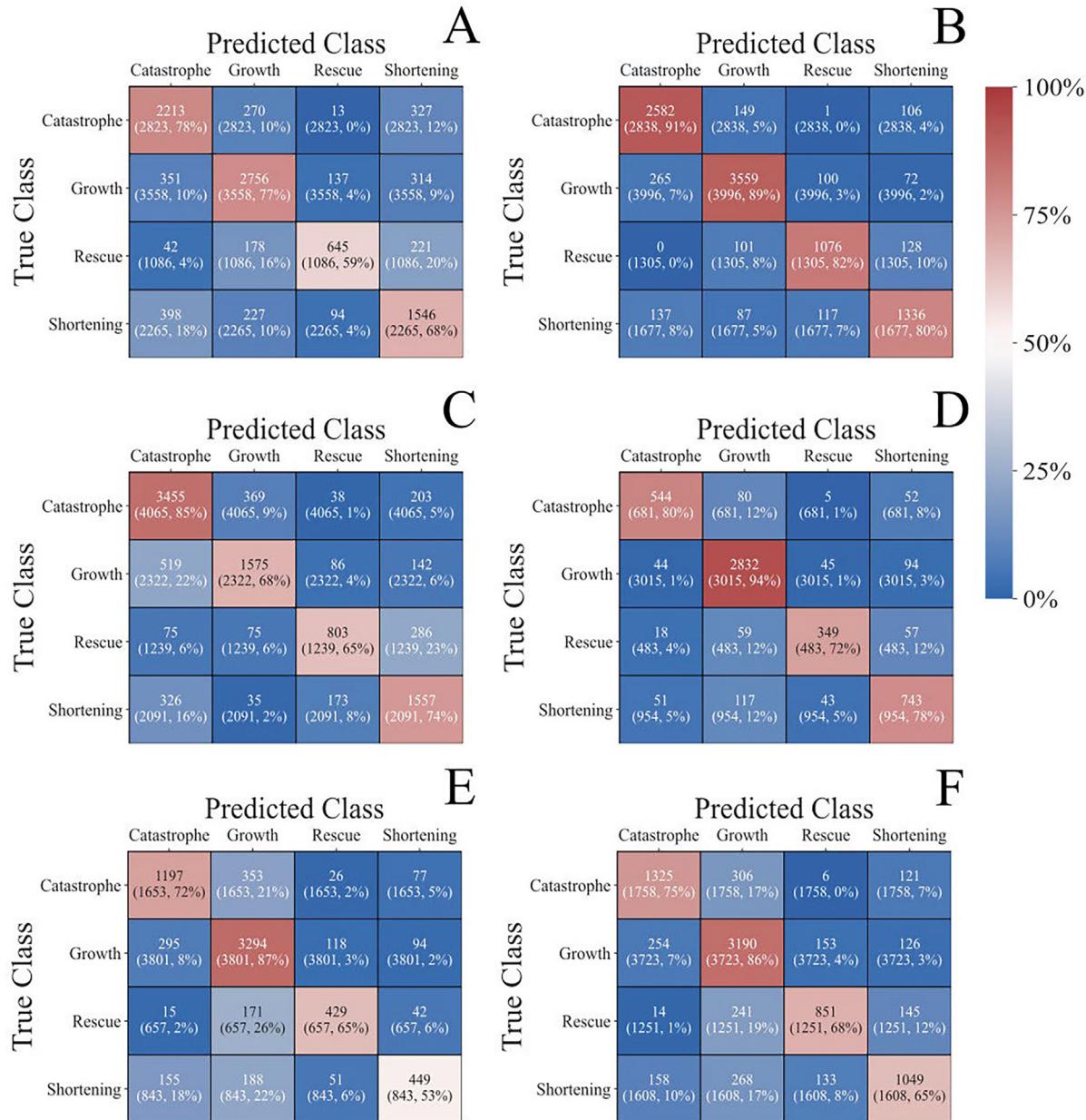
Next, when using XGB for Case Studies 2 and 4 we applied the SHAP analysis [66] (Materials and methods) to determine which factors drive the prediction. Namely, SHAP enables an in-depth view of how the XGB prediction was made by calculating the Shapley values for each feature, which can take both positive and negative values. SHAP determines whether a feature helps with the prediction of a class based on the range and identity of the feature measurements that have positive SHAP values. Using the mean absolute value of all calculated Shapley values, we determined the importance of each feature for MT growth, catastrophe, shortening, and rescue. The plots of feature importance are displayed in Fig. 8 for Case Study 2, and in Figs. S3–S7 for the other Case Studies 1, and 3–6. This method of identifying the most important features



**Fig. 7. Distributions of numerical values of features 1–14 for Case Study 2:** Plots are for each of the following 14 features in statistical inference: (1) MT length  $L$ , (2) number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice  $n_{hyd}$ , (3) total energy of lateral interactions in MT lattice  $U_{lat}$ , (4) total energy of longitudinal interactions in MT lattice  $U_{long}$ , (5) energy of lateral interactions in MT tip  $u_{lat}$ , (6) energy of longitudinal interactions in MT tip  $u_{long}$ , (7) average longitudinal curvature  $\kappa_{long}$ , (8) average lateral curvature  $\kappa_{lat}$ , (9) MT tip length  $l$ , (10) MT tip width  $w$ , (11) average number of longitudinal interactions per protofilament in MT lattice  $n_{long}$ , (12) average number of lateral interactions per helical pitch in MT lattice  $n_{lat}$ , (13) energy of lateral interactions required to complete MT to full cylinder  $U_{lat}^{add}$ , and (14) energy of longitudinal interactions required to complete MT to full cylinder  $U_{long}^{add}$ . These 14 features were chosen based upon current ideas of dynamit instability concerning the control of MT growth and shortening.



**Fig. 8. Machine learning based classification of MT features for Case Study 2:** **A)** Box and whisker plots for comparison of the prediction accuracy for various classification and ensemble methods (LDA, SVM, BAG, RF, ET, and XGB). The minimum value at the bottom of each box represents the first quartile. The third quartile is at the top of the box, which ends at the maximum value. Green lines indicate the median for each method, while red arrowheads indicate the mean values. **B)** Feature importance plots obtained with the XGB method ranking the features that were the most and the least important at predicting the kinetic states using the mean absolute SHAP values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

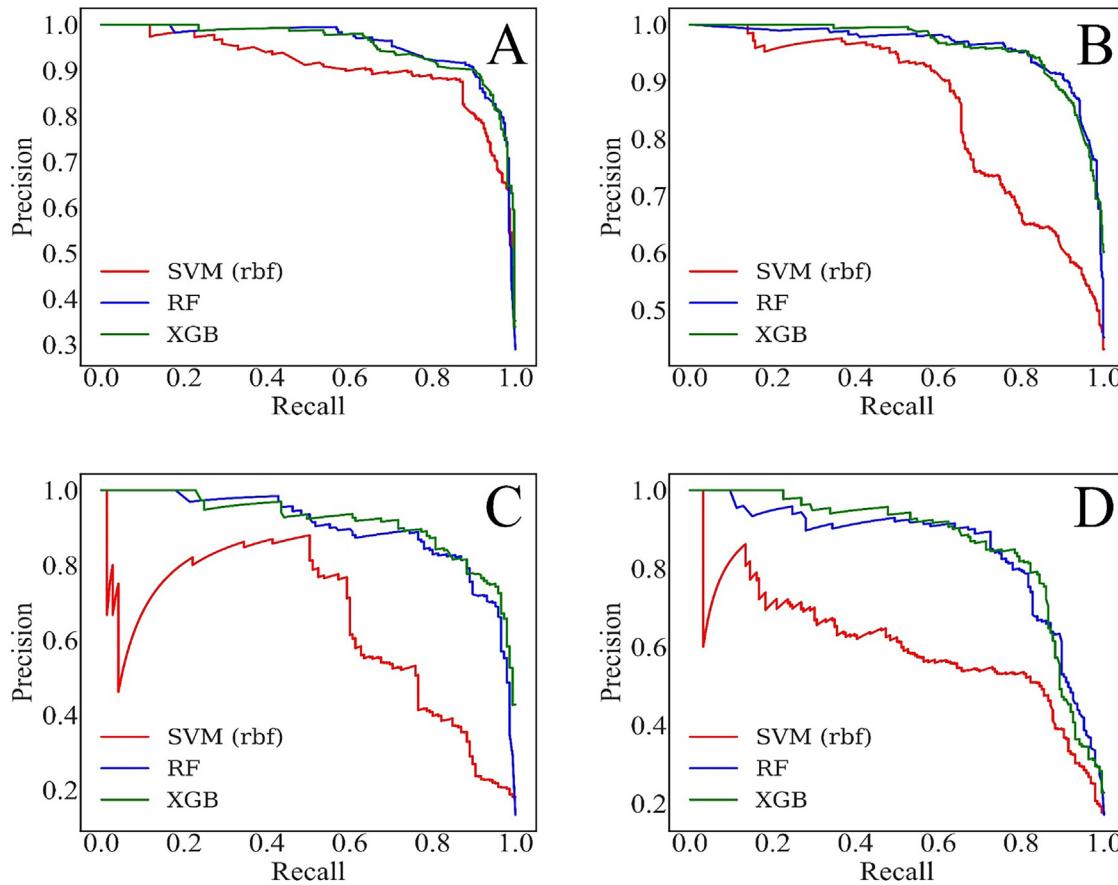


**Fig. 9. Confusion Matrices obtained using XGBoost for all Case Studies 1–6:** Confusion matrices describing the number of predictions made for each of the 5 k-folds (repeated 3 times) using XGB compared to every predictions' true class. Values in the center of each box indicate the number of predictions. In parenthesis, the first value refers to the sum of the number of samples for a class found in each k-fold test set. The second value is the percent of predictions made for a certain class compared to the overall number of true class samples. The heatmap color range is based on this relative percentage for Case Study 1 (**A**), Case Study 2 (**B**), Case Study 3 (**C**), Case Study 4 (**D**), Case Study 5 (**E**), and Case Study 6 (**F**).

was chosen to allow for a consistent and accurate assessment of feature contribution [67]. The most important parts of the analysis are summarized in Table 4, which lists the top most and least important predictors for Case Studies 2 and 4. For both low and high tubulin concentrations, the MT length  $L$  (feature 1) and average longitudinal curvature  $\kappa_{long}$  (feature 7) were among the top important predictors, while the MT tip length  $l$  (feature 9) and average number of longitudinal interactions per protofilament in MT lattice  $n_{long}$  (feature 11) were among the least important features. Other features were also found to be important for Case Study 2, which include the MT tip width  $w$  (feature 10), total energy of longitudinal interactions in MT lattice  $U_{long}$  (feature 4), the energy of longitudinal interactions required to complete MT to full cylinder  $U_{long}^{add}$  (feature 14) and the energy of lateral interac-

tions required to complete MT to full cylinder  $U_{lat}^{add}$  (feature 13). For Case Study 4, we found that the number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice  $n_{hyd}$  (feature 2), average number of lateral interactions per helical pitch in MT lattice  $n_{lat}$  (feature 12), total energy of lateral interactions in MT lattice  $U_{lat}$  (feature 3), and energy of longitudinal interactions in MT tip  $u_{long}$  (feature 6) are the features that had the largest impact in distinguishing between the four kinetic states.

A highly informative output of the SHAP analysis is represented by beeswarm plots, which are depicted for each kinetic state for Case Studies 2 and 4 in Figs. S12 and S13, respectively. Each plot is a dense summary of each feature's contribution in classification of a state by visualizing the distribution and sign of Shapley values for all observations, while also indicating the magnitude of the



**Fig. 10. Classification quality analysis for Case Study 2 with Precision-Recall curves:** PR curves comparing the overall quality value of classification attained using Support Vector Machines with RBF kernel (red curves), Random Forest (blue curves), and XGBoost (green curves) for MT catastrophe (panel A), MT growth (panel B), MT rescue (panel C), and MT shortening (panel D). The closer the curve is to the upper right-hand corner, the better is the model prediction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

**Area under the curve (AUC) for Precision-Recall (PR) curves for Case Studies 2 and 4:** Shown are the AUC values obtained using the SVM (with RBF kernel), RF, and XGB methods. The corresponding PR curves are displayed in Fig. 10 for Case Study 2 and in Fig. S9 for Case Study 4 (AUC value of 1/0 represents a perfect/poor classification).

Kinetic State	Case Study 2			Case Study 4		
	SVM (RBF)	RF	XGB	SVM (RBF)	RF	XGB
Catastrophe	0.91	0.95	0.95	0.43	0.72	0.70
Growth	0.86	0.96	0.96	0.91	0.96	0.96
Rescue	0.64	0.90	0.91	0.34	0.64	0.63
Shortening	0.60	0.84	0.86	0.56	0.73	0.74

**Table 4**

**Feature importance comparison:** Feature importance was calculated based on the results obtained with XGBoost (XGB) methods. The top six most and least important features for each of the Case Studies 2 and 4 are listed with bolded high importance features and italicized low importance features to indicate when the feature appeared in XGB method.

Case Study	Features of High Importance	Features of Low Importance
2	$L (\mathbf{1})$ , $\kappa_{long} (7)$ , $w (10)$ , $U_{long} (4)$ , $U_{lat}^{add} (13)$ , $U_{long}^{add} (14)$	$u_{lat} (5)$ , $u_{long} (6)$ , $l (9)$ , $n_{long} (11)$ , $n_{lat} (12)$ , $U_{lat} (3)$
4	$n_{hyd} (2)$ , $L (\mathbf{1})$ , $n_{lat} (12)$ , $\kappa_{long} (7)$ , $U_{lat} (3)$ , $u_{long} (6)$	$l (9)$ , $n_{long} (11)$ , $U_{long}^{add} (14)$ , $U_{lat}^{add} (13)$ , $u_{lat} (5)$ , $U_{long} (4)$

feature measurements for the predicted class (see Materials and methods). The order of features agrees with the overall ranking seen in Fig. 8 and Fig. S5 described above. By focusing on the

magnitude (color) of the data distribution for a feature in the region corresponding to positive Shapley values in a beeswarm plot of a given class, one can assign the importance of either high or low values of a feature in the prediction of the kinetic class. For example, from the beeswarm plots we found that a longer MT length  $L$  in Case 2 is predictive for catastrophe, while shorter lengths would suggest either growth or rescue. To summarize the main trends found in the beeswarm plots that characterize the kinetic states at low and high soluble tubulin concentrations, we calculated the relative averages of the top six important features using the subset of data found to have positive SHAP values (Materials and methods) for MT growth, catastrophe, shortening, and rescue for Case Studies 2 and 4 (Fig. 11). We defined the relative average for a feature to be the ratio between the mean value and the maximum mean value for that feature for all four kinetic states, with the maximum value shown in Table S2.

For 25  $\mu\text{M}$  tubulin concentration (Case Study 2), the MT catastrophe corresponds to the longest MT length ( $L$ ), largest average longitudinal curvature ( $\kappa_{long}$ ), largest energy of lateral interactions required to complete MT to full cylinder ( $U_{lat}^{add}$ ) and smallest energy of longitudinal interactions required to complete MT to full cylinder ( $U_{long}^{add}$ ). MT shortening was characterized by having the largest MT tip width ( $w$ ) and lowest total energy of longitudinal interactions in MT lattice ( $U_{long}$ ) compared to the other kinetic states. MT growth had exactly the opposite trends compared to MT catastrophe, i.e. the shortest MT length ( $L$ ), smallest average longitudinal curvature ( $\kappa_{long}$ ), and smallest energy of lateral interactions required to complete MT to full cylinder ( $U_{lat}^{add}$ ). Finally, MT rescue had the smallest MT width ( $w$ ) and highest total energy of longitudinal interactions in MT lattice ( $U_{long}$ ). For 250  $\mu\text{M}$  soluble tubulin concentration (Case Study 4), MT catastrophe is characterized by the largest number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice ( $n_{hyd}$ ), longest MT length ( $L$ ), highest average number of lateral interactions per helical pitch in MT lattice ( $n_{lat}$ ), and most stable total energy of lateral interactions in MT lattice ( $U_{lat}$ ), making it clearly distinguishable from the other kinetic states. MT growth is described by the smallest number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice ( $n_{hyd}$ ), lowest average number of lateral interactions per helical pitch in MT lattice ( $n_{lat}$ ), lowest average longitudinal curvature ( $\kappa_{long}$ ), most stable total energy of lateral interactions in MT lattice ( $U_{lat}$ ), and most stable energy of longitudinal interactions in MT tip ( $u_{long}$ ). We note that the ability of the top features in Case Study 4 to classify MT growth better than in Case Study 2 also agrees with the results from confusion matrices (Fig. 9B and 9D). MT rescue is distinguished from the other states by having the shortest MT length ( $L$ ) and least stable energy of longitudinal interactions in MT tip ( $u_{long}$ ). MT shortening is characterized by the largest average longitudinal curvature ( $\kappa_{long}$ ).

### 3. Discussion and conclusion

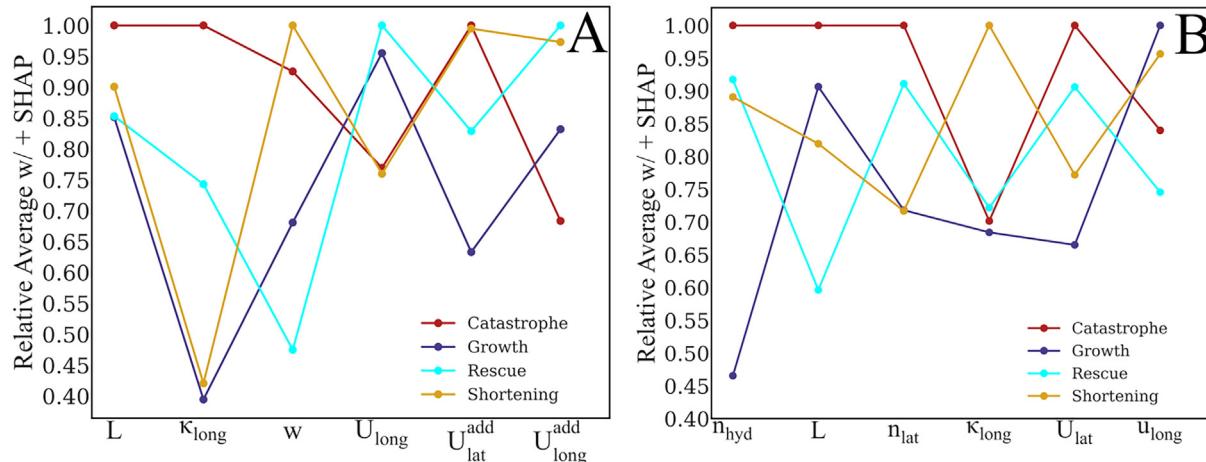
The various important functional roles played by MTs in eukaryotic cells necessitate rigorous quantitative analysis of their physico-chemical, dynamic, and energetic properties. Detailed understanding of the mechanisms of catastrophe and rescue, two of the four kinetic states observed during the dynamic instability behavior exhibited by MTs, is impeded by a lack of quantitative knowledge about the thermodynamic, biomechanical, and chemical factors which control the onset of transitions between these kinetic states associated with MT assembly and disassembly. Here, we developed the molecularly detailed, realistic, computational 3D model of Microtubule Assembly and Disassembly DYnamics (MADDY; Fig. 1). The atomic-level details of the  $\alpha$ - and  $\beta$ -tubulins and of the larger  $\alpha\beta$ -tubulin complexes they form were coarse-grained into the  $\alpha$ - and  $\beta$ -tubulin based MADDY model, which was then used to explore the long timescale dynamics of MT growth and shortening. We utilized this coarse-grained modeling approach because the physico-chemical properties and thermodynamic characteristics of the  $\alpha\beta$ -tubulin dimers, symmetries of the  $\alpha\beta$ -tubulins packing into an MT lattice, as well as the geometry and native topology of MT filaments, rather than the atomic-level detail of the  $\alpha$ - and  $\beta$ -tubulin structure, are what drives the dynamics of cycles of MT growth and shortening. Similar multiscale modeling approaches have been used by researchers to explore the dynamics of complex biomolecular assemblies such as MT filaments and virus shells on very long timescales spanning several decades of biological time [35,68–70]. In addition, when the coarse-grained modeling is implemented on Graphics Processing Units (GPUs), the long second timescales become accessible in

computer simulations [49]. By taking advantage of the computational acceleration on a GPU, we were able to carry out detailed exploration of MT dynamics on a long time scale spanning tens of seconds of biological time.

The MADDY model uniquely combines the implicit description of chemical kinetics (i.e. GTP-to-GDP hydrolysis) described using the Monte Carlo (MC) algorithm, and the explicit description of the dynamics of MT self-assembly and disassembly propagated forward in time using the Langevin Dynamics (LD) in the overdamped limit (Brownian diffusion). The model was mapped into a standard CUDA code, fully implemented on a GPU. Numerical routines for the generation of (pseudo)random numbers (Hybrid Taus method) for MC and LD are described in the previous publications [50,71]. We implemented the first order integration scheme (McCammon algorithm [72]) to propagate forward in time the Langevin equations of motion for all particles ( $\alpha\beta$ -tubulin heterodimers). In Langevin Dynamics, the particle–particle interactions e.g. excluded volume interactions, stretching and bending of  $\alpha\beta$ -tubulin dimers forming the MT lattice, and formation and dissociation of the lateral and longitudinal bonds, are the computational bottleneck. These interactions are described by the same empirical potential energy function (force field; see Eqs. (1)–(4)), and the MT dynamics is obtained by solving numerically the same Langevin equations of motion for all  $\alpha\beta$ -tubulin dimers. Therefore, when running Langevin Dynamics on a GPU, the MADDY program executes the same operation (e.g. generation of random forces, evaluation of molecular forces, integration of equations of motion) for all particles at the same time. By the same token, the Monte Carlo algorithm estimates the probability of GTP hydrolysis for all  $\alpha\beta$ -tubulin dimers in the MT lattice and rejects or accepts the MC move for all  $\alpha\beta$ -tubulin dimers at the same time. Therefore, when mapping the MADDY model into the CUDA code, we took advantage of the GPU architecture to accelerate both the LD and MC components of numerical modeling. The contemporary GPUs helps speed up MADDY based simulations ~50-fold compared to the CPU based implementation. For example, it takes ~7 days of wall-clock time to generate a single ~1-s trajectory of the MT disassembly of a 400-nm long fragment of MT on a GPU, which corresponds to the experimental 417 nm/s (25  $\mu\text{m}/\text{min}$ ) disassembly rate. It takes ~14 days of computational time to generate a single ~3-s trajectory of the MT growth by 30-nm on a GPU, which corresponds to the experimental 42 nm/s (2.5  $\mu\text{m}/\text{min}$ ) rate of MT assembly.

To achieve top performance on a GPU, the LD and MC algorithms were recast into a data-parallel form so that the computational threads run the same instruction stream, but on different data sets (i.e. particles). The tasks were made compute-intensive so that the GPU performs computations rather than reading and writing data. These efforts enabled us to reach the biologically important seconds timescale. The simulation output is organized into the coordinate file for 3D-coordinates ( $x$ ,  $y$  and  $z$ ) and the angle file for 3D-angles ( $\psi$ ,  $\theta$  and  $\phi$ ) of all  $\alpha\beta$ -tubulin dimers in the MT lattice and free tubulin solution, and the energy file for the entire reaction volume (MT lattice plus solution) recorded for each time step. In addition to storing numerical values of various energy terms,  $U_{long}$ ,  $U_{lat}$ ,  $U_{harm}$ ,  $U_\psi$ ,  $U_\theta$ ,  $U_\phi$  and  $U_{rep}$  (see Eqs. (1)–(4)), the energy file also provides information about the identity of GTP-to-GDP-hydrolyzed tubulin dimers and moments of time when these reactions occurred. The coordinate, angle, and energy files can then be used in data visualization and processing, and in analyses of the results of simulations.

To optimize numerical values of parameters defining the potential energy (force field)  $U_{MADDY}$ , (Eqs. (1)–(4)), we used results from the prior molecular-mechanical studies of MT disassembly [28], our study of MT indentation *in silico* [35], experimental kinetic data



**Fig. 11. Summary of average values for top important features most capable of distinguishing between kinetic states of MT dynamics:** Profiled are the relative averages (mean value for a given state divided by the maximal mean value among data that have positive SHAP values, i.e.,  $\geq -0.1$ ) of the most important features identified using the XGBoost method for the MT catastrophe (red data points and lines), MT growth (blue), MT rescue (cyan), and MT shortening (yellow) for Case Study 2 (25  $\mu\text{M}$  tubulin concentration; panel A), and for Case Study 4 (250  $\mu\text{M}$  tubulin concentration; panel B). The mean values and standard deviations are listed in Table S2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on the rates of MT growth and shortening *in vitro* [14], experimental electron crystallography data [53], cryo-EM data [52], as well as our own intuition built step-by-step to span and explore the full parameter space for MT assembly-disassembly. It took  $\sim 2$  years of research to obtain an accurate parameterization of the MADDY model. At this stage, our goals were to obtain the MADDY model parameterization that would result in agreement i) with the experimental 417 nm/s (25  $\mu\text{M}/\text{min}$ ) rate of MT disassembly [14] and ii) with the experimental 17–67 nm/s (1–4  $\mu\text{m}/\text{min}$ ) rate of MT assembly which corresponds to the 7–15.5  $\mu\text{M}$  concentration range of soluble tubulin [14] (Fig. 2). These efforts have resulted in the optimized parameters of the MADDY force-field (Table 1). There is a small discrepancy between the experimental rates of MT growth and MT shortening and rates of these processes simulated with the MADDY model (Fig. 2). This might be due to the fact that *in vitro* experiments might involve accessory proteins and might involve tubulin sequence variants or post translational modifications of tubulin, whereas in the *in silico* experiments we used the drug-stabilized MT structure solved based on X-ray crystallography data. In addition, the MADDY model employs a simplified bead per tubulin monomer-based representation of the MT lattice. Notwithstanding these model limitations, subsequent analysis of the statistics of length of MT protofilaments and curvature of depolymerizing MTs revealed iii) that the experimental histograms of lengths of MT protofilaments [46,47] and simulated distributions of this quantity agree very well, and iv) that the simulated and experimental scatterplots of the protofilaments' profiles show very good agreement (Fig. 3). Moreover, we analyzed the force-generating properties of depolymerizing MTs. The multi-subunit Dam1 protein complex, bound to the MT, is essential for chromosome segregation and it plays a major role in coupling yeast MTs to kinetochores. The Dam1 ring-like complex was demonstrated to transduce force from depolymerizing MTs and was shown to withstand force loads up to 30 pN [19]. The statistics of stalling forces of disassembling MTs upon the Dam1 ring from the MADDY based simulations and experiment [19] agree (Fig. 3). These multiple points of agreement validate the MADDY model.

While exploring the full parameter space, we found out that the rate of MT disassembly is primarily defined by the interplay between the strength of lateral interaction  $D_{\text{lat}}$  and flexural rigidity  $B_\theta$  (Table 1). Indeed, in the depolymerizing MT (with primarily

GDP-bound tubulin dimers), the lateral interactions between neighboring PFs are disrupted as the PFs bend outwards from the MT axis, and so higher flexural rigidity is necessary to disrupt stronger lateral non-covalent bonds. On the other hand, disassembling MTs are known to be efficient force transducers, and so flexural rigidity directly affects pushing forces produced by splaying PFs. Additionally, the strength of longitudinal interactions  $D_{\text{long}}$  affects the length statistics of curled “rams’ horns” at the end of shortening MT polymers. Interestingly, flexural rigidities  $B_\psi$  and  $B_\phi$  define the statistics of singlets and doublets. It is worth mentioning that in a previous modeling study [41], the splaying outward PFs from a depolymerizing MD were restricted to bend in a plane, not in full 3D-space. In the MADDY model, the PFs are capable of bending around all three  $x$ -,  $y$ - and  $z$ -axes (Fig. 1), which corresponds to the bending angles  $\psi$ ,  $\theta$  and  $\varphi$  (Fig. 1) and bending rigidities  $B_\theta$ ,  $B_\psi$  and  $B_\phi$  (Table 1). This enables the PFs to bend out of MTs as monomeric structures (singlets), as dimeric structures (doublets), or even occasionally as trimeric structure bundles (triplets). In the doublets and triplets, the lateral interactions maintain the nearest neighbor associations along the PFs (Fig. 3B). In the polymerizing MT (with the GTP-bound dimers), the strength of longitudinal interactions  $D_{\text{long}}$  primarily affects the rate of MT assembly. We found that the average length of curled protofilaments is determined by the strength of longitudinal interactions  $D_{\text{long}}$ , and that good statistics of curled PF lengths (Fig. 3A) is observed for the 12.0–16.5 kcal/mol range  $D_{\text{long}}$ . The rate of disassembly depends on the ratio between the strength of the lateral potential  $D_{\text{lat}}$  and the value of protofilament stiffness  $B_\theta$ : higher stiffness makes a PF bend outward and disrupt lateral bonds more rapidly. When the protofilament stiffness is in the  $1.6 \times 10^3$ – $1.9 \times 10^3$  pN·nm/rad $^2$  interval, the best match in disassembly rate is found when  $D_{\text{lat}}$  is in the 6.5–8.5 kcal/mol range. Performing simulations in the narrower parameter space allowed us to obtain higher resolution values of parameters for GDP-bound tubulin depolymerization (Table 1). The reasons why the lateral and longitudinal bond energies are different for MT growth and MT shortening are the following: i) we used the drug-stabilized MT structure solved based on X-ray crystallography data, which might create an energy bias; ii) the MADDY model uses a bead per tubulin monomer-based description of the MT lattice and assumes that the  $\alpha$ - and  $\beta$ -tubulin monomers are rigid and,

hence, the model ignores conformational fluctuations within  $\alpha$ - and  $\beta$ -tubulin monomers; in addition, iii) the MADDY model ignores the hydrodynamic interactions, which destabilize the MT lattice when the MT shortening sets in (unpublished data).

Using the MADDY model, we carried out *in silico* experiments of MT growth. One of the highlights from this study (Fig. 4) is the formation of erroneous attachments formed by  $\alpha\beta$ -tubulin dimers on the MT lattice (lattice defects) and empty lattice sites lacking tubulin dimers (lattice vacancies; see Movie S4). These MT lattice defects and lattice vacancies are more manifest at higher concentration of free tubulin, which corresponds to more far-from-equilibrium conditions of the MT growth process. An immediate consequence of these effects is that a large density of accumulated improper attachments and empty lattice sites results in a slowdown or even a complete stoppage of MT growth (*the inset in Fig. 5A*). Once lattice defects and vacancies around the MT tip area are corrected dynamically, normal MT growth resumes. This accumulation of improper incorporations and empty lattice sites followed by their dynamic repair may be a significant contributor to the non-monotonic, variable growth rate behavior observed in the MT assembly kinetics. The driving force for correction of lattice defects and lattice vacancies is the propensity of a growing MT to attain its free-energy minimum state with correct attachments formed by all incorporated  $\alpha\beta$ -tubulin dimers and all MT lattice sites fully occupied, which is a restatement of the Second Law of Thermodynamics. Defects in the microtubule lattice such as voids mean that there are tubulin dimers that neighbor the void that have missing lateral or/and longitudinal bonds, so these dimers are searching for a free dimer to form these missing bonds. We cannot say with certainty that all lattice defects in the growing MT are corrected dynamically. Some defects may be corrected by alternative mechanisms that involve proofreading proteins or enzymes. Permanent MT errors would have deleterious consequences for MT function. Another highlight from *in silico* experiments on MT growth is the large variability of MT tip extensions, which is positively correlated with the variable (i.e. non-monotonic) rate of MT growth (Fig. 5). Another finding from this study is that at lower (higher) tubulin concentration, the tip of a growing MT is more blunt (more sharp), which implies that the MT tip shape is tubulin concentration dependent. The stronger longitudinal bonds form more rapidly than the weaker lateral bonds (Table 1), and so at higher tubulin concentration this strength difference is manifest at the MT tip.

Next, we performed long MADDY based simulations of the dynamics of MT polymerization-depolymerization, in which we observed and analyzed the kinetic states of MT growth, MT shortening, MT catastrophe, and MT rescue (Fig. 6). In these simulations, we set  $D_{long} = 14.2$  kcal/mol and  $D_{lat} = 7.6$  kcal/mol. These values of the strength of longitudinal and lateral non-covalent bonds are arithmetic averages of the values of  $D_{long}$  and  $D_{lat}$  used in the simulations of MT growth and MT shortening (see Table 1). To simulate and study multiple transitions between these kinetic states, i.e., from MT growth to MT shortening (through MT catastrophe) and from MT shortening back to MT growth (through MT rescue), in Case Studies 1–6 (Table 2) we varied the concentration of soluble tubulin between the experimentally relevant 25- $\mu$ M soluble tubulin concentration and 10-fold higher 250- $\mu$ M tubulin concentration, and increased ~4–40 fold the kinetic rate constant for GTP hydrolysis in the 2–20- $s^{-1}$  range. Tubulin concentrations in different cells vary considerably and have been measured numerous times; they range from about 5 to 40  $\mu$ M [73], but it might vary depending on a location in the cell. For example, in *C. elegans* cells, the overall tubulin concentration is about 47  $\mu$ M, and the soluble tubulin within centrosomes could be as high as 470  $\mu$ M [74], and so both 25 and 250  $\mu$ M soluble tubulin concentrations are biologically relevant.

While a higher tubulin concentration is expected to speed up MT growth and a larger GTP hydrolysis rate is expected to promote more rapid MT shortening, accelerating the kinetics in this way does not change the fundamental mechanisms of MT assembly and disassembly, as well as any of the features we have identified through Machine Learning as significant for the different kinetic states, including the MT catastrophe and rescue. The results of MADDY based experiments *in silico* revealed a complex physical picture underlying the dynamics of MT assembly-disassembly, and so in order to clarify these processes we turned to statistical modeling. Using the simulation output, we surveyed 14 different quantitative characteristics (features) of structure and thermodynamics of the MT lattice and MT tip (Fig. 7). These features were chosen based upon the current understanding of MT dynamic instability, since they involve metrics of the MT structure and energetic state and its propensity for a change of state [74].

Extensive Machine Learning analysis of the space of predictors for various kinetic states of MT dynamic instability (classes) revealed, quite surprisingly, that the MT structural and energetic characteristics that discriminate between the classes (kinetic states) depend on the experimental conditions, i.e. sets of top important MT features for lower and higher tubulin concentrations overlap, but only partially (Table 5). For example, the MT length  $L$  and average longitudinal curvature  $\kappa_{long}$  are important for both higher and lower tubulin concentration. However, the total energy of longitudinal interactions on MT lattice  $U_{long}$ , energy of longitudinal interactions required to complete MT to full cylinder  $U_{long}^{add}$ , energy of lateral interactions required to complete MT to full cylinder  $U_{lat}^{add}$ , and MT tip width  $w$  are important at low tubulin concentration, whereas the number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice  $n_{hyd}$ , average number of lateral interactions per helical pitch in MT lattice  $n_{lat}$ , total energy of lateral interactions on MT lattice  $U_{lat}$ , and energy of longitudinal interactions in MT tip  $u_{long}$

**Table 5**

**Summary of the most important MT features:** Listed are the top important features which distinguish between phases of MT growth, MT shortening, MT catastrophe and MT rescue for Case Studies 2 and 4 for low concentration (25  $\mu$ M) and high concentrations (250  $\mu$ M) of soluble tubulin, respectively. These case studies have produced the best overall predictions. The relative importance values ranging from 0 (least important feature) and 1 (most important feature) are calculated using bar plot in Fig. 8B for each feature importance and bar plot in Fig. S5C for each feature importance with respect to the top feature.

Concentration	Feature	Feature Description	Importance
25 $\mu$ M	1	MT length $L$	1
	7	Average longitudinal curvature $\kappa_{long}$	0.76
	10	MT tip width $w$	0.67
	4	Total energy of longitudinal interactions in MT lattice $U_{long}$	0.46
	13	Energy of lateral interactions required to complete MT to full cylinder $U_{lat}^{add}$	0.43
	14	Energy of longitudinal interactions required to complete MT to full cylinder $U_{long}^{add}$	0.43
	2	Number of hydrolyzed $\alpha\beta$ -tubulin dimers in MT lattice $n_{hyd}$	1
250 $\mu$ M	1	MT length $L$	0.73
	12	Average number of lateral interactions per helical pitch in MT lattice $n_{lat}$	0.29
	7	Average longitudinal curvature $\kappa_{long}$	0.27
	3	Total energy of lateral interactions in MT lattice $U_{lat}$	0.26
	6	Energy of longitudinal interactions in MT tip $u_{long}$	0.21

are important at high tubulin concentration. Our finding that the energy of longitudinal interactions in the MT tip  $U_{long}$  is crucial at high concentration is not surprising, since stronger longitudinal bonds form more rapidly than weaker lateral bonds, and at high tubulin concentration this strength difference plays a decisive role. Interestingly, some important features identified for low tubulin concentration, that drop from the list of top important features found for high concentration, move to the list of the least important features at high tubulin concentration. To better understand this trend in predictors' importance, we compared the distributions of these features ( $w$ ,  $U_{long}$ ,  $U_{long}^{add}$ , and  $U_{lat}^{add}$ ) for Case Studies 2 and 4. We found that the distributions of  $U_{long}^{add}$  and  $U_{lat}^{add}$  are broader at low tubulin concentration (Fig. S14), compared to the distributions of these features for high tubulin concentration (Fig. S16 A-C). The overall distributions of  $w$  and  $U_{long}$  are more similar for the two states, however the distributions based on the kinetic states (Fig. S14) showed that the mean feature values for each class differs more for Case Study 2 than for Case Study 4, which explains why these features are important only for Case Study 2. Conversely, the distributions of MT features that are important for high, but not for low tubulin concentration ( $n_{hyd}$ ,  $n_{lat}$ , and  $U_{lat}$ ; Fig. S15) are much broader for high tubulin concentration as compared to low tubulin concentration (Fig. S16 E-G). At the same time, the distributions of the same top important features identified for the lower and higher tubulin concentrations are similarly broad in Case Studies 2 and 4.

Furthermore, analysis of the mean values for each of the features that alter their importance from the lower tubulin concentration case to the higher tubulin concentration case and from the higher tubulin concentration case to the lower tubulin concentration case, in each of the four kinetic states (Tables S2 and S3), revealed that the top important features take on distinctly different mean values in all four kinetic states (Table S2). By contrast, the MT features are no longer important when they reach comparable mean value in all four states (Table S3). As an example, consider the tip width  $w$ , which is found to be important for lower tubulin concentration, but not for higher concentration of free tubulin. As discussed above, this is because for the MT growth to occur, the MT tip should become sharper and, hence, the MT tip extension should be more variable at higher tubulin concentration, where the tip width  $w$  is expected to play an important role in determining the MT kinetic state. The average width of MT tip  $w$  is small (~3–4 dimers) and changes between all four states at lower tubulin concentration; by contrast, at higher tubulin concentration  $w$  is large (~6–7 dimers) and it takes on similar values in all four kinetic states. Thus, while the tip width  $w$  experiences a substantial increase at higher concentration versus lower tubulin concentration, which is expected based on the broadening of the distribution of MT tip extensions, the fact that there is no variation in values of  $w$  between the kinetic states for higher tubulin concentration renders this feature not important for higher tubulin concentration.

We found that both for higher and lower soluble tubulin concentrations, the MT catastrophe is the most distinguishable MT dynamic state from the other kinetic states based on the top important features, while the MT shortening is the least distinguishable kinetic state. We found for MT catastrophe that the majority of top important features are characterized by the maximum mean value among the four kinetic states. The ability of the top important features to distinguish the other two kinetic states, i.e. MT rescue and MT growth, depends on the concentration of soluble tubulin. The MT rescue is easily distinguishable at the lower tubulin concentration, while the MT growth is easily distinguishable at the higher tubulin concentration. Using the mean values of the top important features predicted by the ML analysis for each case (Fig. 11B vs Fig. 11A), we found that, overall, the distinc-

tion between the four kinetic states is more pronounced for low rather than high tubulin concentration. This finding is important since 25  $\mu\text{M}$  is the concentration of soluble tubulin in cytoplasm [75,76].

These ML analysis results, revealing the most important features distinguishing the 4 kinetic states of MTs during dynamic instability, represent significant findings. However, they are limited by the fact that the MADDY model is not a complete representation of the MT's dynamic instability behavior in its cellular environment. As such, the most important features we identified may be thought to represent the features intrinsic to the MT structure itself that can affect kinetic state stability and propensity for change. Thinking more broadly, in the cellular environment the 4 kinetic states of MTs are expected to be affected by additional factors not yet incorporated into this first version of the MADDY model. In fact, experimental results have revealed a number of such cellular factors interacting with MTs that are also thought to affect kinetic state stability and transitions. A variety of MAP proteins have been shown to affect kinetic state behavior through their interaction with MTs. For example, End Binding (EB) proteins were shown to bind *in vivo* to the MT + end GTP cap structure and affect stability [29,77,78]. Furthermore, a number of molecular motors affect kinetic states. For example, two microtubule catastrophe factors have different effects on *in vitro* MT length distributions. Kip3, accumulates at MT + ends [79] and acts as a depolymerase, destabilizing MTs and shortening them more rapidly during catastrophe [80]. Another motor protein, Kif18A, appears to be involved in the *in vivo* shaping of microtubule length distributions [81] and chromosome oscillations, correlating with K-fiber catastrophes [82]. The MCAK motor protein, a Kinesin-13 family member, affects microtubule lengths by a destabilizing interaction, increasing the catastrophe rate [83]. Conserved members of the CLASP protein family have been shown to affect MT tip structure, helping to modulate catastrophe and rescue via the influence of EB proteins [84]. Interestingly, CLASP proteins may serve a nucleating function during rescue [85], potentially facilitating a nucleation event at the MT's remnant GTP islands to initiate rescue. Furthermore, recent studies showed that MT severing proteins, such as spastin and katanin, can amplify MT arrays *in vivo* by catalyzing exchange of GTP-tubulin subunits along a MT filament [86]. Finally, a different type of cellular environment factor affecting MT properties are the varying states of tension that MTs are exposed to during its dynamic instability trajectory. For example, MT – end-directed molecular motors can create mechanical tension acting to promote rescue [87,88]. From the few examples we have just discussed, it is clear that there are many cellular factors that can affect MT kinetic state occupancy observed within the dynamic instability framework. Our point is to emphasize the limitations of using the current MADDY model to describe a complete picture of the most important features differentiating the four kinetic states *in vivo*. Perhaps in future developments of the MADDY model, these types of additional cellular environmental factors can be included as model features. This would allow a more accurate description of *in vivo* MT behavior, thereby allowing a better statistical modeling-based description of the most important features distinguishing the 4 kinetic states in dynamic instability.

To conclude, here we show that experiments *in silico* are capable of providing a complete and high-resolution simulation view of the entire process of MT growth and shortening. We have presented the MADDY model to carry out such high-resolution simulations and using MADDY we observed, for the first time, all four of MT's kinetic states: growth, catastrophe, shortening and rescue. The most important physical characteristics for low tubulin dimer concentration, according to the highest scoring statistical model, are: the MT length, average longitudinal curvature, MT tip width, total energy of longitudinal interaction in MT lattices, and energies

of lateral and longitudinal interactions required to complete MT to full cylinder. The number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice, MT length, average number of lateral interactions per helical pitch in MT lattice, average longitudinal curvature, total energy of lateral interactions in MT lattice, and energy of longitudinal interactions in MT tip were found to be important in identifying the kinetic states of dynamic instability for MTs under high tubulin dimer concentration conditions. The current version of the MADDY model does not take into account the wide range of possible variations in tubulin sequence and post-translational modifications. Since the majority of published *in vitro* assays are performed with tubulin purified from brain tissue, we have parameterized the model using experimental data from the literature sources of published assembly and disassembly kinetics for brain origin tubulin. The structure of a finite-length fragment of the MT lattice was obtained from the 13-subunit ring structure of  $\alpha\beta$ -tubulin dimers for bovine brain tubulin. If we consider another source of tubulin differing significantly in properties, we will have to parameterize the MADDY model again using experimental data for that type of tubulin we want to model. The MADDY model can also be extended to incorporate any and all microtubule associated proteins (MAPs) in order to explore the role of MAPs in the dynamics of MT growth and shortening. One thing that would have to be taken into account in such studies is that MT surface bound MAPs (i.e., motor proteins) could affect the magnitude of the lateral/longitudinal tubulin interactions.

#### 4. Materials and methods

##### 4.1. MADDY force field

In the MADDY model, each  $i$ -th tubulin monomer is described by a spherical bead (of radius  $R = 2$  nm; see Fig. 1A) using three translational coordinates  $x_i, y_i, z_i$ , and three rotational coordinates  $\psi_i, \theta_i, \varphi_i$ ,  $i = 1, 2, \dots, N$ , where  $N$  is the total number of monomer subunits (Fig. 1A). The total potential energy function for a (dis)assembling MT,  $U_{MADDY}$ , depends on the conformations of tubulin monomers forming the MT lattice.  $U_{MADDY}$  involves contributions from the bonding energy, bending energy and repulsive energy,

$$U_{MADDY} = U_{bond} + U_{bend} + U_{rep} \quad (1)$$

In Eq. (1), the potential energy describing formation and dissociation of tubulin-tubulin bonds,  $U_{bond}$ , is the sum of three energy terms, each depending on the distance  $r_{ij}$  between the longitudinal or lateral interaction sites  $i$  and  $j$  in the  $\alpha$ - and  $\beta$ -tubulins in different heterodimers, or in the same dimer (Fig. 1B),

$$\begin{aligned} U_{bond} &= U_{long} + U_{lat} + U_{harm} \\ &= \sum_{ij} D_{long} [(1 - e^{-a_{long}r_{ij}})^2 - 1] \\ &\quad + \sum_{ij} D_{lat} [(1 - e^{-a_{lat}r_{ij}})^2 - 1] + \sum_{ij} Kr_{ij}^2/2 \end{aligned} \quad (2)$$

In Eq. (2), the longitudinal potential  $U_{long}$  describes the non-covalent head-to-tail attachments of tubulin dimers in an MT protofilament, whereas the lateral potential  $U_{lat}$  determines the non-covalent interactions between the protofilaments. For longitudinal and lateral bonds, we used the Morse potential with parameters  $D_{long}$  and  $D_{lat}$ , which determine the strength of interaction, and  $a_{long}$  and  $a_{lat}$ , which define the range of interaction. In  $U_{long}$  and  $U_{lat}$ ,  $r_{ij}$  denotes the distance between adjacent longitudinal or lateral interaction sites  $i$  and  $j$  that belong to different  $\alpha\beta$ -tubulin dimers (Fig. 1B). The harmonic potential  $U_{harm}$  takes into account the non-covalent interactions between the  $\alpha$ - and  $\beta$ -tubulin subunits within the tubulin dimer with the spring constant  $K$  (Fig. 1B).

In Eq. (1), the angle-dependent potential energy for protofilament bending  $U_{bend}$  is given by

$$\begin{aligned} U_{bend} &= U_\psi + U_\theta + U_\varphi \\ &= \sum_{ij} B_\psi [1 - \cos(\psi_{ij} - \psi_0)] + \sum_{ij} B_\theta [1 - \cos(\theta_{ij} - \theta_0)] \\ &\quad + \sum_{ij} B_\varphi [1 - \cos(\varphi_{ij} - \varphi_0)] \end{aligned} \quad (3)$$

Protofilament bending occurs through the  $\alpha$ - and  $\beta$ -subunits' rotations around the  $x$ -,  $y$ -, and  $z$ -axes at the longitudinal interaction sites in the same protofilament (Fig. 1A). In Eq. (3), the bending potential is expressed for each orientation angle in terms of the difference between the angles for the next-neighbor monomers  $\psi_{ij}$ ,  $\theta_{ij}$ , and  $\varphi_{ij}$  and the equilibrium angles  $\psi_0$ ,  $\theta_0$ , and  $\varphi_0$ . We used  $\theta_0 = \theta_0^{GTP} = 0.1$  rad and  $\theta_0 = \theta_0^{GDP} = 0.2$  rad for the GTP-bound and GDP-bound conformations of  $\beta$ -tubulins, respectively, depending on their nucleotide state.  $B_\psi$ ,  $B_\theta$ , and  $B_\varphi$  are the flexural rigidities for the protofilament bending around the angles  $\psi$ ,  $\theta$ , and  $\varphi$ , respectively (Fig. 1D).

In Eq. (1), the repulsive energy term  $U_{rep}$  is given by the Lennard-Jones potential:

$$U_{rep} = \sum_{ij} \varepsilon \left( \frac{\sigma}{r_{ij}} \right)^6 \quad (4)$$

which takes into account the excluded-volume interactions;  $r_{ij}$  is the distance between the centers of any pair of tubulin monomers, and  $\varepsilon = 0.5$  kcal/mol and  $\sigma = 3.5$  nm determine the strength and the range of repulsion (Fig. 1C). Values of all constant parameter of the MADDY model are accumulated in Table 1.

##### 4.2. Langevin dynamics of MT growth and shortening

The dynamics of tubulin subunits are propagated forward in time using the Langevin equations of motion in the overdamped limit (Brownian dynamics). The coordinates  $r_i = \{x_i, y_i, z_i\}$  (translational degrees of freedom) and angles  $\chi_i = \{\psi_i, \theta_i, \varphi_i\}$  (rotational degrees of freedom) for each particle  $i$  ( $\alpha$ - or  $\beta$ -tubulin monomer) at the  $\tau + 1$  simulation step are calculated using the coordinates computed at the previous step  $\tau$  [72], i.e.

$$r_i^{\tau+1} = r_i^\tau - \left( \frac{\partial U_{MADDY}}{\partial r_i} \right) \frac{\Delta t}{\gamma_r} + \sqrt{2k_B T \frac{\Delta t}{\gamma_r}} N(0, 1) \quad (5)$$

$$\chi_i^{\tau+1} = \chi_i^\tau - \left( \frac{\partial U_{MADDY}}{\partial \chi_i} \right) \frac{\Delta t}{\gamma_\chi} + \sqrt{2k_B T \frac{\Delta t}{\gamma_\chi}} N(0, 1) \quad (6)$$

In Eqs. (5) and (6),  $\gamma_r = 6\pi\eta R$  and  $\gamma_\chi = 8\pi\eta R^3$  are the translational and rotational friction coefficients, respectively,  $\eta$  is the solvent viscosity,  $\Delta t$  is the integration time step, and  $T$  is absolute temperature ( $k_B$  is Boltzmann's constant). Also,  $N(0, 1)$  is the zero-average random number sampled from the normal distribution describing random forces for translations and rotations. The derivatives  $(\frac{\partial U_{MADDY}}{\partial \chi_i})$  and  $(\frac{\partial U_{MADDY}}{\partial r_i})$  define, respectively, the translational and rotational components of the molecular forces acting on the  $i$ -th particle. Eqs. (5)–(6) were integrated numerically with the time step  $\Delta t = 200$  ps [89] at  $T = 300$  K temperature using the solvent viscosity (cell cytoplasm)  $\eta = 0.2$  Pa·s [90].

##### 4.3. Modeling GTP-to-GDP hydrolysis

The GTP hydrolysis is modeled using Monte-Carlo algorithm implemented on a CPU. The probability that a hydrolysis reaction in an  $\alpha\beta$ -tubulin dimer occurs at time  $t$  is  $p_{hyd}(t) = \int_0^t k_{hyd} \exp(-k_{hyd}\tau) d\tau = 1 - \exp(-k_{hyd}t)$ , where  $k_{hyd}$  is the

GTP hydrolysis rate constant, and the probability that a reaction occurs in the time interval  $dt$  (Monte Carlo timestep) is given by  $k_{hyd}dt$ . For each value of  $k_{hyd}$  (Table 2), we selected the timestep  $dt$  so that the probability  $k_{hyd}dt = 0.02$  (i.e. 2% probability of transition):  $dt = 0.01$  s, 0.005 s, 0.004 s, 0.003 s, 0.0025 s, 0.002 s and 0.001 s for  $k_{hyd} = 2$  s<sup>-1</sup>, 4 s<sup>-1</sup>, 5 s<sup>-1</sup>, 6 s<sup>-1</sup>, 8 s<sup>-1</sup>, 10 s<sup>-1</sup> and 20 s<sup>-1</sup>, respectively.

#### 4.4. Building MT structures:

The structures of MT polymers used in the simulations of MT growth and shortening were obtained from the 13-subunit ring structure of  $\alpha\beta$ -tubulin dimers modeling a 12-nm (3 monomers) helical pitch [35]. The ring structure uses atomic coordinates of the  $\alpha\beta$ -tubulin dimer (PDB entry: 1JFF; [91]) with the E-site in  $\beta$ -tubulin monomer occupied by GDP. The short and long finite-length fragments of MT polymers with GDP-bound tubulin dimers were constructed by replicating the ring structure 3 and 50 times using the shift distance of 8 nm (length of a tubulin dimer) to obtain an MT polymer of 3 and 50 tubulin dimers in length (MT3/13 and MT50/13). The bead-per-monomer structural models of MT polymers MT3/13 and MT50/13 were obtained by placing soft bead particles of radius 2 nm at the centers-of-mass of tubulin monomers. The bead-per-monomer structure of MT3/13 is shown in Fig. 1A. The short 12-nm MT fragment MT3/13 was used in the *in silico* experiments of MT growth, whereas the 400-nm long MT fragment MT50/13 was used in the simulations of MT shortening. The GTP- and GDP-bound tubulin subunits were modeled similarly, except for the difference in the equilibrium bending angle  $\theta_0$  (Table 1; see also Fig. 1D).

#### 4.5. Performing *in silico* experiments

*In the simulations of MT disassembly*, all the  $\beta$ -tubulin monomers in the MT lattice were in the GDP-bound form and  $\alpha\beta$ -tubulin dimers tended to attain the “bent conformation” ( $\theta_0 = 0.2$  rad). In these simulations, we set the concentration of soluble tubulin to zero. The first helical pitch formed by the  $\alpha\beta$ -dimers at the MT minus end was constrained to prevent rotations of an MT polymer. This mimics connection of an MT cylinder with the centrosome pole. We completed a total of 10 independent simulation runs of 1 s biological time (Fig. 2A, see also Movie S1). *In the simulations of MT growth*, the  $\beta$ -tubulin monomers in free tubulin dimers were in the GTP-bound form and  $\alpha\beta$ -tubulin dimers formed the “straight conformation”. In these simulations, we used the short MT fragment MT3/13 with the tubulins in the GMPCPP-bound nucleotide state. *In silico* tubulin polymerization was carried out in a cylindrical box of 160-nm base diameter and 400-nm length containing free soluble tubulin dimers. The 15, 20, 25 and 30  $\mu\text{M}$  concentrations of free tubulin were maintained constant during each simulation run. In the structure of the short MT fragment MT3/13, we constrained the  $\alpha\beta$ -tubulin dimers in the first three helical pitches’ portion of the MT lattice at the minus end. This mimicked clamping of the MT at the mitotic pole and prevented MT3/13 from rotations around the longitudinal z-axis. A total of 10 simulation runs of 8-s duration were completed for 15, 20, 25 and 30  $\mu\text{M}$  soluble tubulin concentration (see Fig. 2B, see also Movies S2 and S3). *In the simulations of MT assembly-disassembly*, stochastic switches between the GTP-bound form of  $\beta$ -tubulin monomers (straight conformation with equilibrium bending  $\theta_0 = 0.1$  rad) and the GDP-bound form (bent conformation with equilibrium bending angle  $\theta_0 = 0.2$  rad) were modeled using information about the hydrolysis state of  $\beta$ -tubulin monomers. We used the short MT fragment MT3/13 with the tubulins in the GMPCPP-bound nucleotide state. Because parameterization of the MADDY force field for MT growth

and MT shortening resulted in slightly different values of  $D_{long}$  and  $D_{lat}$ , in the simulations of MT assembly-disassembly we set the values of  $D_{long}$  and  $D_{lat}$  to be equal to the arithmetic average of the values used in the simulations of MT growth and MT shortening. For each Case Study 1–6 (Table 2), we performed 5 independent runs (2–4 s duration for Case Studies 1–3 and 1 s duration for Case Studies 4–6; see Fig. 6).

#### 4.6. Force generating properties of MT

Dam1 was modeled as a ring encircling an MT cylinder, with the outer diameter of 48 nm composed by 13 subunits (spherical beads) of 4.5 nm radius (Fig. 3C). Dam1 subunits were described by translational coordinates  $x_i$ ,  $y_i$ ,  $z_i$ , and rotational coordinates  $\psi_i$ ,  $\theta_i$ ,  $\varphi_i$  ( $i = 1, 2, \dots, 13$ ). The total potential energy for Dam1,  $U_{Dam1} = U_{bond} + U_{bend} + U_{rep}$ , is the sum of the bonding energy  $U_{bond}$  due to interactions between a pair of the nearest neighbor Dam1 subunits and between the Dam1 subunits and optical trap (virtual bead), the bending energy  $U_{bend}$  due to bending in any triplet of the Dam1 subunits, and the repulsive energy  $U_{rep}$  due to interaction between Dam1 subunits and tubulin monomers forming the MT lattice. The bonding energy  $U_{bond} = U_{ring} + U_{opt} + U_{opt-ring}$  is described by the harmonic potentials for coupled Dam1 subunits  $U_{ring}$ , for the optical trap  $U_{opt}$ , and for the interaction between the optical trap and subunits of Dam1 ring  $U_{opt-ring}$  (Fig. 3). In  $U_{bond}$ , the spring constants are  $K_{Dam1} = 2.0 \times 10^3$  pN/nm for  $U_{ring}$  and  $U_{opt-ring}$ , and  $K_{opt} = 0.1$  pN/nm for  $U_{opt}$ . Equilibrium distances were  $r_{ring,0} = 9$  nm for  $U_{ring}$  and  $r_{opt-ring,0} = 150$  nm for  $U_{opt-ring}$ . The bending potential  $U_{bend}$  is same as Eq. (3) with the flexural rigidities  $B_\psi = B_\theta = B_\varphi = 5.5 \times 10^5$  pN·nm/rad<sup>2</sup> and equilibrium angles  $\psi_0 = \theta_0 = \varphi_0 = 0$ . The repulsive potential  $U_{rep}$  is the same as in Eq. (4) with  $r_{ij}$  being the distance between any pair of beads;  $\varepsilon = 0.5$  kcal/mol is the repulsion strength, and  $\sigma = 8$  nm and 6 nm are, respectively, the ranges of repulsion between Dam1 subunits and between Dam1 subunits and tubulin monomers (Fig. S1).

#### 4.7. Simulations of MT disassembly stalled by Dam1 ring

The optical trap was represented by a spherical bead of radius  $R_{opt} = 80$  nm (Fig. 3C). A harmonic potential along the MT axis (z-axis)  $U_{opt}$  was applied to the center-of-mass of the optical trap. The initial MT length was 240 nm. Using the output from the simulations of MT disassembly in the presence of the Dam1 ring, we measured the equilibrium displacements of the Dam1 ring from its initial position  $\Delta z$  due to splaying outward of the protofilaments from the depolymerizing MT until the Dam1 ring became stalled. At the stall point, the values of  $\Delta z$  were translated to the values of equilibrium force (stalling force)  $F_{st}$  using the Hook’s law,  $F_{st} = K_{opt}\Delta z$  (Fig. 3). A total of 80 independent 0.2-s trajectories were carried out to generate the distribution of stalling forces  $p(F_{st})$  (Fig. 3D).

#### 4.8. Analysis of output from Langevin simulations of MT assembly-disassembly

We used the energy and coordinate files from MADDY model-based simulations to analyze the time dependence of 14 features which characterize the MT structure and energetics, and the MT tip shape and energetics. The numerical output (i.e., coordinate and energy files from 5 independent runs for each Case Study 1–6) was divided into the data sets for each class – kinetic states associated with the MT growth, MT catastrophe, MT shortening, and MT rescue. Ascending and descending portions of the MT length  $L$  versus time profiles were assigned to the true classes

associated with the MT growth and MT shortening, respectively. The portions of the MT length  $L$  vs time profiles between the ascending and descending parts and between the descending and ascending parts of the profiles were assigned to the true classes associated with the MT catastrophe and MT rescue, respectively. The corresponding numerical values of the other 13 features were calculated for each data point (value of  $L$ ) and for each class using the coordinate and energy files.

The MT length  $L$  (feature 1) was calculated as  $L = l_{\alpha\beta}N_{\alpha\beta}/N_{PF}$ , where  $l_{\alpha\beta} = 8$  nm is the size of a single  $\alpha\beta$ -dimer,  $N_{\alpha\beta}$  is the total number of  $\alpha\beta$ -dimers in the MT lattice and  $N_{PF} = 13$  is the number of protofilaments (PFs). The number of hydrolyzed  $\alpha\beta$ -tubulin dimers in MT lattice  $n_{hyd}$  (feature 2), total energy of lateral interactions in MT lattice  $U_{lat}$  (feature 3) and total energy of longitudinal interactions in MT lattice  $U_{long}$  (feature 4) are directly obtained from the output files. The energies of lateral interactions in MT tip  $u_{lat}$  (feature 5) and longitudinal interactions in MT tip  $u_{long}$  (feature 6) are estimated as  $u_{lat} = D_{lat}N_{lat,tip}$  and  $u_{long} = D_{long}N_{long,tip}$ , respectively, where  $N_{lat,tip}$  and  $N_{long,tip}$  are the numbers of lateral and longitudinal bonds in the MT tip, respectively. Average longitudinal curvature  $\kappa_{long}$  (feature 7) is calculated as  $\frac{1}{l_{\alpha\beta}N_{PF}} \sum_{j=1}^{N_{PF}} \left[ \arccos \frac{(\vec{r}_{1,j} \cdot \vec{r}_{2,j})}{|\vec{r}_{1,j}| |\vec{r}_{2,j}|} \right]$ , where  $\vec{r}_{1,j}$  is a vector that connects the  $\alpha$ - and  $\beta$ -monomers in the last  $\alpha\beta$ -tubulin dimer in the  $j$ -th PF, and  $\vec{r}_{2,j}$  is a vector that connects the  $\alpha$ - and  $\beta$ -monomers in the last  $\alpha\beta$ -tubulin dimer in the  $j$ -th PF. The average lateral curvature  $\kappa_{lat}$  (feature 8) is calculated using the same formula,  $\frac{1}{l_{\alpha\beta}N_{PF}} \sum_{i=1}^{l_{max,tip}} \left[ \arccos \frac{(\vec{r}_{1,i} \cdot \vec{r}_{2,i})}{|\vec{r}_{1,i}| |\vec{r}_{2,i}|} \right]$ , where  $\vec{r}_{1,i}$  is a vector that connects the  $\alpha$ - and  $\beta$ -monomers in the first dimer in the  $i$ -th helical pitch in the MT tip, and  $\vec{r}_{2,i}$  is now a vector that connects the  $\alpha$ - and  $\beta$ -monomers in the second  $\alpha\beta$ -tubulin dimer in the  $i$ -th helical pitch in the MT tip. The MT tip length  $l$  (feature 9) and width  $w$  (feature 10) are calculated, respectively, as  $l = l_{\alpha\beta}N_{\alpha\beta,tip}/N_{PF}$  (or  $l = N_{\alpha\beta,tip}/N_{PF}$  if  $l$  is given in units of tubulin dimers) and  $w = l_{\alpha\beta}N_{\alpha\beta,tip}/l_{max,tip}$  (or  $w = N_{\alpha\beta,tip}/l_{max,tip}$  if  $w$  is in units of tubulin monomers), where  $N_{\alpha\beta,tip}$  is the total number of  $\alpha\beta$ -dimers in the MT tip,  $l_{\alpha} = 4$  nm is the size of a tubulin monomer, and  $l_{max,tip}$  is the number of dimers in the longest PF in the MT tip. The average number of longitudinal interactions per protofilament in MT lattice  $n_{long}$  (feature 11) is calculated as  $n_{long} = N_{long}/N_{PF}$ , where  $N_{long}$  is the total number of longitudinal bonds in the MT lattice. The average number of lateral interactions per helical pitch in MT lattice  $n_{lat}$  (feature 12) is calculated as  $n_{lat} = N_{lat}/l_{max}$ , where  $N_{lat}$  is the total number of lateral bonds in the MT lattice and  $l_{max}$  is the number of dimers in the longest PF. Finally, the energies of lateral interactions required to complete the MT to a full cylinder  $U_{lat}^{add}$  (feature 13) and longitudinal interactions required to complete the MT to a full cylinder  $U_{long}^{add}$  (feature 14) are estimated as  $U_{lat}^{add} = 2l_{max}D_{lat}(N_{PF} - 1) - U_{lat}$  and  $U_{long}^{add} = l_{max}D_{long}N_{PF} - U_{long}$ .

#### 4.9. Machine learning methods

To characterize the four kinetic states of MT dynamic instability (growth, rescue, shortening, and catastrophe), we assessed the ability of 14 features, extracted from the numerical output of the MADDY model-based simulations, to account for factors responsible for state stability and the distinction between states. These features were measured over the course of several simulations for 6 distinct Case Studies 1–6, corresponding to variations in the tubulin concentration and the rate of GTP hydrolysis (see Table 2). The change in length of the MT lattice over the course of each trajectory

was used to manually assign measurements of the 14 features to each particular kinetic state depending on whether the lattice was growing, shortening, or representing a sudden shift from growth to shortening (catastrophe) or from shortening to growth (rescue). We used the supervised classification and ensemble methods briefly described below to assess trends between features and states by comparing each algorithm's prediction accuracy, since each kinetic state was treated as equally important for classification [92]. We used a k-fold cross-validation approach by splitting the data into 5 k-folds and this was repeated 3 times for each method to avoid overfitting and to provide a more generalized prediction model. The implementation of the various Machine Learning methods was based on the Python scikit-learn package. The python script for the ML analysis is freely available (<https://github.com/DimaUClab/MADDY.git>).

**Dummy Classifier:** A baseline accuracy was needed in order to compare the rest of the methods' ability to predict the most accurate kinetic state. The "most frequent" strategy was set to only predict the class that had the largest volume of data for that particular dataset. **Linear Discriminant Analysis (LDA):** Singular value decomposition was used due to the large number of features being studied. No shrinkage was applied because of the solver chosen. The remaining parameters were kept at the default setting. **Support Vector Machines (SVM):** Support vector classifier (SVC) was the algorithm chosen so that the kernel type could be adjusted. The data was normalized before the classifier was applied. Linear, polynomial (degree of 3), sigmoid, and radial basis function (RBF) kernels were tested to better fit the projected data in higher dimensions. **Logistic Regression (LR):** The solver chosen was 'LBFGS' (Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm) since each of the 6 datasets were relatively small (~3000 data points; see Table 2). The multi class parameter was set to 'multinomial' given that there were 4 prediction classes (i.e. growth, shortening, catastrophe and rescue). **K-Neighbors Classifier (KNN):** The number of neighbors set was 3, with all points having uniform weights. The remaining parameters were set to the default. **Bagging Classifier (BAG):** Various ensemble methods were tested in hopes that combining multiple techniques would provide higher prediction accuracies. BAG groups together multiple decision trees when training models on datasets. The number of base estimators was set to 1000, while the remaining parameters were kept at the default. **Random Forest (RF) Classifier:** RF Classifier is a bagging ensemble method that randomly assigns, or bootstraps, observations while training models on datasets. The method included 1000 estimators. **Extra Trees (ET) Classifier:** This is an ensemble method similar to RF, with the difference being that measurements cannot be replaced during training. 1000 estimators were also applied. **XGB Classifier (XGB):** Boosting ensemble method that corrects mislabeled observations throughout the training process with the parameters kept at default. The Synthetic Minority Oversampling Technique (SMOTE) [62] was applied to each classification model to increase the prediction accuracy by creating synthetic data points for the minority classes to compensate for possible skews found in some datasets. Feature importance was plotted for the RF and XGB models by ranking prediction coefficient values. The features found to be the least important for these two models were removed to test whether fewer features could result in increased accuracy.

Imbalanced datasets might result in classification accuracies that represent the most populated class instead of 4 classes (for 4 kinetic states), and so multiclass confusion matrices were constructed for each dataset using either RF or XGB classifier. Confusion matrices visualize the number of predictions made per class while comparing them to their correct class in order to identify

trends in mislabeled predictions. Diagonals in each matrix indicate true positives (TP) for a class, column values beside the unit found in the diagonal were false positives (FP), row values except the diagonal unit were false negatives (FN), and the remaining matrix included all true negatives (TN). Each unit was labeled with the number of predictions, the total number of samples for each class, and the relative percentage of predictions made for a particular case. Along with the confusion matrices, receiver operating characteristic (ROC) and precision-recall (PR) curves were plotted to further quantitate the value of the classification done by applying each method. Here, we compared SVM using the RBF kernel, RF, and XGB for each of the four kinetic states in Case Studies 2 and 4. Both ROC and PR curves highlight the ability of classification methods to make correct predictions. The ROC curves use the true positive rate (TPR)

$$TPR = Recall = TP / (TP + FN) \quad (7)$$

against the false positive rate (FPR)

$$FPR = FP / (FP + TN) \quad (8)$$

to show the ratio of correct prediction versus the rate of predicting any of the negative classes as positive. The PR curve plots the recall, which is equal to TPR, and precision

$$Precision = TP / (TP + FP) \quad (9)$$

focusing on the comparison of methods based on correct predictions. For the imbalanced data, the PR curve and area under the curve (AUC) values associated with each plot were prioritized over the ROC results since the PR plots take into account mainly the minority, or true positive, classes. AUC values increase with classification ability and range from 0 to 1.

The best classification algorithms, along with confusion matrices and PR curves, address how well the four kinetic states can be correctly identified, but how each decision is made and the combination of features that drive the algorithm to make a certain decision are not easy to decipher. To shed light on this “black box” issue, we used SHAP (SHapley Additive exPlanations) as an explainable approach to understand the type and magnitude of MT parameters needed to distinguish among the kinetic states. This concept is used in cooperative game theory, which provides the contribution of each feature towards the classification of each localized observation [93]. Calculating the Shapley value by comparing subsets of the model with and without each feature and summing up the contributions is what drives the prediction of a particular class, and this was done for each observation separately. Feature importance plots were produced using the mean absolute Shapley value of all data entries (observations) for each feature and for each class. The top features that give the highest collective importance for the kinetic states were chosen as features of interest for low and high tubulin concentrations. As customarily done in SHAP analysis, we obtained beeswarm plots for each class per Case Study using the test set of the final k-fold, which depicts the ranking of features in the order of importance for a given class, the distribution of Shapley values on the x-axis for all observations for each feature, and the magnitude of the feature measurements by coloring values either pink or blue depending if the data point represents a high or low feature value, respectively. The color of the distribution corresponding to positive SHAP values indicates the magnitude of the feature values that benefits the classification towards predicting the respective class. Relative averages of feature values within the distribution corresponding to positive SHAP values (defined as values  $\geq -0.1$ ) for each kinetic state were calculated and plotted for the selected features of interest.

## CRediT authorship contribution statement

**Eugenii Kliuchnikov:** Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Eugene Klyshko:** Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Maria S. Kelly:** Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Artem Zhmurov:** Methodology. **Ruxandra I. Dima:** Methodology, Supervision, Writing – original draft, Writing – review & editing. **Kenneth A. Marx:** Conceptualization, Formal analysis, Supervision, Writing – original draft, Writing – review & editing. **Valeri Barsegov:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the NIH grant R01HL148227 and NSF grant MCB-2027530 to VB, and by NSF grant MCB-18179488 to RID.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.01.028>.

## References:

- [1] Dogterom M, Surrey T. Microtubule organization in vitro. *Curr Opin Cell Biol* 2013;25(1):23–9.
- [2] Downing KH, Nogales E. Tubulin and microtubule structure. *Curr Opin Cell Biol* 1998;10(1):16–22.
- [3] Goldyn AM, Kaiser P, Spatz JP, Ballestrem C, Kemkemer R. The kinetics of force-induced cell reorganization depend on microtubules and actin. *Cytoskeleton* 2010;67:241–50.
- [4] Mandelkow E, Mandelkow E-M. Microtubules and microtubule-associated proteins. *Curr Opin Cell Biol* 1995;7(1):72–81.
- [5] Howes SC, Geyer EA, LaFrance B, Zhang R, Kellogg EH, Westermann S, et al. Structural and functional differences between porcine brain and budding yeast microtubules. *Cell Cycle* 2018;17(3):278–87.
- [6] Tilney LG, Bryan J, Bush DJ, Fujiwara K, Mooseker MS, Murphy DB, et al. Microtubules: evidence for 13 protofilaments. *J Cell Biol* 1973;59(2):267–75.
- [7] Li H, DeRosier DJ, Nicholson WV, Nogales E, Downing KH. Microtubule structure at 8 Å resolution. *Structure* 2002;10(10):1317–28.
- [8] Lasser M, Tiber J, Lowry LA. The role of the microtubule cytoskeleton in neurodevelopmental disorders. *Front Cell Neurosci* 2018;12:165.
- [9] Kellogg EH, Hejab NM, Poepsel S, Downing KH, DiMaio F, Nogales E. Near-atomic model of microtubule-tau interactions. *Science* 2018;360 (6394):1242–6.
- [10] Prezel E, Elie A, Delaroche J, Stoppin-Mellet V, Bosc C, Serre L, et al. Tau can switch microtubule network organizations: from random networks to dynamic and stable bundles. *Mol Biol Cell* 2018;29(2):154–65.
- [11] Barlan K, Gelfand VL. Microtubule-based transport and the distribution, tethering, and organization of organelles. *Cold Spring Harb Perspect Biol* 2017;9(5):a025817.
- [12] Mitchison T, Kirschner M. Dynamic instability of microtubule growth. *Nature* 1984;312(5991):237–42.
- [13] Erickson HP, O'Brien ET. Microtubule dynamic instability and GTP hydrolysis. *Annu Rev Biophys Biomol Struct* 1992;21(1):145–66.
- [14] Walker RA, O'Brien ET, Pryer NK, Sobocieno MF, Voter WA, Erickson HP, et al. Dynamic instability of individual microtubules analyzed by video light microscopy: rate constants and transition frequencies. *J Cell Biol* 1988;107 (4):1473–48.
- [15] de Forges H, Bouissou A, Perez F. Interplay between microtubule dynamics and intracellular organization. *Int J Biochem Cell Biol* 2012;44(2):266–74.
- [16] Desai A, Mitchison TJ. Microtubule polymerization dynamics. *Annu Rev Cell Dev Biol* 1997;13(1):83–117.

- [17] Barsegov V, Ross JL, Dima RI. Dynamics of microtubules: highlights of recent computational and experimental investigations. *J Phys: Condens Matter* 2017;29(43):433003.
- [18] McIntosh JR, Volkov V, Ataullakhanov FI, Grishchuk EL. Tubulin depolymerization may be an ancient biological motor. *J Cell Sci* 2010;123:3425–34.
- [19] Volkov VA, Zaytsev AV, Gudimchuk N, Grissom PM, Gintsburg AL, Ataullakhanov FI, et al. Long tethers provide high-force coupling of the Dam1 ring to shortening microtubules. *Proc Natl Acad Sci* 2013;110(19):7708–13.
- [20] Joglekar AP, Bloom KS, Salmon ED. Mechanisms of force generation by end-on kinetochore-microtubule attachments. *Curr Opin Cell Biol* 2010;22(1):57–67.
- [21] Kaverina I, Krylyshkina O, Beningo K, Anderson K, Wang YL, Small JV. Tensile stress stimulates microtubule outgrowth in living cells. *J Cell Sci* 2002;115:2283–91.
- [22] Parker AL, Kavallaris M, McCarroll JA. Microtubules and their role in cellular stress in cancer. *Front Oncol* 2014;1:1–19.
- [23] Tran PT, Joshi P, Salmon ED. How tubulin subunits are lost from the shortening ends of microtubules. *J Struct Biol* 1997;118(2):107–18.
- [24] Howard J, Hyman AA. Dynamics and mechanics of the microtubule plus end. *Nature* 2003;422(6933):753–8.
- [25] VanBuren V, Cassimeris L, Odde DJ. Mechanochemical model of microtubule structure and self-assembly kinetics. *Biophys J* 2005;89(5):2911–26.
- [26] Selzer T, Schreiber G. New insights into the mechanism of protein-protein association. *Proteins Struct Funct Genet* 2001;45(3):190–8.
- [27] Gardner M, Charlebois B, Jánosi I, Howard J, Hunt A, Odde D. Rapid microtubule self-assembly kinetics. *Cell* 2011;146(4):582–92.
- [28] Zakharov P, Gudimchuk N, Voevodin V, Tikhonravov A, Ataullakhanov F, Grishchuk E. Molecular and mechanical causes of microtubule catastrophe and aging. *Biophys J* 2015;109(12):2574–91.
- [29] Gardner MK, Zanic M, Howard J. Microtubule catastrophe and rescue. *Curr Opin Cell Biol* 2013;25(1):14–22.
- [30] Carlier MF, Didry D, Pantaloni D. Microtubule elongation and guanosine 5'-triphosphate hydrolysis. Role of guanine nucleotides in microtubule dynamics. *Biochemistry* 1987;26(14):4428–37.
- [31] Howard J, Hyman AA. Growth, fluctuation and switching at microtubule plus ends. *Nat Rev Mol Cell Biol* 2009;10(8):569–74.
- [32] Brouhard GJ, Bement W. Dynamic instability 30 years later: complexities in microtubule growth and catastrophe. *Mol Biol Cell* 2015;26(7):1207–10.
- [33] Chrétien D, Fuller SD, Karsenti E. Structure of growing microtubule ends: two-dimensional sheets close into tubes at variable rates. *J Cell Biol* 1995;129:1311–28.
- [34] Nogales E, Wang H-W. Structural intermediates in microtubule assembly and disassembly: how and why? *Curr Opin Cell Biol* 2006;18(2):179–84.
- [35] Kononova O, Kholodov Y, Theisen KE, Marx KA, Dima RI, Ataullakhanov FI, et al. Tubulin bond energies and microtubule biomechanics determined from nanoindentation in silico. *J Am Chem Soc* 2014;136(49):17036–45.
- [36] Molodtsov MI, Ermakov EA, Shnol EE, Grishchuk EL, McIntosh JR, Ataullakhanov FI. A molecular-mechanical model of the microtubule. *Biophys J* 2005;88(5):3167–79.
- [37] Bowne-Anderson H, Zanic M, Kauer M, Howard J. Microtubule dynamic instability: a new model with coupled GTP hydrolysis and multistep catastrophe. *BioEssays* 2013;35(5):452–61.
- [38] Chen YD, Hill TL. Monte Carlo study of the GTP cap in a five-start helix model of a microtubule. *Proc Natl Acad Sci USA* 1985;82(4):1131–5.
- [39] Flyvbjerg H, Holy TE, Leibler S. Stochastic dynamics of microtubules: a model for caps and catastrophes. *Phys Rev Lett* 1994;73(17):2372–5.
- [40] Michaels TCT, Feng S, Liang H, Mahadevan L. Mechanics and kinetics of dynamic instability. *Elife* 2020;9:e54077.
- [41] Gudimchuk NB, Ulyanov EV, O'Toole E, Page CL, Vinogradov DS, Morgan G, et al. Mechanisms of microtubule dynamics and force generation examined with computational modeling and electron cryotomography. *Nat Commun* 2020;11(1).
- [42] Strothman C, Farmer V, Arpağ G, Rodgers N, Podolski M, Norris S, et al. Microtubule minus-end stability is dictated by the tubulin off-rate. *J Cell Biol* 2019;218:2841–53.
- [43] Brouhard GJ, Stear JH, Noetzel TL, Al-Bassam J, Kinoshita K, Harrison SC, et al. XMAP215 is a processive microtubule polymerase. *Cell* 2008;132(1):79–88.
- [44] Laan L, Husson J, Munteanu EL, Kerssemakers JWJ, Dogterom M. Force-generation and dynamic instability of microtubule bundles. *Proc Natl Acad Sci USA* 2008;105(26):8920–5.
- [45] van Doorn GS, Tănasă C, Mulder BM, Dogterom M. On the stall force for growing microtubules. *Eur Biophys J* 2000;29(1):2–6.
- [46] Mandelkow EM, Mandelkow E, Milligan RA. Microtubule dynamics and microtubule caps: a time-resolved cryo-electron microscopy study. *J Cell Biol* 1991;114:977–91.
- [47] McIntosh JR, Grishchuk EL, Morphew MK, Efremov AK, Zhudenkov K, Volkov VA, et al. Fibrils connect microtubule tips with kinetochores: a mechanism to couple tubulin dynamics to chromosome motion. *Cell* 2008;135(2):322–33.
- [48] Müller-Reichert T, Chrétien D, Severin F, Hyman AA. Structural changes at microtubule ends accompanying GTP hydrolysis: Information from a slowly hydrolyzable analogue of GTP, guanyllyl ( $\alpha$ ,  $\beta$ )methylene-diphosphonate. *Proc Natl Acad Sci USA* 1998;95:3661–6.
- [49] Zhmurov A, Dima RI, Kholodov Y, Barsegov V. SOP-GPU: accelerating biomolecular simulations in the centisecond timescale using graphics processors. *Proteins Struct Funct Bioinforma* 2010;78(14):2984–99.
- [50] Zhmurov A, Rybnikov K, Kholodov Y, Barsegov V. Generation of random numbers on graphics processors: forced indentation in silico of the bacteriophage HK97. *J Phys Chem B* 2011;115(18):5278–88.
- [51] Alekseenko A, Kononova O, Kholodov Y, Marx KA, Barsegov V. SOP-GPU: influence of solvent-induced hydrodynamic interactions on dynamic structural transitions in protein assemblies. *J Comput Chem* 2016;37(17):1537–51.
- [52] Wang H-W, Nogales E. Nucleotide-dependent bending flexibility of tubulin regulates microtubule assembly. *Nature* 2005;435(7044):911–5.
- [53] Nogales E, Downing KH, Amos LA, Lowe J. Tubulin and FtsZ form a distinct family of GTPases. *Nat Struct Biol* 1998;5(6):451–8.
- [54] Alushin G, Lander G, Kellogg E, Zhang R, Baker D, Nogales E. High-Resolution microtubule structures reveal the structural transitions in  $\alpha\beta$ -tubulin upon GTP hydrolysis. *Cell* 2014;157(5):1117–29.
- [55] General IJ, Dragomirova R, Meirovitch H. Absolute free energy of binding of avidin/biotin, revisited. *J Phys Chem B* 2012;116(23):6628–36.
- [56] Hoefling M, Gottschalk KE. Barnase-Barstar: From first encounter to final complex. *J Struct Biol* 2010;171(1):52–63.
- [57] Höög JL, Huisman SM, Sebö-Lemke Z, Sandblad L, McIntosh JR, Antony C, et al. Electron tomography reveals a flared morphology on growing microtubule ends. *J Cell Sci* 2011;124(5):693–8.
- [58] Zovko S, Abrahams JP, Koster AJ, Galjart N, Mommaas AM, Holzbaur EL. Microtubule plus-end conformations and dynamics in the periphery of interphase mouse fibroblasts. *Mol Biol Cell* 2008;19(7):3138–46.
- [59] Guesdon A, Bazile F, Buey RM, Mohan R, Monier S, García RR, et al. EB1 interacts with outwardly curved and straight regions of the microtubule lattice. *Nat Cell Biol* 2016;18(10):1102–8.
- [60] Melki R, Fievez S, Carlier M-F. Continuous monitoring of Pi release following nucleotide hydrolysis in actin or tubulin assembly using 2-amino-6-mercaptop-7-methylpurine ribonucleoside and purine-nucleoside phosphorylase as an enzyme-linked assay. *Biochemistry* 1996;35(37):12038–45.
- [61] Salvatore M, Warholm P, Shu N, Basile W, Elfösson A, Valencia A. SubCons: a new ensemble method for improved human subcellular localization predictions. *Bioinformatics* 2017;33(16):2464–70.
- [62] Rahman MS, Rahman MK, Kaykobad M, Rahman MS. isGPT: An optimized model to identify sub-Golgi protein types using SVM and Random Forest based feature selection. *Artif Intell Med* 2018;84:90–100.
- [63] Wu Z, Ramsundar B, Feinberg E, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci* 2018;9(2):513–30.
- [64] Hachmann J, Afzal MAF, Haghightari M, Pal Y. Building and deploying a cyberinfrastructure for the data-driven design of chemical systems and the exploration of chemical space. *Mol Simul* 2018;44(11):921–9.
- [65] Wang H, Liu C, Deng L. Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting. *Sci Rep* 2018;8:1–13.
- [66] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2(1):56–67.
- [67] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems. p. 4768–77.
- [68] Kononova O, Zhmurov A, Marx KA, Barsegov V. Mechanics of viruses. Coarse-grained modeling of biomolecules. CRC Press; 2017. p. 367–416.
- [69] Theisen KE, Zhmurov A, Newberry ME, Barsegov V, Dima RI. Multiscale modeling of the nanomechanics of microtubule protofilaments. *J Phys Chem B* 2012;116(29):8545–55.
- [70] Kononova O, Snijder J, Brasch M, Cornelissen J, Dima RI, Marx KA, et al. Structural transitions and energy landscape for cowpea chlorotic mottle virus capsid mechanics from nanomanipulation in vitro and in silico. *Biophys J* 2013;105:1893–903.
- [71] Zhmurov A, Rybnikov K, Kholodov Y, Barsegov V. Generation of Random Numbers on Graphics Processors: Forced Indentation *In Silico* of the Bacteriophage HK97. *The Journal of Physical Chemistry B* 2011;115(18):5278–88.
- [72] Ermak DL, McCammon JA. Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 1978;69(4):1352–60.
- [73] Gard DL, Kirschner MW. Microtubule assembly in cytoplasmic extracts of Xenopus oocytes and eggs. *J Cell Biol* 1987;105:2191–201.
- [74] Baumgart J, Kirchner M, Redemann S, Bond A, Woodruff J, Verbavatz J-M, et al. Soluble tubulin is significantly enriched at mitotic centrosomes. *J Cell Biol* 2019;218(12):3977–85.
- [75] Gard DL, Kirschner MW. A microtubule-associated protein from Xenopus eggs that specifically promotes assembly at the plus-end. *J Cell Biol* 1987;105:2203–15.
- [76] Montecinos-Franjola F, Chaturvedi SK, Schuck P, Sackett DL. All tubulins are not alike: Heterodimer dissociation differs among different biological sources. *J Biol Chem* 2019;294(26):10315–24.
- [77] Maurer SP, Bieling P, Cope J, Hoenger A, Surrey T. GTP $\gamma$ S microtubules mimic the growing microtubule end structure recognized by end-binding proteins (EBs). *Proc Natl Acad Sci* 2011;108:3988–93.
- [78] Maurer S, Fourniol F, Bohner G, Moores C, Surrey T. EBs recognize a nucleotide-dependent structural cap at growing microtubule ends. *Cell* 2012;149(2):371–82.
- [79] Gupta ML, Carvalho P, Roof DM, Pellman D. Plus end-specific depolymerase activity of Kip3, a kinesin-8 protein, explains its role in positioning the yeast mitotic spindle. *Nat Cell Biol* 2006;8(9):913–23.

- [80] Gardner M, Zanic M, Gell C, Bormuth V, Howard J. Depolymerizing kinesins Kip3 and MCAK shape cellular microtubule architecture by differential control of catastrophe. *Cell* 2011;147(5):1092–103.
- [81] Stumpff J, Wagenbach M, Franck A, Asbury C, Wordeman L. Kif18A and chromokinesins confine centromere movements via microtubule growth suppression and spatial control of kinetochore tension. *Dev Cell* 2012;22(5):1017–29.
- [82] Stumpff J, Du Y, English C, Maliga Z, Wagenbach M, Asbury C, et al. A tethering mechanism controls the processivity and kinetochore-microtubule plus-end enrichment of the kinesin-8 Kif18A. *Mol Cell* 2011;43(5):764–75.
- [83] Kline-Smith SL, Walczak CE, Salmon T. The microtubule-destabilizing kinesin XKCM1 regulates microtubule dynamic instability in cells. *Mol Biol Cell* 2002;13(8):2718–31.
- [84] Lawrence EJ, Arpač G, Norris SR, Zanic M, Théry M. Human CLASP2 specifically regulates microtubule catastrophe and rescue. *Mol Biol Cell* 2018;29(10):1168–77.
- [85] Roostalu J, Cade NI, Surrey T. Complementary activities of TPX2 and chTOG constitute an efficient importin-regulated microtubule nucleation module. *Nat Cell Biol* 2015;17:1422–34.
- [86] Vemu A, Szczesna E, Zebr EA, Spector JO, Grigorieff N, Deaconescu AM, et al. Severing enzymes amplify microtubule arrays through lattice GTP-tubulin incorporation. *Science* 2018;361(6404).
- [87] Laan L, Pavlin N, Husson J, Romet-Lemonne G, van Duijn M, López M, et al. Cortical dynein controls microtubule dynamics to generate pulling forces that position microtubule asters. *Cell* 2012;148(3):502–14.
- [88] Hendricks A, Lazarus J, Perlson E, Gardner M, Odde D, Goldman Y, et al. Dynein tethers and stabilizes dynamic microtubule plus ends. *Curr Biol* 2012;22(7):632–7.
- [89] Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des* 1997;2(1):1–22.
- [90] Valberg PA, Feldman HA. Magnetic particle motions within living cells. Measurement of cytoplasmic viscosity and motile activity. *Biophys J* 1987;52(4):551–61.
- [91] Löwe J, Li H, Downing KH, Nogales E. Refined structure of  $\alpha\beta$ -tubulin at 3.5 Å resolution. *J Mol Biol* 2001;313:1045–57.
- [92] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [93] Shapley LS, Kuhn HW, Tucker AW. Contributions to the theory of games. *Ann Math Stud* 1953;28:307–17.
- [94] Bura E, Zhmurov A, Barsegov V. Nonparametric density estimation and optimal bandwidth selection for protein unfolding and unbinding data. *J Chem Phys* 2009;130(1):015102.