Electronic Supplementary Material (ESI) for Soft Matter.

Supplementary Data

Table 1 Parameters

c (μM)	k_h (aut^{-1})	k_{grow} $(\mu M^{-1} aut^{-1})$	k_{short} (aut^{-1})	k_{bond} (aut^{-1})	k_{break} (aut^{-1})	k_{break} seam (aut^{-1})	k_{bond} seam (aut^{-1})	π_{break}
10	0.7	1.25	0.02 T 20 D	100	70 T|T 90 T|D 90 D|T 400 D|D	140 T|T 180 T|D 180 D|T 800 D|D	200	1000

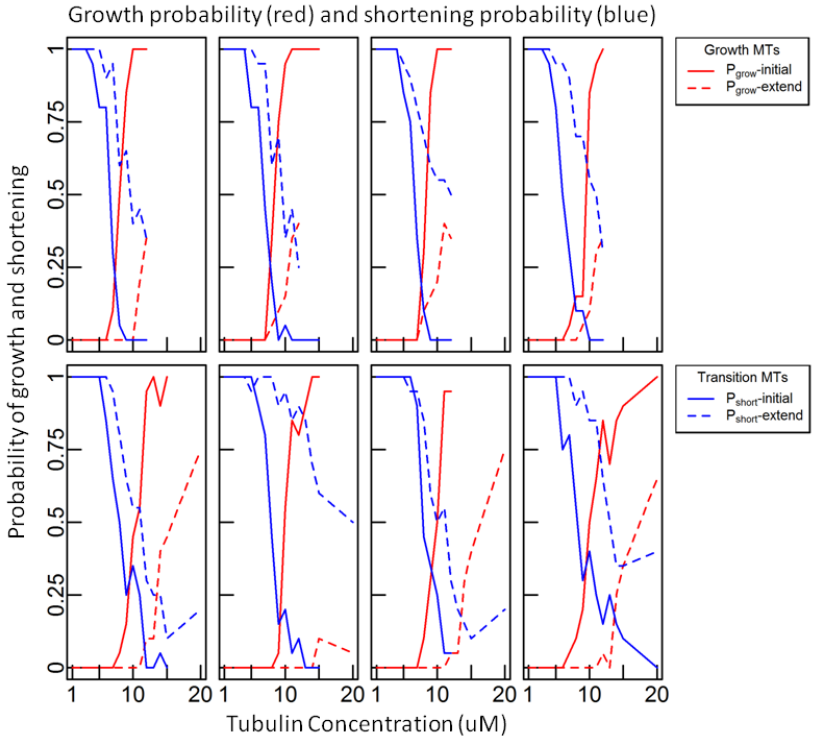

Figure 1. Effect of changing crack depth on dynamic instability. These two rows show the effect of extending all cracks by nine subunit lengths for a single step. Each plot shows the probability of growth (red) or shortening (blue) as a function of tubulin concentration for a given reference structure (solid lines) or its altered version (dashed lines; structure is altered by healing or extended cracks as indicated). The four panels in each row correspond to the four reference structures from each of the three dynamic instability states (growth, shortening, or transition) as shown in Fig. 2 (paper);. These data show that in most cases, there is a strong effect on dynamic instability when crack depth is extended deeper, at least within the parameter ranges tested.

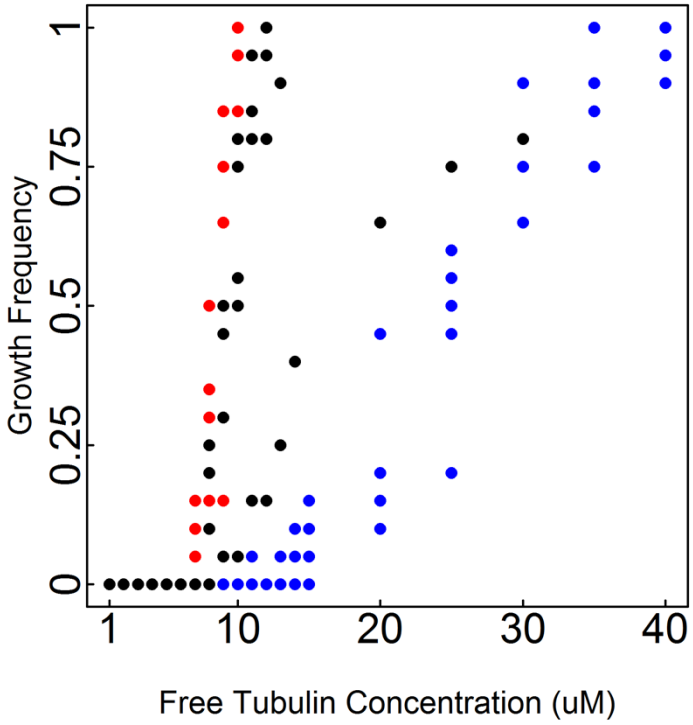

Figure 2. Plot of growth frequency as a function of free tubulin concentration for 12 spontaneous MT structures in Fig. 2 (paper). Red dots are 4 growth MTs. Blue dots are 4 shortening MTs. Black dots are 4 transition MTs. Some dots overlap. Based on this data, we see that MT structure in different states follow different curves. This is why we fit two different logistic regressions, one for growing MTs and one for shortening MTs.

Autocorrelation coefficients

We calculated the autocorrelation coefficients of the three variables used in our study, i.e. the number of GTP-bound subunits at the bottom of the cracks, the average crack depth, and the number of GTP-bound subunits per PF, with 3 seconds time lag. To calculate these autocorrelation coefficients, we ran a simulation with a tubulin concentration 10μM and calculated the number of GTP-bound subunits at the bottom of the cracks, the average crack depth, and the number of GTP-bound subunits per PF, for every spontaneously occurred structure at each time point in the simulation. We then obtained three time series for the three variables. We calculated the autocorrelation coefficient for each variable using the following formula

$$\rho(\tau) = \frac{\frac{1}{N_1}\sum(X[t] - \bar{X})(X[t+\tau] - \bar{X})}{\frac{1}{N_2}\sum(X[t] - \bar{X})(X[t] - \bar{X})}$$

where $\rho(\tau)$ is the autocorrelation coefficient between structures that are τ seconds apart; $X[t]$ is the value of variable *X* at time point *t*; *X* is one of the three variables, i.e. the number of GTP-bound subunits at the bottom of the cracks, the average crack depth, and the number of GTP-bound subunits per PF; $\bar{X}$ is the average of the variable *X*; N_1 is the number of terms $(X[t] - \bar{X})(X[t+\tau] - \bar{X})$; N_2 is the number of terms $(X[t] - \bar{X})(X[t] - \bar{X})$. N_2 is larger than N_1. We obtained the autocorrelation coefficients of 0.42, 0.025, and 0.29 for the number of GTP-bound subunits at the bottom of the cracks, the average crack depth, and the number of GTP-bound subunits per PF, respectively. Because these autocorrelation coefficients are small, we conclude that the structures 3 seconds apart are weakly correlated in some respects (number of GTP-bound subunits) or not at all (average crack depth).

Significance test of the effect

To test statistical significance, we used the p-value calculated using **method 2.3** in the main article. We choose 0.05 as cutoff. If the p-value of a variable (effect) is smaller than 0.05, then this variable (effect) is significant [16]. We calculated that the p-values of X_{ck-GTP}, X_{depth}, X_{GTP} and $X_{[Tu]}$ in Table 1.

Table 1 p-values of X_{ck-GTP}, X_{depth}, X_{GTP} and $X_{[Tu]}$

	Growth state Eqn. (3a)	Shortening state Eqn. (4a)
X_{ck-GTP}	0.02	0.0003
X_{depth}	0.02	0.0002
X_{GTP}	<1e-16	0.0004
$X_{[Tu]}$	<1e-16	<1e-16

Multinomial logistic regression at fixed free tubulin concentration

To study the influence of average number of GTP-bound subunits at the bottom of crack X_{ck-GTP}, average crack depth X_{depth} and average number of GTP-bound subunits per PF X_{GTP} on DI, we performed multinomial logistic regression at fixed tubulin concentration. When tubulin concentration is low (< 8μM), the growth probabilities of MT structures chosen are ~ 0. On the other hand, when tubulin concentration is high (> 11μM), MT almost grow persistently. Therefore, we performed multinomial logistic regression for tubulin concentration 8μM, 9μM, 10μM, 11μM. When we performed this multinomial logistic regression at each fixed tubulin concentration, we use all the 24 MT structures. For example, for multinomial logistic regression at tubulin concentration 8μM, we will use the growing/shortening/transition probabilities of all 8 growing MT structures, 8 shortening MT structures and 8 transition MT structures as our responses. Also, we use the values of X_{ck-GTP}, the values of X_{depth} and the values of X_{GTP} of all 8 growing MT structures, 8 shortening MT structures and 8 transition MT structures as our predictors.

At tubulin concentration 8μM, the regression equation is

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = -2.35 + 1.45X_{ck-GTP} - 0.76X_{depth} + 0.25X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 0.35 - 1.19X_{ck-GTP} - 0.85X_{depth} - 0.58X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A1a)

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = -2.48 + 1.72X_{ck-GTP} - 1.46X_{depth} + 0.06X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 5.32 - 1.48X_{ck-GTP} - 3.68X_{depth} - 0.14X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A1b)

The p-value of X_{ck-GTP} is 0.04. The p-value of X_{depth} is 0.02. The p-value of X_{GTP} is 0.16. X_{ck-GTP} and X_{depth} are significant while X_{GTP} is not.

At tubulin concentration 9μM, the regression equation is

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = -1.43 + 1.20X_{ck-GTP} - 0.76X_{depth} + 0.85X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 0.05 - 2.47X_{ck-GTP} - 0.87X_{depth} + 0.51X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A2a)

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = -2.47 + 1.79X_{ck-GTP} - 0.73X_{depth} + 0.09X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 6.01 - 3.15X_{ck-GTP} - 4.12X_{depth} - 0.03X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A2b)

The p-value of X_{ck-GTP} is 0.0003. The p-value of X_{depth} is 0.0003. The p-value of X_{GTP} is 0.2. X_{ck-GTP} and X_{depth} are significant while X_{GTP} is not.

At tubulin concentration 10μM, the regression equation is

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = 0.0053 + 0.17X_{ck-GTP} - 1.67X_{depth} + 2.36X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 0.78 - 3.14X_{ck-GTP} - 1.34X_{depth} + 1.27X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A3a)

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = 1.02 + 0.95X_{ck-GTP} - 1.89X_{depth} + 0.41X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 9.23 - 4.27X_{ck-GTP} - 6.22X_{depth} - 0.04X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A3b)

The p-value of X_{ck-GTP} is <1e-8. The p-value of X_{depth} is <1e-6. The p-value of X_{GTP} is <1e-8. X_{ck-GTP} and X_{depth} and X_{GTP} are significant.

At tubulin concentration 11μM, the regression equation is

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = 0.14 - 0.80X_{ck-GTP} - 1.42X_{depth} + 3.03X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 2.39 - 3.43X_{ck-GTP} + 0.81X_{depth} + 1.57X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A4a)

$$\log\left(\frac{\pi_{grow}}{\pi_{transition}}\right) = -0.96 - 2.89X_{ck-GTP} + 0.42X_{depth} + 1.09X_{GTP}$$

$$\log\left(\frac{\pi_{shorten}}{\pi_{transition}}\right) = 3.68 - 7.75X_{ck-GTP} + 0.21X_{depth} + 0.72X_{GTP}$$

$$\pi_{transition} = 1 - \pi_{grow} - \pi_{shorten}$$

(A4b)

The p-value of X_{ck-GTP} is 0.0002. The p-value of X_{depth} is 0.04. The p-value of X_{GTP} is 0.0004. X_{ck-GTP} and X_{depth} and X_{GTP} are significant.

From Eqn. (A1) - (A4), we see that the influence of average number of GTP-bound subunits at the bottom of crack X_{ck-GTP} and average crack depth X_{depth} are larger than average number of GTP-bound subunits per PF X_{GTP} on DI. The GTP-bound subunits seems to impact DI mostly through GTP-bound subunits at the bottom of crack. This is consistent with our simulation study.

Table 2 p-values of X_{ck-GTP}, X_{depth}, X_{GTP} at fixed tubulin concentrations

	8μM Eqn. (A1a)	9μM Eqn. (A2a)	10μM Eqn. (A3a)	11μM Eqn. (A4a)
X_{ck-GTP}	0.04	0.0003	<1e-8	0.0002
X_{depth}	0.02	0.0003	<1e-6	0.04
X_{GTP}	0.16	0.2	<1e-8	0.004

Validation of statistical model

To test the prediction power of our regression, we randomly chose 100 new configurations from a computational simulation at 10μM, and compared the predicted growth probability of these structures (as estimated by the regression) to their actual growth probability (as measured by using these structures as the start configuration for new simulations). More specifically, we calculated the average crack depth for each of these 100 structures and found that the range of this average was (0, 1.6). We divided the structures into four bins based on this average crack depth: (0, 0.4), (0.4, 0.8), (0.8, 1.2) and (1.2, 1.6). In addition, we calculated the average number of GTP-bound subunits at the base of the cracks and divided the structures into 4 bins based on this average, (0, 0.5), (0.5, 1), (1, 1.5) and (1.5, 2). Based on these data, we then assigned each of the 100 structures to one of the 16 resulting bins, and then randomly chose one structure from each bin. For each of these 16 structures, we then used the regression formula to calculate the predicted growth probability across a span of tubulin concentrations and compared these predictions to actual growth probability obtained from MT growth simulations started from these structures. Figure 3 is the comparison of growth probabilities from these two methods. These data show that our regression predicts the growth probability reasonably well.

To further test our statistical regression model, we chose an additional 16 MT structures from an independent simulation conducted at 7.5μM tubulin concentration. We found that our statistical model continued to quantitatively describes the fundamental relationship between MT structure and growth/shortening probability (Figure 4). These observations indicate that our statistical model can with reasonable accuracy predict the growth/shortening probability of an arbitrary MT structure based only on the average number of GTP-bound subunits at the base of the cracks and the average crack depth between protofilaments.

For the above new 16 structures obtained at 10μM tubulin concentration and the above new 16 structures obtained at 7.5μM tubulin concentration, we standardized their predictors before we plug their values in regression formula (Eqn. (3a) and Eqn. (4a) in the main article). We assume the distribution of new data is the same as the distribution

of the data in main article which are used to build multinomial logistic regression. Thus, we used the mean and the standard deviation calculated from data in the main article as the mean and standard deviation of the new data and standardized the new data. For example, to standardize the average number of GTP-bound subunits at the bottom of the cracks X_{ck-GTP} in the new data, we use μ_{ck-GTP} and σ_{ck-GTP} to calculate the standardized X_{ck-GTP} as $\frac{X_{ck-GTP}-\mu_{ck-GTP}}{\sigma_{ck-GTP}}$.

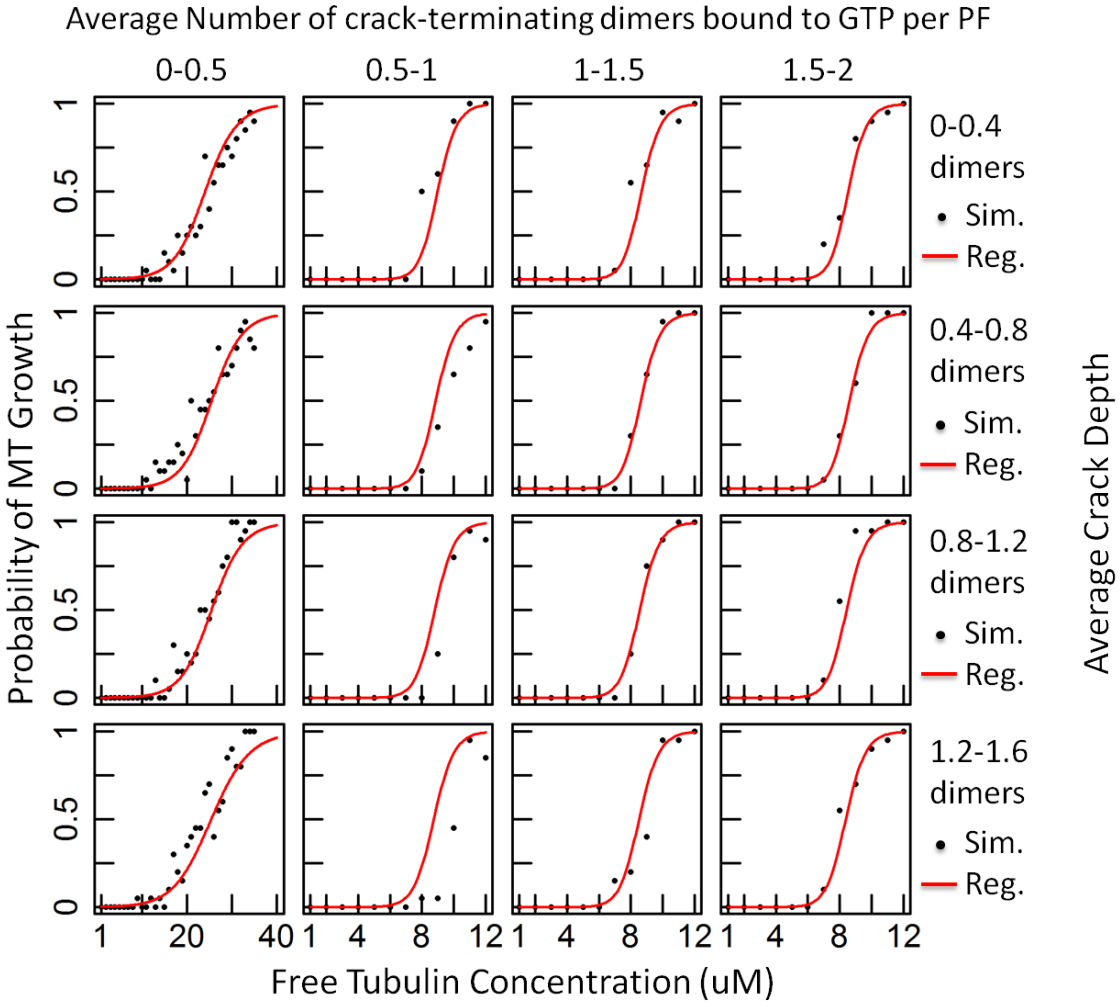

Figure 3 Comparison of growth probabilities calculated from regression (red lines) and computational model (black dots) for 16 MT structures extracted from a simulation at tubulin concentration 10μM. The x-axis in each plot is tubulin concentration in μM. The y-axis in each plot is the growth probability. Each row corresponds to one bin of average crack depth. They are (0, 0.4), (0.4, 0.8), (0.8, 1.2) and (1.2, 1.6) from top to bottom. Each column corresponds to one bin of average number of GTP-bound subunits at the base of the cracks. They are (0, 0.5), (0.5, 1), (1, 1.5) and (1.5, 2) from left to right. The

tubulin concentration in first column is larger because these four structures are in shortening state. Predictions of our regression match results of computational model very well.

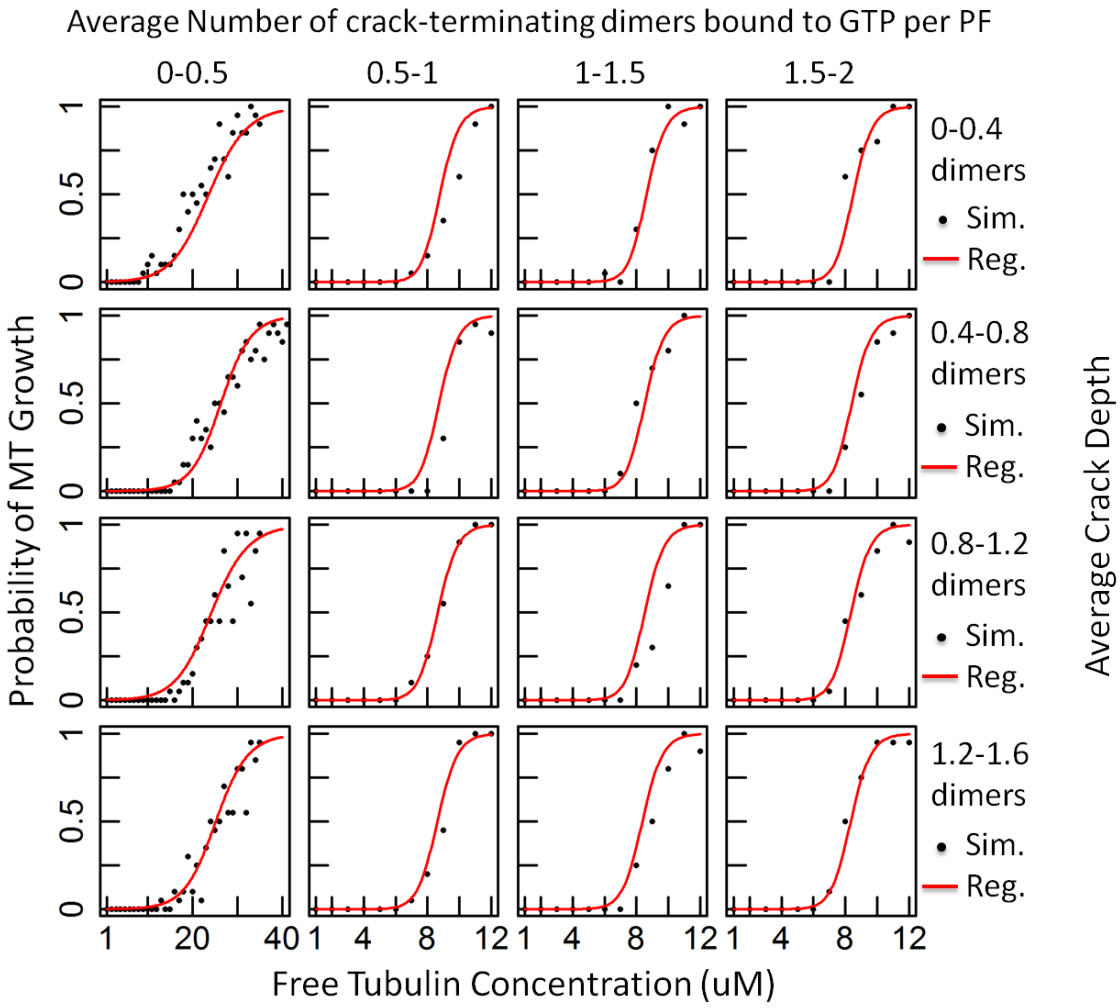

Figure 4 Comparison of growth probabilities calculated from regression (red lines) and computational model (black dots) for 16 MT structures extracted from a simulation at tubulin concentration 7.5μM. The x-axis in each plot is tubulin concentration in μM. The y-axis in each plot is the growth probability. Each row corresponds to one bin of average crack depth. They are (0, 0.4), (0.4, 0.8), (0.8, 1.2) and (1.2, 1.6) from top to bottom. Each column corresponds to one bin of average number of GTP-bound subunits at the base of the cracks. They are (0, 0.5), (0.5, 1), (1, 1.5) and (1.5, 2) from left to right. The tubulin concentration in first column is larger because these four structures are in shortening state. Predictions of our regression match results of computational model very well.

Comparison of accuracy between regression models

We have two different types of multinomial logistic regression models. The first type is the one where we include tubulin concentration as a variable. This model includes Eqn. (3a,b) and Eqn. (4a,b) in the main article. The second type is the one where we fixed tubulin concentration. This model includes Eqn. (A1a,b), (A2a,b), (A3a,b) and (A4a,b). In this type of multinomial logistic regression, we focus on the influence of different factors: average number of GTP-bound subunits at the bottom of crack X_{ck-GTP}, average crack depth X_{depth} and number of GTP-bound subunits per PF X_{GTP}, on DI. For the first type of model, we compares its predictions with simulation results (Figure 3 and 4). We will compare the predictions of second model as well.

To compare the predictions of fixed-tubulin-concentration model (second type) with simulation results, we calculate the difference between prediction of second type model and simulation results for tubulin concentrations used (8μM - 11μM) and sum up the square of difference. Small sum of squares indicates better predictions of the second type model. We perform this comparison for the 16 MT structures used in Figure 3 (chosen from a simulation ran at 10μM tubulin concentration) and also for the 16 MT structures used in Figure 4 (chosen from a simulation ran at 7.5μM tubulin concentration). The result is the red barplot in Figure 5. We observe that, for MT structures chosen from simulation 10μM (left panel in Figure 5), the prediction of growth/shortening probability of fixed-tubulin-concentration model (second model) is better when we make prediction of growth/shortening probability of a MT structure at a lower free tubulin concentration (8μM and 9μM). The same is observed for MT structures chosen from simulation at tubulin concentration 7.5μM (right panel in Figure 5).

We also calculate the sum of squared difference for the first model (where tubulin concentration is variable) at tubulin concentrations 8μM-11μM. The results are summarized as the blue barplot in Figure 5. We observe that, the fixed-tubulin-concentration model (second model) is more accurate than the first model (except for predictions at 11μM for MT structures chosen from a simulation ran at tubulin concentration 10μM, rightmost two bars in left panel of Figure 5). Therefore, if we are interested in predicting growth/shortening probability for an arbitrary MT structure at free tubulin concentration ranging 8μM-11μM, we can use the fixed-tubulin-concentration model (second model). Nevertheless, the first model is more general where tubulin concentration is a variable.

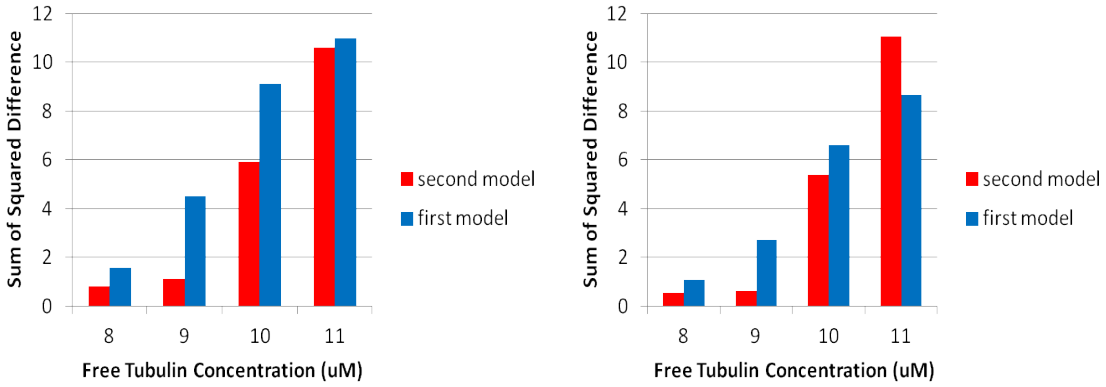

Figure 5. Sum of squared difference between predictions of multinomial logistic regression and simulation results. X-axis is tubulin concentration at which predictions are made. Y-axis is the sum of squared difference between predictions of growth/shortening probability of multinomial logistic regression and growth/shortening probability calculated from simulation. **Left panel**: we use multinomial logistic regression model to calculate growth/shortening probability of 16 MT structures randomly chosen from a simulation ran at 10μM tubulin concentration and take the difference between this probability and probability calculated from simulations at tubulin concentration 8μM-11μM. Then we sum up the squared difference. Red bars use fixed-tubulin-concentration multinomial logistic regression for prediction (second model). Blue bars use multinomial logistic regression where tubulin concentration is variable (first model). **Right panel**: same calculations as left panel with 16 MT structures randomly chosen from a simulation ran at 7.5μM tubulin concentration.