Multivalence cooperativity leading to "all-or-nothing" assembly: the case of nucleation-growth in supramolecular polymers

Elkin Lopez-Fontal, Lilia Milanesi and Salvador Tomas

Supplementary information

General Methods

NMR binding experiments. Samples of **C** (25 μM to 2 mM, depending on the experiment) were prepared by diluting a 10 mM stock solution of **C** in water with the appropriate amount of water, phosphate buffer 200 mM, pH 7.2 , D_2O and were added of aliquots of a stock solution of **B**)so that the final sample contained phosphate buffer of the desired concentration (5 to 100 mM depending of the experiment), 10% of D_2O v/v and the desired concentration of the ligand (0 to 5 mM depending on the experiment). Samples were left to equilibrate for a minimum of 1 hour before the ^1H NMR spectrum was recorded in a Bruker AV600 NMR spectrometer.

NMR diffusion experiments. A sample of **C** and **B** 200 μM each in phosphate buffer 5 mM pH 7.2 and 10 % D2O was prepared by mixing the appropriate quantities of stock solutions of **B** and **C** in water, water, phosphate buffer 200 mM, pH 7.2 and D_2O . The sample was left equilibrate for 6 hours. The experiment was carried in a Bruker AV600 NMR spectrometer.

UV binding experiments. Samples of **C** (1 μM to 2 mM, depending on the experiment) were prepared by diluting a stock solution of 10 mM C in water with the appropriate amount of water, phosphate buffer 200 mM, pH 7.2 and aliquots of the stock solutions of **B** so that the final samples contained phosphate buffer of the desired concentration (5 to 100 mM depending of the experiment) and the desired concentration of **B** (0 to 5 mM depending on the experiment). Samples were left to equilibrate for a minimum of 1 hour before the UV spectrum was recorded in a Cary300 UV spectrophotometer.

EM experiments. For negative stain EM, 5μL of a solution of **C** and **B** ([**B**]=[**C**] = 50 μM) in 100 mM phosphate buffer pH 7.2 were applied to a carbon-coated, glow discharged, 300-mesh copper grid and blotted after 30 seconds. The grids were stained twice with 10 μL of 2% (w/v) uranyl acetate, blotted after 30 seconds and allow to air-dry. Images were collected using minimal electron dose at a nominal magnification of 26000x and 67000x in a Tecnai 12 microscope (FEI, Eindhoven, NL) with a tungsten filament operating at 120 kv. Images were recorded with a Gatan 1K Ultrascan camera (Gatan, USA) between 1.2-2.0 μm underfocus**.**

NMR diffusion experiments.

DOSY experiment was carried out on a sample 200 μM of **C** and **B**, in buffer phosphate 5 mM and pH 7.20. In these conditions the spectrum shows 7 singlet peaks between 8.5 and 9.5 ppm assigned to the beta porphyrin proton (see Fig. 2A and Fig. S1 A). The change in intensity of these peaks was used to derive the corresponding diffusion coefficients (Fig. S1).For comparison, the diffusion coefficient of the corresponding species was calculated using the program Hydropro (see ref. 33) (Table S1). Hydropro computes the hydrodynamic properties of rigid macromolecules from their atomic-level structure, as specified by the atomic coordinates taken from a PDB. The PDB structure of the relevant complexes was generated based on the x-ray data of **CB** complexes in reference 36. There is an excellent correlation between the coefficients calculated and the ones experimentally derived (Table S1).

Table S1. **Diffusion coefficients.**

	CB	**CB_2**	**C_2B**	**C_2B_3**	**C_2B_2**
Experimental	2.62E-06	2.56E-06	1.97E-06	1.88E-06	1.85E-06
Experimental normalized	1.00	0.98	0.75	0.72	0.71
Calculated	2.18E-06	2.12E-06	1.75E-06	1.68E-06	1.71E-06
Calculated normalized	1.00	0.97	0.80	0.77	0.79

The units in the diffusion coefficients are cm^2s^{-1}. The error in the experimental value is of the order of 10%.

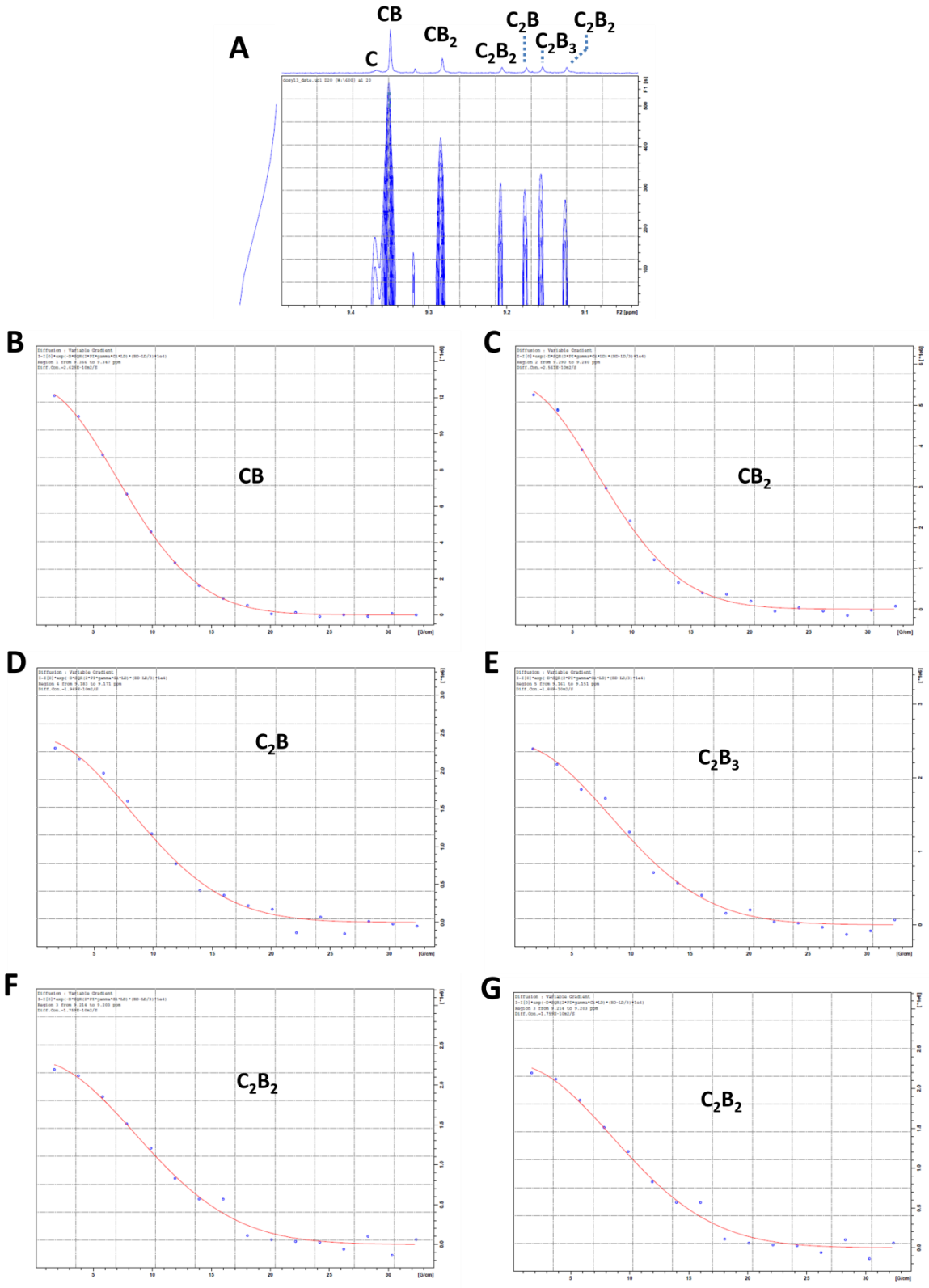

Figure S1. A. Section of the DOSY-NMR experiment showing the changes in intensity with time of the peaks assigned to the β proton of the different species present. The assignment to each specie is also shown. B-G. Fit of the integral of each of the peaks to the diffusion model that yields the diffusion coefficient shown in Table S1. For **C_2B_2**, the diffusion coefficient was calculated as the average from the coefficients derived from fittings in F and G.

UV titration experiments

UV spectroscopy was used to determine the stepwise binding constants for complexes with bipyridine **B** in concentration conditions where the dominant species are of the form **CB** and $\mathbf{CB}_2$ (that is, with concentration of **C** less than 5 μM). The stepwise formation constants K_1 and K_2 for complexes **CB** and $\mathbf{CB}_2$ are

$$K_1 = \frac{[\mathbf{CB}]}{[\mathbf{C}][\mathbf{B}]} \qquad (1)$$

$$K_2 = \frac{[\mathbf{CB}_2]}{[\mathbf{CB}][\mathbf{B}]} \qquad (2)$$

The observed absorbance at a particular wavelength *A* is:

$$A = \varepsilon_{\mathbf{C}}[\mathbf{C}] + \varepsilon_{\mathbf{CB}}[\mathbf{CB}] + \varepsilon_{\mathbf{CB2}}[\mathbf{CB}_2] \qquad (S1)$$

where $\varepsilon_{\mathbf{C}}$, $\varepsilon_{\mathbf{CB}}$ and $\varepsilon_{\mathbf{CB2}}$ are the corresponding extinction coefficients at the wavelength under study. The mass balances are:

$$[\mathbf{C}]_0 = [\mathbf{CB}_2] + [\mathbf{CB}] + [\mathbf{C}] \qquad (S2)$$

$$[\mathbf{B}]_0 = 2[\mathbf{CB}_2] + [\mathbf{CB}] + [\mathbf{B}] \qquad (S3)$$

Were $[\mathbf{C}]_o$ and $[\mathbf{B}]_o$ are the total concentration of **C** and **B**. In order to remove the baseline drifting we performed a graphical derivative of the spectra, by subtracting to the absorbance at a given wavelength the absorbance at a 5 nm greater wavelength (spectra 1 and 2 in Figure S2). Equation (S1) can be used for a difference on absorbances as much as for absorbance. Therefore, the derivative of the spectra was used to determine the stepwise binding constants by fitting the changes in the absorbance derivative at three different wavelengths to the model defined by equations (1), (2), (S1)-(S3). We used the program Micromath Scientist 2.0 (Fig. S2) for the data fitting.

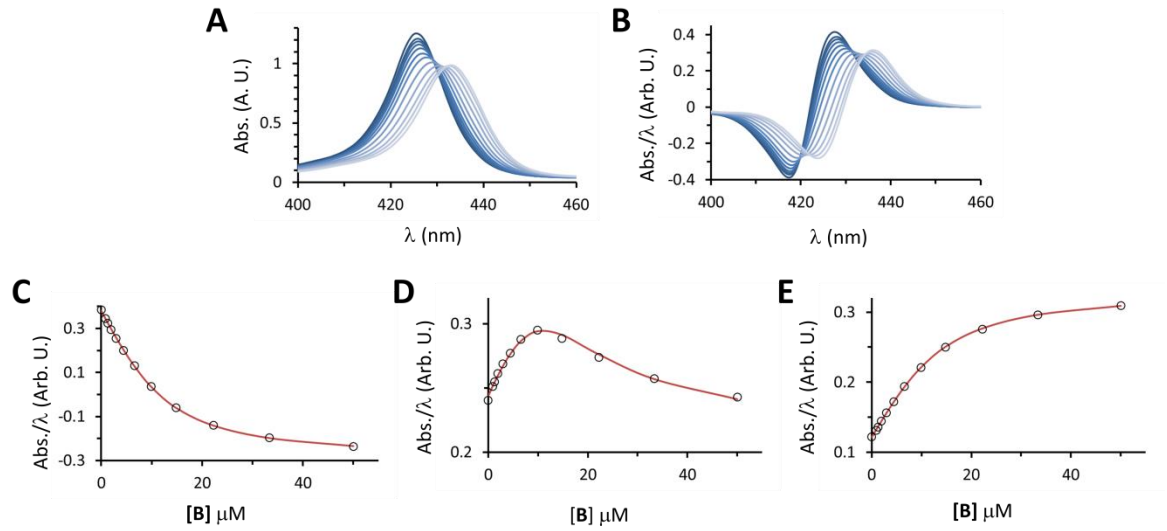

Figure S2. A. Changes in the Soret band of **C** upon addition of increasing amounts of **B**. B. Changes in the first derivative of the spectra shown in A. C. Changes in the intensity of the derivative of the absorbance at 426 (empty circles) and best fit to the model defined by equations (1), (2), (S1)-(S3) (red line). C. Idem for changes in intensity at 432 nm. D. Idem for changes in intensity at 436 nm. The concentration of **C** in this experiment is 9.4 μM with a cell pathlength of 0.4 cm.

Table S2. Binding parameters in sodium phosphate 100 mM, pH 7.20[a]

K_1	K_2	K_3
$9.7 \times 10^5 \pm 8.3 \times 10^4$	$1.1 \times 10^4 \pm 7.5 \times 10^3$	$3.0 \times 10^4 \pm 6.0 \times 10^{3\,(b)}$

(a) The units for the binding constants are M^{-1} in all cases. The quoted error is twice the standard deviation of the mean. (b) Determined by integration of the NMR signal.

X-Ray crystalography-derived oligomer-dimer structure.

The molecular model displayed in Fig. 3C was build up from the crystal structure reported in ref. 36. For Fig. 3C, using the program Discovery Studio 4.0, we took a section of the PDB structure of the crystal containing two complexes of the form **CB** in close proximity (i.e., a dimer of the monomer **CB**, Fig. S3A) and generated the dimer of the oligomer $\mathbf{C_2B_2}$ by duplicating it (Fig. S3B). The distances between the porphyrin rings within the original dimer of monomer **CB** and the new inter-porphyrin distances in the $\mathbf{C_2B_2}$ dimer are virtually the same.

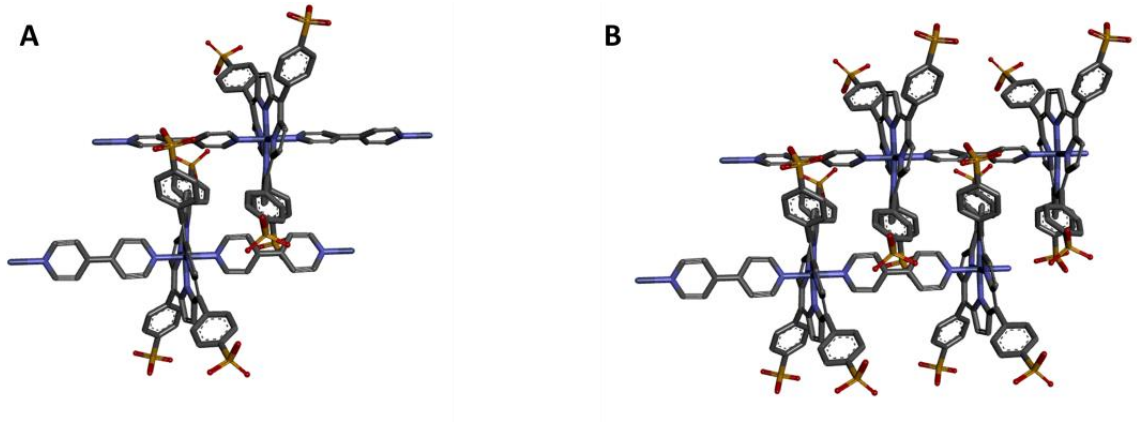

Figure S3. X-ray Crystal structure of the oligomer dimer. A. Section of the crystal structure reported in reference 36. B. Reconstruction of the oligomer based on the structure displayed in A.

TEM data analysis.

We used pictures obtained with a nominal magnification of 67000x to estimate the average width of the fibre-like features found. For this magnification the value of pixels per Angstrom is 1.022, and this value was used to convert the width in pixels, as measured using the program GIMP 2.8, to the corresponding value in Angstroms. From 2 different pictures (Fig. S4 A and B), we chose the 60 features that where the best defined. The distribution of widths is shown in Figure S4 C. Figure 3D is composed of a section of Figure S4 B and a section of a third picture showing an isolated fibrilar feature.

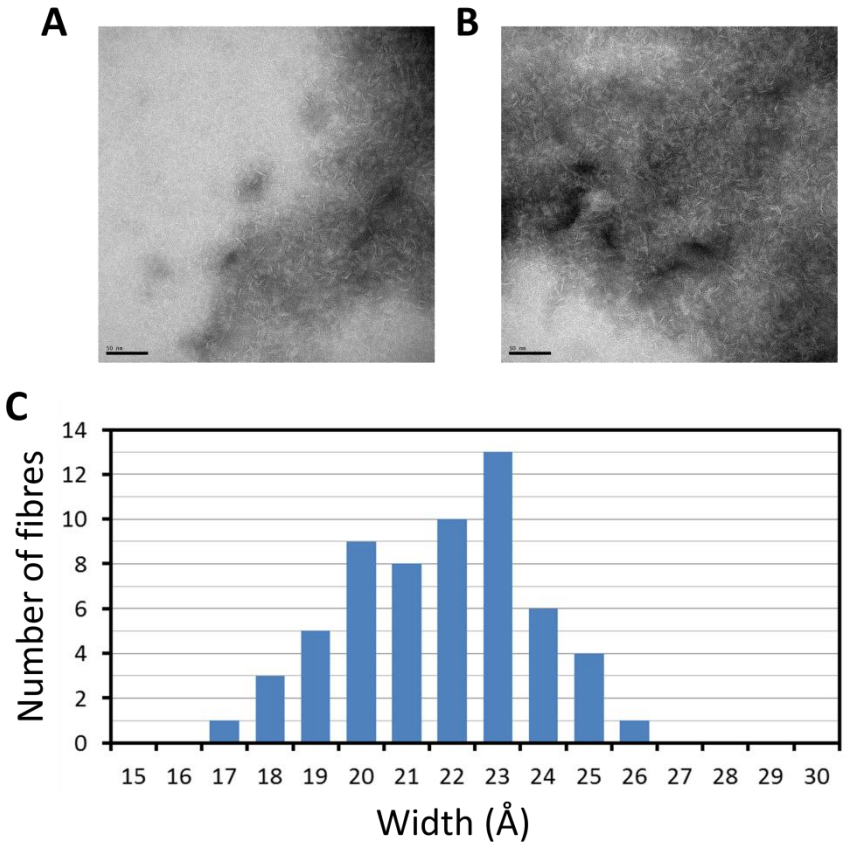

Figure S4. A anb B. TEM pictures obtained at a nominal magnification of 67000x. C. Distribution of the widths of the best defined needle-like features in A and B. The average width obtained from this analysis is 22 ±2 Å

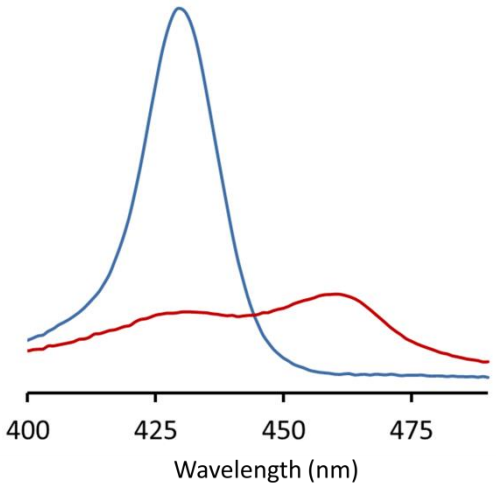

Figure S5. Section of the UV spectrum of a mixture C/B in a ratio 1 to 1 (concentration of C = 1 μM) in buffer phosphate 5 mM, pH 7.20 (blue trace) and in buffer phosphate 500 mM, pH 7.20 (red trace).

Model of the single strand oligomerization.

The expressions for the stepwise binding constants K_1, K_2 and K_3 for complexes **CB**, $\mathbf{CB_2}$ and $\mathbf{C_2B}$, are:

$$K_1 = \frac{[\mathbf{CB}]}{[\mathbf{C}][\mathbf{B}]} \qquad (1)$$

$$K_2 = \frac{[\mathbf{CB_2}]}{[\mathbf{CB}][\mathbf{B}]} \qquad (2)$$

$$K_3 = \frac{[\mathbf{C_2B}]}{[\mathbf{CB}][\mathbf{C}]} \qquad (3)$$

K_1, K_2 and K_3 where calculated by integration of the signals assigned to the beta protons of the corresponding complex. K_1 and K_2 were also calculated by UV spectroscopy methods as described above. The resulting values are the same, within the error, using either technique (Table 1).

Using **C** and **B** as building blocks, three kinds of oligomers can be formed: (i) with the same number of **C** and **B** building blocks, with generic formula $\mathbf{C_nB_n}$, (ii) an oligomer capped with **B** molecules and generic formula $\mathbf{C_nB_{n+1}}$ and (iii) and oligomer capped with **C** molecules with generic formula $\mathbf{C_{n+1}B_n}$. The formation constants for these oligomers, K_{1o}, K_{2o} and K_{3o} can be expressed as:

$$K_{1O} = \frac{[\mathbf{C_nB_n}]}{[\mathbf{C}]^n[\mathbf{B}]^n} \qquad (S4)$$

$$K_{2O} = \frac{[\mathbf{C_nB_{n+1}}]}{[\mathbf{C}]^n[\mathbf{B}]^{n+1}} \qquad (S5)$$

$$K_{3O} = \frac{[\mathbf{C_{n+1}B_n}]}{[\mathbf{C}]^{n+1}[\mathbf{B}]^n} \qquad (S6)$$

It is reasonable to assume that the binding of any **B** unit to a pre-existing oligomer is determined by K_2, while the addition of any **C** unit by K_3, applying the corresponding statistical corrections. For example , the binding constant for the addition of **B** to an oligomer of the form $\mathbf{C_nB_n}$ is K_2:

$$K_2 = \frac{[\mathbf{C_nB_{n+1}}]}{[\mathbf{C_nB_n}][\mathbf{B}]} \qquad (S7)$$

since $\mathbf{C_nB_n}$ has exactly the same number and kind of free binding sites as present in **CB**.

On the other hand, for the addition of **B** over an oligomer of the form $\mathbf{C_nB_{n-1}}$ a statistical correction should be applied accounting for the fact that twice as many binding sites are now available (relative to $\mathbf{C_nB_n}$), thus the corresponding binding constant is:

$$4K_2 = \frac{[\mathbf{C_nB_n}]}{[\mathbf{C_nB_{n-1}}][\mathbf{B}]} \qquad (S8)$$

Similarly, for the addition of **C** over $\mathbf{C_nB_n}$ we have that the binding constant K_3 is:

$$K_3 = \frac{[\mathbf{C_{n+1}B_n}]}{[\mathbf{C_nB_n}][\mathbf{C}]} \qquad (S9)$$

and that for the addition of **C** over $\mathbf{C_{n-1}B_n}$ is:

$$4K_3 = \frac{[\mathbf{C_nB_n}]}{[\mathbf{C_{n-1}B_n}][\mathbf{C}]} \qquad (S10)$$

Therefore, taking into account the statistical corrections, K_{1o}, K_{2o} and K_{3o} can be written as a function of K_1, K_2 and K_3 :

$$K_{1O} = K_1(4K_2K_3)^{n-1} \quad \text{(S11)}$$

$$K_{2O} = K_1K_2(4K_2K_3)^{n-1} \quad \text{(S12)}$$

$$K_{3O} = K_1K_3(4K_2K_3)^{n-1} \quad \text{(S13)}$$

On the other hand, the total concentration of receptor in solution, $[\mathbf{C}]_0$, is the sum of all the **C** species:

$$[\mathbf{C}]_0 = [\mathbf{C}] + \sum_{n=1}^{\infty} n[\mathbf{C_nB_n}] + \sum_{n=1}^{\infty} n[\mathbf{C_nB_{n+1}}] + \sum_{n=1}^{\infty}(n+1)[\mathbf{C_{n+1}B_n}] \quad \text{(S14)}$$

Combining equations (S4)-(S6) with equations (S11)-(S13) we have that:

$$[\mathbf{C_nB_n}] = K_1(4K_2K_3)^{n-1}[\mathbf{C}]^n[\mathbf{B}]^n \quad \text{(S15)}$$

$$[\mathbf{C_nB_{n+1}}] = K_1K_2(4K_2K_3)^{n-1}[\mathbf{C}]^n[\mathbf{B}]^{n+1} \quad \text{(S16)}$$

$$[\mathbf{C_{n+1}B_n}] = K_1K_3(4K_2K_3)^{n-1}[\mathbf{C}]^{n+1}[\mathbf{B}]^n \quad \text{(S17)}$$

Substituting equations (S15)-(S17) in equation (S14) we have that:

$$[\mathbf{C}]_0 = [\mathbf{C}] + \sum_{n=1}^{\infty} nK_1(4K_2K_3)^{n-1}[\mathbf{C}]^n[\mathbf{B}]^n + \sum_{n=1}^{\infty} nK_1K_2(4K_2K_3)^{n-1}[\mathbf{C}]^n[\mathbf{B}]^{n+1} + \sum_{n=1}^{\infty}(n+1)K_1K_3(4K_2K_3)^{n-1}[\mathbf{C}]^{n+1}[\mathbf{B}]^n \quad \text{(S18)}$$

Re-arranging equation (S18) we have:

$$[\mathbf{C}]_0 = [\mathbf{C}] + K_1[\mathbf{C}][\mathbf{B}]\sum_{n=1}^{\infty} n(4K_2K_3[\mathbf{C}][\mathbf{B}])^{n-1} + K_1K_2[\mathbf{C}][\mathbf{B}]^2\sum_{n=1}^{\infty} n(4K_2K_3[\mathbf{C}][\mathbf{B}])^{n-1} + K_1K_3[\mathbf{C}]^2[\mathbf{B}]\sum_{n=1}^{\infty} n(4K_2K_3[\mathbf{C}][\mathbf{B}])^{n-1} + K_1K_3[\mathbf{C}]^2[\mathbf{B}]\sum_{n=1}^{\infty}(4K_2K_3[\mathbf{C}][\mathbf{B}])^{n-1} \quad \text{(S19)}$$

which can be further re-arranged into:

$$[\mathbf{C}]_0 = [\mathbf{C}] + (1 + K_2[\mathbf{B}] + K_3[\mathbf{C}])K_1[\mathbf{C}][\mathbf{B}] \sum_{n=1}^{\infty} n(4K_2K_3[\mathbf{C}][\mathbf{B}])^{n-1} + K_1K_3[\mathbf{C}]^2[\mathbf{B}] \sum_{n=1}^{\infty} (4K_2K_3[\mathbf{C}][\mathbf{B}])^{n-1}$$

(S20)

Taylor formulae for the relevant convergent series state that, for $x < 1$:

$\sum_{n=1}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}$ (S21)

and

$\sum_{n=1}^{\infty} x^{n-1} = \frac{1}{1-x}$ (S22)

Applying equations (S21) and (S22) to the equation (S20) we have:

$$[\mathbf{C}]_0 = [\mathbf{C}] + (1 + K_2[\mathbf{B}] + K_3[\mathbf{C}]) \frac{K_1[\mathbf{C}][\mathbf{B}]}{(1 - 4K_2K_3[\mathbf{C}][\mathbf{B}])^2} + \frac{K_1K_3[\mathbf{C}]^2[\mathbf{B}]}{1 - 4K_2K_3[\mathbf{C}][\mathbf{B}]}$$

(S23)

For the total concentration of bipyridine , $[\mathbf{B}]_0$, we have that:

$[\mathbf{B}]_0 = [\mathbf{B}] + \sum_{n=1}^{\infty} n[\mathbf{C_n B_n}] + \sum_{n=1}^{\infty} (n+1)[\mathbf{C_n B_{n+1}}] + \sum_{n=1}^{\infty} n[\mathbf{C_{n+1} B_n}]$ (S24)

Applying the analogous transformations that have been applied to equation (S14) for $[\mathbf{C}]_0$, we have that:

$$[\mathbf{B}]_0 = [\mathbf{B}] + (1 + K_2[\mathbf{B}] + K_3[\mathbf{C}]) \frac{K_1[\mathbf{C}][\mathbf{B}]}{(1 - 4K_2K_3[\mathbf{C}][\mathbf{B}])^2} + \frac{K_1K_2[\mathbf{C}][\mathbf{B}]^2}{1 - 4K_2K_3[\mathbf{C}][\mathbf{B}]}$$

(S25)

Knowing K_1, K_2 and K_3 the systems of equations (S23) and (S25) can be solved for any set of total concentrations $[\mathbf{C}]_0$ and $[\mathbf{B}]_0$, obtaining the concentrations of free building blocks, [**C**] and [**B**] as the solutions of the system of equations. Knowing [**C**] and [**B**] the concentration of any particular oligomeric species can be calculated, using equations (S15)-(S17). For example, the curves that show the variation of the main species in the NMR experiment (**C**, **CB**, $\mathbf{CB_2}$, $\mathbf{C_2B}$, $\mathbf{C_2B_2}$, and $\mathbf{C_2B_3}$, Figure 2C) as a function of total concentration of bypiridine, $[\mathbf{B}]_0$, at constant $[\mathbf{C}]_0$ were generated solving the system of equations (S23) and (S24) and applying the values of [**C**] and [**B**] found to the equations (S26)-(S30) below:

$[\mathbf{CB}] = K_1[\mathbf{C}][\mathbf{B}]$ (S26)

$[\mathbf{CB_2}] = K_1K_2[\mathbf{C}][\mathbf{B}]^2$ (S27)

$$[\mathbf{C_2B}] = K_1K_3[\mathbf{C}]^2[\mathbf{B}] \quad \text{(S28)}$$

$$[\mathbf{C_2B_2}] = K_1(4K_2K_3)[\mathbf{C}]^2[\mathbf{B}]^2 \quad \text{(S29)}$$

$$[\mathbf{C_2B_3}] = K_1K_2(4K_2K_3)[\mathbf{C}]^2[\mathbf{B}]^3 \quad \text{(S30)}$$

Equations (S26)-(S30) were obtained by replacing the appropriate *n* values in equations (S15)-(S17). We used the program Micromath Scientist 2.0 to generate the curves shown in Figure 2C.

Average number of repeats in the oligomer of the form C_nB_n

In an isodesmic oligomerization the average length of the oligomer corresponds to the average number of monomer repeats, <*N*>, and it can be calculated from the concentration of the free monomer and the oligomerization constant as follows:

$$< N >= \frac{1}{1-K_o[\mathbf{M}]} \quad \text{(S31)}$$

Combining equation (1) with equation (S15) we have that:

$$[\mathbf{C_nB_n}] = K_1(4K_2K_3)^{n-1}\frac{[\mathbf{CB}]^n}{{K_1}^n} \quad \text{(S32)}$$

That can be re-arranged as:

$$\frac{(4K_2K_3)^{n-1}}{{K_1}^{n-1}} = \frac{[\mathbf{C_nB_n}]}{[\mathbf{CB}]^n} \quad \text{(S33)}$$

On the other hand, the oligomerization constant K_o of the monomer **CB** can be expressed as:

$${K_o}^{n-1} = \frac{[\mathbf{C_nB_n}]}{[\mathbf{CB}]^n} \quad \text{(S34)}$$

Combining equations (S33) and (S34) we have that:

$$K_o = \frac{4K_2K_3}{K_1} \quad \text{(S35)}$$

On the other hand, applying equation (S31) to the oligomerization of **CB** we have:

$$< N >= \frac{1}{1-K_o[\mathbf{CB}]} \quad \text{(S36)}$$

Combining equations (1) and (S36) we have:

$$< N >= \frac{1}{1-4K_2K_3[\mathbf{C}][\mathbf{B}]} \quad \text{(5)}$$

Model for the dimerization of the oligomer

In our system, only long oligomers undergo dimerization. For long oligomers, the number of **C** and **B** building blocks can be assumed to be the same. We need therefore only to consider the formation of oligomers of the form $\mathbf{C_nB_n}$, which can be de-assembled by the presence of excess **C** (to yield $\mathbf{C_2B}$) or excess **B** (resulting in $\mathbf{CB_2}$). $\mathbf{C_nB_n}$ can also dimerize to form oligomer dimers $(\mathbf{C_nB_n})_2$. The dimerization constant for an oligomer with n repeats, K_{nl}, can be expressed as:

$$K_{nl} = \frac{[(\mathbf{C_nB_n})_2]}{[\mathbf{C_nB_n}]^2} \qquad (6)$$

K_{nl} can be expressed as a function of the binding affinity per unit repeat, K_d, and the effective molarity between the dimerization sites, *EM*:

$$K_{nl} = {K_l}^n EM^{n-1} \qquad (7)$$

Substituting equation (5) in equation (S35) we have that:

$${K_l}^n EM^{n-1} = \frac{[(\mathbf{C_nB_n})_2]}{[\mathbf{C_nB_n}]^2} \qquad (S37)$$

And combining equations (S15) and (S37) we have that:

$$[(\mathbf{C_nB_n})_2] = {K_l}^n EM^{n-1}(K_1(4K_2K_3)^{n-1}[\mathbf{C}]^n[\mathbf{B}]^n)^2 \qquad (S38)$$

The total concentration of **C**, $[\mathbf{C}]_0$, is therefore

$$[\mathbf{C}]_0 = [\mathbf{C}] + [\mathbf{CB_2}] + 2[\mathbf{C_2B}] + \sum_{n=1}^{\infty} n[\mathbf{C_nB_n}] + \sum_{n=1}^{\infty} 2n[(\mathbf{C_nB_n})_2] \qquad (S39)$$

Combining equations (1), (2), (3), (S15), (S38) and (S39) we have that:

$$[\mathbf{C}]_0 = [\mathbf{C}] + K_1K_2[\mathbf{C}][\mathbf{B}]^2 + 2K_1K_3[\mathbf{C}]^2[\mathbf{B}] + \sum_{n=1}^{\infty} nK_1(4K_2K_3)^{n-1}[\mathbf{C}]^n[\mathbf{B}]^n + \sum_{n=1}^{\infty} 2n{K_l}^n EM^{n-1}(K_1(4K_2K_3)^{n-1}[\mathbf{C}]^n[\mathbf{B}]^n)^2 \qquad (S40)$$

that can be re-arranged into:

$$[\mathbf{C}]_0 = [\mathbf{C}] + K_1K_2[\mathbf{C}][\mathbf{B}]^2 + 2K_1K_3[\mathbf{C}]^2[\mathbf{B}] + K_1[\mathbf{C}][\mathbf{B}]\sum_{n=1}^{\infty} n(4K_2K_3[\mathbf{C}][\mathbf{B}])^{n-1} + 2{K_1}^2[\mathbf{C}]^2[\mathbf{B}]^2K_d\sum_{n=1}^{\infty} n(K_lEM(4K_2K_3)^2[\mathbf{C}]^2[\mathbf{B}]^2)^{n-1} \qquad (S41)$$

Applying equations (S21) to equation (S41) we have that:

$$[\mathbf{C}]_0 = [\mathbf{C}] + K_1K_2[\mathbf{C}][\mathbf{B}]^2 + 2K_1K_3[\mathbf{C}]^2[\mathbf{B}] + \frac{K_1[\mathbf{C}][\mathbf{B}]}{(1-4K_2K_3[\mathbf{C}][\mathbf{B}])^2} + \frac{2K_l{K_1}^2[\mathbf{C}]^2[\mathbf{B}]^2}{(1-K_lEM(4K_2K_3)^2[\mathbf{C}]^2[\mathbf{B}]^2)^2}$$

(S42)

Similarly, for $[\mathbf{B}]_0$ we have:

$$[\mathbf{B}]_0 = [\mathbf{B}] + 2K_1K_2[\mathbf{C}][\mathbf{B}]^2 + K_1K_3[\mathbf{C}]^2[\mathbf{B}] + \frac{K_1[\mathbf{C}][\mathbf{B}]}{(1-4K_2K_3[\mathbf{C}][\mathbf{B}])^2} + \frac{2K_l{K_1}^2[\mathbf{C}]^2[\mathbf{B}]^2}{(1-K_lEM(4K_2K_3)^2[\mathbf{C}]^2[\mathbf{B}]^2)^2}$$

(S43)

The system of equations (S42) and (S43) is rather complex and solving it is not an easy task. A useful simplification can be drawn from the fact that K_1, which is responsible for the formation of the repeating unit in the oligomer (i.e., the specie **CB**) is much larger than all the other constants. We can therefore assume that this is the dominant equilibrium that determines how much of the repeating unit **CB** is available for oligomerization or for interaction with free **B** or **C**. Thus, the first step is to determine the available concentration of **CB** prior of oligomerization, that is, $[\mathbf{CB}]_0$ the excess concentration of **B**, $[\mathbf{B}]_x$, (that is, **B** free after formation of **CB**) and excess concentration of **C**, $[\mathbf{C}]_x$, (that is , **C** free after formation of **CB**) using the following system of equations:

$$K_1 = \frac{[\mathbf{CB}]_0}{[\mathbf{C}]_x[\mathbf{B}]_x} \qquad \text{(S44)}$$

$$[\mathbf{C}]_0 = [\mathbf{C}]_x + [\mathbf{CB}]_0 \qquad \text{(S45)}$$

$$[\mathbf{B}]_0 = [\mathbf{B}]_x + [\mathbf{CB}]_0 \qquad \text{(S46)}$$

The monomer **CB** may also form $\mathbf{CB_2}$ and $\mathbf{C_2B}$ by interaction with **B** or **C**, according to equations (2) and (3). Upon oligomerization or interaction with excess **C** and **B**, the total concentration of monomer, $[\mathbf{CB}]_0$, is:

$$[\mathbf{CB}]_0 = [\mathbf{CB_2}] + [\mathbf{C_2B}] + \sum_{n=1}^{\infty} n[\mathbf{C_nB_n}] + \sum_{n=1}^{\infty} 2n[(\mathbf{C_nB_n})_2] \qquad \text{(S47)}$$

Using the corresponding mass balances:

$$[\mathbf{B}]_x = [\mathbf{B}] + [\mathbf{CB_2}] \qquad \text{(S48)}$$

$$[\mathbf{C}]_x = [\mathbf{C}] + [\mathbf{C_2B}] \qquad \text{(S49)}$$

and equations (2) and (3), the concentrations of species $\mathbf{CB_2}$ and $\mathbf{C_2B}$ can be written as a function of the corresponding constant, the concentration of free monomer **CB** and the total excess **B** or **C**, $[\mathbf{B}]_x$ or $[\mathbf{C}]_x$:

$$[\mathbf{CB_2}] = \frac{K_2[\mathbf{CB}][\mathbf{B}]_x}{(1+K_2[\mathbf{CB}])} \qquad \text{(S50)}$$

$$[\mathbf{C_2B}] = \frac{K_3[\mathbf{CB}][\mathbf{C}]_x}{(1+K_3[\mathbf{CB}])} \qquad \text{(S51)}$$

On the other hand, combining (1) with (S35) and (S15) we have that:

$$[\mathbf{C_n B_n}] = (K_o[\mathbf{CB}])^{n-1} \qquad \text{(S52)}$$

And combining (1) with (S35) and (S38):

$$[(\mathbf{C_n B_n})_2] = (K_l E M K_o^2[\mathbf{CB}]^2)^{n-1} \qquad \text{(S53)}$$

Combining equations (S47) with equations (S50)-(S53) we have:

$$[\mathbf{CB}]_0 = \frac{K_2[\mathbf{CB}][\mathbf{B}]_x}{(1+K_2[\mathbf{CB}])} + \frac{K_3[\mathbf{CB}][\mathbf{C}]_x}{(1+K_3[\mathbf{CB}])} + [\mathbf{CB}]\sum_{n=1}^{\infty} n(K_o[\mathbf{CB}])^{n-1} + 2K_l[\mathbf{CB}]^2\sum_{n=1}^{\infty} n(K_l E M K_o^2[\mathbf{CB}]^2)^{n-1} \qquad \text{(S54)}$$

and applying equation (S21) to the relevant convergent series in equation (S54) we have:

$$[\mathbf{CB}]_0 = \frac{K_2[\mathbf{CB}][\mathbf{B}]_x}{(1+K_2[\mathbf{CB}])} + \frac{K_3[\mathbf{CB}][\mathbf{C}]_x}{(1+K_3[\mathbf{CB}])} + \frac{[\mathbf{CB}]}{(1-K_o[\mathbf{CB}])^2} + \frac{2K_l[\mathbf{CB}]^2}{(1-K_l E M K_o^2[\mathbf{CB}]^2)^2} \qquad (8)$$

which in combination with equations (S45) and (S46) allow us to estimate the concentration of free **CB** for any set of initial **C** and **B** concentrations. Knowing free **CB** it is then possible to determine the concentration of any particular assembly in solution.

Changes in absorbance at wavelength around 460 nm are attributed to changes in concentration of polymer building block **CB** at the of double stranded polymer $(\mathbf{C_nB_n})_2$, [**CB**@dsp], an can be written as:

$$A = \varepsilon[\mathbf{CB}@dsp] \qquad \text{(S55)}$$

where ε is the molar extinction coefficient for the monomer units within the polymer dimer at the wavelength under study.

On the other hand, , [**CB**@dsp] is:

$$[\mathbf{CB}@dsp] = \sum_{n=1}^{\infty} 2n[(\mathbf{C_nB_n})_2] \qquad \text{(S56)}$$

which, substituting in equation (S55) becomes:

$$[\mathbf{CB}@dsp] = 2K_l[\mathbf{CB}]^2 \sum_{n=1}^{\infty} n(K_l EMK_o^2[\mathbf{CB}]^2)^{n-1} \qquad \text{(S57)}$$

and applying equation (S23) becomes:

$$A = \varepsilon \frac{2K_l[\mathbf{CB}]^2}{(1-K_l EMK_o^2[\mathbf{CB}]^2)^2} \qquad (9)$$

To fit the experimental data we first obtained the values of $[\mathbf{CB}]_0$, $[\mathbf{C}]_x$ and $[\mathbf{B}]_x$ from the initial concentration of **C** and **B** used in the experiments, i.e., $[\mathbf{C}]_0$ and $[\mathbf{B}]_0$. $[\mathbf{CB}]_0$, $[\mathbf{C}]_x$ and $[\mathbf{B}]_x$ were then used as the independent variables input in the model described by equations (8) and (9). We used Micromath Scientist 3.0 to fit the data. We then use the ratio **B**/**C** in the *x* axis for clarity in the graphical representation shown in Figure 4B and in Figure S6.

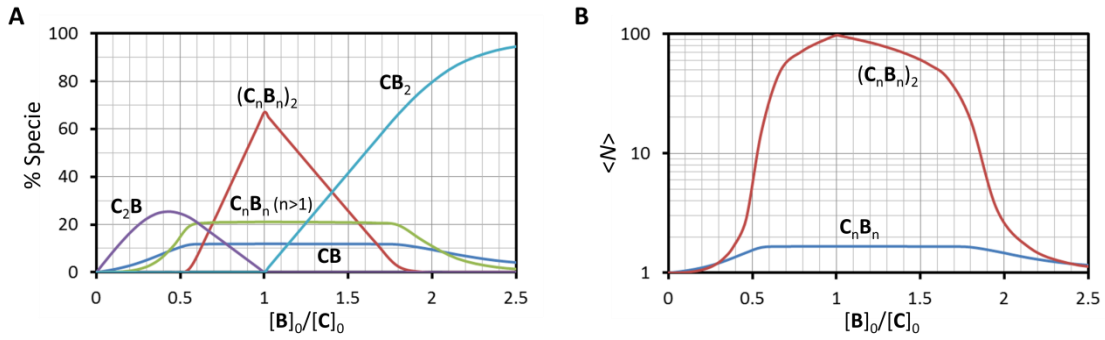

Figure S6. A. Variation of the percentage of **C** present in each of the species with the ratio of concentrations of **B** and **C** during the the titration experiment shown in Figure 3B. B. Changes in the average number of repeats <N> for the single stranded oligomer and the double stranded polymer during the titration experiment shown in Figure 3B.

Average size of the oligomer and the oligomer dimer.

The formation of the polymer dimer can be approached as the dimerization of linear oligomers of the form $\mathbf{C}_n\mathbf{B}_n$ or as the polymerization of the dimer $(\mathbf{CB})_2$, formed by the lateral association of two **CB** complexes. From equation (S37), it can be shown that the dimerization of an oligomer with a single repeat unit can be written as:

$$K_l = \frac{[(\mathbf{CB})_2]}{[\mathbf{CB}]^2} \qquad \text{(S58)}$$

The polymerization of the dimer $(\mathbf{CB})_2$ can be then written as a function of the oligomerization constant of the dimer, K_{2o}:

$$K_{2o} = \frac{[(\mathbf{C_nB_n})_2]}{[(\mathbf{C_{n-1}B_{n-1}})_2][(\mathbf{CB})_2]} \qquad \text{(S59)}$$

K_{2o} can be also written as a function of K_o, the oligomerization constant of the complex **CB**:

$$K_{2o} = {K_o}^2 EM \qquad \text{(S60)}$$

By combining equations (S36), (S59) and (S60), the average number of repeats in the polymer dimer, $<N_2>$, can be written as:

$$< N_2 >= \frac{1}{1-{K_o}^2 EM[(\mathbf{CB})_2]} \qquad \text{(S61)}$$

and substituting in equation (S58) we have:

$$< N_2 >= \frac{1}{1-{K_o}^2 EMK_l[\mathbf{CB}]^2} \qquad \text{(10)}$$

Equation (10) was used to calculate the average number of repeats shown in Figure S6 B.

Modelling of the formation of oligomer bundles.

We assume that the formation of the bundle is isodesmic, that is, pre-formed bundles have the same affinity for an additional strand regardless of the number of bundles already present. Unlike the growth in the longitudinal direction, the number of strands that form the bundle is discrete. The affinity constant between a bundle and an additional strand is K_l. The total concentration of monomer for an oligomer that can form bundles of *m* strands is:

$$[\mathbf{M}]_0 = \sum_{i=1}^{i=m} \sum_{n=1}^{\infty} in[(\mathbf{M}_\mathrm{n})_i] \qquad \text{(S62)}$$

Let's assume that $n = 4$. Equation (S65) becomes:

$$[\mathbf{M}]_0 = \sum_{n=1}^{\infty} n[\mathbf{M_n}] + \sum_{n=1}^{\infty} 2n[(\mathbf{M}_\mathrm{n})_2] + \sum_{n=1}^{\infty} 3n[(\mathbf{M}_\mathrm{n})_3] + \sum_{n=1}^{\infty} 4n[(\mathbf{M}_\mathrm{n})_4] \qquad \text{(S63)}$$

The formation of a linear strand $\mathbf{M}_\mathrm{n}$ can be written as a function of the free monomer concentration [**M**] as follows (See equilibrium 1 in Fig. 5A):

$${K_o}^{n-1} = \frac{[\mathbf{M}_\mathrm{n}]}{[\mathbf{M}]^n} \qquad \text{(S64)}$$

while the formation of a multi-stranded bundle of *m* strands from the assembly linear single strands $\mathbf{M}_\mathrm{n}$ can be written as (equilibrium 3 in Fig. 5A):

$${K_{nl}}^{m-1} = \frac{[(\mathbf{M}_\mathrm{n})_m]}{[\mathbf{M}_\mathrm{n}]^m} \qquad \text{(S65)}$$

which in combination with equation (5) gives:

$$({K_l}^n EM^{n-1})^{m-1} = \frac{[(\mathbf{M}_\mathrm{n})_m]}{[\mathbf{M}_\mathrm{n}]^m} \qquad \text{(S66)}$$

Alternatively, the formation of a bundle of monomers of **M**, $\mathbf{M}_\mathrm{m}$, can be written as (equilibrium 2, Fig. 5A):

$${K_l}^{m-1} = \frac{[\mathbf{M}_\mathrm{m}]}{[\mathbf{M}]^m} \qquad \text{(S67)}$$

while the polymerization of $\mathbf{M}_\mathrm{m}$ (equilibrium 4 in Fig. 5A) can be written as:

$$(K_{mo})^{n-1} = \frac{[(\mathbf{M}_\mathrm{n})_m]}{[\mathbf{M}_\mathrm{n}]^m} \qquad \text{(S68)}$$

which in combination with equation (11) gives:

$$({K_o}^m EM^{m-1})^{n-1} = \frac{[(\mathbf{M}_n)_m]}{[\mathbf{M}_n]^m} \qquad \text{(S69)}$$

Combining either equations (S64) and (S66) or (S67) and (S69) allows writing the concentration of any bundle as a function of the concentration of free monomer:

$$[(\mathbf{M}_n)_m] = ({K_l}^n EM^{n-1})^{m-1}({K_o}^{n-1}[\mathbf{M}]^n)^m \qquad \text{(S70)}$$

Substituting equation (S70) in equation (S63) we have, for *m* = 4:

$$[\mathbf{M}]_0 = \sum_{n=1}^{\infty} n{K_o}^{n-1}[\mathbf{M}]^n + \sum_{n=1}^{\infty} 2n{K_l}^n EM^{n-1}({K_o}^{n-1}[\mathbf{M}]^n)^2 + \sum_{n=1}^{\infty} 3n{K_l}^{2n} EM^{2(n-1)}({K_o}^{n-1}[\mathbf{M}]^n)^3 + \sum_{n=1}^{\infty} 4n{K_l}^{3n} EM^{3(n-1)}({K_o}^{n-1}[\mathbf{M}]^n)^4$$

(S71)

Applying the Taylor convergent series formula (S21) we have that:

$$[\mathbf{M}]_0 = \frac{[\mathbf{M}]}{(1-K_o[\mathbf{M}])^2} + \frac{2K_l[\mathbf{M}]^2}{(1-K_o^2K_lEM[\mathbf{M}]^2)^2} + \frac{3{K_l}^2[\mathbf{M}]^3}{(1-K_o^3{K_l}^2EM^2[\mathbf{M}]^3)^2} + \frac{4{K_l}^3[\mathbf{M}]^4}{(1-K_o^4{K_l}^3EM^3[\mathbf{M}]^4)^2}$$

(S72)

which for the general case where we can have up to *m* strands in a bundle, can be written as:

$$[\mathbf{M}]_o = \sum_{i=1}^{i=m} \frac{i{K_l}^{i-1}[\mathbf{M}]^i}{(1-K_o^i{K_l}^{i-1}EM^{i-1}[\mathbf{M}]^i)^2} \qquad \text{(12)}$$

The average number of repeats in the polymer bundle, $<N_m>$, can be written as a function of the concentration of the corresponding monomer, $\mathbf{M}_m$, with polymerization constant K_{mo}:

$$< N_m >= \frac{1}{1-K_{mo}[\mathbf{M}_m]} \qquad \text{(S73)}$$

Using equation (S67) and (11) in (S73) we have that:

$$< N_m >= \frac{1}{1-{K_o}^m{K_l}^{m-1}EM^{m-1}[\mathbf{M}]^m} \qquad \text{(13)}$$

Equations (12) and (13) where used to generate the plots in Figure 5.

Inspecting equation (12) it becomes clear that the concentration of free monomer is limited by the fact that the denominator has to be a positive number. That is:

$$(1-K_o^m{K_l}^{m-1}EM^{m-1}[\mathbf{M}]^m)^2 > 0 \qquad \text{(S74)}$$

which means that [**M**] has to be:

$$[\mathbf{M}] < \frac{(K_lEM)^{1/m}}{K_lEMK_o} \qquad \text{(S75)}$$

Therefore the maximum concentration of **M** that can be reached, $[\mathbf{M}]_x$, is:

$$[\mathbf{M}]_x = \frac{(K_l EM)^{1/m}}{K_l EMK_o} \qquad \text{(S76)}$$

Before $[\mathbf{M}]_x$ is reached only monomer and single stranded oligomers are present to any meaningful extent. After $[\mathbf{M}]_x$ is reached any additional monomer added is added exclusively to the bundle with the maximum number of strands *m* (Fig. 5B). Therefore, in practice, the total concentration of monomer, $[\mathbf{M}]_0$ is:

$$[\mathbf{M}]_o = \sum_{n=1}^{\infty} nK_o{}^{n-1}[\mathbf{M}]^n + \sum_{n=1}^{\infty} mn[(\mathbf{M}_\mathrm{n})_m] \qquad \text{(S77)}$$

which can be written as a function of the concentration of free monomer as follows:

$$[\mathbf{M}]_o = \frac{[\mathbf{M}]}{(1-K_o[\mathbf{M}])^2} + \frac{mK_l{}^{m-1}[\mathbf{M}]^m}{(1-K_o^m K_l{}^{m-1} EM^{m-1}[\mathbf{M}]^m)^2} \qquad \text{(S78)}$$

We define the nucleation concentration, *NC*, as the total concentration of monomer at which $[\mathbf{M}]_x$ is reached. Since in practice only monomer and single stranded oligomer exist before $[\mathbf{M}]_x$ is reached, the *NC* is the sum of single stranded oligomer and monomer when $[\mathbf{M}]_x$ is reached, that is:

$$NC = \frac{[\mathbf{M}]_x}{(1-K_o[\mathbf{M}]_x)^2} \qquad \text{(S79)}$$

Substituting in equation (S76) we have that:

$$NC = \frac{\frac{(K_l EM)^{1/m}}{K_l EMK_o}}{(1-K_o\frac{(K_l EM)^{1/m}}{K_l EMK_o})^2} \qquad \text{(S80)}$$

that can be simplified to:

$$NC = \frac{(K_l EM)^{1/m}}{K_l EMK_o(1-\frac{(K_l EM)^{1/m}}{K_l EM})^2} \qquad \text{(14)}$$

Combining equation (13) with equation (S78) we have that:

$$[\mathbf{M}]_o = \frac{[\mathbf{M}]}{(1-K_o[\mathbf{M}])^2} + mK_l{}^{m-1}[\mathbf{M}]^m(< N_m >)^2 \qquad \text{(S81)}$$

which, when $[\mathbf{M}]_x$ is reached, becomes:

$$[\mathbf{M}]_o = \frac{[\mathbf{M}]_x}{(1-K_o[\mathbf{M}]_x)^2} + mK_l{}^{m-1}[\mathbf{M}]_x{}^m(< N_m >)^2 \qquad \text{(S82)}$$

Substituting in equation (S79) we have:

$$[\mathbf{M}]_o = NC + mK_l{}^{m-1}[\mathbf{M}]_x{}^m(< N_m >)^2 \qquad \text{(S83)}$$

Substituting equation (S76) in equation (S83) and re-arranging we have that:

$$(< N_m >)^2 = \frac{[\mathbf{M}]_o - NC}{mK_l{}^{m-1}\left(\frac{(K_l EM)^{1/m}}{K_l EMK_o}\right)^m} \qquad \text{(S84)}$$

and with further re-arrangement we have:

$$< N_m >= \sqrt{\frac{{K_o}^m EM^{m-1}}{m}([\mathbf{M}]_o - NC)} \qquad \text{(S85)}$$

which in the logarithmic form is:

$$log < N >= 0.5mlogK_o + 0.5(m-1)logEM - 0.5logm + 0.5log([\mathbf{M}]_o - NC) \qquad \text{(15)}$$

Substituting equation (14) into (15) we have:

$$log < N >= 0.5mlogK_o + 0.5(m-1)logEM - 0.5logm + 0.5log\left([\mathbf{M}]_o - \frac{(K_lEM)^{1/m}}{K_lEMK_o(1-\frac{(K_lEM)^{1/m}}{K_lEM})^2}\right) \qquad \text{(S86)}$$

When the total concentration of **M**, $[\mathbf{M}]_0$, is twice the *NC*, equation (S83) adopts the form:

$$log < N >= 0.5mlogK_o + 0.5(m-1)logEM - 0.5logm + 0.5log\left(\frac{(K_lEM)^{1/m}}{K_lEMK_o(1-\frac{(K_lEM)^{1/m}}{K_lEM})^2}\right) \qquad \text{(S87)}$$

which can be rearranged to:

$$log < N >= 0.5(m-1)\log K_oEM - 0.5logm - 0.5\left(\frac{m-1}{m}\right)\log K_lEM - log(1-(K_lEM)^{(1-m)/m}) \qquad \text{(16)}$$