Stimulated Binding of Polymer Chains in Narrow Tube Confinement

Gillian M. Fraser1, Márcio Sampaio Gomes Filho2,3, and Eugene M. Terentjev2
1Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
2Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge CB3 0HE, UK and
3Instituto de Física, Universidade de Brasília, Brasília-DF, Brazil

In biology, there are several processes in which unfolded protein chains are transported along narrow-tube channels. Normally, unfolded polypeptides would not bind to each other. However, when the chain entropy is severely reduced in the narrow channel, we find that polymer chains have a propensity to bind, even if there is no great potential energy gain in doing so. We find the average length of binding $m^*$ (the number of monomers on overlapping chain ends that form bonds) and the critical tube diameter at which such constrained binding occurs. Brownian dynamics simulations of tightly confined chains, confirm the theoretical arguments and demonstrate chain binding over the characteristic length $m^*$, changing in tubes of different diameter.

PACS numbers: 36.20.Ey, 87.15.hp

Much of the physics of polymer chains is determined by entropic effects; entropic barriers associated with chain configurations are widely recognized to control polymer behaviour in narrow pores [1–4]. Here we consider a related problem of how confinement in a narrow tube could precipitate binding of polymer chains, even though in free unconstrained conditions they would not have a propensity to bind. The problem is motivated by a set of remarkable biological structures that facilitate movement of nucleic acids or polypeptides through narrow channels: (i) contractile tail bacteriophages inject their nucleic acid genome into host bacterial cells through their hollow tail tube [5]; (ii) bacterial type IV secretion systems transfer nucleic acids between bacteria and proteins into eukaryotic host cells [6]; (iii) ribosomes release growing polypeptides through the ribosomal exit tunnel where, for some proteins, the beginning of folding is registered while in confinement [7, 8]. Finally, bacterial type III secretion systems export unfolded polypeptides that assemble into cell-surface nanomachines that inject bacterial proteins into eukaryotic cells [9] or, in the case of bacterial flagella, facilitate cell motility [10]. During flagella biogenesis, individual subunits synthesised inside the cell need to be unfolded and threaded through a narrow channel that runs the full length of the growing structure, to be assembled at the distal tip [9,11].

For flagella assembly, two mechanisms of subunit transport along the narrow channel have been investigated. One relies on single-file diffusion [12], which requires ‘pushing’ a column of subunits through the narrow channel, against tremendous resistance once the crowded regime sets in [13]. The other mechanism suggests that unfolded flagellar subunits form a linked chain along the whole length of the flagellar channel, which is then ‘pulled’ to the site of subunit assembly at the flagellum tip [14]. Here, we propose a mechanism by which unfolded polymers would strongly bind to form head-to-tail links, when constrained in a channel.

The properties of polymer chain ($N$ segments of size $a$) constrained in a narrow channel are very well known, starting from the seminal work of Casassa [15] and Edwards [16,17], and extensively reviewed in many subsequent publications [4,18,19]. The free energy excess contains two contributions: the ‘ideal gas’ motion of the chain centre of mass along the tube, and the reduction in configurational entropy under constraint:

$$
\frac{F_1}{k_BT} \sim -\ln \left( \frac{d}{a} \right)^2 + \left( \frac{R_g}{d} \right)^2 = -2\ln \left( \frac{d}{a} \right) + \frac{\pi^2 N}{3} \left( \frac{a}{d} \right)^2,
$$

(1)

where the second expression is for the ideal Gaussian chain, with the radius of gyration $R_g = N^{1/2}a$. This is good enough for our purposes; in fact, the commonly used model of self-avoiding chain in good solvent might be less appropriate for a case of protein unfolded in a...
When two such chains are in the channel, Fig. 3(a), the free energy is additive: $2F_1$. When these two chains are instead bonded over the length of $m$ overlapping units, see Fig. 1(b), the corresponding free energy has only one centre-of-mass term. The configurational free energy has $2(N - m)$ monomers constrained in a tube of diameter $d$, and strictly speaking, $m$ monomers constrained in a smaller diameter $(d - a)$:

$$
\frac{F_{\text{bond}}}{k_B T} = \frac{-m\Delta}{k_B T} - 2\ln\left(\frac{d}{a}\right) + \frac{2\pi^2}{3}(N - m)\left(\frac{a}{d}\right)^2 + \frac{\pi^2 m}{3}\left(\frac{a}{d - a}\right)^2,
$$

where $\Delta$ is the potential energy gain on making one bond (see 20 and many subsequent studies of this energy for aminoacid residues in contact). We shall be interested in the situation when this gain is small (or non-existent), so that the monomers would not bond in the free-chain conditions. Figure 1(c) shows these two free energies compared, as a function of changing tube diameter. What we discover, is that at large $d$ (when the entropic constraint is weak), the free energy of two independent chains is lower, but as $d$ becomes smaller – the free energy of the bonded pair is lower. In fact, the free energy difference $\Delta F = F_{\text{bond}} - 2F_1$ no longer depends on the overall chain length $N$, and reveals the key equilibrium effect:

$$
\frac{\Delta F}{k_B T} = 2\ln\left(\frac{d}{a}\right) - m\left[\frac{\pi^2 a^2 (d^2 - 4ad + 2a^2)}{3d^2 (d - a)^2} + \beta \Delta\right],
$$

with $\beta = 1/k_B T$. We see that the ideal-gas entropy 'wins' at large $d$, but at $(d - a) \to 0$ the confinement effect enforces the binding via the negative-$m$ contribution. Even when there is no potential gain in binding ($\Delta = 0$) or the monomers repel each other (negative $\Delta$), still in a sufficiently narrow tube they would prefer to bind.

However, the equilibrium analysis leading to the Eq. 3 cannot predict which is the length of bonded segment: obviously, in equilibrium, the larger the overlap $m$ is, the lower is the free energy. The reason why the two chains adopt a particular binding length $m^*$ is entirely kinetic. Two effects compete: the effective rate of the bonding reaction, and the rate of chain reptation along the tube to increase the overlap length $m$. Once any one of the $m$ bonds along the overlapped segment is established, the reptation stops, and the rest of the bonds set rapidly, in sequence, via the process of accelerated zipper 21.

Let us say that the rate of an individual bonding reaction between two monomers is $k_{\text{on}}$, determined by the chemistry involved. The probability of setting an individual bond during a given time interval $\Delta t$ is 22:

$$
p_{\text{on}} = 1 - e^{-k_{\text{on}}\Delta t},
$$

(at small $\Delta t$, $p_{\text{on}} = k_{\text{on}}\Delta t$, while at large $\Delta t$, the reaction occurs with an almost certainty). The probability to establish at least one bond out of a possible $m$ sequence is

$$
P(m) = 1 - [1 - p_{\text{on}}]^m = 1 - e^{-mk_{\text{on}}\Delta t(m)}.
$$

Any one of the $m$ possible bonds along the overlapping segment stops the reptation diffusion, and is rapidly followed by the full bonding of the segment.

We are now left with the second aspect of this kinetic problem: to find the time interval $\Delta t(m)$ for which the sequence of $m$ monomers dwells when two chains slide reptate past each other. The theory of polymer reptation is classical 17, and in our context we need to recall the characteristic time to diffuse the chain of $N$ monomers over a distance $ma$ in a tube of the remaining clearance $(d - a)$, cf. Fig. 1(b):

$$
\tau[m] = \frac{(ma)^2}{D_c}, \quad \text{with} \quad D_c = \frac{\pi^2 k_B T}{N\gamma} \left(\frac{d - a}{a}\right)^2,
$$

where the diffusion constant for the chain centre of mass is given by $D_c$, in which $\gamma$ is the friction constant of just one monomer (so $D_0 = k_B T/N\gamma$ is the diffusion constant of the whole polymer in free space: a value frequently measured for globular proteins in water). Equation 6 is more familiar in the form of reptation time of the whole chain length $(Na)$ in a clear tube $(d)$, when $\tau_d = N^3a^4\gamma/\pi^2k_BTd^2$; see 17 for detail. We now have:

$$
\Delta t(m) = \tau[m+1] - \tau[m] = \frac{Na^4\gamma}{\pi^2 k_B T (d - a)^2}(2m + 1).
$$

All these expressions from reptation theory are only valid for the tightly confined chain, with $d \ll R_g = N^{1/2}a$.

Figure 2 illustrates how the probability $P(m)$ depends on its key parameters. Now the kinetic problem we are addressing becomes that of the mean first-passage time 23 24. The question is: at which $m^*$ the first binding reaction would occur. The probability density that binding occurs at a given value of overlap $m$ is given by $f(m) = dP(m)/dm$, so that the mean first-binding length is $\bar{m} = \int m f(m) dm$. Carrying out the algebra,
we may work with the simpler expression in Eq. (8). Note the logarithmic scale of the m-axis. The dashed line indicates the range of tube diameters \( d \) above which the reptation theory is no longer valid, and the probability of \( m \) monomers to overlap vanishes.

The mean bonded length \( \bar{m} \) is also easily calculated, but its expression is more cumbersome. Figure 3 shows that there is no great difference in values between the mean and the median of \( f(m) \), and for all practical purposes we may work with the simpler expression in Eq. (8).

The mean bonded length \( \bar{m} \) is also easily calculated, but its expression is more cumbersome. Figure 3 shows that there is no great difference in values between the mean and the median of \( f(m) \), and for all practical purposes we may work with the simpler expression in Eq. (8).

It is interesting to assess the critical diameter of the tube, \( d^* \), below which the bonding is likely. It is obtained by solving the transcendent equation \( \Delta F(d) = 0 \), and has the form

\[
d^* = \frac{\sqrt{\pi^2 m^*/3}}{ProductLog(\pi^2 m^* e^{-m^* \beta^2 \Delta}/3)} \approx e^{m^* \beta^2 \Delta^2/2}, \tag{9}
\]

where \( m^* \) is given by Eq. (8), and the approximate form is valid when \( m^* e^{-m^* \beta^2 \Delta} \ll 1 \) (i.e. essentially for all non-negative \( \Delta \)). The exponential approximation of Eq. (9) is prominent for all non-repulsive interactions, while for negative \( \Delta \) (repulsion between monomers) the other limit of ‘ProductLog’ is in force, and \( d^* \sim a \) (in other words, bonding is unlikely). However, there is another upper boundary for the channel diameter that promotes binding: all analysis in Eqs. (6) was based on the chain reptation dynamics, which stops being applicable at \( d \geq N^{1/2}a \) when the random-coil (blob) chain dynamics would take over (the exact crossover is hard to identify). This is schematically labelled in Fig. 3.

Assuming we are in narrow enough channels for the reptation theory to work, we should try estimating the kinetically-set bonding length \( m^* \), which requires the value of parameter \( B \). The friction constant for one monomer (e.g. aminoacid) can be obtained approximately from \( \gamma = 6 \pi \eta a \), with \( \eta \) being the viscosity of water (0.7 mPa.s), and \( a = 0.3 \text{ nm} \) the size of a residue [20]. A flagellin protein with \( N \approx 495 \), at room temperature, should have the diffusion constant: \( D_0 = 2 \cdot 10^{-12} \text{ m}^2/\text{s} \) (in agreement with a typical diffusion constant of amino acids in water [23, 26]: \( k_B T/\gamma \sim 10^{-9} \text{ m}^2/\text{s} \), divided by 500).

It is hard to find values of the bonding reaction rate \( k_{on} \) for a pair of amino acid residues: it strongly depends on a large number of specific chemical factors [27]. In the spirit of our ‘average polymer chain’ with a single characteristic value of bond energy gain \( \Delta \), we may try a basic Kramers estimate: \( k_{on} = \omega_0 e^{\beta \Delta} \), where \( \omega_0 \) is the collision frequency in solution: \( \omega_0 \approx 10^8 \text{ s}^{-1} \) [22, 28]. Then, for a low \( \beta \Delta = 1 \), we have \( k_{on} = 2.7 \cdot 10^8 \text{ s}^{-1} \). Using Eq. (7) with the diameter of flagellar channel \( d = 2 \text{ nm} \), we obtain a value: \( B = 1.63 \), and the preferred length of binding \( m^* = 2.3 \). As a consequence, for the same set of parameters, the free energy of binding is very low: \( \Delta F \approx -k_B T \). A high reaction rate would commit the constrained chains to bond on the first monomer (\( m^* \leq 1 \)), and the bonding free energy is determined purely by the potential energy of binding \( \beta \Delta \). For comparison, a lower reaction rate \( k_{on} = 10^7 \text{ s}^{-1} \) would give: \( B = 8.4 \) and \( m^* = 13 \), with the corresponding binding free energy \( \Delta F \approx -12k_B T \) (maintaining \( \beta \Delta = 1 \)).

In order to test the ideas of the confinement-induced bonding, we carried out simulation of two chains in a reflective channel, using the LAMMPS package. We used the Kremer-Grest model for coarse-grained polymers [29, 30] where bonds between consecutive monomers are modelled with a harmonic-like finitely extensible elastic potential, made of a combination of the attractive FENE part and the repulsive LJ part (the approach used in many studies, old and new [31, 32], and the weak bending potential with the constant \( k_\theta = k_B T \). The non-bonded particles interact through the LJ potential of variable weak attraction strength (see [13, 33]). Two kinds of LJ potential create a subtlety in the simulation setup: one needs to be clear which ‘LJ units’ are used. We kept the average temperature constant, equal to the attractive strength \( \varepsilon \) of the pair LJ potential (which is equivalent to our earlier condition \( \beta \Delta = 1 \)). The LJ time and the damping constant of the simulation were linked to the same energy scale via the fluctuation-dissipation theorem: \( \tau = \sigma \sqrt{M/\varepsilon} \) and \( \text{damp} = 1/\tau \), where \( M \) is the mass and \( \sigma \) the diameter of an individual monomer, equivalent to the parameter \( a \) used in the analysis above (the simulation time-step was \( dt = 0.01\tau \)). On the other hand, in the Kremer and Grest model, the bond potential has the FENE bond strength linked to the repulsive-LJ strength: \( k = 30e^* /\sigma^2 \) [29, 30], which makes the bond length \( r_{bond} \sim 0.96\sigma \).

However, we found that when the LJ repulsion strength...
Overlap number $m^*$ of two long chains in a tube with different values of $d$. We determined $m^*$ as the number of monomers, which were $r \leq 1.1225\sigma$ apart: just above the minimum of the LJ potential at $2^{1/6}\sigma$. The time $t_B$ when the two chains first ‘lock’ is shown in the inset. The most frequent binding time $t_B$ increases rapidly when the tube diameter increases.

$\epsilon^* = \epsilon$ (which, in turn, is equal to $k_B T$), the chains do not bind in the channel. The reason was that the length of the bond between monomers on the chain was fluctuating strongly, and the required ‘matching’ of monomers along the two chains was not possible. When we insisted that the bond along the chain is strongly confined, by taking $\epsilon^* = 100\epsilon$, the confinement-induced binding of two chains was a strong and obvious effect, see Fig. 4. We observed that tightly confined chains eventually come in contact during their reptation diffusion; mostly such a contact was short-lived and the chains separated again – but on rare occasions, when a sequence of several monomers come into contact simultaneously, the chains ‘lock’ in binding. After a long simulation time, the number of overlapping monomers $m^*$ was reaching the equilibrium value, depending on the channel diameter, see Fig. 5. Note that in these simulations the $m^*$ plateau value for $d = 14\sigma$ and $18\sigma$ was the same; this was a systematic result reproduced in a large number of independent simulation runs, for all large diameters where the binding occurred. For $d = 20\sigma$ we did not register any binding event within the given time of simulation.

As the stimulated-binding process is clearly driven by chain kinetics in the confining tube, we have done many simulations (with different random seeds) to establish the statistical distribution of binding time $t_B$. Figure 6 gives an illustration of such a distribution, $f(t_B)$, obtained for $d = 5\sigma$; the plot also gives the result for $d = 10\sigma$ in the background, but we were not able to acquire a sufficient statistics for larger $d$ because of too many instances of very long $t_B$. This distribution has a very characteristic ‘heavy tail’: it decays with a weak power law (our fitting suggests $t^{-0.8}$). One of the consequences is that the average binding time $\langle t_B \rangle = \int f(t_B) dt_B$ is not defined (the integral diverges), and we have to use the median $t_B$ instead. Such distributions are found in the Levy flight processes [34, 35], and in our case the origin is similar: the events leading to an overlap of $m^*$ monomers are rare, but only these lead to the lasting binding of the chains. The key point is clear: two chains with very weak attraction (i.e. not naturally prone to binding on contact in the bulk) would be strongly bound if a large number of their monomers would simultaneously overlap. In a weakly constrained case, the probability of such overlap is very low, and the time $t_B$ to wait for such a rare event is longer than any realistic experimental time.

The theory presented here is deliberately qualitative, demonstrating the key concept of stimulated bonding in narrow channels. Its analysis is essentially scaling, with explicit calculations based on the ideal Gaussian chain limit (however, this limit becomes accurate for polymer chains under strong lateral confinement, such as in dense melts or other ‘tube-model’ situations). The idea of an ‘average’ homopolymer chain is also very limiting: a real protein would have a variety of aminoacid residues with very different potential energy of pair binding $\Delta$, between $-4k_B T$ and $+4k_B T$ (see [20] for detail): a follow-up problem with the broadly distributed quenched $\Delta$ is certainly of interest. Nevertheless, these approximations and simplifications allowed us to expose the physics of entropically stimulated bonding, and obtain analytical expressions for the main points of interest.
It is possible that a similar analysis would predict a much enhanced protein folding, once the conformational entropy is reduced by the tight channel confinement, as in ribosomes [7, 8]. Another interesting corollary of this concept is the need to examine effective reaction rates in micro-volumes. Today the technology of microfluidics and a ‘lab on a chip’ is widespread: the reduction of translational entropy (and when unfolded polymer chains are involved, also the configurational entropy) needs to be taken into account when analysing the reaction rates in very small volumes.

We thank Samuel Bell and Cheng-Tai Lee for very helpful discussions. GMF acknowledges support from the BBSRC (BB/M007197/1). MSGF acknowledges support from Coordena\c c\~ao de Aperfei\c coamento de Pessoal de N\ivel Superior, Brasil (CAPES), Fin. Code 001. EMT acknowledges support from the ERC AdG 786659 (APRA).

[1] M. Muthukumar and A. Baumgartner, Macromolecules 22, 1937 (1989).
[2] P. J. Park and W. Sung, Phys. Rev. E 57, 730 (1998).
[3] E. A. DiMarzio and A. J. Mandell, J. Chem. Phys. 107, 5510 (1997).
[4] M. Muthukumar, J. Chem. Phys. 118, 5174 (2006).
[5] B. Hu, W. Margolin, I. J. Molineux, and L. J., Proc. Natl. Acad. Sci. USA 112, E4919 (2015).
[6] J. E. Galán and G. Waksman, Cell 172, 1306 (2018).
[7] J. A. Faras-Rico, F. Ruud Selin, I. Myronidi, M. Frühhauf, and G. von Heijne, Proc. Natl. Acad Sci. USA 115, E9280 (2018).
[8] J. Marino, G. von Heijne, and R. Beckmann, FEBS Lett. 590, 655 (2016).
[9] L. D. Evans, C. Hughes, and G. Fraser, Trends Microbiol. 22, 566 (2014).
[10] R. M. Macnab, Annu. Rev. Microbiol. 57, 77 (2003).
[11] L. Turner, A. S. Stern, and H. C. Berg, J. Bacteriol. 194, 2437-2442 (2012).
[12] A. S. Stern and H. C. Berg, Biophys. J. 105, 182 (2013).
[13] A. Lappala, A. Zacccone, and E. M. Terentjev, Sci. Reports 3, 3103 (2013).
[14] L. D. B. Evans, S. Poulter, E. M. Terentjev, C. Hughes, and G. M. Fraser, Nature 504, 287 (2013).
[15] E. F. Casassa, J. Polym. Sci. B 5, 773 (1967).
[16] S. F. Edwards and K. F. Freed, J. Phys. A. 2, 145 (1969).
[17] M. Doi and S. F. Edwards, The Theory of Polymer Dynamics (Oxford University Press, Oxford, 1988).
[18] P. G. de Gennes, Scaling Concepts in Polymer Physics (Cornell University Press, Ithaka, NY, 1979).
[19] T. Sakaue and E. Rafael, Macromolecules 39, 2621 (2006).
[20] S. Miyazawa and R. L. Jernigan, Macromolecules 18, 534 (1985).
[21] S. Bell and E. M. Terentjev, Sci. Reports 7, 17272 (2017).
[22] S. Glaesstone, K. J. Laidler, and H. Eyring, The Theory of Rate Processes (McGraw Hill, NY, 1941).
[23] G. H. Weiss, in Advances in Chemical Physics, Vol. 13, edited by I. Prigogine (John Wiley & Sons, The address of the publisher, 1967) pp. 1–18.
[24] P. Talkner, Z. Physik B 68, 201 (1987).
[25] L. G. Longsworth, J. Am. Chem. Soc. 75, 5705 (1953).
[26] Y. Ma, C. Zhu, P. Ma, and K. T. Yu, J. Chem. Eng. Data 50, 1192 (2005).
[27] some good reference, label:rates about K-on 18, 534 (1985).
[28] K. J. Laidler, Reaction Kinetics, Vol.2: Reactions in Solutions (Pergamon Press, Oxford, 1963).
[29] K. Kremer and G. S. Grest, J. Chem. Phys. 92, 5057 (1990).
[30] A. Rosa and R. Everaers, PLoS Comput. Biol. 4, e1000153 (2008).
[31] Y. Jung, J. Kim, S. Jun, and B.-Y. Ha, Macromolecules 45, 3256 (2012).
[32] S. Jun, A. Arnold, and B.-Y. Ha, Phys. Rev. Lett. 98, 128303 (2007).
[33] A. Lappala, L. Sefton, P. W. Fennimore, and E. M. Terentjev, Sci. Reports 9, 3830 (2019).
[34] R. N. Mantegna and H. E. Stanley, Phys. Rev. Lett. 73, 2946 (1994).
[35] R. Metzler and J. Klafter, Phys. Rep. 339, 1 (2000).
[36] S. Miyazawa and R. L. Jernigan, J. Mol. Biol. 256, 623 (1996).