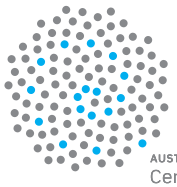
Exact Solution of a Simple Adsorption Model of De-naturating DNA

[†] Aleks Owczarek, [‡] Andrew Rechnitzer, and [†] Rami Tabbara

[†] MASCOS and Department of Mathematics and Statistics, The University of Melbourne

[‡] Department of Mathematics and Statistics, University of British Columbia

June, 2014



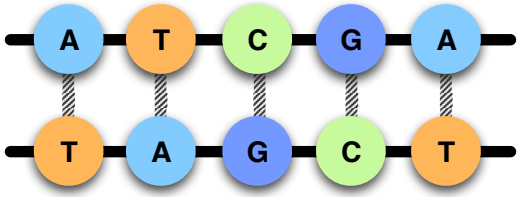
AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics and Statistics of Complex Systems

2014 SIAM Conference on Discrete Mathematics

THE UNIVERSITY OF MELBOURNE

DNA

- DNA is a polymer consisting of four repeating nucleic bases A,C,G,T.
- Two strands entwined with a helix structure
- Denaturation: At high T, strands pulled apart



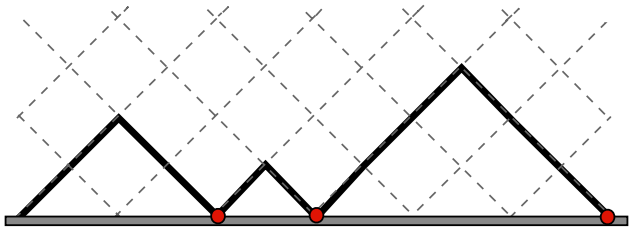
The other physical motivation is the adsorption phase transition where polymers in solution can stick to the surface of a container

- Second order phase transition
- Order parameter is coverage of the surface by the polymer

Let use models in two dimensions and lattice models at that — Proven to be insightful and integrable

ADSORPTION: VERY SIMPLE ONE DIRECTED WALK MODEL

- Single Dyck path in a half space
- Energy $-\varepsilon_a$ for each time (number m_a) it visits the surface
- Boltzmann weight $a = e^{\varepsilon_a / k_B T}$



Consider the coverage, our order parameter (indicator for a phase):

$$\mathcal{A}(a) = \lim_{n \to \infty} \frac{\langle m_a \rangle}{n}$$

There exists two phases, desorbed and adsorbed, and a phase transition at a temperature T_a given by $a = 2$ between these:

- For $T > T_a$ (small a) the walk moves away entropically and $\mathcal{A} = 0$
- For $T < T_a$ (large a) the walk is adsorbed onto the surface and $\mathcal{A} > 0$

Adsorption and Unzipping

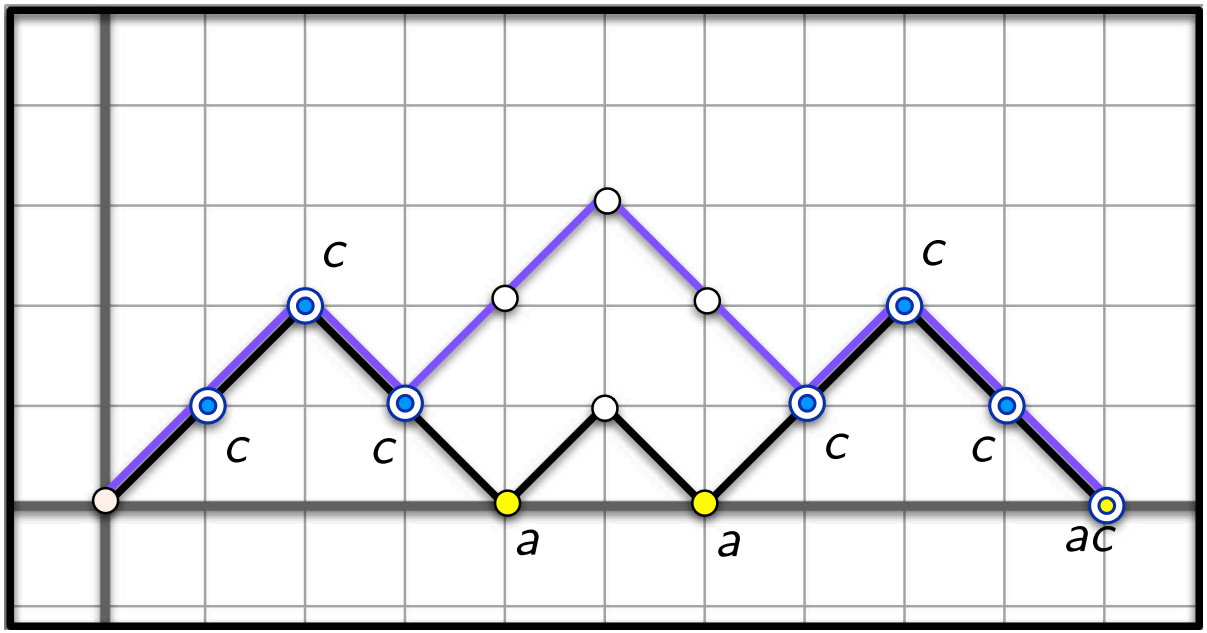
(Our system — Adsorption and Unzipping)

- *double DNA strand in a solvent*
- *near attractive surface*
- *assume aligned base sequence*
- *so expect both adsorption and denaturation (unzipping)*

ALLOWED WALKS

Consider two directed walks along the square lattice.
Let our model contain the class of allowed configs. with n steps as described:

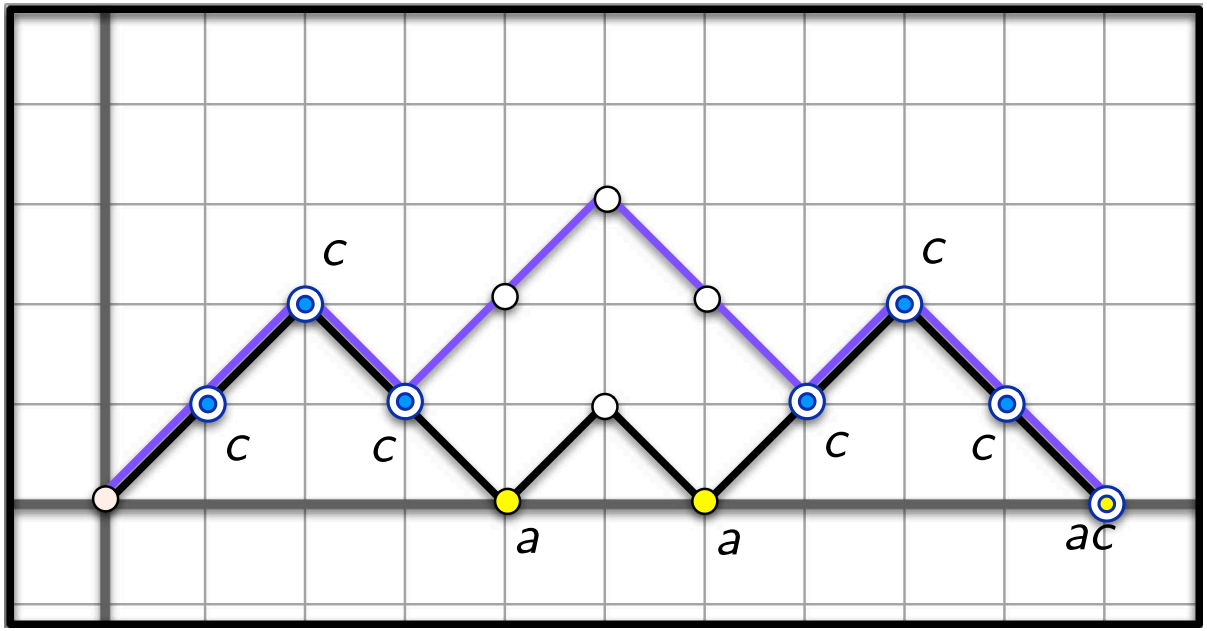
- both walks begin at $(0,0)$, end at $(2n,0)$.
- directed: can only take steps in the $(\pm 1, 0)$ directions.
- (∞) - friendly: walks can share sites, but cannot cross



UNZIPPING ADSORPTION MODEL

Let T be the system temperature, k_B the Boltzmann constant.

- surface visit sites: $a \equiv e^{\varepsilon_a / k_B T}$
- shared site contacts: $c \equiv e^{\varepsilon_c / k_B T}$
- trivial walk consisting of zero steps has weight 1.



An allowed configuration of length 10. The overall weight is $a^3 c^7$

GENERATING FUNCTION

- Partition function: $Z_n(a,c) = \sum_{\widehat{\varphi} \ni |\widehat{\varphi}|=n} a^{m_a(\widehat{\varphi})} c^{m_c(\widehat{\varphi})}$
- Generating function: $G(a,c) \equiv G(a,c;z) = \sum_{n\geq 1} Z_n(a,c) z^n$
- Reduced free energy:

$$\kappa(a,c) = \lim_{n\to\infty} n^{-1} \log Z_n(a,c) = \log z_s(a,c)$$

where $z_s(a,c)$ is dominant singularity of G w.r.t. z

Two order parameters:

$$\mathcal{A}(a,c) = \lim_{n\to\infty} \frac{\langle m_a \rangle}{n} \qquad \text{and} \qquad \mathcal{C}(a,c) = \lim_{n\to\infty} \frac{\langle m_c \rangle}{n},$$

Context

No wall/interaction:

- Vicious dir. walks: *Lindström-Gessel-Viennot thm.* ('85), *Essam & Guttmann* ('95), *Guttmann, Owczarek & Viennot* ('98)
- Friendly walks & Osculating walks: *Brak* ('97), *Guttmann & Vöge* ('02), *Bousquet-Mélou* ('06)

With wall (but no interaction)

- Vicious: *Krattenhaler, Guttmann & Viennot* ('00)

With wall (interaction)

- Vicious: *Brak, Essam & Owczarek* ('99, '01)
- Friendly: *Owczarek, Rechnitzer & Wong* ('12)

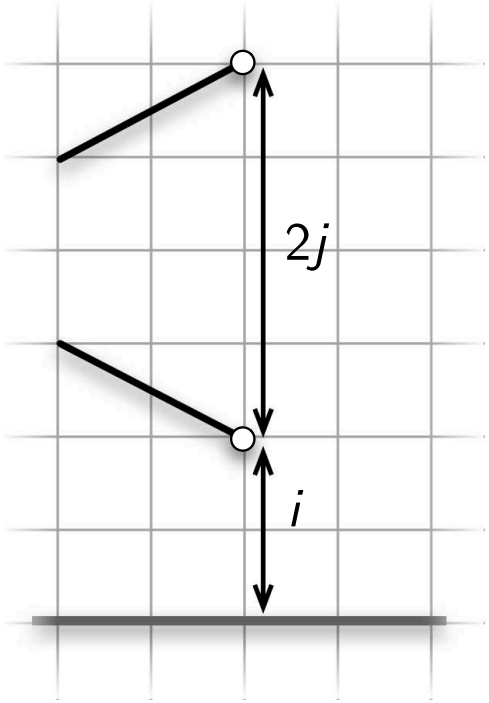
Generalised generating function

We consider walks φ in the larger set, where each walk can end at any possible height.

- To find $G(a,c)$, consider larger class of configs.
- Generalised generating function:

$$F(r,s) \equiv F(r,s,a,c;z) = \sum_{\varphi\in\Omega} a^{m_a(\varphi)} c^{m_c(\varphi)} r^i s^j z^n$$

- $G(a,c) = F(0,0)$



Establishing a functional equation

- By considering the addition of a single column onto a configuration, and the types of walks obtained, we can find a decomposition of all configurations
- Translating back to generating functions we end up with

$$K(r,s)F(r,s) = \frac{1}{ac} + \left(C(c) - \frac{zr}{s}\right) F(r,0) + \left[A(a) - \frac{z}{r}(s+1)\right] F(0,s) - A(a)C(c)F(0,0)$$

where

$$A(a) = \frac{a-1}{a}, \quad C(c) = \frac{c-1}{c}$$

and the kernel $K(r,s)$ is

$$K(r,s) \equiv K(r,s;z) = \left(1 - z\left[r + \frac{s}{r} + \frac{r}{s} + \frac{1}{r}\right]\right).$$

OBSTINATE KERNEL METHOD

- Equation is written as ”bulk = boundary terms” where bulk term is product of kernel and bulk generating function
- Answer needed is one of the boundary generating functions so try to remove bulk by setting the value of a catalytic variable to a value that makes the kernel vanish
- Standard kernel method due to *Knuth* (1968): use values of ”catalytic variable’ to ”kill” kernel
- From $\approx$ early ’00’s applied to a number of dir. walk problems
- Sometimes need multiple values of catalytic variable(s): obstinate kernel method
- More than one catalytic variable requires this
- Earliest combinatorial application of the obstinate kernel method due to *Bousquet-Mélou* (‘02).
- See Bousquet-Mélou *Math. and Comp. Sci 2* (2002)), Bousquet-Mélou, Mishna *Contemp. Math.* **520** (2010)

Symmetries of the Kernel

The kernel is symmetric under the following two transformations, which are involutions:

$$(r,s) \mapsto \left(r, \frac{r^2}{s}\right), \qquad (r,s) \mapsto \left(\frac{s}{r}, s\right)$$

Transformations generate a family of 8 symmetries ('group of the walk')

$$(r,s), \left(r, \frac{r^2}{s}\right), \left(\frac{s}{r}, \frac{s}{r^2}\right), \left(\frac{r}{s}, \frac{1}{s}\right), \left(\frac{1}{r}, \frac{1}{s}\right), \left(\frac{1}{r}, \frac{s}{r^2}\right), \left(\frac{r}{s}, \frac{r^2}{s}\right), \text{ and } \left(\frac{s}{r}, s\right)$$

- *We make use of four of these which only involve positive powers of r.*
- *This gives us four equations.*
- *One can eliminate many of the unknown generating functions by a clever choice of adding these equations*

ROOTS OF THE KERNEL

- The kernel has two roots as function of either r or s
- choose the one which gives a positive term power series expansion in z
- with Laurent polynomial coefficients in s (r):

$$\hat{r}(s;z) \equiv \hat{r} = \frac{s\left(1-\sqrt{1-4\frac{(1+s)^2z^2}{s}}\right)}{2(1+s)z} = \sum_{n\geq 0} C_n \frac{(1+s)^{2n+1}z^{2n+1}}{s^n},$$

where $C_n = \frac{1}{n+1}\binom{2n}{n}$ is a Catalan number.

- *Make the substitution $r \mapsto \hat{r}$*

FINDING THE SOLUTION

Key idea

- Treat K as fn. of r or s to get roots $\hat{r}$ and $\hat{s}$
- Then use subset of $\mathcal{F}$ to get system of eqns. E.g. Using $\hat{r}$:

$(\hat{r}, s)$	$F(\hat{r}, 0)$	$F(0, s)$	$F(0, 0)$
$(\hat{r}, \hat{r}^2/s)$	$F(\hat{r}, 0)$	$F(0, \hat{r}^2/s)$	$F(0, 0)$
$(\hat{r}/s, \hat{r}^2/s)$	$F(\hat{r}/s, 0)$	$F(0, \hat{r}^2/s)$	$F(0, 0)$
$(\hat{r}/s, 1/s)$	$F(\hat{r}/s, 0)$	$F(0, 1/s)$	$F(0, 0)$

- Combine these eqns. to get new fn. eqn

$$N_1^\star(s;z)F(0,1/s)+N_2^\star(s;z)F(0,s) = \left[M^\star(s) - c^2 H^\star(s;z)\right]\left(\frac{1}{ac} - ACF(0,0)\right),$$

- Can do the same using $\hat{s}$!
- Nice things happen when $a = 1$ or $c = 1$ to $N_1^\star(s;z)$ etc

SOLUTION FOR $G(a,1)$

Exact solution for $G(a,1)$ is known and can be found using the method described

- Brak, Essam & Owczarek (1998, 2001): Partition fn. using Lindström-Gessel-Viennot Thm.
- Owczarek, Rechnitzer & Wong (2012): Gen. fn calculated by employing same kernel method techniques.

Specifically:

$$G(a,1) = \sum_{n\geq 0} z^{2n} \sum_{k=0}^{n} a^k \frac{k(k+1)(k+2)}{(2n-k)(n+1)^2(n+2)} \binom{2n-k}{n} \binom{2n}{n}.$$

SOLUTION FOR $G(1,c)$

- No known previous solution for $G(1,c)$

We can write functional equation as

$$G(1,c) = F(0,0,1,c;z) = [r^1]\frac{\hat{s}\left(r^2-1\right)\left[r - cr + cz\left(1+r^2-\hat{s}\right)\right]}{(c-1)\left(\hat{s} - c\hat{s} + crz\right)},$$

expanding RHS as power series in c and so obtain, after some work:

$$G(1,c;z) = 1 + c^2z^2 + c^3\left(1+2z\right)z^4$$
$$+ \sum_{i=3}^{\infty} z^{2i} \sum_{m=3}^{2i} c^m \sum_{k=3}^{m} (-1)^{k+1} \frac{k(k-1)(k-2)(2i-k+1)(i-k+2)}{i^2(i-1)^2(i+1)(i-2)} \binom{m}{k}\binom{2i-k}{i-2}\binom{2i-k-1}{i-3}.$$

SOLUTION FOR $G(1, c)$

- While we have an explicit solution for $G(1, c)$ it is advantageous for analysis to directly read off the singularities
- Alternative — find differential equation satisfied by generating function
- Use Zeilberger-Gosper algorithm: Maple: DETools package, Zeilberger hyperexp. implementation
- Result: DE for $G(1, c)$ is order 6 with poly. coeff of $\deg_z = 12$

FORTUNATE DECOMPOSITION OF $G(a,c)$

Using various combinatorial relationships between the generating functions we can re-write $G(a,c)$ in terms of $G(a,1)$ and $G(1,c)$:

$$G(a,c) = \frac{1}{(a-1)(c-1)} + \frac{p_1(a,c,z)}{p_2(a,c,z) + p_3(a,c,z)G(a,1) + p_4(a,c,z)G(1,c)}$$

where p_i are polynomials in a, c and z: quadratics in z^2.

Key point: With solutions to $G(a,1)$ and $G(1,c)$ we additionally have solved for $G(a,c)$.

SINGULARITIES OF $G(a,1)$ & $G(1,c)$

- Recall, free energy $\kappa(a,c) = \log z_s(a,c)$
- For $G(a,1)$, prev. known:

$$z_s(a,1) = \begin{cases} z_b \equiv 1/4, & a \leq 2 \\ z_a \equiv \frac{\sqrt{a-1}}{2a}, & a > 2 \end{cases}$$

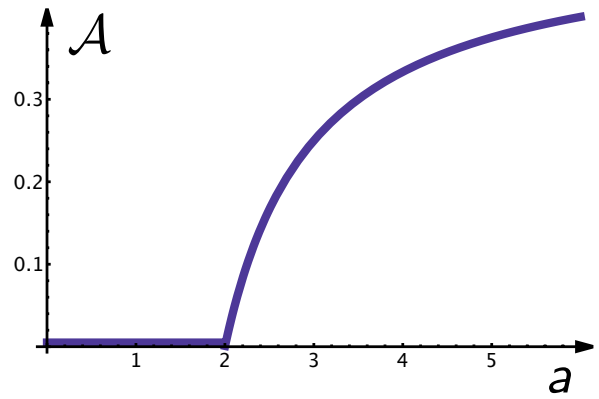
- For $G(1,c)$, we use the DE (roots of leading poly. coeff.):

$$z_s(1,c) = \begin{cases} z_b \equiv 1/4, & c \leq 4/3 \\ z_c \equiv \frac{1-c+\sqrt{c^2-c}}{c}, & c > 4/3 \end{cases}$$

TRANSITIONS OF $G(a,1)$ & $G(1,c)$

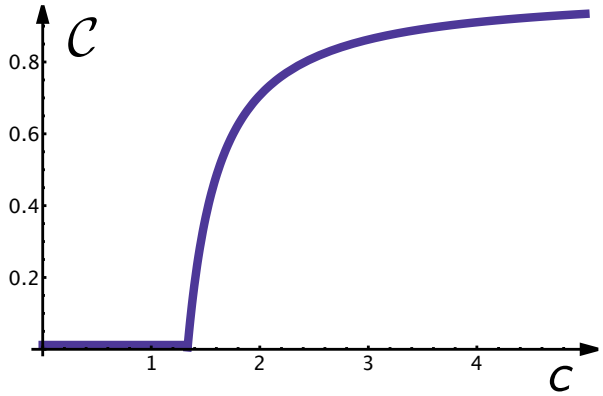
- For $G(a,1)$: the order parameter associated with the phase transition is the surface coverage

$$\mathcal{A}(a,1) = \begin{cases} 0, & a \leq 2 \\ \frac{a-2}{2(a-1)}, & a > 2 \end{cases}$$



- For $G(1,c)$: the order parameter associated with the phase transition is the shared site density

$$\mathcal{C}(1,c) = \begin{cases} 0, & c \leq 4/3 \\ \frac{c-2+\sqrt{c(c-1)}}{2(c-1)}, & c > 4/3 \end{cases}$$



- Second-order adsorption and zipping phase trans. resp.

SINGULARITIES AND PHASES

This leads us to associate the singularities of $G(a,1)$ and $G(1,c)$ with the phases as

- $z_b = 1/4$ with a desorbed phase where $\mathcal{A} = 0$ and $\mathcal{C} = 0$
- $z_a = \frac{\sqrt{a-1}}{2a}$ with an adsorbed phase where $\mathcal{A} > 0$
- $z_c = \frac{1-c+\sqrt{c^2-c}}{c}$ with a zipped phase where $\mathcal{C} > 0$

ORDER PARAMETERS FOR THE FULL MODEL

Four possible phases:

- Desorbed: $\mathcal{A} = \mathcal{C} = 0$
- Adsorbed: (a-rich) $\mathcal{A} > 0, \mathcal{C} = 0$
- Zipped: (c-rich) $\mathcal{A} = 0, \mathcal{C} > 0$
- Zipped & Adsorbed: (ac-rich) $\mathcal{A} > 0, \mathcal{C} > 0$

ANALYSING $G(a,c)$

Recall

$$G(a,c) \sim \frac{p_1(a,c,z)}{p_2(a,c,z) + p_3(a,c,z)G(a,1) + p_4(a,c,z)G(1,c)}$$

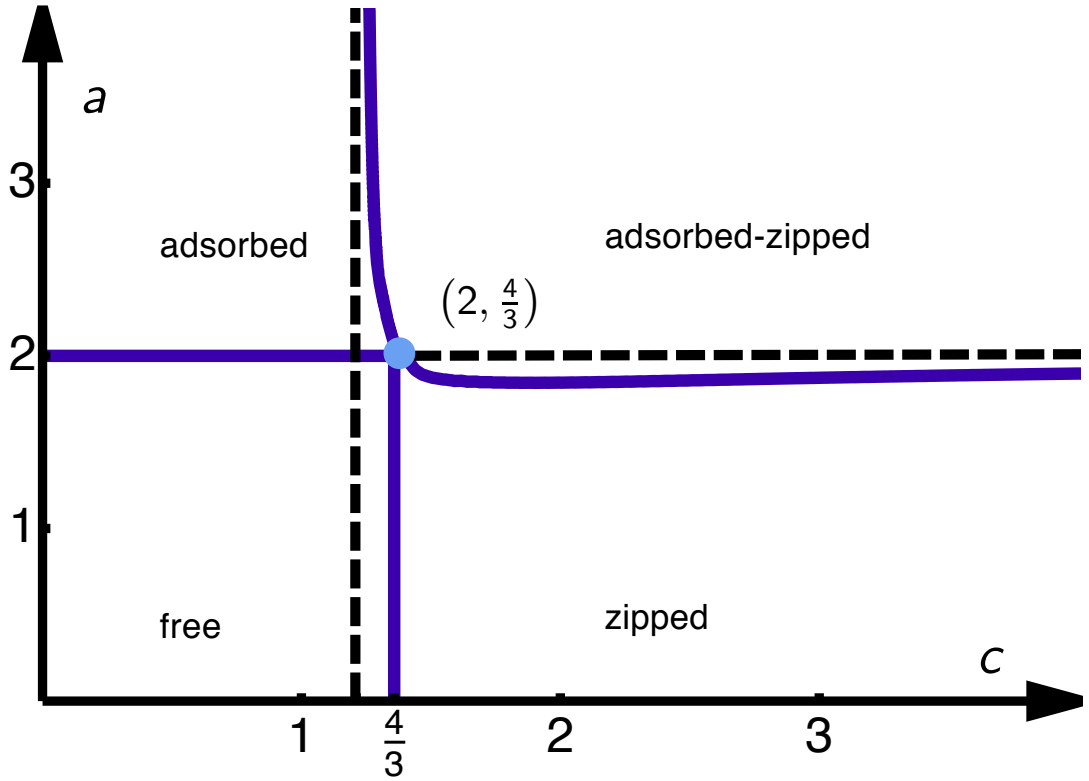
- $\Rightarrow$ Singularities: Look at $G(a,1)$, $G(1,c)$ and root of above denom.
- root of denominator is associated with the zipped-adsorbed phase

The dominant singularity $z_s(a,c)$ of the generating function $G(a,c;z)$ is one of four types associated with the four phases

$$z_s(a,c) = \begin{cases} z_b \equiv 1/4, & a \le 2, c \le 4/3 \\ z_a(a) \equiv \frac{\sqrt{a-1}}{2a}, & a > 2, c \le \alpha(a) \\ z_c(c) \equiv \frac{1-c+\sqrt{c^2-c}}{c}, & a \le \gamma(c), c > 4/3 \\ z_{ac}(a,c), & a > \gamma(c), c > \alpha(a) \end{cases}$$

- $\alpha(a)$ is boundary between adsorbed and zipped-adsorbed phases
- $\gamma(c)$ is the boundary between zipped and zipped-adsorbed phases

PHASE DIAGRAM



All transitions found to be second order

Low-temp argument gives

- $c \to \infty, \gamma(c) \to 2$
- $a \to \infty, \alpha(a) \to \sqrt{5} - 1$

ASYMPTOTICS

Table : The growth rates of the coefficients $Z_n(a,c)$ modulo the amplitudes of the full generating function $G(a,c;z)$ over the entire phase space.

phase region	$Z_n(a,c) \sim$
free	$4^n n^{-5}$
free to adsorbed boundary	$4^n n^{-3}$
free to zipped boundary	$4^n n^{-3}$
$a = 2, c = 4/3$	$4^n n^{-3}$
adsorbed	$z_a(a)^{-n} n^{-3/2}$
zipped	$z_c(c)^{-n} n^{-3/2}$
adsorbed to adsorbed-zipped boundary $(\alpha(a))$	$z_a(c)^{-n} n^{-1/2}$
zipped to adsorbed-zipped boundary $(\gamma(c))$	$z_c(c)^{-n} n^{-1/2}$
adsorbed-zipped	$z_{ac}(a,c)^{-n} n^{-1}$

Conclusion

- Simple model of DNA as two friendly walks near a boundary
- Used combinatorial decomposition to obtain linear functional equation
- Used obstinate kernel method to solve functional equations (using symmetries to provide sufficient information)
- Explicit series solutions for $G(a, 1)$ and $G(1, c)$
- Combined these equations to relate $G(a, c)$ to both $G(a, 1)$ and $G(1, c)$
- Also used Zeilberger-Gosper algorithm to find linear DE for $G(1, c)$
- Full analysis of asymptotics and phase diagram
- R. Tabbara, A. L. Owczarek and A. Rechnitzer, *J. Phys. A.: Math. Theor*, **47**, 015202 (34pp), 2014

FUTURE WORKS

- Combine single, double surface and unzipping interactions: work in progress
- Consider three walks with multiple unzipping interactions: work in progress
- Also in progress is work in a slit: see Thomas Wong's talk
- Extend these to various interactions