

Crosslinking and gelation between linear polymers: DNA-antibody complexes in systemic lupus erythematosus

Frederik W Wiegel,[†] Bernard J Geurts[†] and Byron Goldstein^{‡§}

[†] Centre for Theoretical Physics, Twente University, PO Box 217, Enschede 7500 AE, The Netherlands

[‡] Theoretical Division, Group T-10, Los Alamos National Laboratory, University of California, Los Alamos, NM 87545, USA

Received 12 March 1987, in final form 11 May 1987

Abstract. In the autoimmune disease systemic lupus erythematosus the DNA molecules of an individual are attacked by its own antibodies. As these antibodies are bivalent they can crosslink different DNA molecules which can lead to the formation of DNA-antibody complexes and gels. Statistical properties of these complexes are derived and evaluated analytically in the limit of very long DNA molecules, as well as the concentrations at which a gel is being formed. We also present various numerical results for DNA molecules of intermediate lengths. This work can also be considered as a theory of the crosslinking and gelation of linear polymers.

1. Introduction

Systemic lupus erythematosus appears to be an autoimmune disease in which patients form antibodies against some of their own DNA molecules (Stollar 1981). The basic antibody-DNA geometry is shown in figure 1. A full discussion can be found in Wiegel and Goldstein (1987) in which we started to apply the methods of equilibrium statistical

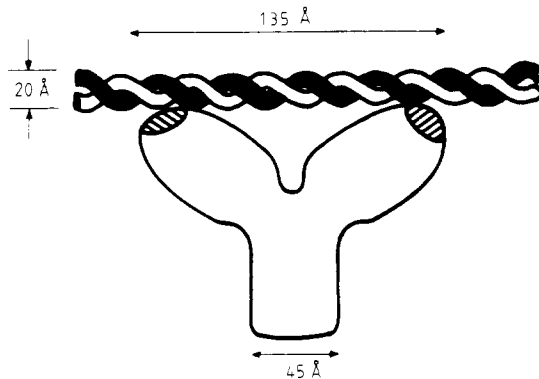


Figure 1. Schematic diagram indicating the dimensions of an IgG molecule bound bivalently to a DNA molecule; from Wiegel and Goldstein (1987).

§ Work supported by the National Institutes of Health Grant no GM35556 and by the United States Department of Energy.

mechanics to the binding of a large number of antibodies to a single very long DNA molecule. In the present paper we consider the full problem in which a large number of antibodies can bind to a large number of DNA molecules. As each antibody has two binding sites it can crosslink different DNA molecules which can lead to the formation of DNA-antibody complexes and gels. Our aim is to calculate the statistical properties of these complexes as well as the DNA-antibody concentrations at which gelation occurs. We shall follow the generating function method in the form in which it is used, for example, by Wiegel and Perelson (1982) to determine the statistical properties of rouleaux consisting of red blood cells.

The model we consider consists of N identical DNA molecules and M identical bivalent antibodies enclosed in a volume V and in thermodynamic equilibrium. Each DNA molecule possesses K regularly spaced possible binding sites for an antibody (cf Stollar (1981) for a brief discussion of the main features of these binding sites on the DNA backbone). An antibody can be in one of four states: (i) it can be free; (ii) it can be bound to DNA with one of its two binding sites, having a weight factor f_1 ; (iii) it can be bound to the same DNA molecule with both binding sites, having a weight f_2 and (iv) it can crosslink two different DNA molecules, having a weight f_3 . In § 2 we determine some generating functions related to a single complex which consists of m antibodies and n DNA molecules. Then, in § 3 expressions for the statistical properties of these complexes are derived and related to these generating functions. Section 4 is devoted to the statistical properties and the gelation if the DNA molecules are very long. Section 5 is devoted to a discussion of the numerical results obtained for DNA molecules of intermediate lengths.

2. A single complex of m antibodies and n DNA molecules

Let $Q_K(m, n)$ denote the configuration sum of a single 'rooted' DNA-antibody complex consisting of m antibodies and n DNA molecules in the case in which the antibodies as well as the DNA molecules are counted as indistinguishable geometrical objects. All DNA molecules are of the same length and consist of K regularly spaced binding sites for an antibody. The complexes are assumed to have the topology of a tree, and a 'root' is assigned to one end of one of the DNA molecules. Using counting variables z for the total number of binding sites per DNA, ξ for the number of antibodies in the complex and η for the number of crosslinks, the generating function is defined as

$$Q(z, \xi, \eta) = \sum_{K=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} Q_K(m, n) z^K \xi^m \eta^{n-1}. \quad (2.1)$$

In the tree approximation, a complex consisting of n DNA molecules has $n - 1$ crosslinks. The use of generating functions to study macromolecular complexes has been reviewed, for example, by one of us (Wiegel 1983).

In figure 2 a diagram representing a DNA molecule with 0, 1, 2, . . . , crosslinks to other DNA molecules is drawn. It shows that if we follow one particular (arbitrarily chosen but otherwise fixed) DNA we either do not encounter any crosslink to another DNA at all, or we do. If we do, the crosslink encountered is given a weight factor $z\xi\eta f_3 A$ where the function $A(\xi, \eta)$ denotes the generating function with respect to m and n of the configuration sum over all possible forms of the 'appendix'. If the generating function of a single DNA with antibody attached to it, but without crosslinks,

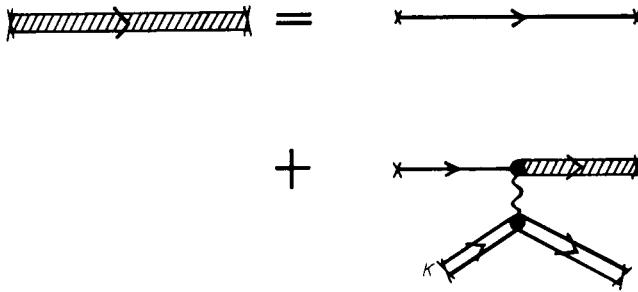


Figure 2. Diagram showing that for any configuration a fixed DNA molecule (hatched double line) either does not contain a crosslink to another DNA (single line) or does contain a crosslink. The first crosslink is indicated by a wavy line and its ‘appendix’ by two double lines. Note that a hatched double line corresponds to a factor $Q(z, \xi, \eta)$, a single line is a factor $Q_0(z, \xi)$, a crosslink is a factor $z\xi\eta f_3$ and a double line with label K is a factor $q_K(\xi, \eta)$.

is denoted by $Q_0(z, \xi)$, then the total generating function can be written as

$$Q(z, \xi, \eta) = Q_0(z, \xi) + Q_0(z, \xi)z\xi\eta f_3 A(\xi, \eta)Q_0(z, \xi) + \dots \tag{2.2a}$$

$$= Q_0(z, \xi)(1 - z\xi\eta f_3 A(\xi, \eta)Q_0(z, \xi))^{-1} \tag{2.2b}$$

$$= Q_0(z, \xi) + Q_0(z, \xi)z\xi\eta f_3 A(\xi, \eta)Q(z, \xi, \eta). \tag{2.2c}$$

The generating function Q_0 was calculated by Wiegel and Goldstein (1987) with $Q_K(0, 1) = 1$, and found to be

$$Q_0(z, \xi) = [1 - z - z\xi f_1 - (1 + \xi f_1)^2 z^4 \xi f_2 (1 - z^2 \xi f_2)^{-1} - (1 + \xi f_1) z^5 \xi^2 f_2^2]^{-1}. \tag{2.3}$$

In deriving (2.3) it was assumed that a doubly bound antibody spans exactly two possible binding sites on the DNA molecule. In view of the definition of $A(\xi, \eta)$, one may write

$$A(\xi, \eta) = \sum_{l=0}^{K-1} q_{K-1-l}(\xi, \eta)q_l(\xi, \eta). \tag{2.4}$$

The functions q_{K-1-l} and q_l which appear here are given by

$$q_L(\xi, \eta) = \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} Q_L(m, n)\xi^m \eta^{n-1} \quad L = K - 1 - l \text{ or } L = l. \tag{2.5}$$

Note that in the last two equations the length of the DNA is fixed at K binding sites (one of which is involved in a crosslink) whereas in (2.1) and (2.2a, b, c) the generating function appears with respect to the length of the DNA. In this sense our generating function formalism treats the DNA in the trunk of the tree differently from the DNA in the branches of the tree. This is indicated in figure 2 as explained in the caption. In (2.4) the sum of the indices on q is equal to $K - 1$, representing the fact that all DNA molecules have exactly K possible binding sites each, one of which is blocked by the crosslink. Of course, the function $Q(z, \xi, \eta)$ determines both $q_K(\xi, \eta)$ and $Q_K(m, n)$, since

$$q_K(\xi, \eta) = \frac{1}{2\pi i} \int_{\Gamma_1} \frac{Q(z, \xi, \eta)}{z^{K+1}} dz \tag{2.6}$$

$$Q_K(m, m) = \frac{1}{(2\pi i)^2} \int_{\Gamma_2} d\xi \int_{\Gamma_3} d\eta \frac{q_K(\xi, \eta)}{\xi^{m+1} \eta^m} \tag{2.7}$$

where $\Gamma_1, \Gamma_2, \Gamma_3$ are closed contours around the origin, for example, circles with radii small enough for the series (2.1) to converge.

Combination of (2.2b) with (2.4) and (2.6) in principle determines the generating function $Q(z, \xi, \eta)$ and hence the configuration sums $Q_K(m, n)$. The rest of this section is devoted to the study of the function $q_K(\xi, \eta)$ which also determines the statistical properties of the DNA-antibody complexes. The contour integration in (2.6) can be performed using Cauchy's theorem. One straightforwardly obtains

$$q_K(\xi, \eta) = \alpha(\xi) \sum_{i=1}^7 c_i z_i^{-K-1} \tag{2.8}$$

where

$$\alpha(\xi) = -[\xi^3 f_2^3 (1 + \xi f_1)]^{-1} \tag{2.9}$$

and

$$c_1 = \frac{1 - z_1^2 \xi f_2}{(z_1 - z_2) \dots (z_1 - z_7)}$$

$$\vdots$$

$$c_7 = \frac{1 - z_7^2 \xi f_2}{(z_7 - z_1) \dots (z_7 - z_6)} \tag{2.10}$$

and where z_1, \dots, z_7 are the roots of

$$1 - z(1 + \xi f_1 + \xi \eta f_3 A(\xi, \eta)) - z^2(\xi f_2) + z^3[\xi f_2(1 + \xi f_1) + \xi^2 \eta f_2 f_3 A(\xi, \eta)] - z^4[\xi f_2(1 + \xi f_1)^2] - z^5[\xi^2 f_2^2(1 + \xi f_1)] + z^7[\xi^3 f_2^3(1 + \xi f_1)] = 0. \tag{2.11}$$

The left-hand side still contains the unknown function A . However, combination of (2.8) and (2.4) gives the following expression for A in terms of the roots z_1, \dots, z_7 :

$$A(\xi, \eta) = \alpha^2(\xi) \left(K \sum_{i=1}^7 c_i^2 z_i^{K-1} + 2 \sum_{i=1}^6 \sum_{l=i+1}^7 c_i c_l \frac{z_i^{-K} - z_l^{-K}}{z_l - z_i} \right). \tag{2.12}$$

Insertion of (2.12) into (2.11) then yields an equation in which only the roots z_1, \dots, z_7 appear as unknowns. The determination of these roots is a complicated matter, which can be treated numerically, leading to a self-consistent calculation of z_1, \dots, z_7 as functions of ξ and η .

In § 4 we consider the problem in the case of very long DNA molecules. The resulting equation simplifies greatly if $K \gg 1$ and can be treated analytically in an approximate way. In § 5 results based on a numerical evaluation of the above scheme will be shown graphically. First, we turn to the determination of the configuration sum of the system and derive expressions for the statistical properties of the complexes.

3. Statistical properties of DNA-antibody complexes

In order to calculate the statistical properties of DNA-antibody complexes we need the configuration sum of the system. This calculation involves several steps.

Let ν_f be the number of free antibodies, ν_c the number of 'clean' DNA molecules, i.e. DNA molecules with no antibody attached to them, and $\nu(m, n)$ the number of DNA-antibody complexes consisting of m antibodies and n DNA molecules. A macrostate of the system can then be represented by the set of numbers $\{\nu_f, \nu_c, \nu(m, n)\}$, provided these numbers satisfy

$$\nu_f + \sum_{m=1}^M \sum_{n=1}^N m\nu(m, n) = M \tag{3.1}$$

$$\nu_c + \sum_{m=1}^M \sum_{n=1}^N n\nu(m, n) = N. \tag{3.2}$$

The number of microstates compatible with a given macrostate $\{\nu_f, \nu_c, \nu(m, n)\}$ will be denoted by $\Omega\{\nu_f, \nu_c, \nu(m, n)\}$. It equals the number of different ways to divide M antibodies and N DNA molecules in groups such that there are ν_f free antibodies, ν_c clean DNA molecules and $\nu(m, n)$ complexes of m antibodies and n DNA molecules. Following a reasoning similar to that found in Wiegand and Perelson (1982) one obtains

$$\Omega\{\nu_f, \nu_c, \nu(m, n)\} = \frac{M! N!}{\nu_f! \nu_c! \prod_{m=1}^M \prod_{n=1}^N (m! n!)^{\nu(m, n)} (\nu(m, n))!}. \tag{3.3}$$

Each complex, free antibody or clean DNA molecule can be moved around through the volume V ; this leads to a combinatorial factor V/v_0 , where v_0 is some fixed elementary volume, roughly equal to the average excluded volume of a single complex.

Finally, we notice that $Q_K(m, n)$ was defined to be the configuration sum of a rooted DNA-antibody complex fixed in space. Since the complexes are physically without roots and the antibodies and DNA molecules are indistinguishable, the physically correct configuration sum for the complexes is $[(m! n!)/2n]Q_K(m, n)$. Hence, for the total configuration sum $Q\{\nu_f, \nu_c, \nu(m, n)\}$ one obtains

$$\begin{aligned} Q\{\nu_f, \nu_c, \nu(m, n)\} &= \frac{\Omega\{\nu_f, \nu_c, \nu(m, n)\}}{M! N!} \left(\frac{V}{v_0}\right)^{\nu_f} \left(\frac{V}{2v_0} Q_K(0, 1)\right)^{\nu_c} \\ &\quad \times \prod_{m=1}^M \prod_{n=1}^N \left(\frac{V}{2nv_0} m! n! Q_K(m, n)\right)^{\nu(m, n)} \\ &= \frac{1}{\nu_f!} \left(\frac{V}{v_0}\right)^{\nu_f} \frac{1}{\nu_c!} \left(\frac{V}{2v_0} Q_K(0, 1)\right)^{\nu_c} \\ &\quad \times \prod_{m=1}^M \prod_{n=1}^N [\nu(m, n)!]^{-2} \left(\frac{V}{2nv_0} Q_K(m, n)\right)^{\nu(m, n)}. \end{aligned} \tag{3.4}$$

The equilibrium distribution $\{\nu_f^*, \nu_c^*, \nu^*(m, n)\}$ is found by determining the maximum of $Q\{\nu_f, \nu_c, \nu(m, n)\}$ subject to the two constraints (3.1) and (3.2). Introducing Lagrange multipliers μ and λ for these constraints one finds that this problem is equivalent to determining the unconstrained maximum of

$$\begin{aligned} L\{\nu_f, \nu_c, \nu(m, n)\} &= \ln(Q\{\nu_f, \nu_c, \nu(m, n)\}) - \mu \left(\nu_f + \sum_{m=1}^M \sum_{n=1}^N m\nu(m, n)\right) \\ &\quad - \lambda \left(\nu_c + \sum_{m=1}^M \sum_{n=1}^N n\nu(m, n)\right). \end{aligned} \tag{3.5}$$

Using Stirling's approximation (i.e. $\ln(x!) = x \ln(x) - x$ if $x \gg 1$) one finds for the equilibrium distribution

$$\nu_f^* = \frac{V}{v_0} e^{-\mu} \quad (3.6a)$$

$$\nu_c^* = \frac{V}{2v_0} Q_K(0, 1) e^{-\lambda} \quad (3.6b)$$

$$\nu^*(m, n) = \frac{V}{4nv_0} Q_K(m, n) e^{-\mu n - \lambda n}. \quad (3.6c)$$

The requirement that $\{\nu_f^*, \nu_c^*, \nu^*(m, n)\}$ has to satisfy both (3.1) and (3.2) leads to the following system of equations for μ and λ :

$$e^{-\mu} + \frac{1}{4} \sum_{m=1}^M \sum_{n=1}^N \frac{m}{n} Q_K(m, n) e^{-\mu m - \lambda n} = v_0 M / V \quad (3.7a)$$

$$\frac{1}{2} Q_K(0, 1) e^{-\lambda} + \frac{1}{4} \sum_{m=1}^M \sum_{n=1}^N Q_K(m, n) e^{-\mu m - \lambda n} = v_0 N / V. \quad (3.7b)$$

This system of equations can be rewritten in terms of the function $q_K(\xi, \eta)$ studied in the previous section. Indeed, since

$$\xi \frac{\partial}{\partial \xi} \int_0^\eta q_K(\xi, \eta') d\eta' = \sum_{m=1}^M \sum_{n=1}^N \frac{m}{n} Q_K(m, n) \xi^m \eta^n \quad (3.8)$$

one may write (3.7a, b) as

$$\frac{1}{4} \xi \left(4 + \frac{\partial}{\partial \xi} \int_0^\eta q_K(\xi, \eta') d\eta' \right) = v_0 M / V \equiv C_M \quad (3.9a)$$

$$\frac{1}{4} \eta (Q_K(0, 1) + q_K(\xi, \eta)) = v_0 N / V \equiv C_N. \quad (3.9b)$$

Given numerical values of C_M and C_N the last two equations determine the corresponding values of ξ and η , from which μ and λ follow through the identification

$$\xi = e^{-\mu} \quad \eta = e^{-\lambda}. \quad (3.10)$$

In (3.9a, b) C_M and C_N are the volume fractions of antibody and DNA molecules, respectively. Knowledge of the function $q_K(\xi, \eta)$ not only allows a solution of (3.9a, b), but also gives some statistical properties of the complexes. For example, the total number of complexes C at given C_M and C_N values is determined by q_K since

$$\begin{aligned} C &= \sum_{m=1}^M \sum_{n=1}^N \nu^*(m, n) = \frac{V}{4v_0} \sum_{m=1}^M \sum_{n=1}^N \frac{1}{n} Q_K(m, n) e^{-\mu m - \lambda n} \\ &= \frac{V}{4v_0} \int_0^\eta (q_K(\xi, \eta') - Q_K(0, 1)) d\eta' \end{aligned} \quad (3.11)$$

provided (ξ, η) are solutions of (3.9a, b). Hence, the average number of antibodies per complex $\langle m \rangle$ is given by

$$\langle m \rangle = \frac{M - \nu_f^*}{C} = \frac{C_M - \xi}{v_0 C / V} \quad (3.12)$$

and the average number of DNA molecules per complex $\langle n \rangle$ by

$$\langle n \rangle = \frac{N - \nu_c^*}{C} = \frac{2C_N - Q_K(0, 1)\eta}{2v_0 C / V}. \quad (3.13)$$

In the next section the calculational scheme of § 2 will be pursued further for the case in which the DNA molecules are very long. We shall demonstrate that the system shows gelation and we shall estimate the range of C_N and C_M values for which gelation occurs.

4. Asymptotic statistical properties of DNA-antibody complexes

Expression (2.8) for the function $q_K(\xi, \eta)$ shows that, for $K \gg 1$, q_K can very well be approximated by

$$q_K(\xi, \eta) = \alpha(\xi)c_1z_1^{-K-1} \tag{4.1}$$

where z_1 is the root with the smallest modulus (it can easily be seen that z_1 has to be real for real ξ and η). Hence, the generating function $A(\xi, \eta)$ can be approximated by

$$A(\xi, \eta) = K\alpha^2(\xi)c_1^2z_1^{-K-1}. \tag{4.2}$$

Inserting (4.2) into (2.11) and rearranging the terms leads to the following equation for z_1 :

$$F(z, \xi) = G(z, \xi, \eta) \tag{4.3}$$

where

$$F(z, \xi) = 1 - z(1 + \xi f_1) - (1 + \xi f_1)^2 \frac{z^4 \xi f_2}{1 - z^2 \xi f_2} - (1 + \xi f_1)z^5 \xi^2 f_2 \tag{4.4a}$$

$$G(z, \xi, \eta) = K\alpha^2(\xi)c_1^2 \xi \eta f_3 z^{-K}. \tag{4.4b}$$

Notice that both F and G depend on the model parameters f_1, f_2 and f_3 . In view of (3.10) we restrict ourselves to positive real ξ, η values. In figure 3 both sides of (4.3)

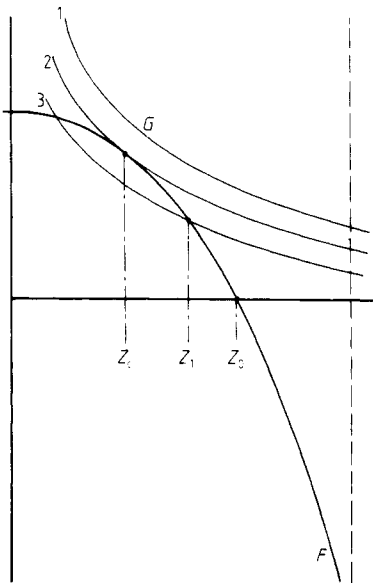


Figure 3. Both sides of equation (4.3) as functions of z , for $\eta > \eta_0$ (curve 1), $\eta = \eta_0$ (curve 2) and $\eta < \eta_0$ (curve 3).

are shown as functions of z , at some fixed ξ value and for a number of η values. It is obvious from the figure that there is a solution if $0 \leq \eta \leq \eta_0(\xi)$, and that there is no solution if $\eta > \eta_0(\xi)$. Here $\eta_0(\xi)$ is the value of η at which both curves osculate, say at $z = z_c(\xi)$, i.e.

$$F(z_c(\xi), \xi) = G(z_c(\xi), \xi, \eta_0(\xi)) \tag{4.5a}$$

$$F_z(z_c(\xi), \xi) = G_z(z_c(\xi), \xi, \eta_0(\xi)) \tag{4.5b}$$

where the subscript z denotes differentiation with respect to z . For η close to but smaller than $\eta_0(\xi)$, the physically relevant solution of (4.3) can be written as

$$z_1 = z_c(\xi) + \left(\frac{-2G_\eta(z_c(\xi), \xi, \eta_0(\xi))}{F_{zz}(z_c(\xi), \xi)} \right)^{1/2} (\eta_0(\xi) - \eta)^{1/2}. \tag{4.6}$$

The factor in front of $(\eta_0(\xi) - \eta)^{1/2}$ will be denoted by $B(\xi)$. Here, as in (4.5a, b) a subscript z or η implies differentiation with respect to that variable.

Both functions $z_c(\xi)$ and $\eta_0(\xi)$ can be related to $z_0(\xi)$, the positive root of F . Since F does not contain any large parameters it can be approximated by a straight line in the vicinity of $z_0(\xi)$

$$F(z, \xi) = F_z(z_0(\xi), \xi)(z - z_0(\xi)). \tag{4.7}$$

Assuming that $z_c(\xi) \approx z_0(\xi)$, the osculation relations (4.5a, b) can be written in the form

$$F_z(z_0(\xi), \xi)(z_c(\xi) - z_0(\xi)) = K\alpha^2(\xi)c_1^2\xi\eta_0(\xi)f_3z_c^{-K} \tag{4.8a}$$

$$F_z(z_c(\xi), \xi) = F_z(z_0(\xi), \xi) = -K^2\alpha^2(\xi)c_1^2\xi\eta_0(\xi)f_3z_c^{-K-1} \tag{4.8b}$$

and thus, dividing (4.8a) by (4.8b),

$$z_c(\xi) = (1 + 1/K)^{-1}z_0(\xi). \tag{4.9}$$

Hence, from (4.8b) and (4.9) we find

$$\eta_0(\xi) = -\frac{F_z(z_0(\xi), \xi)}{K^2\alpha^2(\xi)c_1^2\xi ef_3} z_0^{K+1}(\xi) \tag{4.10}$$

where we used

$$\lim_{K \rightarrow \infty} (1 + 1/K)^{K+1} = e = 2.71828 \dots \tag{4.11}$$

From (4.9) one notices that indeed $z_c(\xi) \approx z_0(\xi)$ if $K \gg 1$ and that osculation always occurs to the left of $z_0(\xi)$.

In order to derive some properties of z_c and η_0 one has to study the root z_0 in more detail. It is implicitly given by

$$F(z, \xi; f_1, f_2) = 0 \tag{4.12}$$

where the parameters f_1, f_2 are included for completeness. These parameters can only take positive values. As $F_z < 0$ if f_1, f_2 and ξ are positive and if $0 \leq z < (\xi f_2)^{-1/2}$, there is only one root in this interval. Using the implicit function theorem, it can be shown that z_0 is a decreasing function of ξ and of the parameters. Finally one has for all allowed parameters and ξ values that $0 < z_0 \leq 1$.

Through (4.9) these results also hold for $z_c(\xi)$ and since $K \gg 1$ the function $\eta_0(\xi)$ is essentially proportional to $z_0^{K+1}(\xi)$. Hence $\eta_0(\xi)$ also decreases sharply as ξ or any of the parameters f_1, f_2 increases. Moreover, combination of (4.1) and (4.6) gives for q_K

$$q_K(\xi, \eta) = \alpha(\xi)c_1z_c^{-K-1}(\xi) \exp\left(-\frac{(K+1)B(\xi)}{z_c(\xi)}(\eta_0(\xi) - \eta)^{1/2}\right) \tag{4.13}$$

showing that q_K is an increasing function of ξ and η and bounded from above by

$$q_K(\xi, \eta_0(\xi)) = \alpha(\xi)c_1z_c^{-K-1}(\xi) = \alpha(\xi)c_1e z_0^{-K-1}(\xi). \tag{4.14}$$

Furthermore, since $G(z, \xi, \eta) \rightarrow 0$ as $\eta \rightarrow 0$, one has

$$q_K(\xi, 0) = \alpha(\xi)c_1z_0^{-K-1}(\xi) \tag{4.15}$$

so

$$q_K(\xi, \eta_0(\xi)) = e q_K(\xi, 0) \quad \text{if } K \gg 1.$$

From (4.10), (4.13) and (4.14) one finds for the left-hand side of (3.9b)

$$\eta(1 + q_K(\xi, \eta)) = \eta_0(\xi)q_K(\xi, \eta_0(\xi)) = -\frac{F_z(z_0(\xi), \xi)}{\alpha(\xi)c_1\xi f_3 K^2} \tag{4.16}$$

for $K \gg 1$, provided ξ is large enough such that $q_K(\xi, \eta) \gg 1$ for all values of η between 0 and $\eta_0(\xi)$.

We now estimate the values of ξ for which the above approximation is valid. This will result in an estimate of the region of DNA-antibody concentrations for which a gel is formed. By definition $q_K(0, 0) = 1$ and hence $\alpha(\xi)c_1 \rightarrow 1$ if $\xi \rightarrow 0$, since $z_0(0) = 1$. For small ξ values one may thus write

$$z_0(\xi) = 1 + \xi \partial_\xi z_0(0) \tag{4.17}$$

where, by the implicit function theorem,

$$\partial_\xi z_0(0) = \lim_{\xi \rightarrow 0} \frac{F_\xi(z_0(\xi), \xi)}{F_z(z_0(\xi), \xi)} = -(f_1 + f_2). \tag{4.18}$$

Thus, for small ξ values

$$q_K(\xi, 0) = \exp[(K+1)(f_1 + f_2)\xi] \tag{4.19}$$

which, for $K \gg 1$, is much larger than unity for small ξ values, say for $\xi \geq \xi_c$ where $\xi_c = O\{1/[(K+1)(f_1 + f_2)]\}$.

If $K \gg 1$ then $\xi_c \ll 1$ and since $F_z(z_0(\xi), \xi) \rightarrow -1$ and $\alpha(\xi)c_1 \rightarrow 1$ as $\xi \rightarrow 0$ we can approximate the upper bound (4.16) for the left-hand side of (3.9b) for $\xi > \xi_c$ by

$$-\frac{F_z(z_0(\xi_c), \xi_c)}{4\alpha(\xi_c)c_1\xi_c f_3 K^2} = \frac{f_1 + f_2}{4f_3} \frac{1}{K}. \tag{4.20}$$

Likewise we can obtain an estimate for an upper bound for (3.9a). Indeed the left-hand side of (3.9a) is bounded from above by

$$H(\xi) = \frac{1}{4\xi} \left(4 + \frac{d}{d\xi} \int_0^{\eta_0(\xi)} q_K(\xi, \eta') d\eta' \right). \tag{4.21}$$

Approximating the integral by a one-step Simpson method one obtains, using (4.14)–(4.16) for $\xi \ll 1$

$$H(\xi) = \frac{1}{4}\xi \left[4 + \frac{e+1}{2K^2 e f_3} \frac{d}{d\xi} \left(\frac{F_c(z_0(\xi), \xi)}{\alpha(\xi) c_1 \xi} \right) \right] \quad (4.22)$$

resulting up to first order in K^{-1} in the estimate

$$H(\xi_c) = \frac{8ef_3 + (e+1)(f_1 + f_2)^2}{8Kef_3(f_1 + f_2)}. \quad (4.23)$$

Thus, if $C_M \geq C_{M,c}$, where $C_{M,c}$ is given by $H(\xi_c)$ approximately, the inverse of (3.9a) giving ξ as a function of η and C_M will be such that $\xi > \xi_c$. If, in addition, $C_N \geq C_{N,c}$, where $C_{N,c}$ is given by (4.20) essentially, there is no solution to (3.9a, b). This is interpreted to indicate gelation in the system, i.e. there are C_N and C_M values such that not all DNA molecules and antibodies can be accommodated in complexes of a finite size, hence macroscopic aggregates of DNA-antibody are formed.

We notice that both $C_{N,c}$ and $C_{M,c}$ decrease as K and/or f_3 increase. In both situations the probability of crosslinks being formed increases, so gelation will occur at lower concentrations. Likewise, if f_1 and/or f_2 increase, $C_{N,c}$ and $C_{M,c}$ increase provided $f_1 + f_2 > 8ef_3/(e+1)$. This corresponds to situations where more antibodies are bound to one and the same DNA molecule, thus decreasing the possibility for crosslinks to form, and a gel is only formed at higher concentrations.

In § 5 we discuss numerical results obtained on the basis of (2.11) in combination with (2.12) for the statistical properties introduced in § 3 for DNA molecules of intermediate lengths. Also the asymptotic results obtained above will be verified and extended by these numerical results.

5. Numerical results for the statistical properties of DNA-antibody complexes

In this section we first consider the numerical procedure to evaluate (3.9a, b) and to calculate such properties as $\langle n \rangle$, $\langle m \rangle$ and C . Next some results will be shown for (i) the boundary function $\eta_0(\xi)$ and the generating function $q_K(\xi, \eta)$, (ii) some statistical properties as a function of C_N and C_M at constant model parameters and (iii) various sets of model parameters.

The central problem in obtaining statistical properties of DNA-antibody complexes is to solve the system of equations (3.9a, b) for ξ and η at various C_N and C_M values. This requires a knowledge of the generating function q_K which was expressed in terms of the roots z_1, \dots, z_7 of (2.11). The roots were determined in a self-consistent way at some fixed set of model parameters for a large number of (ξ, η) pairs on a variable self-adjusting grid. In this way the generating function q_K was obtained and the left-hand sides of (3.9a, b) could be determined numerically. The integral in (3.9a) was approximated using Simpson's rule and the derivative was determined by taking finite differences.

The rate of convergence of this self-consistent approach was extremely high for (ξ, η) pairs with η away from $\eta_0(\xi)$ and decreased rapidly if η approached $\eta_0(\xi)$. The function $\eta_0(\xi)$ was determined with a simple bisection method for a large number of ξ values, using the convergence-divergence of the method as the criterion for selecting new intervals.

Once the left-hand sides of (3.9a, b) were determined, (3.9b) was inverted at constant C_N to obtain a solution curve $\eta(\xi; C_N)$ and along this curve the corresponding C_M values were determined from the left-hand side of (3.9a), such that $(\xi, \eta(\xi; C_N))$ was a solution to (3.9a, b). This procedure was repeated for a number of C_N values, thus yielding the statistical properties as a function of C_N and C_M for a given set of model parameters.

An example of the generating function q_K is shown in figure 4. It is typical for all sets of model parameters and shows that q_K is an increasing function of both ξ and η , bounded from above by $q_K(\xi, \eta_0(\xi))$ which itself is a sharply increasing function of ξ . Notice that even at this small value of K , $q_K(\xi, \eta_0(\xi))$ is roughly proportional to $q_K(\xi, 0)$ as remarked below (4.14) and (4.15) with a proportionality constant roughly equal to e .

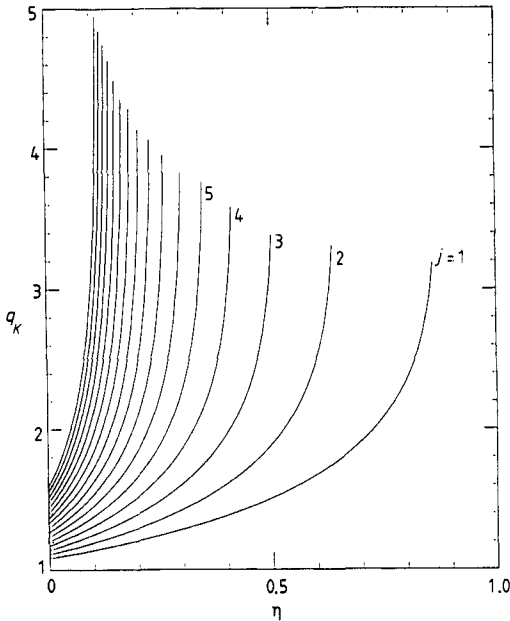


Figure 4. The function $q_K(\xi, \eta)$ as a function of η for $\xi = j\Delta\xi$ with $\Delta\xi = 5 \times 10^{-4}$ and $j = 1, 2, \dots$, as indicated on the corresponding curves. $K = 10$ and $f_1 = f_2 = f_3 = 1$.

The boundary curve $\eta_0(\xi)$ decreases very rapidly as a function of ξ , which is mainly a consequence of the fact that $z_0(\xi)$ decreases as ξ is increased and essentially $\eta_0 \approx z_0^{K+1}$. As K or any of the model parameters increases η_0 decreases, which can be traced back to the fact that z_0 is a decreasing function of all model parameters, as was discussed in the previous section. The dependence of η_0 on f_3 or K is quite strong, as can be seen from (4.10). The dependence on f_1 and f_2 is less dominant. The function $\eta_0(\xi)$ is plotted in figure 5.

We now turn to the results obtained for the statistical properties $\langle n \rangle$, $\langle m \rangle$ and $C_V \equiv Cv_0/V$, the volume fraction occupied by the complexes. The results obtained for a reference case will first be considered and the influences of variations of the parameters on these quantities will then be investigated.

In figures 6 and 7 the average number of DNA molecules per complex $\langle n \rangle$ and the average number of antibodies per complex $\langle m \rangle$ as a function of C_N and C_M are shown

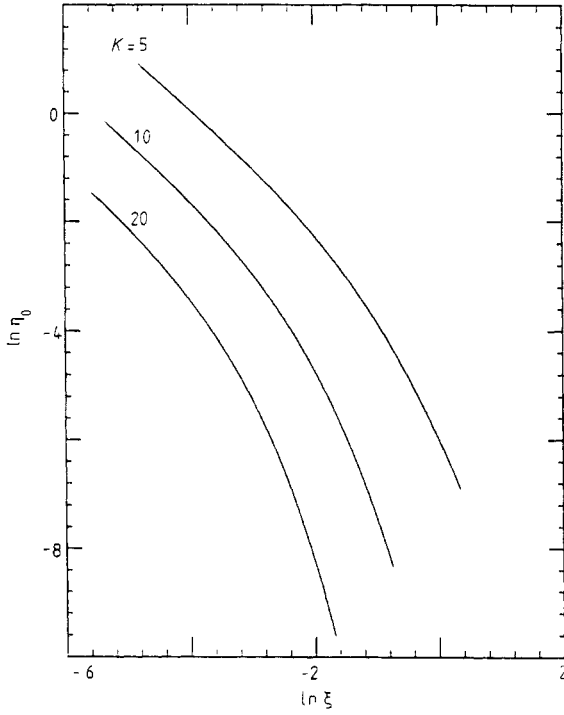


Figure 5. Plot of $\ln \eta_0$ as a function of $\ln \xi$ for $K = 5, 10, 20$. In all cases $f_1 = f_2 = f_3 = 1$.

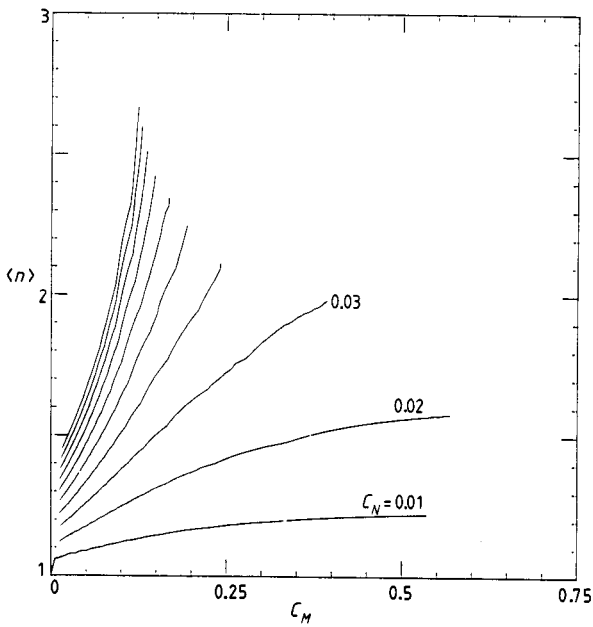


Figure 6. The average number of DNA molecules per complex as a function of C_M . The values of C_N start at 0.01 and increase with steps of 0.01. In all cases $K = 10$ and $f_1 = f_2 = f_3 = 1$.

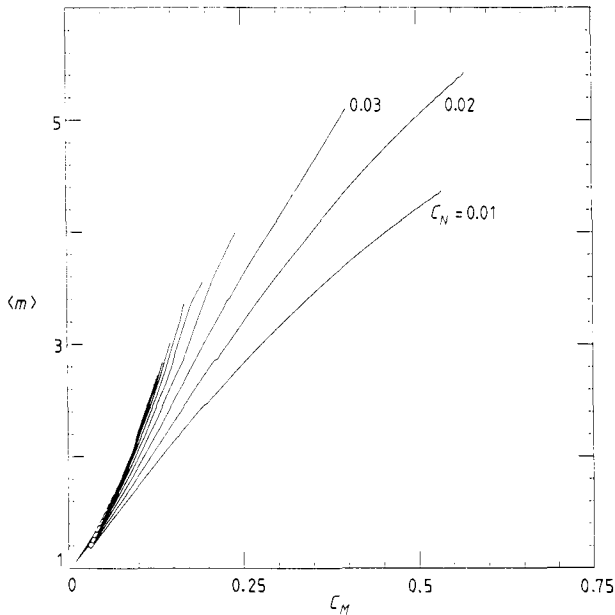


Figure 7. The average number of antibody molecules per complex as a function of C_M , with parameters as in figure 6.

for situations in which no gel is formed. From them, one recognises that both $\langle n \rangle$ and $\langle m \rangle$ are increasing functions of C_N and C_M . This also shows that the resulting gels will have a variety of crosslink densities which are essentially determined by C_N . Taking $\langle m \rangle / \langle n \rangle$ at the gelation boundary as a measure for the crosslink density in the gel, one sees that this density decreases as C_N increases.

A numerical evaluation of C_V as a function of C_M and C_N shows that for small C_M values the number of complexes increases rapidly as C_M increases and the increase is sharper as C_N is larger. For C_M approaching the corresponding gelation value we find that the number of complexes decreases. Apparently, in combination with figures 6 and 7, close to gelation there are fewer but larger complexes.

We have also studied the influence of variations in the model parameters on the statistical properties $\langle n \rangle$, $\langle m \rangle$ and C_V . First, the influence of varying the number of possible binding sites K on the DNA molecule was considered. Our results for $\langle n \rangle$ and $\langle m \rangle$ indicate that, as the number of possible binding sites K increases, $\langle n \rangle$ and $\langle m \rangle$ increase at the same C_M and C_N values. The influence of varying K on C_V shows that in situations in which there is a surplus of DNA molecules the number of complexes increases as the number of possible binding sites on the DNA increases. However, since for larger K gelation occurs at lower C_N and C_M values (cf (4.20) and (4.23)) the maximal number of complexes decreases rapidly as K is increased.

Next we varied the model parameters f_1 , f_2 and f_3 keeping the number of possible binding sites on the DNA molecule fixed. We found that as f_1 and f_2 are increased, more, but smaller complexes with more densely packed DNA molecules occur, and as f_3 was increased, the complexes become larger (i.e. more DNA molecules on the average per complex) but with less antibodies per DNA molecule.

Finally we considered the gelation boundary C_N as a function of C_M which is a set of C_N , C_M pairs such that the system is just on the transition boundary between

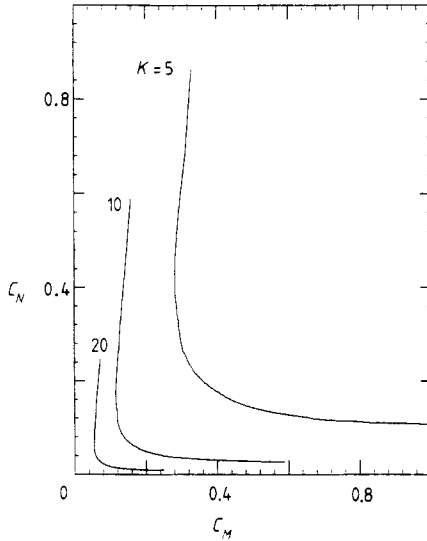


Figure 8. The gelation boundary for $K = 5, 10, 20$. In all cases $f_1 = f_2 = f_3 = 1$. The gel region is on the upper right-hand side of the boundary.

the fluid phase, where a large number of complexes exists, and the gel phase, where macroscopic DNA-antibody aggregates are formed. In the previous section this phenomenon was investigated in the asymptotic limit. Here we present the data obtained from the full numerical evaluation of the computational scheme discussed in § 2. The influence of varying the number of possible binding sites K is characterised in figure 8. It shows that, as K is increased, gelation occurs at lower C_M values for constant C_N and vice versa. In the previous section it was shown that $C_{N,c}$ and $C_{M,c}$ are proportional to K^{-1} . There is a remarkably good agreement between (4.20) and (4.23) and the numerical results, even though the values of K are quite small. This is due to the fact that the roots z_1, \dots, z_7 of (2.11) are very well separated from each other, so the approximations made in (4.1) and (4.2) for the generating functions q_K and A are already very accurate for relatively low values of K .

A new aspect which did not show up in the asymptotic considerations of § 4 is the fact that the gelation boundary curve is two-valued for C_M close to $C_{M,c}$. This implies that for suitable C_M the addition of extra DNA molecules to an already existing gel causes the gel to disintegrate. The influence of variations in f_1, f_2 and f_3 can be summarised as follows. $C_{M,c}$ and $C_{N,c}$ turn out to decrease as f_3 is increased, and increase as f_1 and/or f_2 are increased. This is in full agreement with the asymptotic results of § 4.

References

- Stollar B O 1981 *Arthritis Rheum.* **24** 1010
 Wiegel F W 1983 *Phase Transitions and Critical Phenomena* vol 7, ed C Domb and J L Lebowitz (New York: Academic) p 101
 Wiegel F W and Goldstein B 1987 *Biopolymer.* **26** 297
 Wiegel F W and Perelson A S 1982 *J. Stat. Phys.* **29** 813