Mathematical analysis on certain physical properties applied on random coil model

Jaydip Datta[1], Chiranjib Bhattacharjee[1,*], T. K. Bhattacharyya[2] and B. R. De[3]

[1]Department of Chemical Engineering, Jadavpur University, Kolkata 700 032, India
[2]Division of Pharmaceutical and Fine Chemicals Technology, Department of Chemical Technology, University of Calcutta, 92, A.P.C. Road, Kolkata 700 009, India
[3]Department of Chemistry and Chemical Technology, Vidyasagar University, Midnapore 721 102, India

The random coil model is analysed to establish the mathematical support behind it. In this approach Stirling's approximation, weight factor and improper integral (gamma function) are analysed in detail. A correlation between the important modelling parameters like radius of gyration, root mean square separation, number of monomers and length of each monomer unit is established. Finally this model is validated by a standard object-oriented programming tool for a protein like serum albumin.

Keywords: Hydrodynamic properties, improper integral, random coil, serum albumin, Stirling's approximation.

THE random coil model is a helpful starting point for estimating the orders of magnitude of the hydrodynamic properties of polymers and denatured protein in solution. This is the most fundamental model based on free rotation of any single bond at any angle with respect to the preceding one.

In their pioneer work, Bloomfield *et al.*[1] adapted the Kirkwood–Riseman[2] hydrodynamic theory, initially conceived for rod-like or chain-like molecules modelled as string beads, to calculate the hydrodynamic properties of macromolecules of arbitrary shape. In the work of one of the two authors, with Bloomfield and other co-workers, the methodology was improved in regard to theoretical aspects, model-building and computational aspects. A procedure was devised by García de la Torre and Carrasco[3], for calculation of hydrodynamic properties of rigid macromolecules compared to the revolution of cylinders. A computer program HYDROSUB[4] was developed and this methodology was presented in the case of solution properties of the human antibody molecule immunoglobulin G3 (IgG3).

The present work recently reviewed alternative strategies for hydrodynamic modelling and hinted at a general procedure to correlate the physical properties like radius of gyration R_g, root mean square separation R_{rms}, and total number of bonds N or monomers in random coil model[5].

*For correspondence. (e-mail: cbhattacharyya@chemical.jdvu.ac.in)

An attempt has been made to establish the mathematical support or basis behind these physical properties, and it is finally validated through a standard software programming language (C^{++}) for a protein molecule like serum albumin.

Input model equations[5]: To start with, we highlight the main equations which are analysed by their general mathematical procedure and model building.

The probability P of the random coil model at the end-to-end separation is determined by the following equations:

$$P = \frac{W}{2^N}, \tag{1}$$

where W is called the weight of the configuration and depends on binomial coefficients.

$$W = \frac{N!}{N_R!\, N_L!}, \tag{2}$$

where N is the total number of bonds and is equal to $(N_R + N_L)$, where N_R and N_L are the number of bonds pointing to the right and left respectively. Defining n as the number of bonds for end-to-end separation, equal to $(N_R - N_L)$, one can obtain the following equations for N_R and N_L:

$$N_R = \left(\frac{N+n}{2}\right) \text{ and } N_L = \left(\frac{N-n}{2}\right).$$

Now the general expression for the mean nth power of end-to-end separation is:

$$\langle R^n \rangle = \int_0^\infty R^n \left[4\pi \left(\frac{a}{\sqrt{\pi}}\right)^3 \exp(-a^2R^2) R^2 \right] dR, \tag{3}$$

$$a = \left(\frac{3}{2Nl^2}\right)^{1/2}, \tag{4}$$

where N is the total number of monomers/bond and l is the length of each monomer unit.

For mean separation $n = 1$,

$$\langle R^n \rangle = \int_0^\infty 4\pi \left(\frac{a}{\sqrt{\pi}}\right)^3 \exp(-a^2R^2) R^3 \, dR$$

$$= 4\pi \left(\frac{a}{\sqrt{\pi}}\right)^3 \int_0^\infty \exp(-a^2R^2) R^3 \, dR. \tag{5}$$

For mean of the square separation $n = 2$,

$$\langle R^n \rangle = 4\pi \left(\frac{a}{\sqrt{\pi}}\right)^3 \int_0^\infty \exp(-a^2R^2) R^4 \, dR. \tag{6}$$

Analysis[6]: Let us analyse model eq. (1). Taking logarithm on both sides of eq. (1), we get

$$\ln P = \ln N! - (\ln N_L! + \ln N_R! + N \ln 2). \quad (7)$$

The reason for introducing logarithm is that it is easier to make approximations. In particular, we can simplify the factorials using Stirling's approximation in the form:

$$y! = (2\pi)^{1/2} y^{(y+\frac{1}{2})} \exp(-y), \quad (8)$$

with an error ε less than 1.0% when y is greater than 10. A larger value of y can also be handled by the above equation. The factorials can be simplified using Stirling's approximation in the form:

$$\ln N! = \ln(2\pi)^{1/2} + \left(N + \frac{1}{2}\right)\ln N - N, \quad (9)$$

$$\ln\left(\frac{N-n}{2}\right)! = \ln(2\pi)^{1/2} + \frac{1}{2}(N-n+1)\ln\frac{N}{2} + \frac{1}{2}(N-n+1)\ln\left(1-\frac{n}{N}\right) - \left(\frac{N-n}{2}\right), \quad (10)$$

$$\ln\left(\frac{N+n}{2}\right)! = \ln(2\pi)^{1/2} + \frac{1}{2}(N+n+1)\ln\frac{N}{2} + \frac{1}{2}(N+n+1)\ln\left(1+\frac{n}{N}\right) - \left(\frac{N+n}{2}\right). \quad (11)$$

Thus $\ln P$ can be calculated as:

$$\ln P = \ln W - \left[\ln\left(\frac{N-n}{2}\right)! + \ln\left(\frac{N+n}{2}\right)!\right] - N\ln 2 = \ln\left(\frac{2}{\pi N}\right)^{1/2} - \frac{1}{2}(N-n+1)\ln\left(1-\frac{n}{N}\right) - \frac{1}{2}(N+n+1)\ln\left(1+\frac{n}{N}\right). \quad (12)$$

Substituting $n/N = x$ and expanding,

$$\ln(1+x) = x + \frac{x^2}{2} + \cdots \qquad (0 < x < 1),$$

$$\ln(1-x) = -x + \frac{x^2}{2} - \cdots \qquad (0 < x < 1).$$

We get the simplified form by applying the above $\ln(1 \pm x)$ series.

$$\ln P = \ln\left(\frac{2}{\pi N}\right)^{1/2} - \frac{n^2}{2N},$$

$$P = \left(\frac{2}{\pi N}\right)^{1/2} \exp\left(-\frac{n^2}{2N}\right) = A\left[\exp\left(-\frac{n^2}{2N}\right)\right], \quad (13)$$

where A is $\left(\frac{2}{\pi N}\right)^{1/2}$.

Now, let us analyse model eq. (3).

$$\langle R^n \rangle = \int_0^\infty R^n f \, dR,$$

where

$$f = 4\pi\left(\frac{a}{\sqrt{\pi}}\right)^3 R^2 \exp(-a^2R^2),$$

and

$$a = \left(\frac{3}{2Nl^2}\right)^{1/2}.$$

For $n = 1$,

$$\langle R \rangle = \int_0^\infty R f \, dR = 4\pi\left(\frac{a}{\sqrt{\pi}}\right)^3 \int_0^\infty R^3 \exp(-a^2R^2)\, dR. \quad (14)$$

The above is an improper integral of the first kind, since in

$$\int_a^b f(x)\, dx, \quad a = 0, b = \infty.$$

So the standard integration of this type

$$\int_0^\infty x^3 \exp(-a^2x^2)\, dx$$

is solved. By applying methods of substitution and integration by parts, the integration becomes

$$\frac{1}{2a^4}\left[\operatorname*{Lt}_{\lambda\to\infty} \int_0^\lambda z e^{-z}\, dz\right].$$

Assuming z as the first function and e^{-z} as the second function, the expression becomes

$$\frac{1}{2a^4}\left(\operatorname*{Lt}_{\lambda\to\infty}\left[e^{-z}\right]_0^\lambda\right) = \frac{1}{2a^4}.$$

Substituting this value in eq. (14), we get

$$\langle R \rangle = \left(\frac{8}{3\pi}\right)^{1/2} N^{1/2} l. \quad (15)$$

For

$$n = 2, \ \langle R^n \rangle = \int_0^\infty R^2 f \, dR. \quad (16)$$

Here another improper integration of the first kind is solved by the Eulerian integral using gamma function. By definition, the gamma function,

$$\Gamma n = \int_0^\infty \exp(x) x^{n-1}\, dx = (n-1)\,\Gamma(n-1) = (n-1)!$$

Table 1. Output model equations from analysed equations

Output model equations[5,7]	Description of physical properties
$N = \left(\frac{R_{rms}}{l}\right)^2$	R_{rms} is the root mean square separation of end-to-end and is described by $\int_0^\infty R^2 f\,dR$. N is the number of monomers or total number of bonds of random walk.
$R_g = \frac{R_{rms}}{\sqrt{6}}$	R_g is the radius of gyration of freely jointed polymer chain.
N (modified) = $\frac{1}{2}\left(\frac{R_{rms}}{l}\right)^{1/2}$	For constrained chains, introducing degrees of freedom $F = \sqrt{2}$ for tetrahedral angle.
R_g(modified) = $\frac{R_{rms}}{\sqrt{6}}$	$R_{rms} = \int_0^\infty R^2 f\,dR$; R_g is the radius of gyration of freely jointed polymer chain.

Table 2. Output of the program

Enter data Index	:	01
Enter the value of radius of gyration (R_g) in nm	:	2.98
Enter the value of bond length (l) in nm	:	0.154
The value of R_{rms}	:	47.409092 ± 2.1
The value of N	:	172.994812 ± 8.6

Equation (16) becomes

$$\int_0^\infty R^2 f\,dR = 4\pi\left(\frac{a}{\sqrt{\pi}}\right)^3 \int_0^\infty R^4 \exp(-a^2R^2)\,dR. \quad (17)$$

Equation (17) is solved in the standard form

$$\int_0^\infty x^4 \exp(-a^2x^2)\,dx.$$

Now by substituting $a^2x^2 = z$, the integration is changed to the form

$$\int_0^\infty z^{3/2}\exp(-z)\,dz = \frac{1}{2a^5}\int_0^\infty z^{(\frac{5}{2}-1)}\exp(-z)\,dz$$

$$= \frac{1}{2a^5}\Gamma\left(\frac{5}{2}\right) = \frac{1}{2a^5}\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)(\pi)^{1/2}$$

$$= \frac{3}{8a^5}(\pi)^{1/2}$$

$$= Nl^2. \quad (18)$$

Equation (18) is obtained by applying the method of substitution and putting the standard value of

$$\Gamma\left(\frac{1}{2}\right) = (\pi)^{1/2}.$$

Output model equations[5,7] from analysed equations (eqs (7–12) and (14–18)) are given in Table 1.

Computational validation[8]: The output model equations given in Table 1 are validated using an object-oriented programing language (C^{++}) software, which has been developed in this study, using a standard protein like serum albumin. The C–C single bond length l = 0.154 nm and radius of gyration R_g = 2.98 nm, from Corey–Pauling rules for serum albumin (l = 0.154 nm, R_g = 2.98 nm and M = 66 kg/mol). The values for l, R_g and M (molecular weight of biopolymer) were taken from the literature[5,9].

The output of the program is shown in Table 2.

The developed software can be executed by multiple data index in case of other biomolecules like DNA, myosin and tobacco mosaic virus with the supplied data of C–C single bond length and value of radius of gyration.

1. Bloomfield, V. A., Dalton, W. O. and Van Holde, K. E., Hydrodynamic modelling of rod-like or chain-like macromolecules. *Biopolymers*, 1967, **5**, 135–148.
2. Kirkwood, J. G. and Riseman, J., Hydrodynamic theory of string bead molecules. *J. Chem. Phys.*, 1948, **16**, 565–573.
3. García de la Torre, J. and Carrasco, B., Hydrodynamic properties of rigid macro-molecules composed of ellipsoidal and cylindrical subunits. *Biopolymers*, 2002, **63**, 163–167.
4. Baumketner, A. and Hiwatari, Y., Influence of the hydrodynamic interaction on kinetics and thermodynamics of minimal protein models. *J. Phys. Soc. Jpn.*, 2002, **71**, 3069–3079.
5. Atkins, P. and Paula, D. J., Macromolecules and aggregates. In *Atkins' Physical Chemistry* (ed. Khan, M.), Oxford University Press, New Delhi, 2003, pp. 719–835.
6. Murray, S. R., Improper integral. In *Theory and Problems of Advanced Calculus* (ed. Hill, M. G.), New York, 1974, pp. 285–289.
7. Van Holde, K. E. and Johnson, W., Structures in motion. In *Principle of Physical Biochemistry* (ed. Ho, P. S.), Prentice Hall, 1998, pp. 800–810.
8. Keith, W., *Array Variables in the Complete C^{++} Primer* (ed. Bryton, F.), Academic Press, New York, 1994, pp. 151–172.
9. Tanford, C. (ed.), Macromolecular aggregates. In *Physical Chemistry of Macro-molecules*, Wiley, New York, 1961, pp. 200–210.

ACKNOWLEDGEMENT. We thank Dr Prasanta Kumar Seth, Mahadevananda Mahavidyalaya, Barrackpore for valuable suggestions.

Received 5 July 2006; revised accepted 21 March 2007