7th Jagna International Workshop (2014)
International Journal of Modern Physics: Conference Series
Vol. 36 (2015) 1560018 (7 pages)
© The Authors
DOI: 10.1142/S2010194515600186

World Scientific
www.worldscientific.com

Int. J. Mod. Phys. Conf. Ser. 2015.36. Downloaded from www.worldscientific.com
by 73.144.208.122 on 04/15/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

On the diffusion of alpha-helical proteins in solvents

Wilson I. Barredo

Department of Physics
Mindanao State University (MSU)-Iligan Institute of Technology
Iligan City, 9200, Philippines
Department of Physics, Mindanao State University
Marawi City, 9700, Philippines
barredo1234@yahoo.com

Jinky B. Bornales

Department of Physics
Mindanao State University (MSU)-Iligan Institute of Technology
Iligan City, 9200, Philippines
jbornales@gmail.com

Christopher C. Bernido

Research Center for Theoretical Physics
Central Visayan Institute Foundation
Jagna, Bohol, 6308, Philippines
cbernido@mozcom.com

Henry P. Aringa

Department of Physics, Mindanao State University
Marawi City, 9700, Philippines
henparinga@yahoo.com

Published 2 January 2015

The winding probability function for a biopolymer diffusing in a crowded cell is obtained with the drift coefficient $f(s)$ involving Bessel functions of general form $f(s) = kJ_{2p+1}(\nu s)$. The variable s is the length along the chain and ν is a constant which can be used to simulate the frequency of appearance of a certain type of amino acid. Application of a particular case $p = 3$ to protein chains is carried out for different alpha helical proteins found in the Protein Data Bank (PDB). Analysis of our results leads us to an empirical formula that can be used to conveniently predict k/D and ν, where D is the diffusion coefficient of various α-helical proteins in solvents.

Keywords: Diffusion coefficient; α-helical proteins; winding probability.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 3.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

Int. J. Mod. Phys. Conf. Ser. 2015.36. Downloaded from www.worldscientific.com
by 73.144.208.122 on 04/15/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

1. Introduction

In this paper, we show that from the results obtained earlier [1-3] for myoglobin and ferritin using the white noise functional approach in modeling the alpha-helical secondary structure of the protein, other α-helical proteins can also be simulated using Brownian paths. In particular, the Fokker-Plank equation is solved to obtain the probability density function from which winding probabilities $W(n, L)$ are calculated. The various helical protein conformations are then viewed as arrays of diffusion paths modulated by a drift coefficient involving Bessel functions of general order $f(s) = J_{2p+1}(\text{vs})$, in particular the case for $p = 3$. We then apply this to investigate the diffusion of proteins consisting mainly of α-helical conformations.

Development of reliable methods of predicting diffusion coefficients for proteins and other macromolecules is of interest since diffusion is involved in a number of biochemical processes such as protein aggregation, transport in intercellular media and the protein folding process [4]. Hence, how proteins diffuse inside the cell has been the subject of recent experimental and theoretical studies. Simple scaling and distribution relationships have been derived from recent databases to describe some of the physical properties of proteins in cellular proteomes [5]. The results show that many properties of proteins, including their sizes, stabilities, folding rates and diffusion coefficients depend simply on the chain length N. For instance, molecular dynamics simulations have been done to predict diffusion coefficients of four proteins: Cytochrome c (1HRC), lysozyme (1BWI), α-chymotrypsinogen-A (1EX3), and ovalbumin (1OVA) in aqueous solution. J. Wang and T. Hou [5] have also compared molecular diffusion coefficients with experimental values.

2. Modeling of the Polymer

In investigating helical structures of proteins, we use the circular cylindrical coordinates, $\mathbf{r} = (\rho, \theta, z)$, and take a winding polymer oriented along the z-axis. The polymer conformation can be viewed as random walk consisting of N steps which starts at point $\mathbf{r}_0$ and ends at some point $\mathbf{r}_1$. We can simplify the study of the winding behavior of a biopolymer by projecting the paths on the ρ–θ plane and consider a polymer chain which lies on the plane with endpoints at $\rho_0 = (\rho_0, \theta_0)$ and $\rho_1 = (\rho_1, \theta_1)$. However, for a typical α-helical conformation, the radius of a helix is known and, hence, we can fix the radial variable at $\rho = R$. From this scenario, a polymer which winds around the z-axis projects a circular structure on the ρ–θ plane. The probability density function can then be written as [3],

$$P(\theta_1, \theta_0; L) = \int exp\left\{-\frac{1}{l}\int_0^L \left[R\frac{d\theta}{ds} - \frac{l}{2D}f(s)\right]^2 ds\right\} D[\theta]. \tag{1}$$

Here, $L = Nl$ is the length of the polymer with l the length of each monomer, D is the diffusion constant and $f(s)$ the drift coefficient.

To reflect the varying interactions of the different amino acids in an aqueous environment, the value of the drift coefficient $f(s)$, with $0 \leq s \leq L$, can also vary

at each length segment along the chainlike molecule. Corresponding to the path, θ can be parameterized as:

$$\theta(s) = \theta_0 + (\sqrt{l}/R)B(s) \tag{2}$$

where θ_0 is an initial value and B the Brownian fluctuation. Eq. (2) deals with paths confined to a circular topology. The paths can be classified topologically and characterized by winding numbers [6-11] $n = 0, \pm 1, \pm 2, \ldots$ where, $n > 0$ signifies n turns counterclockwise around the origin; $n < 0$ means $|n|$ turns clockwise, and $n = 0$ signifies no winding. With Eq. (2), an evaluation of Eq. (1) using white noise calculus yields the result [3],

$$P(\theta_1, \theta_0) = \sum_{n=-\infty}^{+\infty} P_n, \tag{3}$$

where,

$$P_n = \sqrt{\frac{R^2}{\pi l L}} \exp\left[-\frac{R^2}{lL}\left(\theta_0 - \theta_1 + 2\pi n + \frac{l}{2DR}\int_0^L f\,(s)\,ds\right)^2\right]. \tag{4}$$

Equation (4) is the probability function for an n-times winding of a path around the z-axis. The probability that a helical conformation has a polypeptide winding n-times about the z-axis is given by, $W(n, L) = P_n/P$. For an arbitrary initial point, we let $\theta_0 = \theta_1$, and we obtain [3],

$$W(n, L) = \sqrt{\frac{4\pi}{lL}} \frac{exp\left[-\frac{R^2}{lL}\left(2\pi n + \frac{l}{2DR}\int_0^L f\,(s)\,ds\right)^2\right]}{\boldsymbol{\theta_3}\left(\frac{1}{4DR}\right)\int_0^L f\,(s)\,ds}, \tag{5}$$

where $\boldsymbol{\theta_3}(u)$ is the theta function [12]. We note that Eq. (5) is an exact result obtained by evaluating Eq. (1). The interaction of each amino acid with the aqueous environment as well as with other monomers would be reflected in the drift coefficient $f(s)$, as s ranges from 0 to L along the length of a biopolymer. The $f(s)$ in turn serves as a modulating function affecting the winding probability $W(n, L)$ that describes a specific winding conformation. The particular drift coefficient used in this study is described in the next section.

3. Besselian Drift Coefficient of Order $2p + 1$

The drift coefficient, $f(s) = kJ_{2p+1}(\nu s)$, where $J_{2p+1}(\nu s)$ is a Bessel function, can be integrated over ds,

$$\int_0^L f\,(s)\,ds = (k/v)\left[1 - J_0\,(vL) - 2\sum_{m=1}^{p} J_{2m}\,(vL)\right], \tag{6}$$

Int. J. Mod. Phys. Conf. Ser. 2015.36. Downloaded from www.worldscientific.com
by 73.144.208.122 on 04/15/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

with $p \geq 1$ [12]. Eq. (6) is then used in Eq. (5) to get the winding probability,

$$W(n,L) = R\sqrt{\frac{4\pi}{Ll}} \exp \times \left[\theta_3 \left(\frac{kl}{4DRv}\left[1 - J_0(vL) - 2\sum_{m=1}^{p} J_{2m}(vL)\right]\right)\right]^{-1}. \quad (7)$$

For long polymers $L = Nl >> 1$, $\boldsymbol{\theta_3}(u) \approx 1$, and the probability for a helical conformation with a winding number n becomes:

$$W(n,L) \approx R\sqrt{\frac{4\pi}{Ll}} \exp\left\{-\frac{R^2}{Ll}\left[2\pi n + \frac{kl}{2DRv} \times \left(1 - J_0(vL) - 2\sum_{m=1}^{p} J_{2m}(vL)\right)\right]^2\right\}. \quad (8)$$

4. Application

We take the case $p = 3$ of the drift coefficient $f(s) = kJ_{2p+1}(\nu s) = kJ_7(\nu s)$ [1] and Eq. (6) reduces to:

$$\int f(s)ds = (k/\nu)\{1 - J_0(\nu L) - 2[J_2(\nu L) + J_4(\nu L) + J_6(\nu L)]\}. \quad (9)$$

The winding probability $W(n,\ L)$ is the same as Eq. (8), but with only $m = 1$, 2, 3 contributing to the summation term in the exponential. We use the general properties of proteins, $R = 0.25\,\text{nm}$, $l = 0.15\,\text{nm}$, and 3.6 residues per helical turn for different alpha helical proteins. The graphs of $W(n,L)$ versus length L are simulated in order to find the values of k/D and ν that will mimic their experimentally verified features.

For example, myoglobin (4MBN) has only one chain with total length of 153 residues, with alpha helical segments about 80% of its length or about 123 residues, and with 11 helices based on PDB. Therefore this protein has about 123 residues divided by 3.6 residues/turns or about 34 helical turns and a length of $L = 0.15(153)\,\text{nm} \approx 23\,\text{nm}$. Plotting the winding probability $W(-n,L)$ versus length L using the above data, the values of $\nu = 1.93/\text{nm}$ and $k/D = 1420/\text{nm}$ were found giving 11 peaks (Figure 1). The peaks apparently correspond to 11 helices of Myoglobin (4MBN), and the negative n signifies that this protein is right-handed.

The method above was used for other alpha-helical proteins presented in Table 1, with the values of k/D and ν for each alpha-helical protein giving a good one-to-one correspondence between the number of peaks in the graph of $W(-n,L)$ versus length L and the number of helical segments based on data from PDB for each of the proteins.

One can observe in Table 1 that longer proteins have larger k/D values and shorter proteins have smaller values of k/D in aqueous solvent. These results agree with experimental observations [14]. In general, larger proteins diffuse slower and smaller ones diffuse faster in aqueous solvents. The results also agree with theoretical results based on the Stokes-Einstein theory [4]. The model therefore, has the

Int. J. Mod. Phys. Conf. Ser. 2015.36. Downloaded from www.worldscientific.com
by 73.144.208.122 on 04/15/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

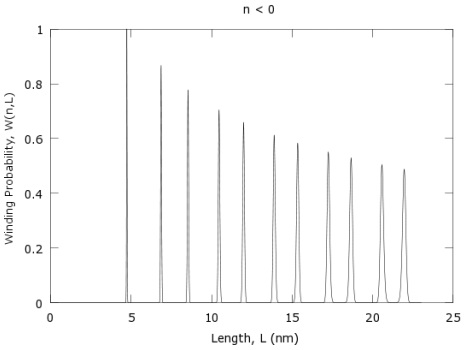

Fig. 1. 4MBN: Graph of $W(-n, L)$ versus length for $\nu = 1.93$/nm and $k/D = 1420$/nm.

Table 1. Properties of alpha helical proteins with the simulated values of k/D and ν.

Protein (PBD code)	**Length (# of Residues)**	**% alpha (# of Residues)**	**# of helices**	**# of turns**	**ν (1/nm)**	**k/D (1/nm)**
2K9J	42	57% (24)	1	7	1.30	490
3IA3-A^{+}	91	71% (65)	3	18	1.30	490
2JUW	80	76% (61)	4	17	1.76	640
2I15	135	59% (80)	5	23	1.29	680
2HMZ	113	69% (79)	6	22	1.65	758
3IA3-D^{++}	145	64% (93)	9	26	1.68	800
2MHB	141	73% (104)	9	29	1.79	1120
4MBN	153	80% (123)	11	34	1.93	1420
2O9D	234	71% (167)	12	47	1.36	1290
4E4V	485	63% (310)	31	86	1.48	2690
2YNS	490	64% (314)	33	87	1.53	2759
4BA3	496	64% (319)	34	89	1.55	2839

Note: $^{+}$3IA3 chain A, $^{++}$3IA3 chain D, and the rest of the above proteins are chain A if there are more than one chain.

potential for describing the general properties of a protein in aqueous solvent. The discouraging feature of this model, however, is that to obtain the values of k/D and ν for the number of peaks to correspond to the helices for each protein being investigated, one had to resort to trial and error scheme. In the graph of $W(-n, L)$ versus length, the desired number of peaks can only be obtained after adjusting several times the two parameters, ν and k/D. Therefore, an empirical formula has been developed to avoid this difficulty.

4.1. *Construction of an Empirical Formula*

In developing the empirical formula, the linear dependence of the values of k/D of the proteins presented in Table 1 to the number of helices was taken into account.

Int. J. Mod. Phys. Conf. Ser. 2015.36. Downloaded from www.worldscientific.com
by 73.144.208.122 on 04/15/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

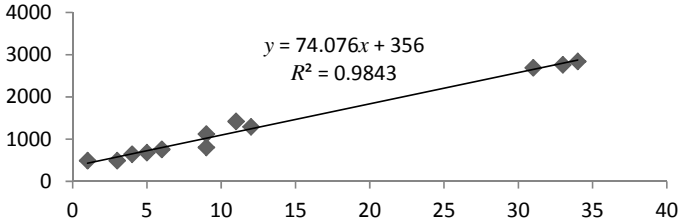

Fig. 2. Plot of k/D versus number of helices.

Table 2. Simulated values compared with predicted values of ν and k/D.

Protein (PBD code)	Length (# of Residues)	% alpha (# of Residues)	Simulated ν(1/nm)	Simulated k/D(1/nm)	Predicted ν(1/nm)	Predicted k/D(1/nm)
2K9J	42	57% (24)	1.30	490	1.49	395.67
3IA3-A^{+}	91	71% (65)	1.30	490	1.415	612.92
2JUW	80	76% (61)	1.76	640	1.83	724.06
2I15	135	59% (80)	1.29	680	1.29	682.80
2HMZ	113	69% (79)	1.65	758	1.67	832.47
3IA3-D^{++}	145	64% (93)	1.68	800	1.78	1012.46
2MHB	141	73% (104)	1.79	1120	1.78	1104.50
4MBN	153	80% (123)	1.93	1420	1.91	1346.46
2O9D	234	71% (167)	1.36	1290	1.36	1319.61
4E4V	485	63% (310)	1.48	2690	1.47	2599.31
2YNS	490	64% (314)	1.53	2759	1.55	2772.50
4BA3	496	64% (319)	1.55	2839	1.56	2845.84

Note: $^{+}$3IA3 chain A, $^{++}$3IA3 chain D, and the rest of the above proteins are chain A if there are more than one chain.

The diffusion coefficient (k/D) is then plotted versus the number of helices (Figure 2) which then gave the best fit linear equation:

$$y = 74.076x + 356 \quad (R^2 = 0.9843) \tag{10}$$

where y is the diffusion coefficient (k/D) and x is the number of helices.

From Figure 2 and from the data in Table 1, proteins with about 65% alpha-helical segments fit closely with the plot of Eq. (10) for the diffusion coefficient of alpha-helical protein in aqueous solvents. The empirical formula,

$$\begin{aligned} k/D &\approx y + (\%\text{alpha} - 65)y \\ &= 74.076x + 356 + (\%\text{alpha} - 65)(74.076x + 356), \end{aligned} \tag{11}$$

seems handy in predicting the diffusion coefficient. The simulated values and the predicted values for the diffusion coefficient k/D of proteins are given in Table 2.

5. Conclusion

In modeling α-helical proteins via the winding probability, Eq. (5), it was shown that an empirical formula Eq. (10) facilitates the determination of ν and k/D. Using

Int. J. Mod. Phys. Conf. Ser. 2015.36. Downloaded from www.worldscientific.com
by 73.144.208.122 on 04/15/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

Eq. (5) and the simulation method, it was also shown that as the length of the biopolymer increases, the values of the diffusion coefficient D decreases which agrees with experimental data.

Acknowledgments

W. I. B. acknowledges support from the Department of Science and Technology-ASTHRDP.

Int. J. Mod. Phys. Conf. Ser. 2015.36. Downloaded from www.worldscientific.com
by 73.144.208.122 on 04/15/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

References

1. H. P. Aringa, Analytical Modeling of Biopolymer Conformations: A White Noise Functional Approach, Ph.D. Dissertation, Mindanao State University-Iligan Institute of Technology (MSU-IIT, 2011).
2. H. P. Aringa, C. C. Bernido, M. V. Carpio-Bernido, and J. B. Bornales, Stochastic Modelling of Helical Biopolymers, *Int. J. Mod. Phys. Conf. Ser.* **17**, 73–76 (2012).
3. C. C. Bernido and M. V. Carpio-Bernido, White Noise Analysis: Some Applications in Complex Systems, Biophysics, and Quantum Mechanics, *Int. J. Mod. Phys. B* **26**, 1230014 (2012).
4. K. Dill, K. Ghosh and J. Schmit, Physical Limits of Cells and Proteomes, *Proc. Natl. Acad. Sci.* **44**, 108, (2011).
5. J. Wang and T. Hou, Application of Molecular Dynamics Simulations in Molecular Property Prediction II: Diffusion Coefficient. *J. Compu Chem.* **32**, 16 (2011).
6. S. F. Edwards, *Proc. Phys. Soc. London*, **91**, 513–519 (1967);
7. S. Prager and H. L. Frisch, *J. Chem. Phys.* **46**, 1475 (1967).
8. L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
9. M. G. Laidlaw and C. M. DeWitt, *Phys. Rev. D* **3**, 1375 (1971).
10. C. C. Bernido and M. V. Carpio-Bernido, *J. Math. Phys.* **43**, 1728–1736 (2002).
11. C. C. Bernido and A. Inomata, *J. Math. Phys.* **22**, 715 (1981).
12. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5^{th} ed. (Academic Press, San Diego, 1994).
13. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (National Bureau of Standards, Washington D. C., 1972).
14. See, e.g. Table 2, S. Papadopoulos, K.D. Jurgens and G. Gros, Protein Diffusion in Living Skeletal Muscle Fibers: Dependence on Protein Size, Fiber Type, and Contraction, *Biophys. J.*, 79, 2084–2094 (2000).