

DYNAMICS AND DYNAMICAL TRANSITIONS IN PROTEINS

by

Derya Vural

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics

Summer 2014

© 2014 Derya Vural
All Rights Reserved

UMI Number: 3642366

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3642366

Published by ProQuest LLC (2014). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

DYNAMICS AND DYNAMICAL TRANSITIONS IN PROTEINS

by

Derya Vural

Approved: _____

Edmund R. Nowak, Ph.D.
Chair of the Department of Physics

Approved: _____

George H. Watson, Ph.D.
Dean of the College of Art and Science

Approved: _____

James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Henry R. Glyde, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Michael K. Crawford, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Yi Ji, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Barry Walker, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Edward R. Lyman, Ph.D.

Member of dissertation committee

ACKNOWLEDGEMENTS

I would first like to gratefully acknowledge the enthusiastic supervision of Prof. Dr. Henry R. Glyde during my time in University of Delaware. His integrated view on science made my experience rich and fruitful. This research project would not have been possible without the support of him.

The molecular dynamics (MD) simulations in this thesis were done using the cluster of Center of Molecular Biophysics in Oak Ridge National Laboratory (ORNL). I would like to thank Prof. Dr. Jeremy C. Smith and Dr. Liang Hong at ORNL for their great help on the MD simulations. Without help from them, this work can not be done.

It is a pleasure to acknowledge valuable discussions with Mark Johnson, Giuseppe Zaccai, Dominique Bicout, Jeremy C. Smith, Liang Hong, and Edward Lyman. This work was supported by the DOE, Office of Basic Energy Sciences, under contract No. ER46680.

I would like to sincerely thank all the faculty members of the Department of Physics and Astronomy for their help and encouragement.

I would also like to thank my parents, and brother. They were always supporting me and encouraging me with their best wishes.

Finally, I would like to thank my friends for their understandings and supports on me in completing this thesis.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
ABSTRACT	xix
 Chapter	
1 INTRODUCTION	1
1.0.1 Neutron Scattering	12
2 INTRINSIC MEAN SQUARE DISPLACEMENTS IN PROTEINS	17
2.1 The Resolution Broadened Incoherent DSF $S_R(\mathbf{Q}, \omega)$	18
2.2 The Model for the Incoherent Resolution Broadened DSF $S_R(\mathbf{Q}, \omega = 0)$	20
2.3 Results	25
2.3.1 The Application of the model of $S_R^N(\mathbf{Q}, \omega = 0)$	25
2.3.2 The Application of the Expression of $\langle r^2 \rangle_R$ and Dynamical Transition	33
2.3.3 Sensitivity of $\langle r^2 \rangle$ to the model $C(t)$	37
3 LONG-TIME MEAN SQUARE DISPLACEMENTS IN PROTEINS	40
3.1 MD Simulation and MD Simulation System	40
3.2 The model of the ISF $I(Q, t)$	43
3.3 Application of the model of the ISF $I(\mathbf{Q}, t)$	44
3.3.1 100 ns MD simulation	44

3.3.2	1 μ s MD simulation	48
3.4	Mean Square Displacement	52
3.4.1	Simulated Mean Square Displacement	52
3.4.1.1	Center of mass (CM) effect on simulated data	56
3.4.2	Resolution Broadened Mean Square Displacement $\langle r^2 \rangle_R$	58
3.4.2.1	Dynamical transition	63
3.5	Sensitivity of the model to $C(t)$	65
3.6	Discussion	66
4	MOTIONAL DISPLACEMENT IN PROTEINS, THE ORIGIN OF Q DEPENDENT VALUES	67
4.1	Possible origins of Q-dependent MSD	69
4.2	The impact of the Gaussian approximation for all H in proteins	72
4.3	The impact of the Gaussian approximation and dynamical diversity, heterogeneity for an individual H in proteins	75
4.4	Discussion	80
5	VIBRATIONAL DYNAMICS OF HYDROGEN IN PROTEINS	82
5.1	Dynamical Model	83
5.2	Potential Models	85
5.2.1	Harmonic Potentials	85
5.2.2	Gaussian Potentials	87
5.2.3	Symmetric Potentials	89
5.2.4	Asymmetric Potentials	91
5.3	Suggested Potential Models in Literature	95
5.3.1	Two State Model	95
5.3.2	Potential for H	96
5.4	Potentials in Proteins	98
5.5	Discussion	103
6	CONCLUSION	104

BIBLIOGRAPHY	108
------------------------	-----

Appendix

A THE OBSERVATION TIME τ_R IN THE INSTRUMENTS IN NEUTRON SCATTERING	117
B INTERMEDIATE SCATTERING FUNCTION AT INFINITE TIME LIMIT	119
C THE MODEL OF $S_R^N(\mathbf{Q}, \omega = 0)$ FOR GAUSSIAN CORRELATION AND RESOLUTION FUNCTIONS	122
D COMPARISON OF INTRINSIC MSD AND SIMULATED MSD	124
E THE IMPORTANCE OF FOURTH CUMULANT TERM	125
F CALCULATING THE FULL ISF AND GAUSSIAN ISF	126
G THE SELF CONSISTENT HARMONIC THEORY	128

LIST OF TABLES

2.1	The energy resolution width, W in units μeV and THz , the observation time, $\tau_R = (8 \ln 2)^{1/2} \hbar / W$ and $\tau_R \propto W^{-1}$ of neutron scattering instruments.	21
-----	---	----

LIST OF FIGURES

1.1	MSD of hydrogen in proteins observed on neutron scattering instruments having different energy resolution widths, W : (a) in hydrated Ribonuclease A observed by Wood et. al [105] on IN16 ($W = 1 \mu\text{eV}$) and IN5 ($W = 100 \mu\text{eV}$), (b) in hydrated heparan sulphate (HS-0.4) observed by Jasnin et. al [50] on IN16 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$) and IN6 ($W = 100 \mu\text{eV}$), and (c) in hydrated Staphylococcal Nuclease (SNase) observed by Nakagawa et. al [73] on IN10 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$), HER ($W = 100 \mu\text{eV}$) and GP-TAS ($W = 1 \text{meV}$).	5
1.2	Type of motion in proteins with some examples in different time scales	7
1.3	Upper : Elastic component ($\omega = 0$) of the dynamical structure factor $S(\mathbf{Q}, \omega = 0)$, as a function of wave vector Q of glutamate dehydrogenase observed by Daniel et al. [17] (e.g. reproduced in Becker et al. [4]). The solid lines are a guide to the eye. Lower: An MSD, obtained from fitting to a simulated value of $S(\mathbf{Q}, \omega = 0)$ in lysozyme by Calandrini et al., which shows a strong Q dependence [8]. The solid circles and open squares represent the MSD for $p = 0.1 \text{MPa}$ and $p = 300 \text{MPa}$	10
1.4	The neutron scattering model where E_i is the initial energy of neutrons, k_i is the state vector of neutrons, θ is the scattering angle of neutrons, E_f is the energy of neutrons after scattering, and k_f is the state vector of neutrons after scattering.	13
1.5	Scattering triangles for (Upper) an elastic scattering in which the neutron is deflected but does not gain or lose energy and (Lower) an inelastic scattering in which the neutron either loses energy or gains energy.	15
2.1	(a) Elastic DSF, $S_R^N(\mathbf{Q}, \omega = 0)$, observed by Paciaroni et al. [77] on IN13 (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$. (b)-(c) The best fit values of $\langle r^2 \rangle$ and λ	25

2.2	MSD, $\langle r^2 \rangle_{exp}$, in lysozyme observed Paciaroni et al. [77] on IN13 (open squares). The intrinsic $\langle r^2 \rangle$ (solid circles) are the intrinsic MSD obtained by fitting the model Eq. (2.16) to data shown in Fig. 2.1. The solid line is a guide to the eye through the intrinsic $\langle r^2 \rangle$. The $\langle r^2 \rangle_R$ (open circles) are the MSD calculated from Eq. (2.17) which should be similar to $\langle r^2 \rangle_{exp}$	26
2.3	Upper frame: Elastic DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$, observed by Wood et al. [106] on IN16 and IN5 (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$. Lower frames: the best fit values of $\langle r^2 \rangle$ and λ . The solid line is a guide to the eye to the best fit $\langle r^2 \rangle$, the same line for both instruments.	27
2.4	MSD, $\langle r^2 \rangle_{exp}$, in hydrated Ribonuclease A observed by Wood et al. [106] on IN16 and IN5 (open squares). The intrinsic MSD $\langle r^2 \rangle$ (solid circles) are obtained by fitting the model Eq. (2.16) to the data shown in Fig. 2.3. The solid line is a guide to the eye to the intrinsic MSD, the same for both instruments. The MSD $\langle r^2 \rangle_R$ (open circles) is calculated from Eq. (2.17).	28
2.5	Upper frame: Elastic DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$, of H in heparan sulphate observed by Jasnin et al. [50] (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$. Lower frames: the best fit values of the intrinsic $\langle r^2 \rangle$ and λ . The solid line is a guide to the eye for $\langle r^2 \rangle$, the same line for all three instruments.	29
2.6	MSD, $\langle r^2 \rangle_{exp}$, in hydrated HS-0.4 observed Jasnin et al. [50] on IN16, IN13 and IN6 (open squares). The intrinsic $\langle r^2 \rangle$ (solid circles) are the intrinsic MSD obtained by fitting the model Eq. (2.16) to data shown in Fig. 2.6. The solid line is a guide to the eye through the intrinsic $\langle r^2 \rangle$, the same for all instruments. The $\langle r^2 \rangle_R$ (open circles) are the MSD calculated from Eq. (2.17) which should be similar to $\langle r^2 \rangle_{exp}$	30
2.7	Upper frame: Elastic DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$, of H in Staphylococcal Nuclease (SNase) by Nakagawa et al. [73] (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$	31
2.8	The best fit values of the intrinsic $\langle r^2 \rangle$ and λ obtained from the fits in Fig. 2.7.	32

2.9	MSD, $\langle r^2 \rangle_{exp}$, of the hydrated SNase observed by Nakagawa et al. [73] on IN10 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$), HER ($W = 100 \mu\text{eV}$) and GP-TAS ($W = 1000 \mu\text{eV}$) (open squares). The intrinsic MSD $\langle r^2 \rangle$ (solid circle) is obtained the fit in Fig. 2.7. The solid line is a guide to the eye to the $\langle r^2 \rangle$, the same for all instruments. The $\langle r^2 \rangle_R$ (open circles) is calculated from Eq. (2.17).	33
2.10	(LHS) MSD, $\langle r^2 \rangle_{exp}$, of glutamate dehydrogenase in CD_3OD/D_2O observed by Daniel et. al [17] on IN16 ($W = 1 \mu\text{eV}$) (triangles) and IN6 ($W = 100 \mu\text{eV}$) (squares). Eq. (2.17) for $\langle r^2 \rangle_R$ is fitted to the observed $\langle r^2 \rangle_{exp}$ and the fitted $\langle r^2 \rangle_R$ is shown as a solid line. (RHS) The intrinsic $\langle r^2 \rangle$ and λ obtained from the fit. The extracted intrinsic $\langle r^2 \rangle$ decreases above $T = 220 \text{ K}$	35
2.11	The MSD as in Fig. 2.9. In this case the intrinsic MSD $\langle r^2 \rangle$ is arbitrarily held constant at temperatures $T > 200\text{K}$. The solid lines show the values of $\langle r^2 \rangle_R$ obtained for IN16 and IN6 when the $\langle r^2 \rangle$ is held constant for $T > 200\text{K}$. The expected $\langle r^2 \rangle_R$ continues to increase above 200 K.	35
2.12	The resolution broadened MSD, $\langle r^2 \rangle_R$, for different resolution widths, W obtained from Eq.(2.17) illustrating the dependence of the apparent T_D on W . The intrinsic $\langle r^2 \rangle$ and λ of glutamate dehydrogenase, obtained from the fits to data by Daniel et al. [17] shown in Fig. 2.10, are used in Eq.(2.17).	36
2.13	MSD, $\langle r^2 \rangle_{exp}$, of the hydrated SNase observed by Nakagawa et al. [73] on IN10 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$), HER ($W = 100 \mu\text{eV}$) and GP-TAS ($W = 1000 \mu\text{eV}$)(open squares). The intrinsic MSD $\langle r^2 \rangle$ (solid circle) is obtained by fitting the observed DSF equation, where the Gaussian $C(t) = \exp(-\lambda^2 t^2/2)$ and $R(\omega) = \exp(-\omega^2/2W^2)/\sqrt{2\pi W^2}$ are used, to the elastic intensity data observed Nakagawa et al. [73]. The solid line is a guide to the eye to the $\langle r^2 \rangle$, the same for all instruments. The $\langle r^2 \rangle_R$ (open circles) is calculated from the Q^2 dependence of the observed DSF.	38
3.1	Two lysozyme molecules of random relative orientation selected by GROMACS.	41

3.2	The intermediate scattering function (ISF), $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 1$ ns of hydrated lysozyme ($h = 0.4$) obtained from a 100 ns MD simulation (open red circles) and fits of the model $I(\mathbf{Q}, t)$ in Eq. (3.1) (blue solid lines) to the $I_{inc}(\mathbf{Q}, t)$ at 100 K, 150 K, 200 K, 250 K and 300 K. From top to bottom, Q : 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4 and 1.6 \AA^{-1}	45
3.3	Parameters of the model $I(\mathbf{Q}, t)$ of Eq. (3.1) obtained from fits of the model to the simulations shown in Fig. 3.2: (a) The intrinsic MSD, $\langle r^2 \rangle$ and (b) the relaxation parameter, λ , versus Q at temperatures 100 K to 300 K.	46
3.4	As Fig. 3.3 for (a) the relaxation parameter λ and (b) the stretched exponential parameter β versus temperature. From bottom to top, Q : 0.4, 0.6, 0.8, 1, 1.2, 1.4 and 1.6 \AA^{-1}	47
3.5	The intrinsic MSD, $\langle r^2 \rangle$, defined in Eq. (2.11), obtained from fits of the model $I(\mathbf{Q}, t)$, Eq. (3.1) to simulations of $I_{inc}(\mathbf{Q}, t)$ at $Q = 0.2 \text{ \AA}^{-1}$. The intrinsic $\langle r^2 \rangle$ shows a break in slope at $T \simeq 140$ K and $T \simeq 220$ K.	48
3.6	The calculated $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 10$ ns of hydrated lysozyme obtained from a $1 \mu\text{s}$ MD simulation (open circles) and fits of the model $I(\mathbf{Q}, t)$ in Eq. (3.1) to the data (solid lines) at 100 K, 150 K, 200 K, 250 K and 300 K. From top to bottom, Q : 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4 and 1.6 \AA^{-1}	49
3.7	Parameters of the model ISF, Eq. (3.1), obtained from the fits of $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 10$ ns shown in Fig. 3.6: (a) the intrinsic MSD, $\langle r^2 \rangle$, and (b) the relaxation parameter, λ , versus Q at five temperatures: 100 K, 150 K, 200 K, 250 K and 300 K.	50
3.8	Parameters of the model ISF, Eq. (3.1), obtained from the fits of $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 10$ ns shown in Fig. (3.6): (a) the relaxation parameter, λ , and (b) the stretched exponential parameter, β , versus temperatures. The Q values from bottom to top is 0.4, 0.6, 0.8, 1.0, 1.2, 1.4 and 1.6 \AA^{-1}	50
3.9	The intrinsic MSD $\langle r^2 \rangle$ versus temperature obtained from fits to the $I_{inc}(\mathbf{Q}, t)$ at $Q = 0.2 \text{ \AA}^{-1}$ obtained from (1) 100 ns (solid circles) and (2) $1 \mu\text{s}$ MD simulations (solid squares). The $\langle r^2 \rangle$ is largely independent of the simulation time fitted.	51

3.10	The MSD $\Delta^2(t)/2$ defined in Eq. (3.2) of non-exchangeable hydrogen versus time at five different temperatures, 100 K to 300 K (a) for 100 ns MD simulation (b) 1 μ s MD simulation. The dots are the corresponding intrinsic MSD $\langle r^2 \rangle$ at each temperature.	53
3.11	Comparison of the intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ and the MSD $\Delta^2(t)/2$ out to times (a) 0.01, 0.1, 0.5 and 1 ns obtained from the 100 ns MD simulation and (b) out to times 0.1, 1, 5 and 10 ns, obtained from the 1 μ s MD simulation.	54
3.12	The MSD $\Delta^2(t)/2$ of non-exchangeable hydrogen versus time up to 1 ns as calculated from MD simulations of 100 ns (red solid squares) and 1 μ s (blue solid triangles) at (a) 300 K and (b) 200 K.	55
3.13	The incoherent ISF $I(\mathbf{Q}, t)$ with and without the CM motion subtracted after $t = 10$ ns at 300 K and at $Q = 1.6 \text{ \AA}^{-1}$, obtained from 1 μ s MD simulation.	56
3.14	The MSD $\Delta^2(t)/2$, with and without the CM motion subtracted after $t = 1$ ns at 300 K, (a) 100 ns MD simulation and (b) 1 μ s MD simulation.	57
3.15	The MSD $\Delta^2(t)/2$, with and without the CM motion subtracted after $t = 1$ ns (100 ns MD simulation) and $t = 10$ ns (1 μ s MD simulation).	58
3.16	(Upper) The Gaussian resolution function $R(t) = \exp(-t^2/2\tau_R^2)$ for $\tau_R = 15, 5, 1.5, 0.15, 0.015$ ns corresponding to $W = 0.1, 0.3, 1, 10, 100 \mu\text{eV}$. (Lower) $S_R(\mathbf{Q}, t) = I(\mathbf{Q}, t) \times R(t)$ is shown in the lower frame in Fig. 3.16 for $W = 0.1, 1, 100 \mu\text{eV}$ at 300 K.	59
3.17	$S_R(\mathbf{Q}, \omega = 0)$ for $W = 0.1, 1, 100 \mu\text{eV}$ at 100, 150, 200, 250 and 300 K.	60
3.18	The present intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ and resolution broadened MSD $\langle r^2 \rangle_R$ calculated from the same model for the energy resolution widths, $W, 0.1, 0.3, 1, 10$ and $100 \mu\text{eV}$	61
3.19	The present intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ (obtained from fits to (1) 100 ns (solid squares) and (2) 1 μ s MD simulations (solid circles)) and the present resolution broadened MSD $\langle r^2 \rangle_R$ ($W = 1 \mu\text{eV}$) (open triangles) for lysozyme at $h = 0.40$ compared with simulated $\langle r^2 \rangle_R$ at $W = 1 \mu\text{eV}$ by Roh et al. 2006 [87] for lysozyme at $h = 0.43$ (solid triangles).	62

3.20	The present intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ for lysozyme ($h = 0.4$), as in Fig. 3.13, compared with the experimental resolution broadened MSD $\langle r^2 \rangle_{exp}$ for $W = 1 \text{ \mu eV}$ for lysozyme at different hydration levels (h) observed by Roh et al. 2006 [87].	63
3.21	The intrinsic MSD ($W = 0$) and resolution broadened MSD $\langle r^2 \rangle_R$ for $W = 1 \text{ \mu eV}$ and $W = 100 \text{ \mu eV}$ with T_D identified. The MSD are obtained from fits to experiment in Ref. [102] using a model $I(\mathbf{Q}, t)$ that has a simple exponential decay function $C(t) = \exp[-\lambda t]$.	64
3.22	Comparison of relaxation functions: a simple exponential (blue dashed dotted line), a stretched exponential with $\beta = 0.5$ (green dashed line), a stretched exponential with $\beta = 0.21$ (black dotted line) and the Mittag-Leffler function (red solid line).	65
4.1	LHS : $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ calculated for all H in lysozyme for $0 < t < 1$ ns obtained from a 100 ns MD simulation at 300 K. Comparison of the Full $I_{inc}(\mathbf{Q}, t)$ (red solid lines) containing all cumulants and the Gaussian $I_{iG}(\mathbf{Q}, t)$ (blue solid lines) containing only second order cumulant term. RHS : The percentage of the difference between the Full $I_{inc}(\mathbf{Q}, t)$ and the Gaussian $I_{iG}(\mathbf{Q}, t)$. From top to bottom on LHS and from bottom to top on RHS, Q values are 0.4, 0.8, 1.2, 1.6 and 2 \AA^{-1}	73
4.2	(Upper) The fit of $I(Q, t)$ to (a) $I_{inc}(\mathbf{Q}, t)$ and (b) $I_{iG}(\mathbf{Q}, t)$ for $0 < t < 1$ ns and for all H with $\beta = 0.24$. (Lower)(c) The intrinsic MSD $\langle r^2 \rangle$ and (b)relaxation parameter λ . (Blue solid circles) the parameter obtained from the fit in (a). (Red solid square) for the fit in (b). The solid lines are a guide to the eye.	74
4.3	Chosen hydrogens in two lysozyme molecules.	75
4.4	(Upper) The comparison of (red solid line) $I_{inc}(\mathbf{Q}, t)$ and (blue solid line) $I_{iG}(\mathbf{Q}, t)$. (Lower) The difference between $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$. H in ILE55 and VAL109.	76
4.5	(Upper) The parameters, (LHS) $\langle r^2 \rangle$ and (RHS) λ , obtained from fits to the $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ in Fig. 4.4a. The MSD remains Q dependent when obtained from a fit to the Gaussian approximation $I_{iG}(Q, t)$. The blue solid circles represent the parameters obtained from fits to the $I_{inc}(\mathbf{Q}, t)$. The blue solid squares represent the parameters obtained from fits to the $I_{iG}(\mathbf{Q}, t)$. (Lower) Same as Upper frame for $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ in Fig. 4.4b.	77

- 4.6 (LHS) A Q independent $\langle r^2 \rangle$ is obtained from simulations of a single H (no dynamical diversity). The solid circle is H in ALA122, the solid diamond is H in THR40 and the open star is H in LYS33. RHS : If the $\langle r^2 \rangle$ of a single H is large enough some Q dependence can remain but there are few such H . The solid square is H in LYS116 and the solid triangle is H in VAL109. 78
- 4.7 The kurtosis term γ along x (red open squares), y (open black squares) and z (blue open triangle) for a single H in lysozyme for $0 < t < 1$ ns obtained from a 100 ns MD simulation at 300 K. H in ILE55, THR40, ARG68 and VAL109. 79
- 4.8 LHS : The $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ to the Full $I_{inc}(\mathbf{Q}, t)$ (solid squares) and the Gaussian approximation $I_{iG}(\mathbf{Q}, t)$ (solid circles) for an individual H in ALA122 and THR40. RHS : The distribution of the MSD of individual H in lysozyme calculated using $\Delta^2(t)/2 = \langle (r(t) - r(0))^2 \rangle / 2$ at time $t = 1$ ns at 300 K. The MSD for a single H , $\Delta^2(t)/2 = \langle (r(t) - r(0))^2 \rangle / 2$, out to time 1 ns is calculated from a 100 ns MD simulation. 80
- 5.1 (a) The mean square vibrational displacement, (MSD), $\langle u^2 \rangle$, given by (Eq. (5.7)) of an atom in a harmonic well. The slope of $\langle u^2 \rangle$ vs. T in the classical limit, $T \geq \theta_E$, is given by $\langle u^2 \rangle = k_B T / \phi_L$. We choose the low temperature force constant ϕ_L in our model so that it reproduces the observed slope of myoglobin at low temperature, i.e. $r_0^2 = \langle r^2 \rangle_{exp} = k_B T_0 / \phi_L$ where $u_0^2 = 0.1 \text{ \AA}^2$ is the observed MSD at temperature $T_0 = 240$ K. (b) The vibrational distribution $\rho(x)$ in harmonic well $v(x) = V(u) / k_B T_0$, $x = u / u_0$ at three temperatures. 86
- 5.2 (a) The MSD, $\langle u^2 \rangle$, of a particle in a Gaussian well, $V(x)$, where $x = u / u_0$. The slope of $\langle u^2 \rangle$ with temperature shows a marked increase at a $T \sim 250$ K simulating a dynamical transition at $T_D \sim 250$ K. (b) The vibrational distribution $\rho(x)$ in the Gaussian well at three temperatures. 87
- 5.3 (a) The calculated MSD $\langle u^2 \rangle$ for a Gaussian potential well (solid line). The low temperature force constant ϕ_L of the Gaussian is set to reproduce the low temperature slope of $\langle u^2 \rangle$ vs. T in purple membrane (PM). The calculated $\langle u^2 \rangle$ at higher temperature for the Gaussian $V(u)$ reproduces the observed $\langle u^2 \rangle$ for PM (dashed-dotted line) [106] well at higher temperature ($T \gtrsim 200$ K). The vibrational distribution $\rho(x)$ is shown in part (b). 88

5.4	(a) The $\langle u^2 \rangle$ in a potential well composed of two harmonic components characterized by force constants ϕ_L and ϕ_H . The slope crosses over gradually from a low (L) temperature to high (H) temperature value. (b) The $\rho(x)$ for a potential composed of two harmonic components.	90
5.5	(a)The $\langle u^2 \rangle$ for an atom in a potential well composed of two harmonic components with a barrier separating the two components (see inset). (b) The $\rho(x)$ for a potential composed of two harmonic components with a barrier.	90
5.6	(a) The MSD, $\langle u^2 \rangle$, for a double well potential. The height of the barrier (at $x_B = 0.75$) between the two wells (see inset) is relatively small, $V(x_B)/k_B T_0 \simeq 0.2$. (b) The $\rho(x)$ for a particle in a double well potential. The center of $\rho(x)$, Δ moves from one well ($\Delta = 0$) at $T = 0$ K to the midpoint between the two wells ($\Delta = x_B$) at a relatively low temperature. The slope at high temperature is set by the harmonic well force constant (H) at larger x	92
5.7	(a) The MSD $\langle u^2 \rangle$ for a well that has a hard wall component. The $\langle u^2 \rangle$ vs. T shows a large increase in slope at $T \simeq 250$ K, an increase arising from the center of $\rho(x)$ moving rapidly away from the wall as T increases as shown in part (b).	93
5.8	(a)The MSD $\langle u^2 \rangle$ for a well that has a hard wall component. The well parameters (ϕ_L , ϕ_H , barrier height and well center) are adjusted to reproduce the $\langle u^2 \rangle$ vs. T observed for myoglobin by Doster et al. at low and high temperature. The observed values of Doster et al. [25] as presented in Ref.[6]. The calculated $\langle u^2 \rangle$ clearly reproduce the observed $\langle u^2 \rangle$ for myoglobin well.(b)The distribution function $\rho(x)$ versus $x = u/u_0$	94
5.9	(a) The MSD, $\langle u^2 \rangle$, for a particle in the two state potential well. (b) The $\rho(x)$ for a particle in the two state potential.	96
5.10	(a)Comparison of the potential of H obtained from numerical data and modelled potential function. (b) The unitless potential $V(x)/k_B T_0$	97
5.11	(a) The MSD, $\langle u^2 \rangle$, for a particle in the potential on RHS. (b) The vibrational distribution $\rho(x)$	98

5.12	<i>H</i> -bond potential $V(r)$ in a formamide dimer versus H^+ to acceptor (O^-) separation, r , from Fig. 4 of Morozov and Kortemme [72]. The potential is shown in units $x = r/u_0$ where $u_0 = 0.33 \text{ \AA}$ and $V(x)/k_B T_0$ where $T_0 = 240 \text{ K}$ and $k_B T_0 = 0.476 \text{ kcal/mole}$. Also shown is the present Gaussian model potential in these units. . . .	100
5.13	Coarse-Grained potential $V(r)$ between two TRP3 amino acids in a protein from Fig. 3 of Basdevant et al. [2]. The potential is in units of $x = r/r_0$ and $V(x)/k_B T_0$ as in Fig. 5.12 and compared with the present Gaussian model potential.	101
5.14	Coarse-Grained potential shown in Fig. 5.13 compared with the present Hard Wall model potential.	102
F.1	The comparison of the ISF, $I_{inc}(\mathbf{Q}, t)$, for all H in lysozyme (red solid lines) calculated from Eq. (F.1) and $I_{inc}(\mathbf{Q}, t)$, for whole lysozyme (black dashed lines) calculated by using SASENA.	126

ABSTRACT

Neutron scattering experiments and molecular dynamics simulations are the most effective tools to explore the dynamics of hydrogen in proteins. The mean square displacement (MSD) of hydrogen (H) in proteins has been extensively measured using neutron scattering and calculated using molecular dynamics simulations. A small MSD is observed at low temperatures and the slope of the MSD significantly increases at a specific temperature T_D . This increase in the slope of the MSD is identified as a dynamical transition, and the temperature it takes place at a specific temperature which is denoted a dynamical transition temperature T_D . The observed MSD in neutron scattering experiments depend on the energy resolution of the instrument.

In this thesis, we first focus on the resolution dependent of the observed MSD $\langle r^2 \rangle_{exp}$ in neutron scattering experiments. We propose a method for obtaining the intrinsic MSD $\langle r^2 \rangle$ of H , which is independent of the resolution of the instrument employed, in proteins. The intrinsic MSD is defined as the infinite time value of MSD which appears in the well-known Debye-Waller factor. In this method, a model of the resolution broadened elastic incoherent dynamic structure factor $S_R(\mathbf{Q}, \omega = 0)$ is developed to extract the intrinsic MSD from the resolution dependent data. The model contains the intrinsic MSD, the instrument resolution width and a relaxation frequency characterizing the motions of H in proteins. The model of $S_R(\mathbf{Q}, \omega = 0)$ is fitted to the resolution broadened DSF data already published in the recent literature and the intrinsic MSD in three proteins was successfully obtained.

Later, we constructed a model for the incoherent intermediate scattering function $I(\mathbf{Q}, t)$ to obtain the intrinsic, long-time MSD of H in proteins from finite time molecular dynamics simulations. In the literature, the simulated MSD increases with increasing time and does not reach a certain limiting value at even 10 ns. The infinite

time MSD, $\langle r^2 \rangle$, is the long time value of the simulated MSD. The model $I(\mathbf{Q}, t)$ fits to the simulated $I_{inc}(\mathbf{Q}, t)$ to obtain the intrinsic long-time MSD $\langle r^2 \rangle$ which is a parameter in the model $I(\mathbf{Q}, t)$. The intrinsic MSD $\langle r^2 \rangle$ of hydrated lysozyme powder ($h = 0.4 \text{ g water/g protein}$) over a temperature range between 100 K and 300 K obtained from data out to 1 ns and to 10 ns is found to be the same. The intrinsic $\langle r^2 \rangle$ is approximately twice the value of the MSD that is reached in simulations after times of 1 ns. The simulated MSD in 1 ns corresponds to the observed MSD measured by neutron scattering instruments having $1 \mu\text{eV}$ energy resolution width.

The observed MSD $\langle r^2 \rangle_{exp}$ extracted from an elastic dynamical structure factor measured in neutron scattering experiments is also found to be Q-dependent. In the second part of this thesis, we analyse the possible origins of the Q-dependence of the observed MSD $\langle r^2 \rangle_{exp}$. We show that this dependence does not arise from the Gaussian approximation, commonly used in the analysis of the neutron scattering data. The Q-dependence of the MSD is the artificial consequence of neglecting the dynamical diversity in the model that are used to analyse the data.

Finally, we illustrate the dynamical transition in proteins by reproducing the change in the slope of MSD versus temperature at T_D within a simple model of vibration, a single particle in an anharmonic potential. Using Self-Consistent-Harmonic theory, we investigate dynamics of a particle in different potential models. A simple Gaussian potential or a potential containing a hard wall and a soft wall is particularly effective to reproduce the change in the slope of MSD versus temperature. The MSD data of myoglobin and purple membrane in the literature is reproduced well using a potential containing hard wall and a Gaussian potential.

Chapter 1

INTRODUCTION

Proteins are large biological molecules comprised of twenty species of amino acids. Their study is a major research area for physics, chemistry and biology. Each protein has a unique tertiary structure, which can be determined using X-ray crystallography and nuclear magnetic resonance (NMR). In the last decades, many researches have been devoted to the dynamics of proteins which is of essential interest in biophysics. It is commonly accepted that the function of proteins is related to their three-dimensional structure and dynamics. The internal motions in proteins is essential for their dynamics [25, 35, 65, 91]. The dynamics of proteins is investigated as a function of temperature, pressure and hydration levels and it is strongly affected by the viscosity of the solvent [11, 25, 33, 34, 76, 77, 79, 83, 100, 110]. Relaxation processes which describe the internal motion in biomolecules have been analyzed using NMR, Mossbauer spectroscopy, neutron scattering, and molecular dynamics (MD) simulations [8, 9, 16, 36, 81, 89]. Despite significant efforts in experimental and computational biophysics, our knowledge of dynamics of proteins remains limited even on a qualitative level.

Incoherent neutron scattering and MD simulation are useful and effective techniques for studying the dynamics of proteins in the picosecond to nanosecond time scale [8, 9, 21, 22, 25, 26, 33, 43, 42, 41, 40, 50, 73, 90, 91, 88, 109]. The mean square displacement (MSD) in proteins is a fundamental dynamical quantity which is commonly investigated in neutron scattering experiments [20, 25, 26, 23, 36, 66, 109]. The average MSD in proteins is typically determined from the elastic component of the incoherent dynamic structure factor (DSF). At low temperatures, a protein is essentially harmonic and a small MSD is observed. The MSD sharply increases at a specific

temperature T_D . The marked increase in the slope of the MSD taking place at T_D is denoted as the dynamical transition, and T_D as the dynamical transition temperature [3, 6, 7, 25, 26, 33, 35, 47, 54, 84, 88, 92, 109, 111]. Above T_D , additional motions appear, and a large MSD is observed. The large MSD suggests that proteins are more flexible and that nuclei in proteins sample new environments, for example; hydrogens make contact with different parts of a protein at large distances. The significant increase in MSD at T_D is related to the onset of function in proteins. The dynamical transition in myoglobin at around 200 K was first observed using Mössbauer spectroscopy by Parak *et al* [80]. Since then, it has also been observed by neutron scattering for several hydrated protein powders [25, 77, 88, 87, 50] and membranes [33, 106]. Several studies have investigated the relation between dynamics of proteins and enzymatic activity in proteins [16, 25, 33, 35, 79, 109]. It is strongly believed that the onset of biological activity in proteins takes coincides with the dynamical transition at T_D [7, 25, 33, 35, 76, 79, 84, 109, 111]. However, some exceptions have been observed [18, 27]. It is noted that the dynamics of proteins has been found to depend on the enzyme and its catalytic mechanism [33, 84].

The increase in the slope of the MSD at T_D has been observed by experiments and predicted by computer simulations for hydrated proteins, DNA and RNA [10, 25, 33, 34, 76, 79, 81, 99, 100]. Typical examples are hydrated lysozyme [54], myoglobin [25] and purple membrane bacteriorhodopsin [88, 107]. The impact of the hydration water on the dynamics of proteins is significant. These studies have revealed that MSD varies linearly with temperature up to 300 – 350 K in dry proteins and up to 200 – 230 K with a dynamical transition at $T \simeq 200$ K in hydrated proteins [16, 33, 94, 100]. The possible explanations for the dynamical transition in proteins have been investigated in the literature. The sudden change in the slope of the MSD at T_D has been interpreted using several different models, jumping between minima [25, 26, 28, 35, 60, 93], including diffusion [25, 53, 56] and “effective force constant” [6, 109]. According to these models, the dynamical transition in MSD at T_D has been attributed to a “glass” transition in the protein [25, 26, 93], to the onset of thermally

activated transitions and diffusion that modify the trapping cages [25, 53] and to a change in the “effective elasticity” of the protein [109]. Notwithstanding, many studies have broadly expressed the view that the dynamical transition of protein is strongly coupled to the onset of translational motions of hydration water [98, 99, 100]; there is an ongoing discussion about the dynamical transition [3, 13, 15, 17, 31, 32, 54]. As discussed in these studies, two possible origins of the dynamical transition in proteins are (1) a consequence of the coupling of the dynamics of water to the dynamics of the proteins [13, 15, 58] and (2) a consequence of a “time-window” effect which is defined an artificial result of the energy resolution of the employed instrument [3, 18, 17, 31, 32, 54]. In the latter, it is believed that the dynamical transition does not arise from any intrinsic biological transition in proteins.

In neutron scattering experiments, the observed data $S_R^{exp}(\mathbf{Q}, \omega)$ is the elastic resolution broadened DSF. The scattering cross section of the hydrogen nucleus is larger than that of other nuclei in proteins. Hence the main contribution to the observed DSF data $S_R^{exp}(\mathbf{Q}, \omega)$ comes from hydrogens in proteins. Hydrogens in proteins effectively reflect the global motions of whole protein in the time and length scale of the experiments. In the data analyses, the observed MSD $\langle r^2 \rangle_{exp}$ is calculated from the slope of the observed DSF $S_R^{exp}(\mathbf{Q}, \omega)$

$$\langle r^2 \rangle_{exp} = -3 \frac{d \ln S_R^{exp}(\mathbf{Q}, \omega = 0)}{dQ^2}. \quad (1.1)$$

The observed $S_R^{exp}(\mathbf{Q}, \omega)$ only includes motions up to $\tau_R \sim 1/W$ where W is the energy resolution width of the instrument employed. Hence the observed MSD $\langle r^2 \rangle_{exp}$ depends on the energy resolution width of the instrument used to measure the observed DSF $S_R^{exp}(\mathbf{Q}, \omega)$. Different MSDs $\langle r^2 \rangle_{exp}$ are extracted from $S_R^{exp}(\mathbf{Q}, \omega)$ observed on different instruments. In order to illustrate the impact of the energy resolution, some groups have measured the MSD of the same protein under the same conditions using neutron scattering instruments having different energy resolutions [50, 73, 106]. The $\langle r^2 \rangle_{exp}$ is the MSD including motions over a limited time, $0 < t < \tau_R$, where τ_R is set by the width of the energy resolution of instrument. Hence, a high energy resolution corresponding to

a small energy resolution width is needed to observe all the motions that contribute to the MSD, including the slow, long time motions. Neutron scattering instruments which are typically used at present have an energy resolution in the range $100 \mu\text{eV} > W > 1 \mu\text{eV}$ which corresponds to the evolution times $10 \text{ ps} < \tau_R < 1 \text{ ns}$. Over this time range (up to τ_R), the observed $\langle r^2 \rangle_{exp}$ is still increasing. This indicates that the intrinsic, long time ($\tau_R \rightarrow \infty$) value of $\langle r^2 \rangle$ has not been observed [16, 50, 73, 106]. The dependence of the observed MSD $\langle r^2 \rangle_{exp}$ on the energy resolution width of the instrument is shown in Fig. 1.1. The apparent temperature T_D of the dynamical transition also depends on the resolution width of the employed instrument. The smaller the energy resolution width, the larger the observed MSD and the lower the apparent dynamical transition temperature T_D . A large MSD including the contribution of all the motions is observed with an instrument which has a high energy resolution. $W = 1 \mu\text{eV}$ is the smallest energy resolution width available.

Our goal in this thesis is to clarify what is measured in neutron scattering experiments. We also aim to obtain instrument independent MSD values that are intrinsic to the protein. We present a method for analysis of the incoherent observed DSF $S_R^{exp}(\mathbf{Q}, \omega = 0)$ and for extracting the intrinsic MSD of hydrogen in the proteins from the resolution dependent data in Chapter 2 [102]. This method includes a fit of the model of the resolution broadened DSF $S_R(\mathbf{Q}, \omega = 0)$ to the observed DSF $S_R^{exp}(\mathbf{Q}, \omega = 0)$. We constructed a simple model of the resolution broadened DSF $S_R(\mathbf{Q}, \omega = 0)$ which corresponds to the measured data in neutron scattering experiments. The model of the $S_R(\mathbf{Q}, \omega = 0)$ consists of the intrinsic MSD $\langle r^2 \rangle$, the relaxation frequency λ , which describes the motional processes, and the energy resolution width W . The most significant advantage of this method is that it allows us to obtain the intrinsic MSD which is independent of the energy resolution width, W of the neutron scattering instrument. The intrinsic MSD $\langle r^2 \rangle$ appears in the full Debye-Waller factor, and includes motions up to $t = \infty$, $\langle r^2(t = \infty) \rangle = \langle r^2 \rangle$. It is the MSD that would be observed on an instrument with zero energy resolution width, $W \rightarrow 0$. The model of the resolution broadened DSF $S_R(\mathbf{Q}, \omega = 0)$ is fitted to the observed DSF

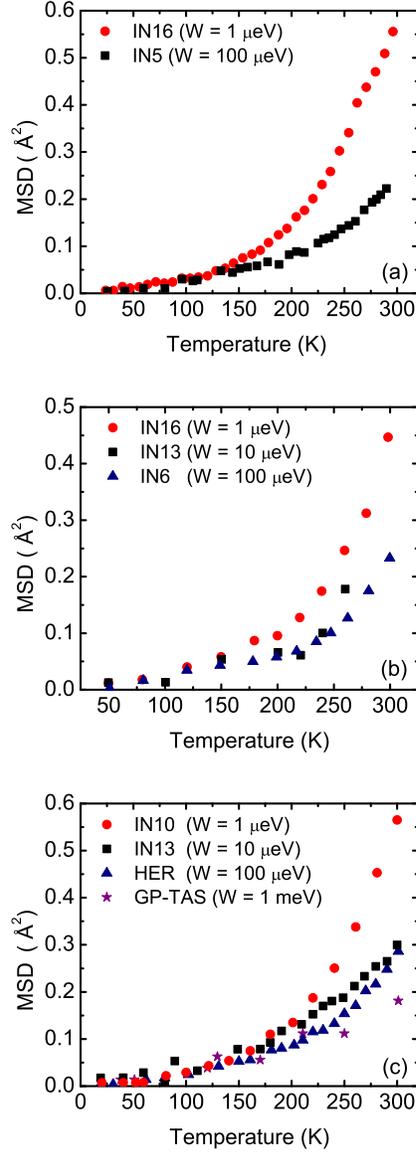


Figure 1.1: MSD of hydrogen in proteins observed on neutron scattering instruments having different energy resolution widths, W : (a) in hydrated Ribonuclease A observed by Wood et. al [105] on IN16 ($W = 1 \mu\text{eV}$) and IN5 ($W = 100 \mu\text{eV}$), (b) in hydrated heparan sulphate (HS-0.4) observed by Jasnin et. al [50] on IN16 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$) and IN6 ($W = 100 \mu\text{eV}$), and (c) in hydrated Staphylococcal Nuclease (SNase) observed by Nakagawa et. al [73] on IN10 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$), HER ($W = 100 \mu\text{eV}$) and GP-TAS ($W = 1 \text{meV}$).

data $S_R^{exp}(\mathbf{Q}, \omega = 0)$ already published in the literature to obtain the intrinsic MSD $\langle r^2 \rangle$ and the relaxation frequency λ as free fitting parameters. The observed DSF $S_R^{exp}(\mathbf{Q}, \omega = 0)$ measured by neutron scattering instruments that have different energy resolution widths is chosen from the literature [50, 73, 77, 106]. The intrinsic MSD for four proteins, lysozyme, heparan sulphate (HS-0.4), Ribonuclease A and Staphylococcal Nuclease (SNase), was successfully obtained.

An expression for the resolution broadened MSD $\langle r^2 \rangle_R$ is also obtained by substituting the model $S_R(\mathbf{Q}, \omega = 0)$ in Eq. (1.1) to clarify the impact of a finite energy resolution width of instruments on a MSD in proteins. The $\langle r^2 \rangle_R$ depends on the resolution width W and is equivalent to the observed MSD $\langle r^2 \rangle_{exp}$ in neutron scattering experiments. The intrinsic MSD $\langle r^2 \rangle$ for glutamate dehydrogenase are successfully obtained from the fit of the model $\langle r^2 \rangle_R$ to the observed MSD $\langle r^2 \rangle_{exp}$ measured by instruments having different energy resolution widths. For a finite energy resolution width, the resolution broadened MSD $\langle r^2 \rangle_R$ always lies below the intrinsic MSD $\langle r^2 \rangle$. The fitting method can be used by anyone to obtain the intrinsic MSD using their data and presents an approach to obtain uniform and consistent values of MSDs and T_D in proteins. We show that the dynamical transition does not arise from the “time-window” issue and T_D is influenced by the energy resolution of instruments. T_D shifts to the higher temperatures with increasing energy resolution width.

MD simulation is effective for studying protein dynamics and is complementary to neutron scattering. Similar to neutron scattering experiments, some MD studies have been performed to understand the dynamics of proteins and its relation as a function of temperature and pressure, and hydration [8, 9, 21, 22, 40, 63, 69, 70]. The energy barriers between conformational substates do not change with pressure, however the curvature of the local harmonic potential increases. This suggests that the vibrations in substates is pressure-dependent, but dynamical transition is not [70]. It is also noted that the dynamics of proteins slows with increasing pressure [8, 9] and with increasing viscosity of the solution [21, 22, 63].

Molecular dynamics is concerned with time dependent motions in molecular

systems. Each motion (i.e., dynamical process) has a characteristic time-scale. Proteins particularly display a broad range of characteristic motions. Fig. 1.2 shows type of motion with some examples and time scale. Local motions, such as atomic fluctuations or side chain motions, are very fast and very localized and observed in femtosecond (fs)- picosecond (ps) time scale. Medium motions, such as loop motions are observed in nanosecond (ns)- microsecond (μs) time scale. The length scale is less than 1\AA^{-1} for motions in fs - ps time scale and around $1-5\text{\AA}^{-1}$ for motions in ns - μs time scale. Slow, long-time motions such as folding or unfolding occur on the scale of the whole molecule.

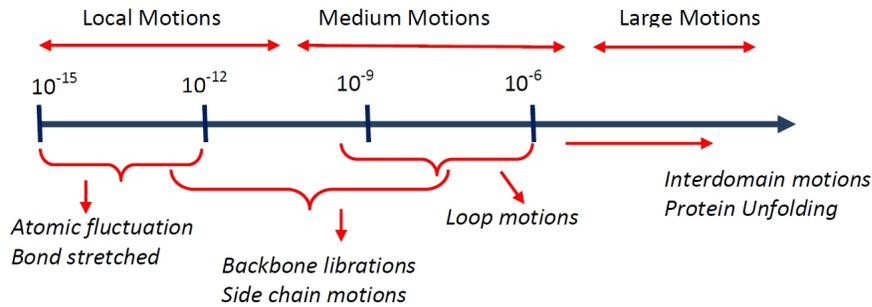


Figure 1.2: Type of motion in proteins with some examples in different time scales

The data measured in neutron scattering experiments such as the incoherent ISF, incoherent DSF and susceptibility spectrum can be reproduced well from MD simulation trajectories [54, 47, 46, 87, 91, 97, 108]. The contributions of time dependence, of non-Gaussian behavior and of dynamical heterogeneity to elastic scattering on the ps-ns timescales have been also investigated in MD studies. It is used to clarify the origin of the measured data in neutron scattering experiments [4, 3, 43, 42, 41, 71, 68, 108]. One of the important quantities calculated from MD trajectories is the average MSD of hydrogen in proteins. The MSD is proportional to the average over time of the position difference of the hydrogen in proteins between reference time t and 0, $\Delta^2(t) = \langle (r(t) - r(0))^2 \rangle$. Although these studies have made a significant contribution to the

understanding of the dynamics of proteins in different environmental conditions, simulations are limited to simulation times of no longer than 10^{-9} s. The simulated MSD $\Delta^2(t)$ is found to be strongly time-dependent and does not reach a converged value at even 1 ns [108]. In contrast to these results, it is expected that folded proteins in well-defined three-dimensional structures have a well defined long time, intrinsic MSD. That is, the internal MSD arising from internal motions in proteins should converge to a plateau as a function of time. An intriguing question has arisen as to whether simulated intrinsic, long time $\langle r^2 \rangle$ can be obtained by fitting a model to incoherent intermediate scattering function (ISF) $I_{inc}(\mathbf{Q}, t)$ calculated finite time simulation.

In MD studies, the resolution broadened MSD $\langle r^2 \rangle_R$ can be calculated in the same way as it is obtained from experiment [42, 87, 96, 106]. The resolution broadened incoherent DSF $S_R(\mathbf{Q}, \omega = 0)$ is calculated using the simulated $I_{inc}(\mathbf{Q}, t)$ and a resolution function, which could be chosen as a Gaussian function or a Lorentzian function, for an energy resolution width, W . After $S_R(\mathbf{Q}, \omega = 0)$ has been evolved for a specific time τ_R set by W , the MSD is obtained using the calculated resolution broadened $S_R(\mathbf{Q}, \omega = 0)$ in Eq. 1.1. At small Q , the resolution broadened MSD $\langle r^2 \rangle_R$ is effectively compared with the observed MSD $\langle r^2 \rangle_{exp}$. Simulated MSD obtained in this way strongly agrees with observed MSD [87, 102]. This supports the question mentioned above: Could a model be constructed to obtain the intrinsic-long time MSD $\langle r^2 \rangle$ in finite time simulations?

In this case, we modify the model of incoherent ISF $I(\mathbf{Q}, t)$ constructed in the method which is used to obtain the intrinsic MSD $\langle r^2 \rangle$ from observed DSF $S_R^{exp}(\mathbf{Q}, \omega = 0)$ in neutron scattering experiments [103]. In modified model of the $I(\mathbf{Q}, t)$, the motions in proteins are defined with a simple representation including several diffusion processes instead of a single one. Our aim in this part of thesis is to propose a method for obtaining an intrinsic, long time, $t \rightarrow \infty$, value of the MSD from finite-time simulations. The intrinsic MSD appears as a parameter in the model $I(\mathbf{Q}, t)$. We fit the model of the ISF $I(\mathbf{Q}, t)$ to the $I_{inc}(\mathbf{Q}, t)$ calculated from the finite time simulations. Explicitly, the $I_{inc}(\mathbf{Q}, t)$ observed in neutron scattering experiments and calculated

from simulations are in a good agreement each other [86]. We treat the $\langle r^2 \rangle$ in the model as a free fitting parameter. In this way we obtain an infinite time value of the MSD $\langle r^2 \rangle$ from fits to simulation data at finite t .

In Chapter 3, we obtain the intrinsic long-time MSD of lysozyme by fitting the model $I(\mathbf{Q}, t)$ to the $I_{inc}(\mathbf{Q}, t)$ calculated up to $t = 1$ ns and 10 ns from simulations in 100 ns and 1 μ s time lengths. The intrinsic MSD $\langle r^2 \rangle$ obtained from two simulations having different time scale is the same. This result is consistent with $\langle r^2 \rangle$ representing a time-independent, long time MSD. The resolution broadened $\langle r^2 \rangle_R$ is calculated using the calculated $S_R(\mathbf{Q}, \omega = 0)$ in Eq. (1.1) for different energy resolution widths. The intrinsic $\langle r^2 \rangle$ is approximately twice the resolution broadened MSD $\langle r^2 \rangle_R$ that is calculated for motions up to 1.5 ns corresponding to $W = 1$ μ eV. The intrinsic MSD $\langle r^2 \rangle$ of lysozyme has a break in the slope at 140 K due to activation of the methyl group rotations and a second dynamical transition at $T_D = 220$ K due to a dynamical transition, as has been observed and calculated for proteins experimentally. According to this model, the dynamical transitions are intrinsic properties of proteins. The transitions are not a consequence of the experimental time window, however the dynamical transition temperature T_D may be modified by it.

It is important to clarify the measurements and calculations involved with the MSD in proteins. Hydrogens in proteins which are dominant in the data measured in neutron scattering experiments has a wide distribution in proteins. The model of the elastic DSF which is commonly used to analyse the experimental data is $S_R(\mathbf{Q}, \omega = 0) = \exp(-1/3Q^2\langle r^2 \rangle)$ corresponding to the $\langle r^2 \rangle_{exp}$ in Eq. (1.1). The model $S_R(\mathbf{Q}, \omega = 0)$ is identified within the Gaussian approximation and neglects the dynamical diversity of hydrogens in proteins. The impact of the distribution of hydrogens in proteins has been investigated by using different models, including three scattering centers [29], including continuum distribution (a Weibull distribution function) [68], including a correction to fourth order in Q that can be used to extract the variance of the distribution of mean-square displacements [4, 108].

MSD in proteins determined from substituting the elastic component of the

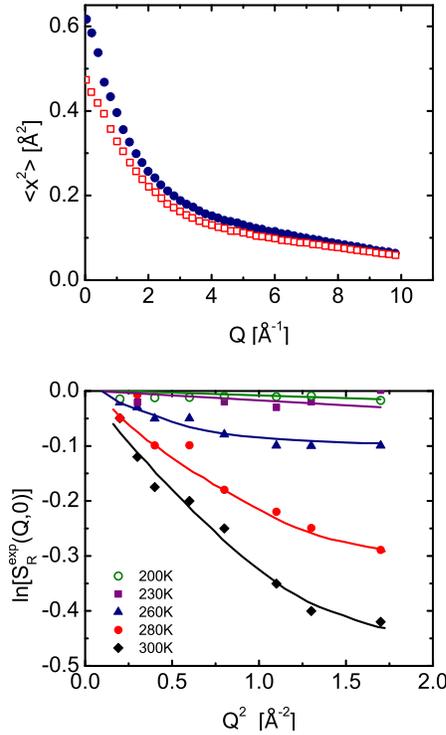


Figure 1.3: Upper : Elastic component ($\omega = 0$) of the dynamical structure factor $S(\mathbf{Q}, \omega = 0)$, as a function of wave vector Q of glutamate dehydrogenase observed by Daniel et al. [17] (e.g. reproduced in Becker et al. [4]). The solid lines are a guide to the eye. Lower: An MSD, obtained from fitting to a simulated value of $S(\mathbf{Q}, \omega = 0)$ in lysozyme by Calandrini et al., which shows a strong Q dependence [8]). The solid circles and open squares represent the MSD for $p = 0.1$ MPa and $p = 300$ MPa.

incoherent DSF $S_R^{exp}(\mathbf{Q}, \omega = 0)$ in Eq. (1.1) is found to depend on the wave vector Q . In the literature, it is generally found that $\ln S_R^{exp}(Q, \omega = 0)$ versus Q^2 is not a straight line and that the extracted MSD, even from simulations, is Q dependent [8, 25, 62, 73, 77, 78]. An example for $\ln S_R^{exp}(\mathbf{Q}, \omega = 0)$ versus Q^2 and for Q -dependent MSD are shown in Fig. 1.3. The values of the observed MSD at the low Q and high Q values are significantly different. The lower the Q -values, the larger that MSD $\langle r^2 \rangle_{exp}$ observed. It is currently believed that the Q -dependence of MSD is an artificial consequence of neglecting the heterogeneity or of using Gaussian approximation in the

analysis.

The aim, broadly stated, is to investigate possible origins of the Q -dependence of the global MSD in Chapter 4. The incoherent ISF $I_{inc}(\mathbf{Q}, t)$ is a collection of scattering from a large number of dynamically diverse scatters (H) in the protein. In the current analysis of the neutron scattering experiments, the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ is represented by a global $I(\mathbf{Q}, t)$ in which a single scattering center represents whole protein. The Q -dependence of the MSD could arise from (i) using Gaussian approximation in the model of global ISF $I(\mathbf{Q}, t)$; the whole expression of the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ includes all cumulant terms, whereas the model of $I(\mathbf{Q}, t)$ includes only the first cumulant term, and (ii) neglecting dynamical diversity, heterogeneity; whole expression of the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ is calculated as sum over hydrogens in proteins, the scattering from different hydrogens is represented by a single scatter in the model of $I(\mathbf{Q}, t)$. The Q -value is the length scale of the motions in proteins, hence the low Q values lead to observe long-distance motions. In addition to these commonly discussed two reasons, there could be an intrinsic Q dependence in the MSD, i.e. at low Q motions over a longer distance are included which leads to a larger MSD at low Q . The impact of the higher cumulants which may contribute significantly to ISF and DSF is investigated by comparing the MSD obtained by fitting the model of $I(\mathbf{Q}, t)$ to the whole expression of the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ and the incoherent ISF including only the first cumulant term, $I_{iG}(\mathbf{Q}, t)$. As stated above, the dynamical diversity studied by different groups in the literature [4, 43, 68, 74, 106, 108] makes a significant contribution to the ISF $I_{inc}(\mathbf{Q}, t)$. Neglecting the dynamical diversity in the model of $I(\mathbf{Q}, t)$ could lead to a Q -dependence. The model of the $I(\mathbf{Q}, t)$ is applied to the simulated ISF data $I_{inc}(\mathbf{Q}, t)$ for individual hydrogen to investigate the impact of the heterogeneity on the MSD.

As discussed before, the dynamical transition could have arisen with “effective elasticity” which is investigated with “effective force constant” model [6]. In this model, the increase in the slope of the MSD at T_D suggests the different force constants at low and high temperatures. In the classical limit and in a harmonic approximation,

the MSD arising from vibration is defined as $\langle u^2 \rangle = k_B T / \phi$ where k_B is the Boltzmann constant, T is the temperature and ϕ is a harmonic force constant. The MSD linearly increases with temperature at low temperatures, then the sharp increase in MSD is observed at a specific temperature T_D . This suggests that the force constant at low temperatures is large and decreases at T_D . The linear increase in MSD continues with lower force constant at higher temperatures.

In this context, a specific goal of this part of the thesis is to present a simple model to illustrate that a marked change in the slope of $\langle u^2 \rangle$ with T can arise from vibrational dynamics of a particle in an anharmonic potential well [101]. The self-consistent harmonic (SCH) theory is used to obtain the MSD of a single particle in 1D well [14, 37, 38, 57]. In SCH theory, the harmonic force constant ϕ is obtained as the second derivative of the potential averaged over the vibrational distribution of a particle in the well. We construct our model by starting from a harmonic potential to obtain the force constant at low temperatures, then apply our model to different potentials, a Gaussian potential, symmetric potentials and asymmetric potentials, to reproduce the $\langle u^2 \rangle$ versus temperature data in the literature. In a Gaussian potential well, the force constant decreases with increasing temperature and this provides that the increase in slope of MSD with T is gradual. Using a Gaussian potential model, the $\langle u^2 \rangle$ versus temperature data for purple membrane is successfully reproduced [106, 107]. In an anharmonic potential with hard wall and soft wall components, the sharp change in $\langle u^2 \rangle$ with temperature can be obtained. As the $\langle u^2 \rangle$ increases, the force constant, ϕ , decreases. This decrease leads to an increase in slope of $\langle u^2 \rangle$ with temperature at T_D as is observed. The $\langle u^2 \rangle$ versus temperature data for myoglobin is also illustrated by using an anharmonic potential with hard wall and soft wall components [101, 25]. The details of this discussion is presented in Chapter 5.

1.0.1 Neutron Scattering

Neutron Scattering is useful tool to determine the the relative positions and motions of atoms in a sample. Neutrons have no charge, and their electric dipole

moment is too small to measure. For these reasons, neutrons can penetrate matter far better than charged particles. Furthermore, neutrons interact with atoms via nuclear rather than electrical forces. Hence, neutrons can travel large distances through most materials without being scattered or absorbed.

Neutrons used in scattering experiment has the wavelength, L , which is usually between 0.1 nm and 1 nm. Often, studies have been done in terms of the neutron wavevector k which is a vector of magnitude $2\pi/L$ that points along the neutrons trajectory. The magnitude of the wavevector, k , is related to the neutron velocity, $k = 2\pi mv/h$, where h is Planck's constant and m is the mass of the neutron. The scattering of a neutron by a free nucleus can be described in terms of a cross section, σ , that is equivalent to the effective area presented by the nucleus to the passing neutron. If the neutron hits this area, it is scattered isotropically, that is, with equal probability in any direction. The scattering is isotropic because the range of the nuclear interaction between the neutron and the nucleus is tiny compared with the wavelength of the neutron, so the nucleus essentially looks like a point scatterer.

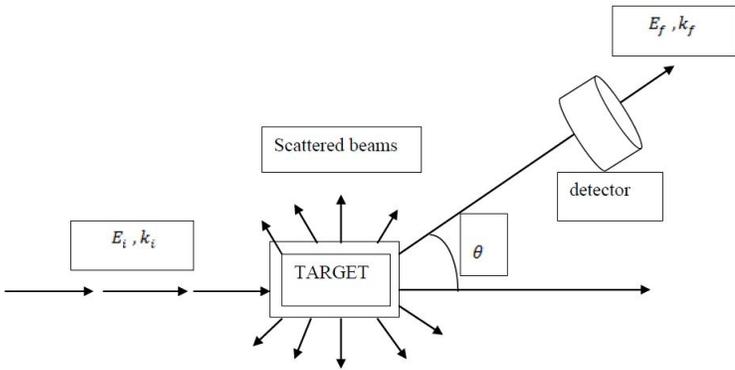


Figure 1.4: The neutron scattering model where E_i is the initial energy of neutrons, k_i is the state vector of neutrons, θ is the scattering angle of neutrons, E_f is the energy of neutrons after scattering, and k_f is the state vector of neutrons after scattering.

An idealized model of neutron scattering is shown in Fig. 1.4, where E_i is the initial energy of neutrons, k_i is the initial wave vector of neutrons, θ is the scattering angle of neutrons, E_f , is the energy of neutrons after scattering, and k_f is the wave vector of neutrons after scattering. As can be seen in Fig. 1.4, neutrons are scattered by atomic nuclei. The scattering intensity is the sum of intensities scattered by individual nuclei. It is analysed by a neutron spectrometer in terms of energy, $\hbar\omega$, and of momentum $\hbar\vec{k}$, (in terms of frequency and scattering state vector). Energy transfer ($\hbar\omega = \hbar\omega_i - \hbar\omega_f$) and momentum transfer ($\hbar\vec{Q} = \hbar\vec{k}_i - \hbar\vec{k}_f$) give information about the structural and dynamic properties of the sample (protein). The quantity $\mathbf{Q} = \mathbf{k}_i - \mathbf{k}_f$ is known as the scattering vector, and the vector relation between \mathbf{Q} , \mathbf{k}_i and \mathbf{k}_f is shown in Fig 1.5. The scattering is elastic for $\omega = 0$ and energy transfer does not occur between incident neutrons and the target. This means that incident and scattering neutrons have the same energy. For ($\omega \neq 0$), inelastic neutron scattering occurs, and there are energy transfers between incident neutrons and target. Elastic scattering changes direction but not the magnitude of the wave vector, but both direction and magnitude of the neutron wave vector in inelastic scattering.

There are two type of neutron scattering. These are coherent and incoherent neutron scattering. In coherent scattering, neutron waves scattered from different nuclei interfere with each other. This type of scattering depends on the distances between atoms and on the scattering vector \mathbf{Q} , and it thus gives information about the structure of a sample. In incoherent scattering, waves scattered by different nuclei do not interfere with each other, so elastic incoherent neutron scattering reflects the displacement of individual atoms in the picosecond to nanosecond time range, depending on the energy resolution of the spectrometer. In biological macromolecules, $\sim 50\%$ the atoms are hydrogen. The cross-section for incoherent scattering by nuclei of hydrogen atoms is much larger than the coherent or incoherent scattering by nuclei of most elements present in biological matter. Thus, incoherent neutron scattering is used to analyse the motion hydrogen atoms which have a large incoherent area [25]. This method helps explain the dynamics of proteins because the distribution of hydrogens in protein

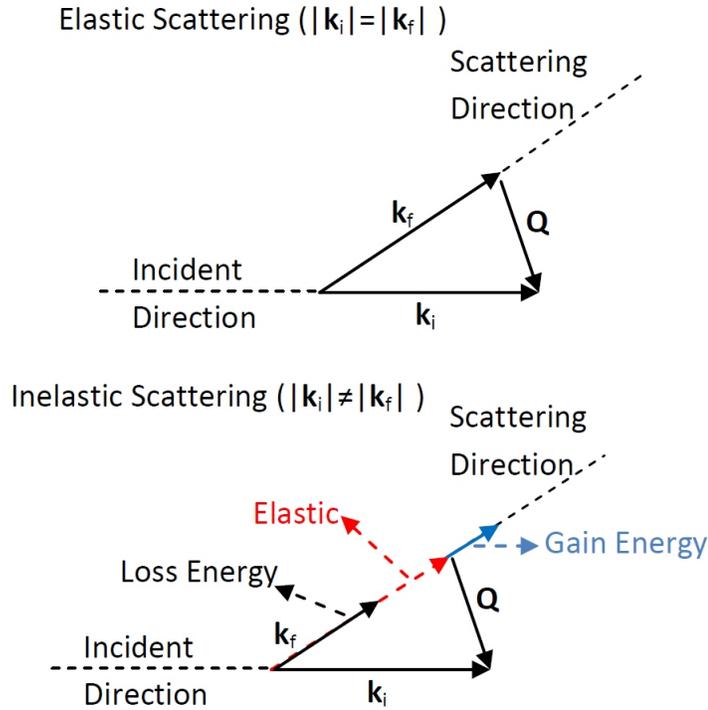


Figure 1.5: Scattering triangles for (Upper) an elastic scattering in which the neutron is deflected but does not gain or lose energy and (Lower) an inelastic scattering in which the neutron either loses energy or gains energy.

consistent.

Neutron scattering experiments measure the number of neutrons scattered by a sample as a function of the wave vector and the energy change of the neutron. Neutron for scattering experiments can be produced either by nuclear fission in a reactor or by The commonly used neutron scattering instruments to work on biological samples are IN16, IN13, IN10, IN6 and IN5. IN16, IN13 and IN10 are backscattering spectrometers, and IN5 and IN6 time of flight spectrometer. Neutron scattering instruments characterized by their energy resolution width W and Q -range. Q -range defines the "space window" with $1/Q$. The energy resolution of instruments, which is the energy width of the elastic peak, defines the "time window" $1/W$. The resolution function of the neutron scattering instrument is a Gaussian, Lorentzian and triangle energy resolution as depending on the instruments type. For example, the resolution function for

IN5 is a Gaussian resolution function.

In neutron scattering, sample has to be large around ~ 100 mg. In incoherent neutron scattering, it is preferred a sample does not include crystalline structure. Because, a sample including crystalline structure give large coherent peak to the observed intensity. In the analysis of biological samples, powder sample is commonly used material. Powder samples are non-crystalline aggregates of low water content. In neutron scattering experiments, hydration values for powder sample is around $h = 0.4 - 0.5$ gwater/gprotein which covers small global proteins by one hydration level.

Chapter 2

INTRINSIC MEAN SQUARE DISPLACEMENTS IN PROTEINS

The global, observed mean square displacement (MSD), $\langle r^2 \rangle_{exp}$, of hydrogen (H) in proteins increases linearly with temperature at low temperatures. At a specific temperature, the temperature dependence increases markedly. The marked increase in slope is denoted as the dynamical transition (DT) and the temperature of the change, the DT temperature T_D . The MSD observed with neutrons is found to depend on the energy resolution of the neutron scattering instrument. The higher the instrument resolution, the larger is the observed MSD, $\langle r^2 \rangle_{exp}$.

In this chapter, we introduce a method to extract the intrinsic MSD $\langle r^2 \rangle$ from a resolution broadened dynamic structure factor (DSF). The intrinsic MSD defined in the full Debye-Waller factor includes slow (long-time) and fast (short-time) motions, and it is independent of the energy resolution of the instrument. It is the MSD that would be observed on an instrument which has perfect resolution (zero energy resolution width, $W \rightarrow 0$). Here, we construct a simple model for the elastic resolution broadened DSF $S_R(\mathbf{Q}, \omega = 0)$ that is equivalent to what is measured in neutron scattering experiments. The model $S_R(\mathbf{Q}, \omega = 0)$ is fitted to the elastic resolution broadened incoherent DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$, to obtain the intrinsic MSD $\langle r^2 \rangle$. Then, an expression for the resolution broadened MSD $\langle r^2 \rangle_R$ is obtained using the model of the $S_R(\mathbf{Q}, \omega = 0)$ to clarify the difference between the intrinsic MSD $\langle r^2 \rangle$ and the resolution broadened MSD $\langle r^2 \rangle_R$. The resolution broadened MSD $\langle r^2 \rangle_R$ is equal to the observed MSD $\langle r^2 \rangle_{exp}$. The $\langle r^2 \rangle$ can be also obtained by fitting the model $\langle r^2 \rangle_R$ to the observed $\langle r^2 \rangle_{exp}$. This fitting method provides us a way to extract the resolution independent MSD from the resolution dependent data.

2.1 The Resolution Broadened Incoherent DSF $S_R(\mathbf{Q}, \omega)$

In neutron scattering experiments, neutrons interact with and scatter from the nuclei in the proteins. The hydrogen nucleus has a large, incoherent scattering cross-section for neutrons, 82 barns, much larger than other nuclei. Because of this, scattering from hydrogen dominates on the observed scattering intensity. The observed scattering intensity is proportional to the incoherent DSF, $S_{inc}(\mathbf{Q}, \omega)$,

$$S_{inc}(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t) I_{inc}(\mathbf{Q}, t). \quad (2.1)$$

In Eq. (2.1), $\hbar\mathbf{Q}$ and $\hbar\omega$ are the momentum and energy, respectively, transferred from the neutron to the protein in the scattering. In Eq. (2.1), the incoherent intermediate scattering function (ISF) is

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^N b_i^2 \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}_i(t)) \exp(i\mathbf{Q} \cdot \mathbf{r}_i(0)) \rangle, \quad (2.2)$$

where $\mathbf{r}_i(\cdot)$ represents the position of nucleus i at reference time t and 0 , and b_i is the incoherent scattering length of nucleus i . In Eq. (2.2), $\langle \cdot \rangle$ is the ensemble average and N is the number of scatterers in the sample. As stated above, the b_i of hydrogen is more than 20 times larger than the b_i of other nuclei in proteins. The incoherent ISF, which contains the contribution from N different scattering centers, is represented as an average over the N self correlation functions of each individual nuclei. In the analysis of data, it is common to represent the ISF, $I_{inc}(\mathbf{Q}, t)$, in Eq. (2.2) summed over all nuclei by a global $I(\mathbf{Q}, t)$,

$$I(\mathbf{Q}, t) = \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(t)) \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle, \quad (2.3)$$

in which a single scatterer center represents the whole protein. The self correlation function of $\mathbf{r}(t)$ in Eq. (2.3) is an average correlation function containing all time scales, short and long, over nuclei in the protein. The global incoherent DSF $S(\mathbf{Q}, \omega)$ of protein is

$$S(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t) I(\mathbf{Q}, t). \quad (2.4)$$

The instruments used in neutron scattering experiments have an energy resolution function, which is denoted here by $R(\omega)$. The observed function $S_R(\mathbf{Q}, \omega)$ is the resolution broadened DSF. This is a convolution of the incoherent DSF $S(\mathbf{Q}, \omega)$ and the resolution function of instrument $R(\omega)$,

$$\begin{aligned} S_R(\mathbf{Q}, \omega) &= \int_{-\infty}^{\infty} d\omega' S(\mathbf{Q}, \omega') R(\omega - \omega') \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t) I(\mathbf{Q}, t) R(t). \end{aligned} \quad (2.5)$$

A perfect resolution function would be $R(\omega) = \delta(\omega)$. For perfect resolution, the energy resolution width is zero ($W = 0$) and the observed data only includes the elastic component of the scattering intensity ($\omega = 0$). For a finite energy resolution width W , the resolution function $R(\omega)$ could be represented as a Lorentzian function

$$R(\omega) = \frac{1}{\pi} \frac{W}{W^2 + \omega^2}. \quad (2.6)$$

Fourier transform of $R(\omega)$ in Eq. (2.6) is $R(t) = \exp(-Wt)$. Another representation for the resolution function $R(\omega)$ with a finite resolution width is a Gaussian function. For the resolution function in Eq. (2.6), the resolution broadened DSF becomes

$$S_R(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t) I(\mathbf{Q}, t) \exp(-Wt). \quad (2.7)$$

In neutron scattering experiments, the elastic component of the incoherent $S_R(\mathbf{Q}, \omega)$, which includes zero energy transfer $\omega = 0$, is commonly measured data. The elastic incoherent DSF $S_R(\mathbf{Q}, \omega = 0)$, is written as

$$S_R(\mathbf{Q}, \omega = 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt I(\mathbf{Q}, t) \exp(-Wt). \quad (2.8)$$

In Eq. (2.8), long time processes in $I(\mathbf{Q}, t)$ are not observable in $S_R(\mathbf{Q}, \omega = 0)$ because it is cut off by the resolution function after at time $\tau_R \sim 1/W$. The energy resolution width of instruments determines which motions are observed in $S_R(\mathbf{Q}, \omega = 0)$. The resolution broadened DSF $S_R(\mathbf{Q}, \omega = 0)$ includes only processes observed up to time τ_R . The higher the instrument resolution, the longer the time in $I(\mathbf{Q}, t)$ that can be

observed in $S_R(\mathbf{Q}, \omega = 0)$. For a perfect resolution (no resolution width, $W = 0$) (e.g. $R(\omega) = \delta(\omega)$), all long-time and short-time motions in $I(\mathbf{Q}, t)$ are observed in $S_R(\mathbf{Q}, \omega = 0)$.

In the neutron scattering experiments, the observed resolution broadened DSF $S_R^{exp}(\mathbf{Q}, \omega)$ includes the desired elastic component $S_R^{exp}(\mathbf{Q}, \omega = 0)$ plus inelastic components around $\omega = 0$ covering an energy range $W/2 < \omega < W/2$. It is often assumed that the observed resolution broadened DSF is given by $S_R^{exp}(\mathbf{Q}) = \exp(-\frac{1}{3}Q^2\langle r^2 \rangle)$. Specifically, the observed MSD $\langle r^2 \rangle_{exp}$ is typically obtained from the observed incoherent DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$. The observed MSD $\langle r^2 \rangle_{exp} = -3d \ln S_R^{exp}(\mathbf{Q}, \omega = 0)/dQ^2$ is obtained as a function of wave vector transfer, \mathbf{Q} . If the data is fitted to or analyzed as if only the purely elastic component i.e. $S_R(\mathbf{Q}, \omega = 0) = I(\mathbf{Q}, t = \infty)$ is observed, then the observed MSD $\langle r^2 \rangle_{exp}$ depends on the resolution width, W , of the instrument. The resolution dependence of the $\langle r^2 \rangle_{exp}$ has been extensively investigated [50, 73, 106]. In Table 2.1, the resolution width and the observation time of the commonly used neutron scattering instruments are shown. The observation time τ_R for a Gaussian resolution function is calculated in *Appendix A*.

2.2 The Model for the Incoherent Resolution Broadened DSF $S_R(\mathbf{Q}, \omega = 0)$

We aim to construct a simple model for the elastic resolution broadened incoherent DSF $S_R(\mathbf{Q}, \omega = 0)$. With this model, our goal is to extract the intrinsic MSD $\langle r^2 \rangle$ from the data measured by neutron scattering. The intrinsic MSD consists of the contributions from all motions, including long time motions and is defined in the Debye-Waller factor. The Debye-Waller factor is the infinite time value of the incoherent ISF $I_{inc}(\mathbf{Q}, t = \infty)$. The model $S_R(\mathbf{Q}, \omega = 0)$ is composed of the model $I(\mathbf{Q}, t)$ and a resolution function $R(\omega)$. The model of $S_R(\mathbf{Q}, \omega = 0)$ can fit to data to obtain the intrinsic MSD $\langle r^2 \rangle$.

We start with the physical properties of the incoherent ISF to construct a simple model for $I(\mathbf{Q}, t)$. $I(\mathbf{Q}, t)$ is separated into two parts,

$$I(\mathbf{Q}, t) = I(\mathbf{Q}, t = \infty) + I'(\mathbf{Q}, t), \quad (2.9)$$

Table 2.1: The energy resolution width, W in units μeV and THz , the observation time, $\tau_R = (8 \ln 2)^{1/2} \hbar / W$ and $\tau_R \propto W^{-1}$ of neutron scattering instruments.

Instrument	$W(\mu\text{eV})$	$W(\text{THz})$	$\tau_R = (8 \ln 2)^{1/2} \hbar / W$ (ns)	$\tau_R \propto W^{-1}$ (ns)
IN16	~ 1	~ 0.00025	~ 1.5	~ 4
IN10	~ 1	~ 0.00025	~ 1.5	~ 4
IN13	~ 10	~ 0.0025	~ 0.15	~ 0.4
IN6	~ 100	~ 0.025	~ 0.015	~ 0.04
IN5	~ 100	~ 0.025	~ 0.015	~ 0.04
HER	~ 100	~ 0.025	~ 0.015	~ 0.04
GP-TAS	~ 1000	~ 0.25	~ 0.0015	~ 0.004

where $I'(\mathbf{Q}, t)$ is a time dependent part and $I_\infty(\mathbf{Q}) = I(\mathbf{Q}, t = \infty)$ is a time independent part. $I_\infty(\mathbf{Q})$ is a constant and time independent value that $I(\mathbf{Q}, t)$ reaches at time $t = \infty$. From Eq. (2.3), $I_\infty(\mathbf{Q}) = I(\mathbf{Q}, t = \infty)$ is written as (see *Appendix B*)

$$\begin{aligned}
I_\infty(\mathbf{Q}) = I(\mathbf{Q}, t = \infty) &= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(\infty)) \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\
&= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(\infty)) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\
&= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\
&= \exp\left(-\frac{1}{3}Q^2 \langle r^2 \rangle + O(Q^4)\right). \tag{2.10}
\end{aligned}$$

In order to obtain the expression in Eq. (2.10), we make three assumptions. These are (1) $\mathbf{r}(\infty)$ and $\mathbf{r}(0)$, the position of global scatterer center at $t = \infty$ and $t = 0$, are completely uncorrelated so that the averages of them are independent (for the second line), (2) the system is translationally invariant in time (no center of mass (CM) motion) so that $\mathbf{r}(\infty) = \mathbf{r}(0)$ (for the third line) and (3) in a cumulant expansion of $\langle \exp(-i\mathbf{Q} \cdot \mathbf{r}) \rangle$, cumulants beyond the second cumulant are negligible (for the last line).

The last assumption is valid if \mathbf{Q} is small or if the distribution over \mathbf{r} is approximately a Gaussian distribution. The cumulants beyond the second term are small for small \mathbf{Q} and are negligible. They vanish exactly for all \mathbf{Q} if the distribution over \mathbf{r} is exactly Gaussian.

We take $\langle r^2 \rangle$ in $I(\mathbf{Q}, \infty)$ in Eq. (2.10) as the definition of the intrinsic, long time value of MSD in the protein that includes all motional processes. The intrinsic MSD $\langle r^2 \rangle$ is equal to the $\langle r^2 \rangle_{exp}$ that would be observed with an infinitely high resolution instrument, $R(\omega) = \delta(\omega)$. The definition of the intrinsic MSD $\langle r^2 \rangle$ is the well-known Debye-Waller factor;

$$I_\infty(\mathbf{Q}) = I(\mathbf{Q}, t = \infty) = \exp(-\frac{1}{3}Q^2\langle r^2 \rangle). \quad (2.11)$$

The ISF $I(\mathbf{Q}, t)$ has the limits

$$I(\mathbf{Q}, t) = \begin{cases} 1 & t = 0 \\ I_\infty & t = \infty. \end{cases}$$

These limits allow us to determine the limits for the time dependent part $I'(\mathbf{Q}, t)$ in $I(\mathbf{Q}, t)$,

$$I'(\mathbf{Q}, t) = I(\mathbf{Q}, t) - I_\infty = \begin{cases} 1 - I_\infty & t = 0 \\ 0 & t = \infty. \end{cases}$$

In respect to these limits, $I'(\mathbf{Q}, t)$ is modelled as

$$I'(\mathbf{Q}, t) = (1 - I_\infty)C(t), \quad (2.12)$$

where the limits of $C(t)$ are

$$C(t) = \begin{cases} 1 & t = 0 \\ 0 & t = \infty. \end{cases}$$

As an example, $C(t)$ is chosen as a simple exponential decay function,

$$C(t) = \exp(-\lambda t), \quad (2.13)$$

where $\lambda = 1/\tau$ represents the decay constant of correlations in the protein, and τ is the relaxation time of the protein. The model of $I(\mathbf{Q}, t)$ is constructed as

$$I(\mathbf{Q}, t) = I_\infty(\mathbf{Q}) + (1 - I_\infty(\mathbf{Q}))C(t), \quad (2.14)$$

which has the correct limits at $t = 0$ and $t = \infty$. Eq. (2.14) is a conceivable representation of the incoherent ISF with a single motional decay process. An example for a single decay process could be diffusion.

Substituting the model of $I(\mathbf{Q}, t)$ in the observed elastic DSF $S_R(\mathbf{Q}, \omega = 0)$ in Eq. (2.8), the elastic resolution broadened DSF $S_R(\mathbf{Q}, \omega = 0)$ becomes

$$\begin{aligned} S_R(\mathbf{Q}, \omega = 0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt I(\mathbf{Q}, t) \exp(-Wt) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt I(\mathbf{Q}, \infty) \exp(-Wt) \\ &+ \frac{1}{2\pi} \int_{-\infty}^{\infty} dt (1 - I(\mathbf{Q}, \infty)) \exp(-\lambda t) \exp(-Wt) \\ &= I_\infty \frac{1}{\pi W} + (1 - I_\infty) \frac{1}{\pi(W + \lambda)}. \end{aligned} \quad (2.15)$$

The data measured in neutron scattering experiments in the literature is generally presented as the “normalized” $S_R(\mathbf{Q}, \omega = 0)$. The model of the elastic DSF $S_R(\mathbf{Q}, \omega = 0)$ in Eq. (2.15) is divided to $S_R(\mathbf{Q} = 0, \omega = 0)$ at $\mathbf{Q} = 0$ to obtain the expression for the “normalized” $S_R(\mathbf{Q}, \omega = 0)$. From Eq. (2.15), the $S_R(\mathbf{Q} = 0, \omega = 0) = (\pi W)^{-1}$ and “normalized” $S_R(\mathbf{Q}, \omega = 0)$ is

$$S_R^N(\mathbf{Q}, \omega = 0) = I_\infty + (1 - I_\infty) \frac{W}{(W + \lambda)} + A. \quad (2.16)$$

Here, A is a constant. A is added to Eq. (2.16) since the data for different \mathbf{Q} values are sometimes separated from one another in a figure by a constant for clarity.

The simple model for the observed elastic DSF $S_R^N(\mathbf{Q}, \omega = 0)$ includes three fitting parameters; the intrinsic MSD $\langle r^2 \rangle$, the relaxation frequency λ and a constant A . The model of $S_R^N(\mathbf{Q}, \omega = 0)$ in Eq. (2.16) is fitted to data for a given W to determine the best fit value of the $\langle r^2 \rangle$, λ and A . The parameter A plays no role. By

fitting the model $S_R^N(\mathbf{Q}, \omega = 0)$ to data, we determine the intrinsic long time value of $\langle r^2 \rangle$ and λ . It is expected that the intrinsic $\langle r^2 \rangle$ is found to be independent of W . The intrinsic MSD $\langle r^2 \rangle$ can be compared with the observed MSD $\langle r^2 \rangle_{exp}$ which depends on the resolution width of instrument employed.

A model for the resolution broadened MSD $\langle r^2 \rangle_R$ can be obtained using the model of the $S_R^N(\mathbf{Q}, \omega = 0)$ in Eq. (2.16) instead of $S_R^{exp}(\mathbf{Q}, \omega = 0)$ in the slope expression in Eq. (1.1),

$$\begin{aligned} \langle r^2 \rangle_R &= -3 \frac{d \ln S_R^N(\mathbf{Q}, \omega = 0)}{dQ^2} \\ &= \langle r^2 \rangle / [1 + \frac{W}{I_\infty \lambda}]. \end{aligned} \quad (2.17)$$

In Eq. (2.17), the $\langle r^2 \rangle_R$ depends on the $\langle r^2 \rangle$, λ and W . The $\langle r^2 \rangle_R$ is the model that is equivalent to $\langle r^2 \rangle_{exp}$. This depends on the instrument resolution width W since $S_R^N(\mathbf{Q}, \omega = 0)$ depends on W . From Eq. (2.17), it is shown that the $\langle r^2 \rangle_R$ depends on the ratio of W/λ . If $W \ll \lambda$, then the instrument resolution is high enough to catch all the decay of $I(\mathbf{Q}, t)$, and the resolution broadened MSD reaches to the intrinsic long time value ($\langle r^2 \rangle_R \rightarrow \langle r^2 \rangle$). The good fit of the model $S_R^N(\mathbf{Q}, \omega = 0)$ to data shows that λ and $\langle r^2 \rangle$ in Eq. (2.16) are well determined. In this case, we expect that the $\langle r^2 \rangle_R$, obtained using the best fit values of the $\langle r^2 \rangle$ and λ in Eq. (2.17), reproduces well the observed $\langle r^2 \rangle_{exp}$.

For zero energy resolution width W , the scattering intensity at $\omega = 0$ is observed. In this case, the observed MSD $\langle r^2 \rangle_{exp}$ given by Eq. (1.1) is equal to the actual $\langle r^2 \rangle$ of the protein, and $S_R(\mathbf{Q}, \omega = 0)$ reduces to $S(\mathbf{Q}, \omega = 0)$. $S_R(\mathbf{Q}, \omega = 0)$ is well approximated by its time independent part $I_\infty(\mathbf{Q})$ given by Eq. (2.10) for $W = 0$. When W is finite, the scattered intensity at $\omega = 0$ and around $\omega = 0$ is observed in $S_R(\mathbf{Q}, \omega = 0)$. Hence, the $\langle r^2 \rangle_{exp}$ is found to be dependent on the resolution width W . More precisely, the ratio of W and λ controls the motions in the protein contributing to the $\langle r^2 \rangle_{exp}$. For a finite energy resolution width, the $\langle r^2 \rangle_{exp}$ extracted from Eq. (1.1) is smaller than the intrinsic MSD $\langle r^2 \rangle$.

2.3 Results

The simple models of the normalized resolution broadened DSF $S_R^N(\mathbf{Q}, \omega = 0)$ and of the resolution broadened MSD $\langle r^2 \rangle_R$ are defined in Sec. 2.2. In order to test the method in Sec. 2.2, we fit the observed, resolution broadened elastic incoherent DSF model $S_R^N(\mathbf{Q}, \omega = 0)$ given by Eq. (2.16) to data in the literature. The goal is to determine the intrinsic MSD $\langle r^2 \rangle$ in specific proteins and obtain a value for the relaxation parameter λ which characterizes the motions contributing to the intrinsic MSD $\langle r^2 \rangle$.

In this chapter, the observation time $\tau_R \sim 1/W$ shown in Table-1 is used in applications of the method.

2.3.1 The Application of the model of $S_R^N(\mathbf{Q}, \omega = 0)$

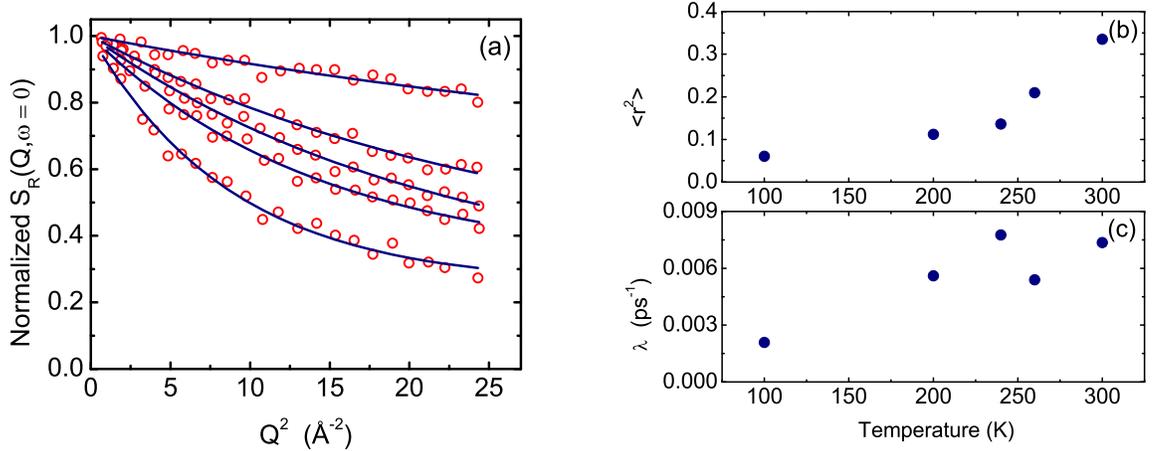


Figure 2.1: (a) Elastic DSF, $S_R^N(\mathbf{Q}, \omega = 0)$, observed by Paciaroni et al. [77] on IN13 (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$. (b)-(c) The best fit values of $\langle r^2 \rangle$ and λ .

In Fig. 2.1a, the model $S_R^N(\mathbf{Q}, \omega = 0)$ is fitted to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$ observed for lysozyme in pure glycerol by Paciaroni et al. [77] to test the model. $S_R^{exp}(\mathbf{Q}, \omega = 0)$ is observed on IN13 instrument which has an energy resolution, $W = 10 \mu\text{eV}$. The open circles represent data for $S_R^{exp}(\mathbf{Q}, \omega = 0)$ at five temperatures.

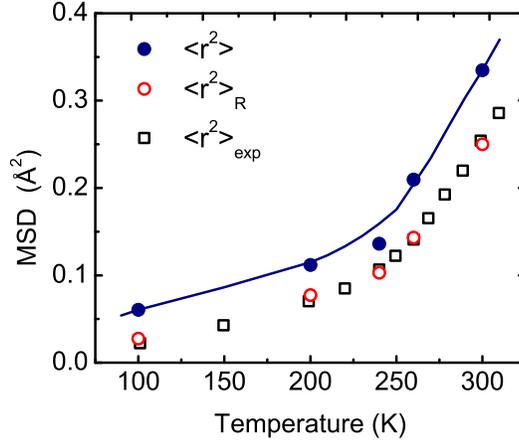


Figure 2.2: MSD, $\langle r^2 \rangle_{exp}$, in lysozyme observed Paciaroni et al. [77] on IN13 (open squares). The intrinsic $\langle r^2 \rangle$ (solid circles) are the intrinsic MSD obtained by fitting the model Eq. (2.16) to data shown in Fig. 2.1. The solid line is a guide to the eye through the intrinsic $\langle r^2 \rangle$. The $\langle r^2 \rangle_R$ (open circles) are the MSD calculated from Eq. (2.17) which should be similar to $\langle r^2 \rangle_{exp}$.

Temperatures from top to bottom are 100, 200, 240, 260 and 300 K. The solid lines represent the fit of the model $S_R(\mathbf{Q}, \omega = 0)$ to data and the fits for all temperatures are good. Fig. 2.1b and c show the values of the $\langle r^2 \rangle$ and λ in the model $S_R^N(\mathbf{Q}, \omega = 0)$ that give the best fit. The best fit value of the $\langle r^2 \rangle$ increases with temperature, and marked change is observed at $T \simeq 250$ K. The relaxation frequency λ is not well determined as shown in Fig. 2.1c. Fig. 2.2 compares the intrinsic MSD $\langle r^2 \rangle$, the resolution broadened model MSD $\langle r^2 \rangle_R$ and the observed MSD $\langle r^2 \rangle_{exp}$ in lysozyme. The solid points are the intrinsic $\langle r^2 \rangle$ obtained from the fits in Fig. 2.1a and the solid line is a guide to the eye through the intrinsic MSD $\langle r^2 \rangle$. The $\langle r^2 \rangle_{exp}$ is the MSD extracted from the DSF data $S_R^{exp}(\mathbf{Q}, \omega = 0)$, measured by Paciaroni et. al [77], using Eq. (1.1), and it is represented by open squares. The $\langle r^2 \rangle_R$ is obtained using the best fit values of the $\langle r^2 \rangle$ and λ in Eq. (2.17). The good agreement between the $\langle r^2 \rangle_R$ and $\langle r^2 \rangle_{exp}$ shows that the fit to the data is precise.

In order to show that the intrinsic MSD is obtained to be independent of the

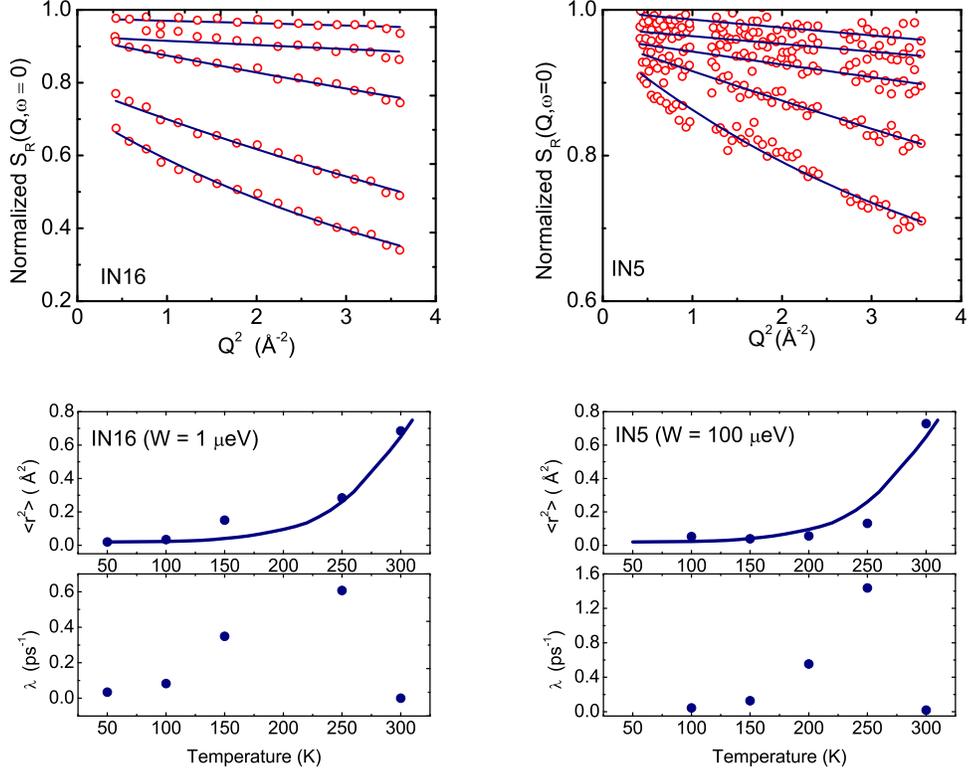


Figure 2.3: Upper frame: Elastic DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$, observed by Wood et al. [106] on IN16 and IN5 (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$. Lower frames: the best fit values of $\langle r^2 \rangle$ and λ . The solid line is a guide to the eye to the best fit $\langle r^2 \rangle$, the same line for both instruments.

resolution of the instrument, the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$ observed on two instruments, IN16 and IN5, by Wood et al. [106] are chosen. The upper frame of Fig. 2.3 shows the fit of the model $S_R^N(\mathbf{Q}, \omega = 0)$ in Eq. (2.16) to the measured normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$ in Ribonuclease A observed by Wood et. al [106]. The fits are good except that there is significant scatter in the data taken on IN5. The best fit values of the $\langle r^2 \rangle$ and λ are shown in the middle and lower frames of Fig. 2.3. The best fit values of the intrinsic $\langle r^2 \rangle$ is independent of the instrument energy resolution width W , and increases with temperature. The marked change in the $\langle r^2 \rangle$ is observed at temperatures above $T \simeq 250$ K. The solid line, which is a guide to the eye through the

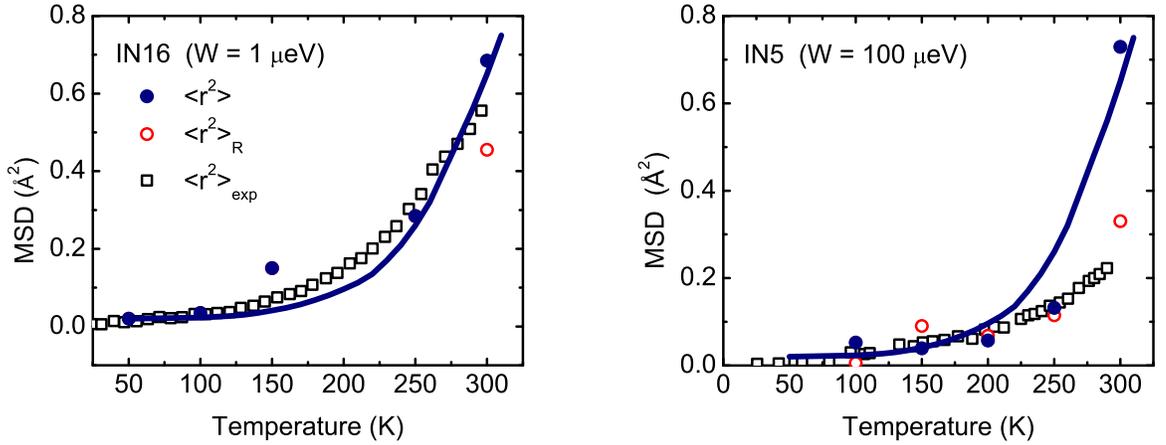


Figure 2.4: MSD, $\langle r^2 \rangle_{exp}$, in hydrated Ribonuclease A observed by Wood et al. [106] on IN16 and IN5 (open squares). The intrinsic MSD $\langle r^2 \rangle$ (solid circles) are obtained by fitting the model Eq. (2.16) to the data shown in Fig. 2.3. The solid line is a guide to the eye to the intrinsic MSD, the same for both instruments. The MSD $\langle r^2 \rangle_R$ (open circles) is calculated from Eq. (2.17).

intrinsic $\langle r^2 \rangle$, is the same for both instruments. In Fig. 2.4, the intrinsic $\langle r^2 \rangle$ is compared with the resolution dependent observed $\langle r^2 \rangle_{exp}$ obtained from the slope of the measured data with Q^2 and the resolution broadened MSD $\langle r^2 \rangle_R$. The $\langle r^2 \rangle_{exp}$ is very different for the two instruments displaying the dependence of $\langle r^2 \rangle_{exp}$ on the resolution width W . The $\langle r^2 \rangle_{exp}$ and $\langle r^2 \rangle_R$ are agree well for IN5 and IN16. In addition to this, the $\langle r^2 \rangle_{exp}$ observed on IN16 agrees well with the intrinsic $\langle r^2 \rangle$. This agreement could arise from that the observed MSD $\langle r^2 \rangle_{exp}$ could reach its final, equilibrium value in a time period, $\tau \sim 1/W$ ns, observable on IN16.

In a similar way, we fit the model $S_R^N(\mathbf{Q}, \omega = 0)$ to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$ observed in heparan sulphate (HS-0.4) by Jasnin et al. [50]. The top frame of Fig. 2.5 shows the fit of the model $S_R^N(\mathbf{Q}, \omega = 0)$ to the observed normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$. $S_R^{exp}(\mathbf{Q}, \omega = 0)$ is observed on three instruments which have different energy resolutions, $W = 1, 10$ and $100 \mu\text{eV}$. There is data for $S_R^{exp}(\mathbf{Q}, \omega = 0)$ at seven temperatures, the lowest temperature at the top. The fits are generally good but some data points lie

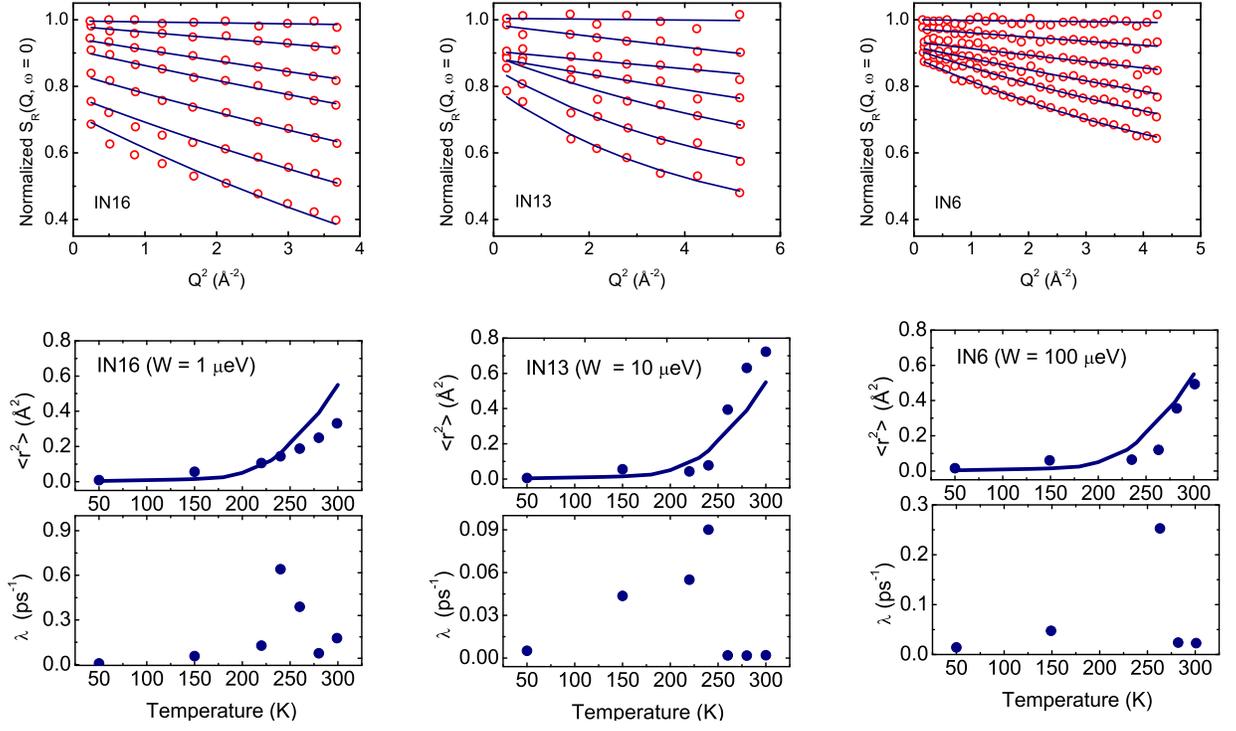


Figure 2.5: Upper frame: Elastic DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$, of H in heparan sulphate observed by Jasnin et al. [50] (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$. Lower frames: the best fit values of the intrinsic $\langle r^2 \rangle$ and λ . The solid line is a guide to the eye for $\langle r^2 \rangle$, the same line for all three instruments.

off the fitted line. The middle and bottom frames in Fig. 2.5 show the values of the $\langle r^2 \rangle$ and λ in the model $S_R^N(\mathbf{Q}, \omega = 0)$ that give the best fit. The best fit value of the $\langle r^2 \rangle$ increases with temperature, markedly for temperatures above $T \simeq 230$ K. There is some scatter in the $\langle r^2 \rangle$ which arises from the uncertainty in the fit. Similar to Fig. 2.1c and the lower frame in Fig. 2.3, the scatter or uncertainty of λ is particularly large showing that the data is relatively insensitive to the value of λ . On the other hand, the data is not very discriminating or sensitive to the time dependence of $C(t)$ in the model. In spite of these scatters in the $\langle r^2 \rangle$, it is important that the $\langle r^2 \rangle$ obtained from the fit is independent of the resolution width, W . The solid black line in the middle frame of Fig. 2.5 is a guide to the eye through all the $\langle r^2 \rangle$, the same line for

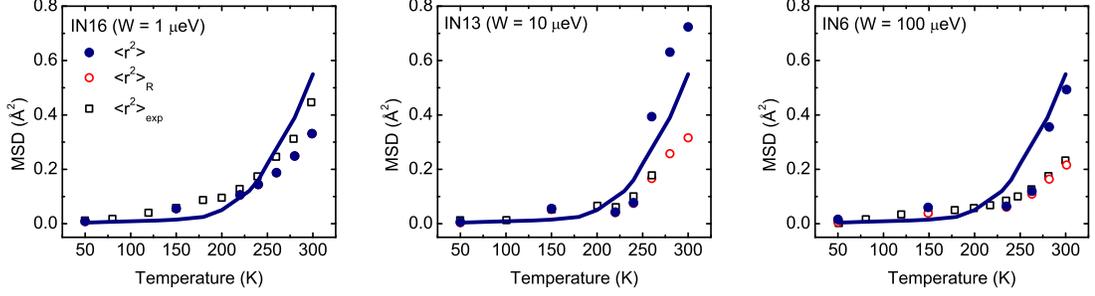


Figure 2.6: MSD, $\langle r^2 \rangle_{exp}$, in hydrated HS-0.4 observed Jasnin et al. [50] on IN16, IN13 and IN6 (open squares). The intrinsic $\langle r^2 \rangle$ (solid circles) are the intrinsic MSD obtained by fitting the model Eq. (2.16) to data shown in Fig. 2.6. The solid line is a guide to the eye through the intrinsic $\langle r^2 \rangle$, the same for all instruments. The $\langle r^2 \rangle_R$ (open circles) are the MSD calculated from Eq. (2.17) which should be similar to $\langle r^2 \rangle_{exp}$.

all instruments. This indicates that the average $\langle r^2 \rangle$ is obtained as independent of the instrument resolution width W by using the model Eq. (2.16), even if there is substantial fluctuations in the individual values of the $\langle r^2 \rangle$.

Similarly, Fig. 2.6 compares the intrinsic $\langle r^2 \rangle$, the model $\langle r^2 \rangle_R$ and the observed $\langle r^2 \rangle_{exp}$ in hydrated HS-0.4. The solid points are the intrinsic $\langle r^2 \rangle$ emerging from the fits in Fig. 2.5 and the solid line is again a guide to the eye through these $\langle r^2 \rangle$. The $\langle r^2 \rangle_{exp}$ is the MSD obtained by Jasnin et al. from the slope of their data using Eq. (1.1). The $\langle r^2 \rangle_R$ is obtained using Eq. (2.17) and the best fit values $\langle r^2 \rangle$ and λ in Fig. 2.5. If the fit to the data is precise, the $\langle r^2 \rangle_R$ and $\langle r^2 \rangle_{exp}$ should agree. This is the case for IN13 and IN6 but less so for the IN16. When the instrument resolution W is small, we expect $\langle r^2 \rangle_R$ to coincide with the intrinsic $\langle r^2 \rangle$. This is the case for IN16 where the $\langle r^2 \rangle_R$ (open circles) lie on top of the $\langle r^2 \rangle$ (black dots). For IN16, the $\langle r^2 \rangle_R$ and $\langle r^2 \rangle_{exp}$ differ somewhat for $T > 250$ K. This indicates that there is not a good fit at higher temperatures, as can be seen in the upper frame of Fig. 2.5. The essential point of Fig. 2.6 is that the intrinsic $\langle r^2 \rangle$ is independent of W and that $\langle r^2 \rangle$ shows a marked increase at temperature, $T_D \simeq 230$ K, the intrinsic DT temperature of HS-0.4.

As a final example, we fit the model $S_R^N(\mathbf{Q}, \omega = 0)$ to the normalized $S_R^{exp}(\mathbf{Q}, \omega =$

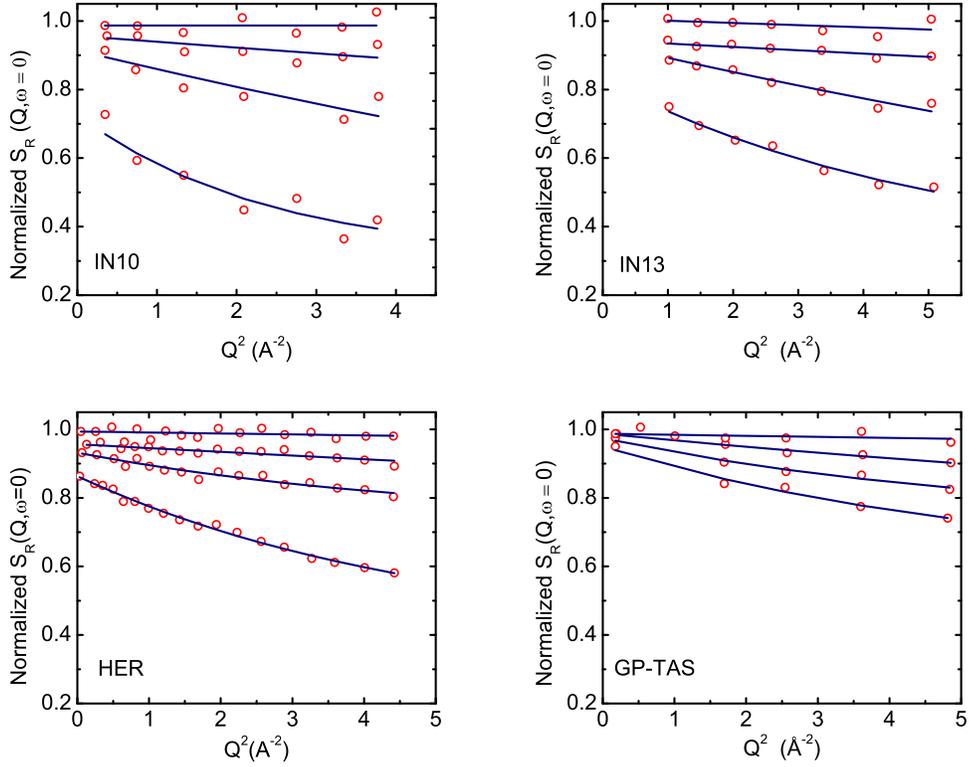


Figure 2.7: Upper frame: Elastic DSF, $S_R^{exp}(\mathbf{Q}, \omega = 0)$, of H in Staphylococcal Nuclease (SNase) by Nakagawa et al. [73] (open circles) and fit of model Eq. (2.16) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$.

0) observed Staphylococcal Nuclease (SNase) by Nakagawa et al. [73] in Fig. 2.7. $S_R^{exp}(\mathbf{Q}, \omega = 0)$ is observed on four instruments which have different energy resolutions, $W = 1, 10, 100$ and $1000 \mu\text{eV}$. There is data for $S_R^{exp}(\mathbf{Q}, \omega = 0)$ at temperatures, 50, 130, 210 and 300 K, the highest temperature at the bottom. Fig. 2.8 shows the best fit values of the $\langle r^2 \rangle$ and λ in the model $S_R^N(\mathbf{Q}, \omega = 0)$. The best fit $\langle r^2 \rangle$ increases with temperature, and a significant increase is observed at $T \simeq 200$ K. There is substantial scatter in the $\langle r^2 \rangle$, specially for the $\langle r^2 \rangle$ obtained from the fit of the $S_R^N(\mathbf{Q}, \omega = 0)$ to data measured by HER ($W = 1000 \mu\text{eV}$) and GP-TAS ($W = 1000 \mu\text{eV}$). In spite of scatter in $\langle r^2 \rangle$, an intrinsic $\langle r^2 \rangle$, which is independent of the instrument resolution width W , can be obtained by fitting the model to the observed $S_R^N(Q, \omega = 0)$. Similar to λ in Fig. 2.6,

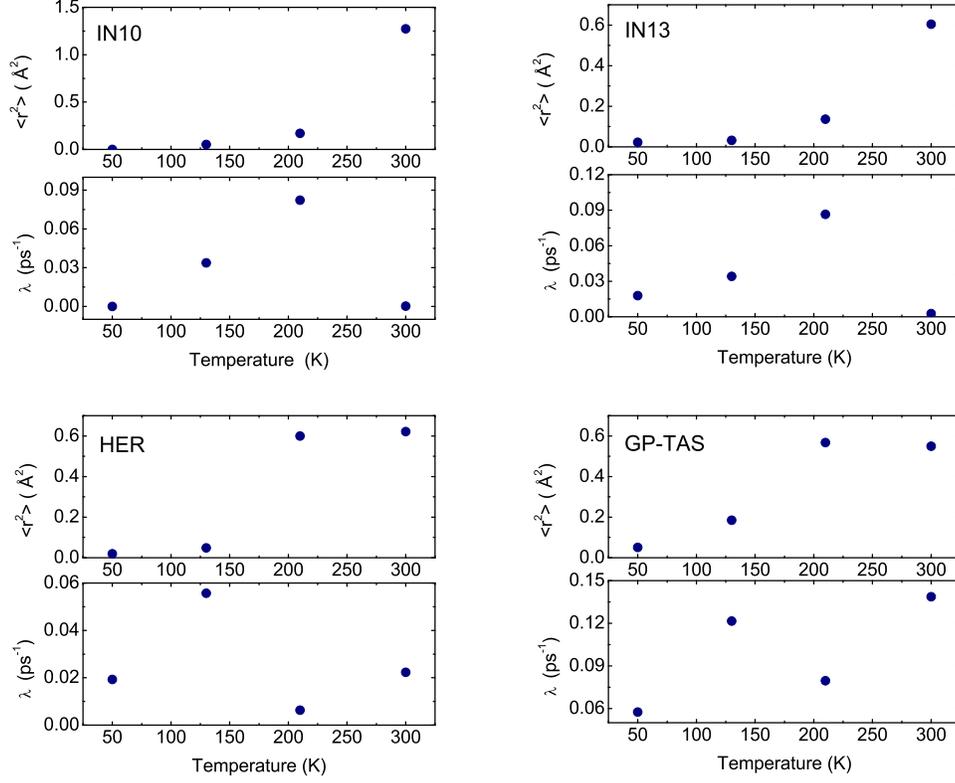


Figure 2.8: The best fit values of the intrinsic $\langle r^2 \rangle$ and λ obtained from the fits in Fig. 2.7.

the scatter or uncertainty of λ is particularly large in Fig. 2.8.

Fig. 2.9 compares the intrinsic $\langle r^2 \rangle$, the resolution broadened MSD model $\langle r^2 \rangle_R$ and the observed $\langle r^2 \rangle_{exp}$ in SNase. The solid points are the intrinsic $\langle r^2 \rangle$ emerging from the fits in Fig. 2.7 and the solid line is again a guide to the eye through all the $\langle r^2 \rangle$. The $\langle r^2 \rangle_{exp}$ is the MSD obtained by Nakagawa et al. from the slope of their data using Eq. (1.1). The $\langle r^2 \rangle_R$ is obtained from the expression in Eq. (2.17) using fitting parameters, $\langle r^2 \rangle$ and λ . Although there is substantial scatter in the $\langle r^2 \rangle$, $\langle r^2 \rangle_R$ and $\langle r^2 \rangle_{exp}$ agree well for all four instruments (four different resolution widths).

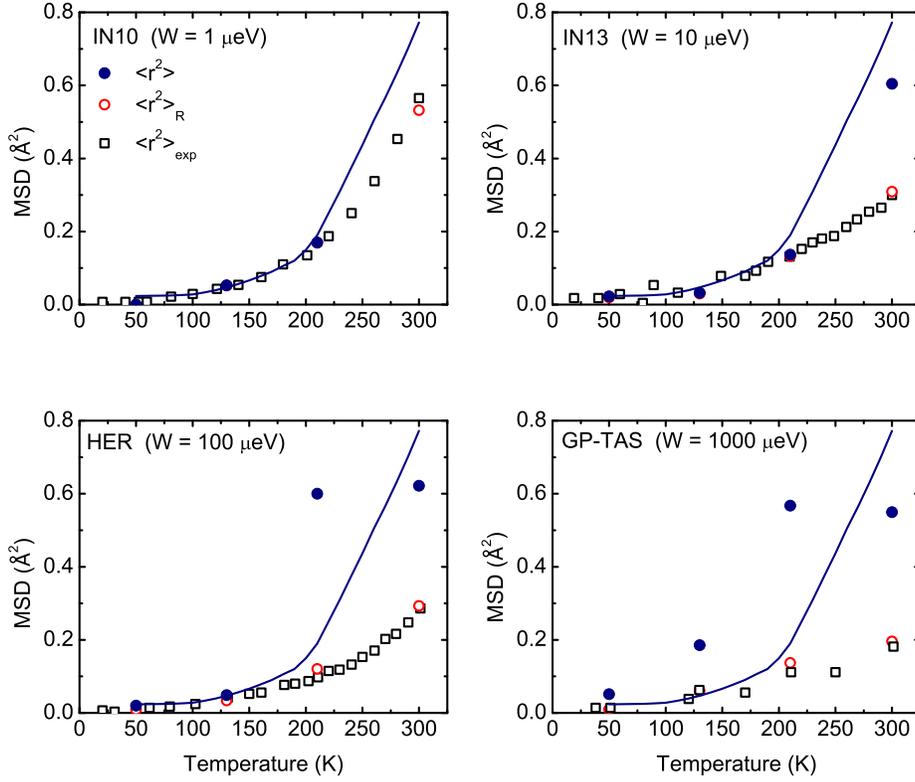


Figure 2.9: MSD, $\langle r^2 \rangle_{exp}$, of the hydrated SNase observed by Nakagawa et al. [73] on IN10 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$), HER ($W = 100 \mu\text{eV}$) and GP-TAS ($W = 1000 \mu\text{eV}$) (open squares). The intrinsic MSD $\langle r^2 \rangle$ (solid circle) is obtained the fit in Fig. 2.7. The solid line is a guide to the eye to the $\langle r^2 \rangle$, the same for all instruments. The $\langle r^2 \rangle_R$ (open circles) is calculated from Eq. (2.17).

2.3.2 The Application of the Expression of $\langle r^2 \rangle_R$ and Dynamical Transition

The origin of the DT has been investigated by experimental and theoretical studies during decades. It has been interpreted as a sudden change of effective elasticity in proteins [109], the onset of motions of specific side groups [61], a glass transition or a phase transition in the hydration water [23, 13], and an artificial effect of a finite instrument resolution width [17, 3, 54, 32]. It has been still debated.

In this part of Sec. 2.3, we focus on the energy resolution dependence of the DT temperature, T_D . The DT temperature, T_D , is the temperature at which $\langle r^2 \rangle_{exp}$

increases markedly with temperature. It appears to be dependent on the energy resolution of the instrument employed. Using the model of the $\langle r^2 \rangle_R$ given by Eq. (2.17), an intrinsic $\langle r^2 \rangle$ and T_D are obtained to be independent of instrument resolution. The observed MSD $\langle r^2 \rangle_{exp}$ of H in glutamate dehydrogenase observed on IN16 and IN6 is shown on the LHS in Fig. 2.10. The $\langle r^2 \rangle_{exp}$ is obtained as usual from the slope of the observed, elastic DSF $S_R^{exp}(\mathbf{Q}, \omega = 0)$ versus Q^2 by using the expression in Eq. (1.1). In this example, on IN16 the apparent T_D is $T_D \simeq 150$ K while on IN6 that is $T_D \simeq 230$ K. The T_D obtained from $S_R^{exp}(\mathbf{Q}, \omega = 0)$ data measured by an instrument which has a larger energy resolution width is higher. The model $\langle r^2 \rangle_R$ in Eq. (2.17) is fitted to the observed MSD $\langle r^2 \rangle_{exp}$ in Fig. 2.10 to determine the intrinsic $\langle r^2 \rangle$ and T_D for this protein. The $\langle r^2 \rangle_R$ in Eq. (2.17) is the model which is equal to $\langle r^2 \rangle_{exp}$ in Eq. (1.1). Specially, at each temperature we determine the two parameters $\langle r^2 \rangle$ and λ by setting $\langle r^2 \rangle_R$ in Eq. (2.17) equal to the $\langle r^2 \rangle_{exp}$ of Fig. 2.10 for each instrument. The fitting values of $\langle r^2 \rangle_R$ are shown as the solid lines through the data points on the LHS in Fig. 2.10. The best fit values of the $\langle r^2 \rangle$ and λ are on the RHS in Fig. 2.10.

The intrinsic MSD $\langle r^2 \rangle$ is greater than the $\langle r^2 \rangle_{exp}$ measured by IN16 and IN6. The $\langle r^2 \rangle_{exp}$ measured by IN6 is equal to the half of the $\langle r^2 \rangle$, and the $\langle r^2 \rangle$ is 20% larger than the $\langle r^2 \rangle_{exp}$ measured by IN16. The intrinsic T_D for glutamate dehydrogenase is observed to be $T_D \simeq 150$ K. The temperature dependence of the intrinsic MSD $\langle r^2 \rangle$ is similar to that of the observed MSD $\langle r^2 \rangle_{exp}$ on IN16. The intrinsic T_D is also very close to the T_D observed on IN16 where $W = 1 \mu\text{eV}$. At temperatures higher than 200 K, an unexpected temperature dependence in the $\langle r^2 \rangle$ is observed. The $\langle r^2 \rangle$ decreases with increasing temperature for $T \gtrsim 200$ K. The value of the intrinsic MSD $\langle r^2 \rangle$ at for $T \gtrsim 200$ K is larger than that for $T \gtrsim 250$ K. This is unphysical. In order to analyse this effect, the intrinsic $\langle r^2 \rangle$ is kept constant at its 200-K value for temperatures $T \gtrsim 200$ K. This $\langle r^2 \rangle$ is shown on the RHS in Fig. 2.11. Using the $\langle r^2 \rangle$ on the RHS in Fig. 2.11, we again set the $\langle r^2 \rangle_R$ to the observed MSD $\langle r^2 \rangle_{exp}$ to obtain the appropriate λ . On the LHS in Fig. 2.11, the solid line represents the $\langle r^2 \rangle_R$ obtained using $\langle r^2 \rangle$ held constant for $T > 200$ K. In this case, the $\langle r^2 \rangle_R$ for $W = 1 \mu\text{eV}$ (IN16) continues to increase

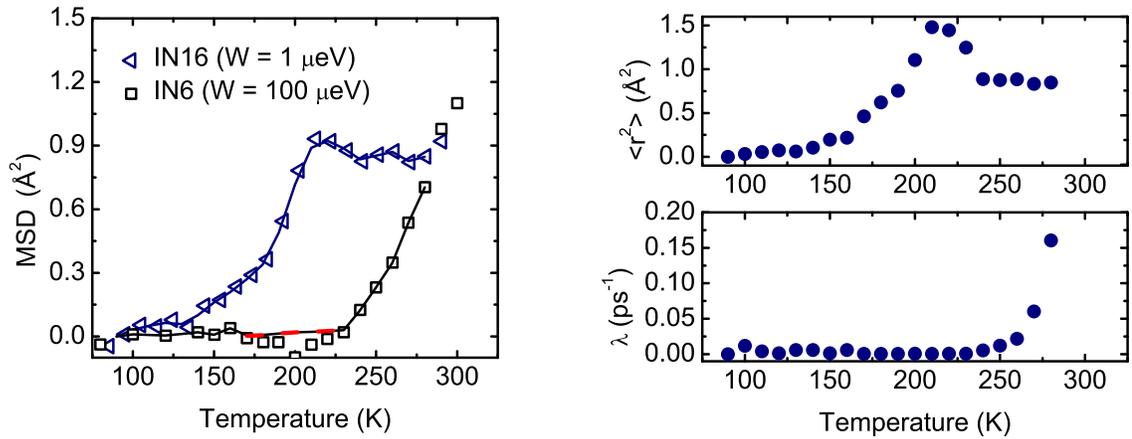


Figure 2.10: (LHS) MSD, $\langle r^2 \rangle_{exp}$, of glutamate dehydrogenase in CD_3OD/D_2O observed by Daniel et. al [17] on IN16 ($W = 1 \mu\text{eV}$) (triangles) and IN6 ($W = 100 \mu\text{eV}$) (squares). Eq. (2.17) for $\langle r^2 \rangle_R$ is fitted to the observed $\langle r^2 \rangle_{exp}$ and the fitted $\langle r^2 \rangle_R$ is shown as a solid line. (RHS) The intrinsic $\langle r^2 \rangle$ and λ obtained from the fit. The extracted intrinsic $\langle r^2 \rangle$ decreases above $T = 220$ K.

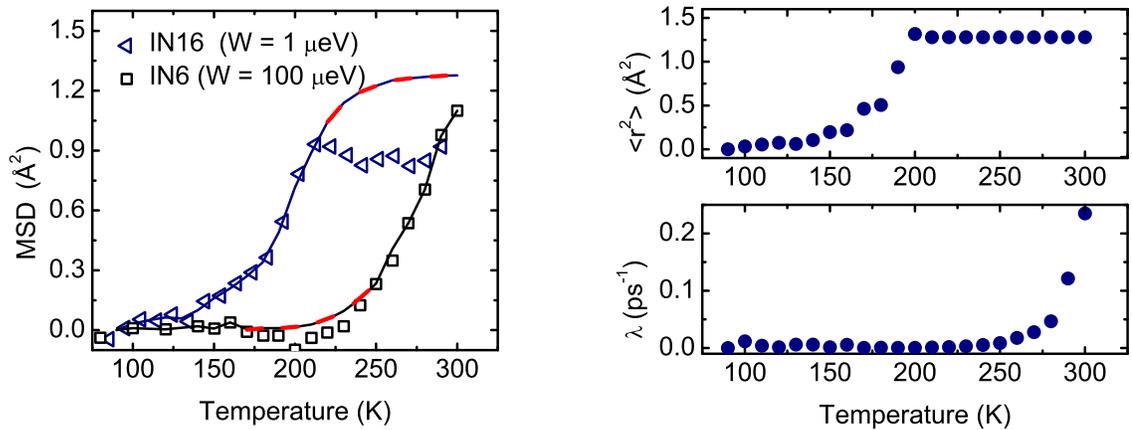


Figure 2.11: The MSD as in Fig. 2.9. In this case the intrinsic MSD $\langle r^2 \rangle$ is arbitrarily held constant at temperatures $T > 200\text{K}$. The solid lines show the values of $\langle r^2 \rangle_R$ obtained for IN16 and IN6 when the $\langle r^2 \rangle$ is held constant for $T > 200\text{K}$. The expected $\langle r^2 \rangle_R$ continues to increase above 200 K.

above $T = 200$ K, and reaches a plateau value above 250 K. This plateau of the $\langle r^2 \rangle$ at higher temperature appears to be unique for H in glutamate dehydrogenase. This result suggests that there is an issue with the sample or data on IN16 for $T > 200$ K.

The apparent T_D seen in the $\langle r^2 \rangle_{exp}$ is significantly affected by the energy resolution of the instrument employed. As shown in Figs. 2.10 and 2.11, it moves to higher temperatures for larger W . This property of the T_D is also shown in Sec. 2.3.1. The data in Fig. 2.9 shows that the larger the W , the higher the temperature that is needed to observe a marked increase in $\langle r^2 \rangle_{exp}$. The apparent T_D is $T_D \simeq 220$ K on IN10 while that on HER is $T_D \simeq 260$ K and no dynamical transition on by GP-TAS. In respect to these results, the intrinsic $\langle r^2 \rangle$ is always greater than the $\langle r^2 \rangle_R$ at a given temperature. For the large W , the $\langle r^2 \rangle_R$ is suppressed below the intrinsic $\langle r^2 \rangle$. Similarly, the intrinsic T_D obtained from the intrinsic $\langle r^2 \rangle$ is always lower than the observed T_D . For a finite energy resolution width W , the apparent T_D is shifted to a higher temperature above the intrinsic T_D . This suggests that the higher temperatures is needed to observe the DT on the MSD for a large energy resolution width W .

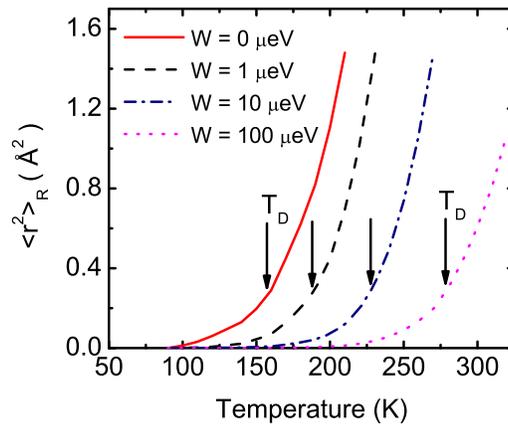


Figure 2.12: The resolution broadened MSD, $\langle r^2 \rangle_R$, for different resolution widths, W obtained from Eq.(2.17) illustrating the dependence of the apparent T_D on W . The intrinsic $\langle r^2 \rangle$ and λ of glutamate dehydrogenase, obtained from the fits to data by Daniel et al. [17] shown in Fig. 2.10, are used in Eq.(2.17).

The model $\langle r^2 \rangle_R$ in Eq. (2.17) may be used to clarify the dependence of the $\langle r^2 \rangle_R$ and the T_D on the instrument resolution width, W . In order to illustrate, the resolution dependence of the $\langle r^2 \rangle_R$ and T_D , the $\langle r^2 \rangle_R$ is calculated using by Eq. (2.17). Fig. 2.12 shows the $\langle r^2 \rangle_R$ versus temperature for different energy resolution widths, $W = 1, 10, 100 \mu\text{eV}$. The intrinsic $\langle r^2 \rangle$ of glutamate dehydrogenase shown in Fig. 2.11 is extrapolated to the higher temperature values and used as input in Eq. (2.17). A reasonable λ similar to that Fig. 2.11 is also used. Solid line in Fig. 2.12 represents the $\langle r^2 \rangle_R$ which is equal to the intrinsic MSD $\langle r^2 \rangle$. This intrinsic $\langle r^2 \rangle$ has an intrinsic $T_D \simeq 150 \text{ K}$. The larger the energy resolution width W , the smaller the $\langle r^2 \rangle_R$ obtained and the higher the T_D observed. For a finite energy resolution width, the $\langle r^2 \rangle_R$ always lies below the intrinsic MSD $\langle r^2 \rangle$. Similar to results in Figs. 2.10 and 2.11, the T_D moves to higher temperatures with increasing energy resolution width W in Fig. 2.12. Equivalently, the shift in the T_D could be explained as a $\langle r^2 \rangle_R$ that lies below $\langle r^2 \rangle$ at a given temperature. Hence, the reduction in $\langle r^2 \rangle_R$ and the apparent increase in T_D with increasing W are the same effect.

2.3.3 Sensitivity of $\langle r^2 \rangle$ to the model $C(t)$

In order to test the impact of the model of $C(t)$ and of the shape of the instrument resolution, we changed $C(t)$ from $C(t) = \exp(-\lambda t)$ to a Gaussian $C(t) = \exp(-\lambda^2 t^2/2)$, and the resolution function from $R(\omega) = \frac{1}{\pi} \frac{W}{W^2 + \omega^2}$ to a Gaussian $R(\omega) = \exp(-\omega^2/2W^2)/\sqrt{2\pi W^2}$. Using Gaussian $C(t)$ and $R(\omega)$, an analytic expression for $S_R^N(\mathbf{Q}, \omega = 0)$ is again obtained in *Appendix C*. Fig. 2.13 shows the comparison of the intrinsic $\langle r^2 \rangle$, obtained from the fit of the $S_R^N(\mathbf{Q}, \omega = 0)$ model in Eq. (D.5) to the normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$ data of hydrated SNase observed by Nakagawa et al. [73], with the resolution broadened MSD $\langle r^2 \rangle_R$ and the observed MSD $\langle r^2 \rangle_{exp}$. The solid dots represent the intrinsic MSD $\langle r^2 \rangle$ and this is greater than the $\langle r^2 \rangle_R$ and $\langle r^2 \rangle_{exp}$. The agreement between the $\langle r^2 \rangle_R$ and $\langle r^2 \rangle_{exp}$ shows that the fit to data is precise. The $\langle r^2 \rangle$ in Figs. 2.9 and 2.13 are barely distinguishable. This indicates that intrinsic $\langle r^2 \rangle$, which is obtained from the fit of the model to the experimental data, is not sensitive

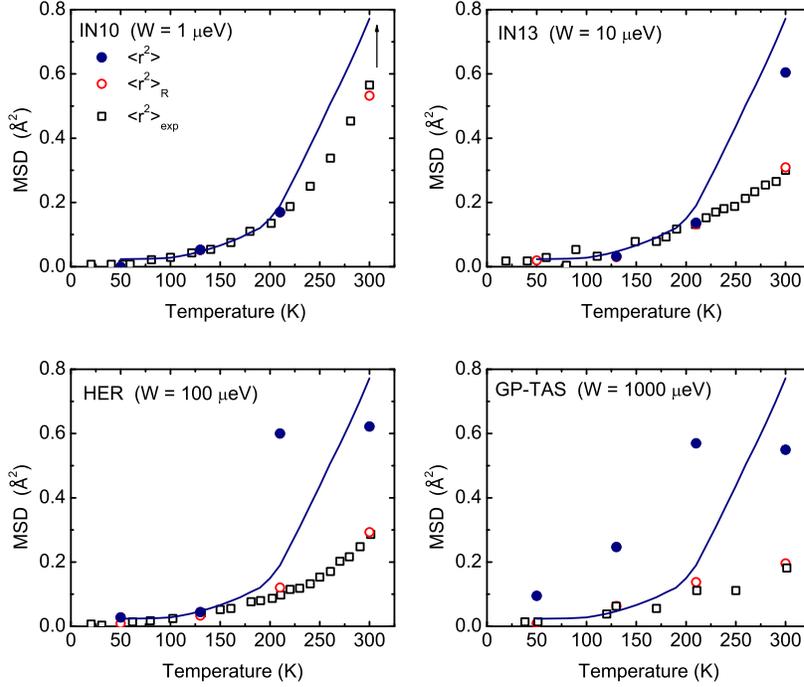


Figure 2.13: MSD, $\langle r^2 \rangle_{exp}$, of the hydrated SNase observed by Nakagawa et al. [73] on IN10 ($W = 1 \mu\text{eV}$), IN13 ($W = 10 \mu\text{eV}$), HER ($W = 100 \mu\text{eV}$) and GP-TAS ($W = 1000 \mu\text{eV}$) (open squares). The intrinsic MSD $\langle r^2 \rangle$ (solid circle) is obtained by fitting the observed DSF equation, where the Gaussian $C(t) = \exp(-\lambda^2 t^2/2)$ and $R(\omega) = \exp(-\omega^2/2W^2)/\sqrt{2\pi}W^2$ are used, to the elastic intensity data observed Nakagawa et al. [73]. The solid line is a guide to the eye to the $\langle r^2 \rangle$, the same for all instruments. The $\langle r^2 \rangle_R$ (open circles) is calculated from the Q^2 dependence of the observed DSF.

to the form of $C(t)$ and $R(\omega)$ used in the model at the present level of precision of the data.

As stated in Sec 2.2, the impact of a finite resolution width W is also seen in $I(\mathbf{Q}, t)$ given by Eqs. (2.8) and (2.16). The ratio of W to λ determines which motions contribute to the integral in the expression of $I_{inc}(\mathbf{Q}, t)$ in Eq. (2.8). The impact of the ratio of W to λ could be analysed in two limits : (i) $\lambda \gg W \Rightarrow \tau = 1/\lambda \ll \tau_R \sim 1/W$; the intrinsic correlations in the protein (in $C(t)$) decay rapidly in shorter time than τ_R , and all correlations in $C(t)$ are included in the integral expression of $I_{inc}(\mathbf{Q}, t)$

in Eq. (2.8), (ii) $\lambda \ll W \Rightarrow \tau = 1/\lambda \gg \tau_R \sim 1/W$; the integral in the expression of $I_{inc}(\mathbf{Q}, t)$ in Eq. (2.8) is cut off after time τ which is shorter than τ_R . In the case $\lambda \ll W$, the intrinsic correlations in the protein (in $C(t)$) decay slowly, then only the short time correlations in $C(t)$ are included in Eq. (2.8). Hence, $I_{inc}(\mathbf{Q}, t)$ is still dependent on W . The relaxation time $\tau = 1/\lambda$ changes from protein to protein, so that we expect the impact of finite resolution width W to vary from protein to protein.

The resolution broadened DSF $S_R^N(\mathbf{Q}, \omega = 0)$ in Sec. 2.2 is a very simple model. In this model, we neglect the contribution comes from ballistic propagation or vibration in $C(t)$, and a simple representation $C(t) = \exp(-\lambda t)$ is used to describe the relaxation of correlations in the protein. This $C(t)$ is appropriate for a single diffusion process. In order to improve the model, a stretched exponential $C(t) = \exp(-(\lambda t)^\beta)$ which represents several diffusion mechanisms contributing to $C(t)$ is used instead of a simple decay function $C(t) = \exp(-\lambda t)$. Then, we fit the model of $S_R^N(\mathbf{Q}, \omega = 0)$ with a stretched exponential $C(t) = \exp(-(\lambda t)^{0.5})$ to the observed normalized $S_R^{exp}(\mathbf{Q}, \omega = 0)$ in literature. However, we found that the experimental data was not significantly precise to distinguished between a simple exponential $C(t) = \exp(-\lambda t)$ and a stretched exponential $C(t) = \exp(-(\lambda t)^\beta)$. For these two $C(t)$ functions, which are a stretched exponential function and a simple exponential function, the intrinsic $\langle r^2 \rangle$ obtained are the same.

Chapter 3

LONG-TIME MEAN SQUARE DISPLACEMENTS IN PROTEINS

Building upon our work on proteins in Chap. 2, we expand our research to obtain the intrinsic long-time MSD from molecular dynamics (MD) simulations. This is the intrinsic long-time MSD that is independent of the simulation time. MD simulation is an important technique used to investigate the dynamics of proteins in ps-ns time-scale. Several MD simulations have been performed to understand the origin of elastic neutron scattering [3, 4, 41, 42, 43, 68, 71, 108]. As stated in Chap. 1, the contribution of time-dependent inelastic scattering to elastic scattering is one of the issues studied using MD simulation. The studies of the time-dependence of the elastic scattering opens the question whether an intrinsic long-time MSD, which is independent of the simulation time, can be obtained from simulations.

Our goal in this chapter is to propose a method to obtain the intrinsic long-time value of the MSD from finite-time simulations. The method used in this chapter is the same as that in Chap. 2. However, we modified the model for the global ISF $I(\mathbf{Q}, t)$ which is constructed in Chap. 2. The intrinsic long-time MSD $\langle r^2 \rangle$ is defined in the model $I(\mathbf{Q}, t)$. We fit the model $I(\mathbf{Q}, t)$ to the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ which is observed in neutron scattering experiments and calculated from simulations.

3.1 MD Simulation and MD Simulation System

MD simulation is used to study systems which have a large number of particles such as proteins, polypeptides and membranes. The MD simulation calculates the time-dependent behaviour of a molecular system. The particles are allowed to interact for a given time period. These interactions provide a view of the motions of particles and gives the configuration of the system in different time steps for a period of time.

In MD simulations, the trajectories of particles are determined by numerically solving Newton's equations of motions for a system of interacting particles. A force field is required to determine the forces between particles and the potential energy. Force fields are defined as a summation of bonded and non-bonded forces. Bonded forces are chemical bonds, bond angles and bond dihedrals, non-bonded forces are van der Waals forces and electronic charges.

It is not possible to find the properties of a large system analytically, but MD simulations eliminate this problem by using numerical methods. Hence, MD simulation is one of the principal tools in the theoretical study of a large system like a protein. It is also important to understand the origin of experimental results. Calculating experimental data using MD simulation is useful to understand what is measured in neutron scattering experiments. Although MD simulation does not have the limitation of resolution which exists in neutron scattering experiments, it has a time limitation.

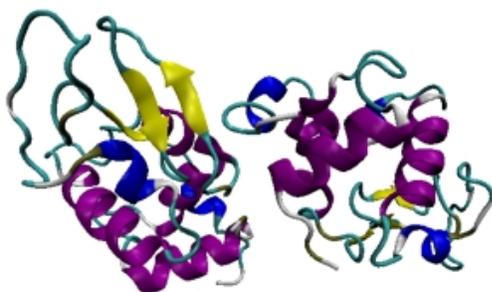


Figure 3.1: Two lysozyme molecules of random relative orientation selected by GRO-MACS.

We simulated a hydrated powder of the protein lysozyme. A configuration of a system which includes molecular positions and a potential function is required for a MD simulation. The system was simulated using GROMACS 4.5.1, the MD engine [45]. In GROMACS, `pdb2gm`, `genbox`, `editconf` and `genion` commands are used to generate the topology and coordinates files. The `pdb` file for lysozyme is obtained from Protein Data Bank (protein databank code: 1AKI[1]). As shown in Fig.3.1, the system

simulated consisted of two randomly oriented lysozyme molecules designed by using `pdb2gm` command.

However, the configuration of the system obtained from `pdb2gm` command does not include any solvent (water). The system including two lysozyme molecules was placed in a simulation box of dimensions $6.5 \text{ nm} \times 3.4 \text{ nm} \times 3.6 \text{ nm}$ and surrounded by 636 water molecules, which corresponds to the hydration level $h = 0.4 \text{ g water/g protein}$, by using `genbox` command. `Editconf` command edits the simulation system. Periodic boundary conditions can not handle with a system which is electrically unstable. Hence, `genion` command is used to add chlorides (CL^-) ions to make the simulation system which has zero ionic states. For our system, eighteen (16) CL^- are substituted with water molecules to the simulation system.

The box was replicated using periodic boundary conditions to mimic the environment of an experimental powder sample. Hence, every molecule can move and cross over from top to bottom and from left to right. Similar simulation systems are discussed in the literature [47, 46, 59, 75, 21, 97]. After the configuration of the system is designed, two commands, `grompp` and `mdrun`, are used to run simulations. `Grompp` command makes a binary file as a run input file from the input files of configuration and topology file. The main calculation procedure in GROMACS is run by `mdrun` command. It performs a simulation based on the binary file generated from `grompp`. Output file obtained from `mdrun` command is trajectory file.

The `pdb` file, `1AKI.pdb`, includes only the information about protein, but not any potential and energy related information. The OPLS-AA force field [52] was used for the protein and the TIP4P force field [49] for the water. These force fields are used to reproduce the experimental self-diffusion coefficient of protein hydration water by Lagi et al. [59]. To have a higher computational speed, a set cutoff radii is available in MD simulations. The van der Waals interaction was truncated at 1.4 nm, and the electrostatic interaction was represented using the Particle Mesh Ewald method [30] with a real-space cutoff of 0.9 nm. All bonds including hydrogen bonds were constrained with a linear constraints solver algorithm (LINCS) [44]. The energy of

the system was first minimized using 50000 steepest descent steps. The system was then equilibrated in the NVT (number of particles-volume-temperature) ensemble at each temperature investigated for 10 ns and in the NPT (number of particles-pressure-temperature) ensemble at 1 bar for 10 ns. The Nose-Hoover algorithm [48] with a coupling time $\tau = 1$ ps and the Parrinello-Rahman algorithm [82] with a coupling time $\tau = 3$ ps were used for the temperature coupling and pressure coupling, respectively.

Simulations of 100 ns length were performed at 18 different temperatures between 80 K and 300 K. Simulations of 1 μ s were made at 5 temperatures, at 100 K and then in steps of 50 K to 300 K. The data was collected every 10 ps at each temperature for both simulations. Simulation trajectories at each temperature includes the coordinates of nuclei in proteins at each 10 ps. In the data presented here, the protein coordinates were superimposed onto the starting structure before analysis to remove the global translation of the protein molecule in the 1 μ s simulation data, but not in the 100 ns simulation data presented in this chapter.

3.2 The model of the ISF $I(Q, t)$

In neutron scattering experiments, the observed data was not sufficiently precise to distinguish between a simple and a more sophisticated model. In contrast, data calculated in MD simulation is discriminative. This requires some changes in the time-dependent part in the model. In Chap. 2, the model of the incoherent ISF $I(\mathbf{Q}, t)$ is constructed. Here, we modify the model of $I(\mathbf{Q}, t)$ as defined in Eq. (2.14). A stretched exponential function $C(t) = \exp(-(\lambda t)^\beta)$ is used instead of a simple decay function $C(t) = \exp(-\lambda t)$ to improve the model. A stretched exponential function $C(t) = \exp(-(\lambda t)^\beta)$ represents several diffusion mechanisms contributing to $C(t)$. The model of $I(\mathbf{Q}, t)$ in Eq. (2.14) becomes

$$I(\mathbf{Q}, t) = I_\infty(\mathbf{Q}) + (1 - I_\infty(\mathbf{Q})) \exp(-(\lambda t)^\beta), \quad (3.1)$$

which is a representation of several motional decay processes and has the correct limits at $t = 0$ and $t = \infty$. Similar to the model $I(\mathbf{Q}, t)$ in Chap. 2, the intrinsic, long-time

MSD $\langle r^2 \rangle$ is defined in the infinite time limit of the incoherent ISF, $I_\infty(\mathbf{Q}) = I(\mathbf{Q}, t = \infty) = \exp(-\frac{1}{3}Q^2\langle r^2 \rangle)$.

Then, we calculate the incoherent ISF, $I_{inc}(\mathbf{Q}, t)$, defined in Eq. (2.2). The positions $r_i(t)$ of each nucleus i in the protein are used as inputs in Eq. (2.2). The $r_i(t)$ is generated in the two MD simulations described in Sec. 3.1, one of length 100 ns and the other 1 μ s. Using the $r_i(t)$ as an input parameter in SASSENA [www.sassena.org], the $I_{inc}(\mathbf{Q}, t)$ is calculated directly for times out to 100 ns and 1 μ s, respectively. To improve statistics, each simulation trajectory is divided into segments. The maximum segment number is 9000 for the 100 ns MD simulation and 90000 for the 1 μ s MD simulation. Then, $I_{inc}(\mathbf{Q}, t)$ is recalculated as an average over $I_{inc}(\mathbf{Q}, t)$ in these segments. In this way, $I_{inc}(\mathbf{Q}, t)$ was calculated out to 1 ns from the 100 ns MD simulation data and out to 10 ns from the 1 μ s MD simulation data.

3.3 Application of the model of the ISF $I(\mathbf{Q}, t)$

In this section, our goal is to apply a method proposed in Sec. 3.2 to obtain the intrinsic, long-time MSD in proteins from finite time simulations. The method consists of calculating the ISF $I_{inc}(\mathbf{Q}, t)$ from a simulation and fitting a model $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$. The model of $I(\mathbf{Q}, t)$ includes three fitting parameters: the intrinsic MSD, $\langle r^2 \rangle$, defined in Eq. (2.11), the relaxation parameter λ and the stretched exponential parameter β which are both defined in Eq. (3.1). In Sec. 3.3.1 and Sec. 3.3.2, our goal is to determine the intrinsic MSD $\langle r^2 \rangle$ of H in lysozyme and to obtain values for the relaxation parameters λ and β . Thus, we fit the model $I(\mathbf{Q}, t)$ given by Eq. (3.1) to the ISF data, $I_{inc}(\mathbf{Q}, t)$, calculated using the 100 ns and 1 μ s simulation data.

3.3.1 100 ns MD simulation

We calculate the ISF, $I_{inc}(\mathbf{Q}, t)$, in the time range $0 < t < 1$ ns using the 100 ns simulation data as an input in Eq. (2.2). Fig. 3.2 shows the fit of the model of $I(\mathbf{Q}, t)$ to

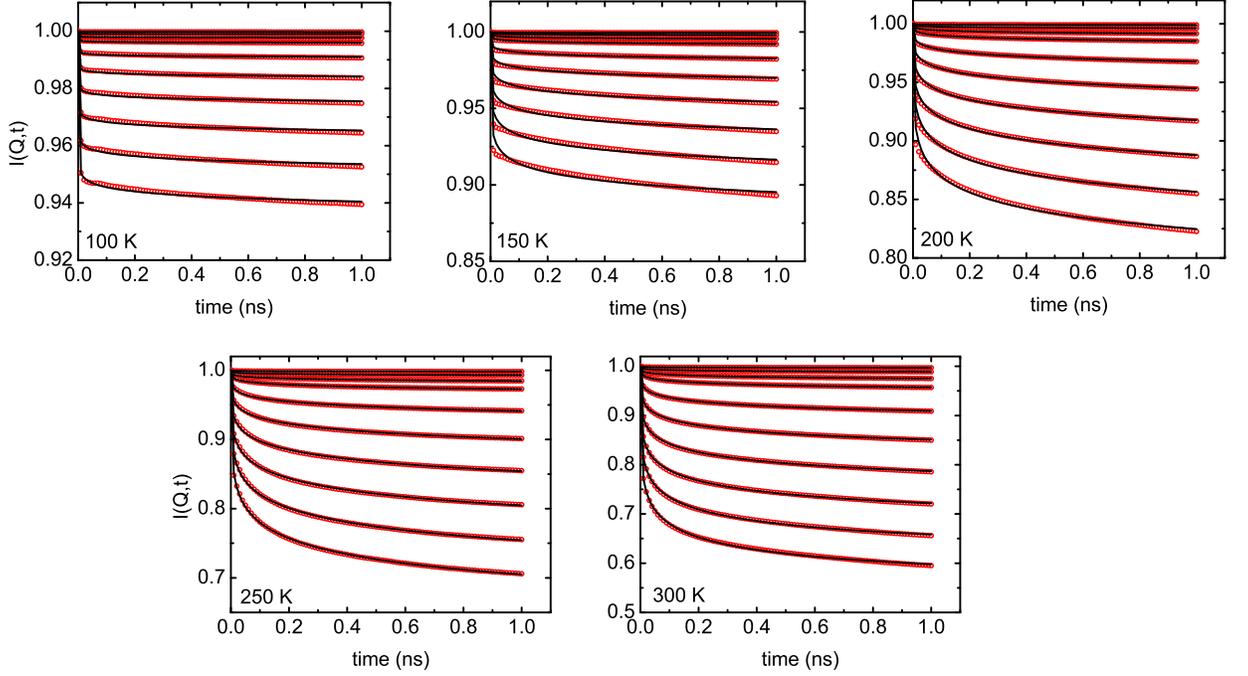


Figure 3.2: The intermediate scattering function (ISF), $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 1$ ns of hydrated lysozyme ($h = 0.4$) obtained from a 100 ns MD simulation (open red circles) and fits of the model $I(\mathbf{Q}, t)$ in Eq. (3.1) (blue solid lines) to the $I_{inc}(\mathbf{Q}, t)$ at 100 K, 150 K, 200 K, 250 K and 300 K. From top to bottom, Q : 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4 and 1.6 \AA^{-1} .

the calculated ISF, $I_{inc}(\mathbf{Q}, t)$. The red open circles represent the $I_{inc}(\mathbf{Q}, t)$ data. Although $I_{inc}(\mathbf{Q}, t)$ was calculated at 18 temperatures, only 5 temperatures are shown in Fig. 3.2. $I_{inc}(\mathbf{Q}, t)$ data is presented for 10 different Q values for each temperature. The Q values from top to bottom are 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4 and 1.6 \AA^{-1} . The solid lines are the fits of the model $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$. The fits are good, except low temperatures. At low temperatures, the $I_{inc}(\mathbf{Q}, t)$ changes slightly with time so that the small β values is used to have better fits. Because of the poor fits, the parameters λ and β that appear in the relaxation function $C(t)$ are not well determined at temperatures below approximately 170 K.

Fig. 3.3a shows that the best fit values of the fitting parameter $\langle r^2 \rangle$ versus Q

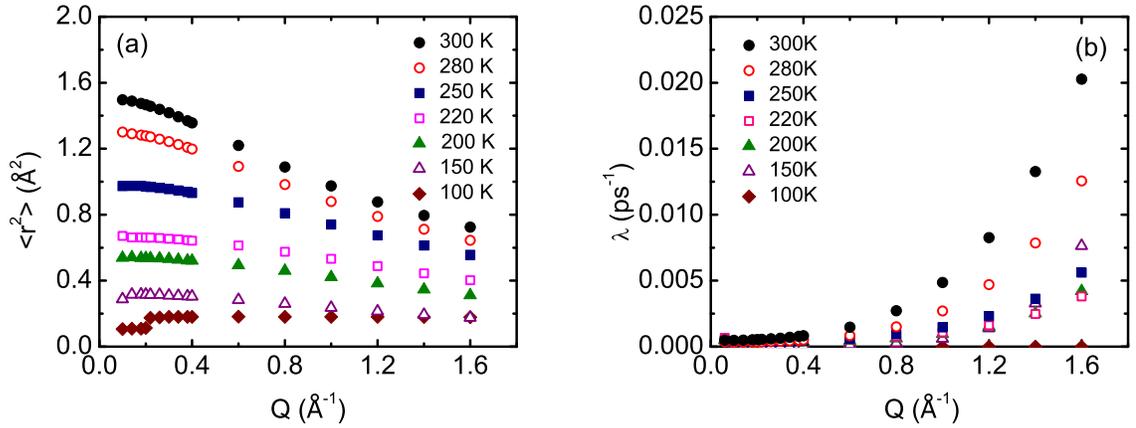


Figure 3.3: Parameters of the model $I(\mathbf{Q}, t)$ of Eq. (3.1) obtained from fits of the model to the simulations shown in Fig. 3.2: (a) The intrinsic MSD, $\langle r^2 \rangle$ and (b) the relaxation parameter, λ , versus Q at temperatures 100 K to 300 K.

for six different temperatures. Temperatures from bottom to top are 100, 150, 200, 220, 250, 280 and 300 K. The intrinsic MSD $\langle r^2 \rangle$ is found to be Q -dependent. The Q -dependence of the $\langle r^2 \rangle$ is much stronger at higher temperatures. The $\langle r^2 \rangle$ is larger and approximately independent of Q at low Q . This Q dependence is similar to the Q -dependence of the observed $\langle r^2 \rangle_{exp}$, which arises from the Q -dependence of the observed data $S_R^{exp}(Q, \omega = 0)$. The origin of the Q -dependence of the MSD will be discussed in Chap. 4.

Figs. 3.3b and 3.4a show the best fit values of the fitting parameter λ versus Q and T , respectively. In Fig. 3.3b, we see that λ is also Q dependent and increases with increasing Q . λ is approximately proportional to Q^2 , and is similar to the Q -dependence of λ that is observed in other simulations [12, 97]. The stretched exponential parameter β defined in Eq. (2.1) is weakly dependent on Q . For the fits in Fig. 3.2, β is averaged over several Q values with some adjustments to obtain smooth behavior as a function of temperature. Fig. 3.4b shows the β values which we used versus temperature. As

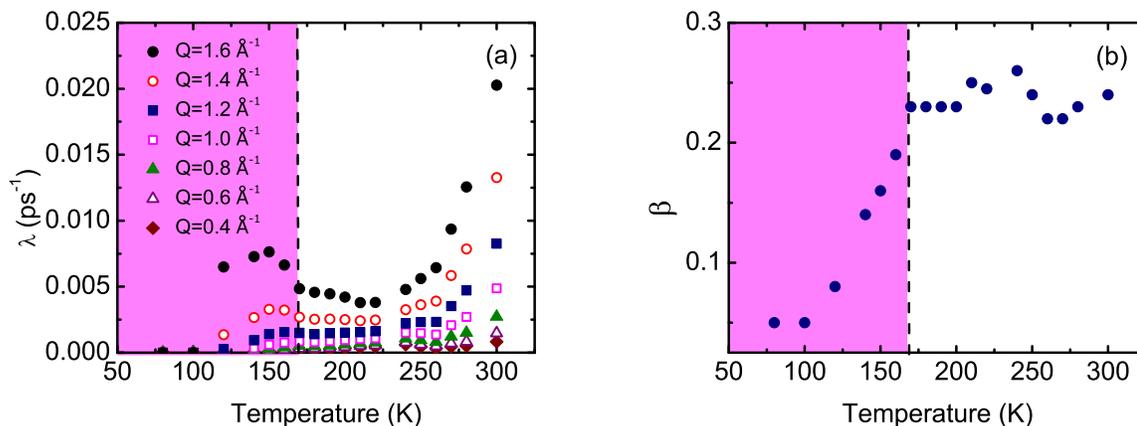


Figure 3.4: As Fig. 3.3 for (a) the relaxation parameter λ and (b) the stretched exponential parameter β versus temperature. From bottom to top, Q : 0.4, 0.6, 0.8, 1, 1.2, 1.4 and 1.6 \AA^{-1} .

stated above, the λ and β are not well determined at temperatures below approximately 170 K. The reason is the time dependence of the $I_{inc}(\mathbf{Q}, t)$ at low temperatures; $I_{inc}(\mathbf{Q}, t)$ decreases rapidly over a short time t and thereafter changes slowly. In spite of the uncertainties in the λ and β , the $\langle r^2 \rangle$ is well determined at low temperatures. The $\langle r^2 \rangle$ represents the value of $I_{inc}(\mathbf{Q}, t)$ at long t .

The intrinsic MSD $\langle r^2 \rangle$ obtained from the fits in Fig 3.2 versus temperature at $Q = 0.2 \text{\AA}^{-1}$ is shown Fig. 3.5. The red solid circles represent the $\langle r^2 \rangle$ and the dashed line guides the eye through the $\langle r^2 \rangle$. The $\langle r^2 \rangle$ versus temperature shows a clear break in slope at around $T = 140$ K and $T = 220$ K. The break at around 140 K has been seen previously in simulations. This break arises from the activation of the dynamics of hydrophobic groups, i.e. the onset of proline puckering and the rotation of methyl groups at around 140 K [87, 71]. A break in slope of $\langle r^2 \rangle$ versus temperature near 140 K has also been observed in several proteins. A second break in slope indicates a dynamical transition (DT), at $T_D \simeq 220$ K, associated with the onset of new larger amplitude motions of hydrophilic groups in which the hydration water plays a determining role. The intrinsic, long time $\langle r^2 \rangle$ shows the onset of both

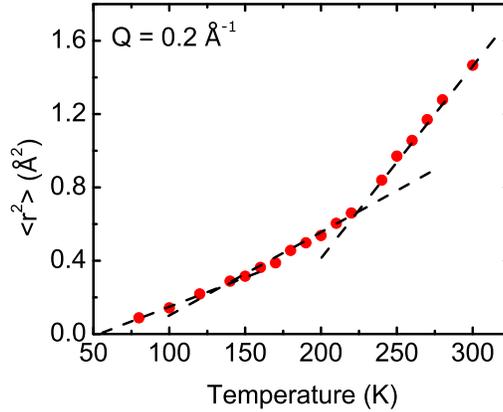


Figure 3.5: The intrinsic MSD, $\langle r^2 \rangle$, defined in Eq. (2.11), obtained from fits of the model $I(\mathbf{Q}, t)$, Eq. (3.1) to simulations of $I_{inc}(\mathbf{Q}, t)$ at $Q = 0.2 \text{\AA}^{-1}$. The intrinsic $\langle r^2 \rangle$ shows a break in slope at $T \simeq 140$ K and $T \simeq 220$ K.

hydrophobic and hydrophilic (DT) motions.

3.3.2 1 μ s MD simulation

Next, we focus on the MD simulation which is ten times longer. Similarly the previous section, the incoherent ISF, $I_{inc}(\mathbf{Q}, t)$, is calculated using the $r_i(t)$ generated in the 1 μ s simulations in Eq. (2.2). The 1 μ s simulation is performed at five different temperatures; 100, 150, 200, 250 and 300 K. Fig. 3.6 shows the fit of the model of $I(\mathbf{Q}, t)$ to the calculated ISF, $I_{inc}(\mathbf{Q}, t)$, $0 < t < 10$ ns. The red open circles represent the $I_{inc}(\mathbf{Q}, t)$ data at five different temperatures. For each temperature, there are $I_{inc}(\mathbf{Q}, t)$ data for 10 different Q values; 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4 and 1.6 \AA^{-1} . The solid lines in Fig. 3.6 are again the fits of the model $I(\mathbf{Q}, t)$ given by Eq. (3.1) to determine $\langle r^2 \rangle$ and λ as free fitting parameters. β values is chosen as similar to the β in Fig. 3.4. The fits are better at the higher temperatures, specially at 200 and 250 K. In the fits to $I_{inc}(\mathbf{Q}, t)$ for times out to 10 ns, we see that the fit is better at 200 K than at 300 K.

Fig. 3.7 shows the best fit values of the intrinsic MSD $\langle r^2 \rangle$ and λ versus Q ,

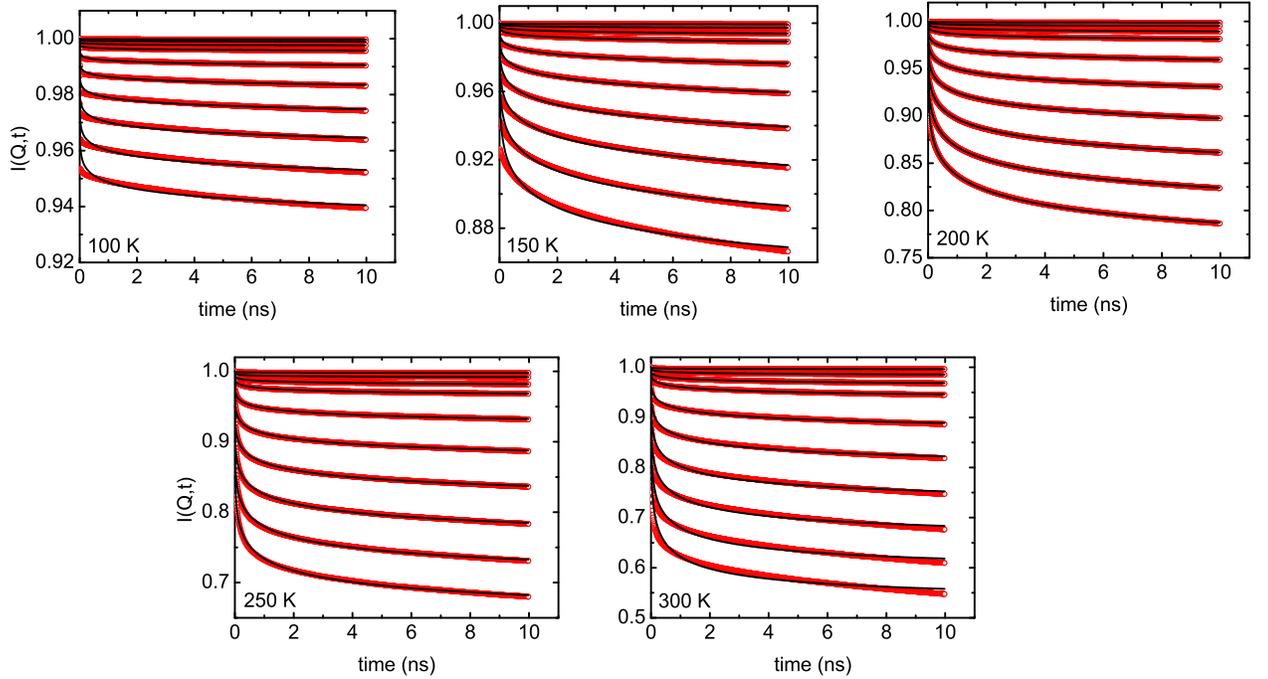


Figure 3.6: The calculated $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 10$ ns of hydrated lysozyme obtained from a $1 \mu\text{s}$ MD simulation (open circles) and fits of the model $I(\mathbf{Q}, t)$ in Eq. (3.1) to the data (solid lines) at 100 K, 150 K, 200 K, 250 K and 300 K. From top to bottom, Q : 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4 and 1.6 \AA^{-1} .

which we obtained from the fits in Fig. 3.6. Similar to Fig. 3.3a, the $\langle r^2 \rangle$ decreases with increasing Q . At 300 K, the $\langle r^2 \rangle$ at $Q = 0.1 \text{ \AA}^{-1}$ is almost twice of the $\langle r^2 \rangle$ at $Q = 1.6 \text{ \AA}^{-1}$, however this difference decreases with increasing temperature. It is also noted that the Q -dependence significantly decreases at low Q -values, except at 300 K. λ is approximately proportional to Q^2 as found in the shorter simulation. For the $1 \mu\text{s}$ simulation, one of the important differences which we observed is that the absolute values of λ obtained from the longer time (10 ns) are significantly smaller than those obtained from the shorter time (1 ns) (compare Figs. 3.3b and 3.7b). For example, the value of λ at $Q = 1.6 \text{ \AA}^{-1}$ and at 300 K is 0.008 ps^{-1} in Fig. 3.7b and 0.02 ps^{-1} in Fig. 3.3b.

The best fit values of the fitting parameter λ and β versus temperature are shown

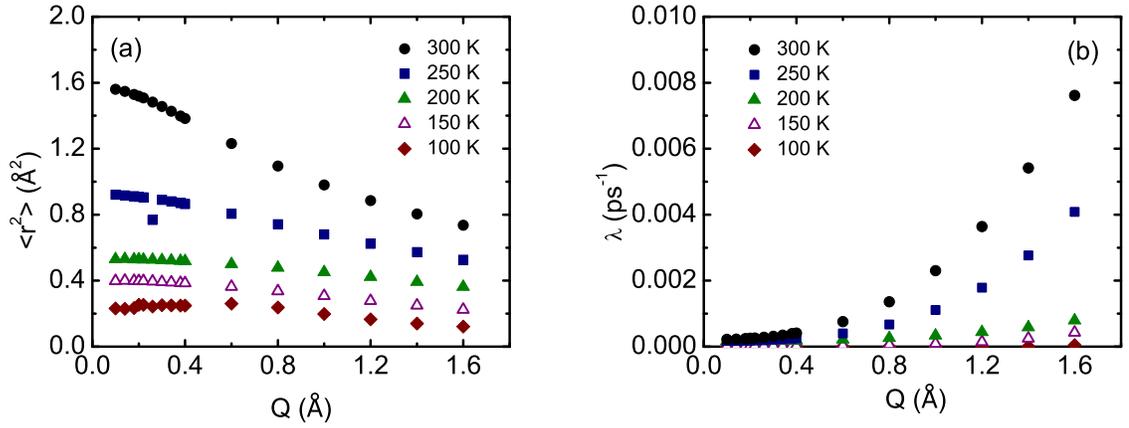


Figure 3.7: Parameters of the model ISF, Eq. (3.1), obtained from the fits of $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 10$ ns shown in Fig. 3.6: (a) the intrinsic MSD, $\langle r^2 \rangle$, and (b) the relaxation parameter, λ , versus Q at five temperatures: 100 K, 150 K, 200 K, 250 K and 300 K.

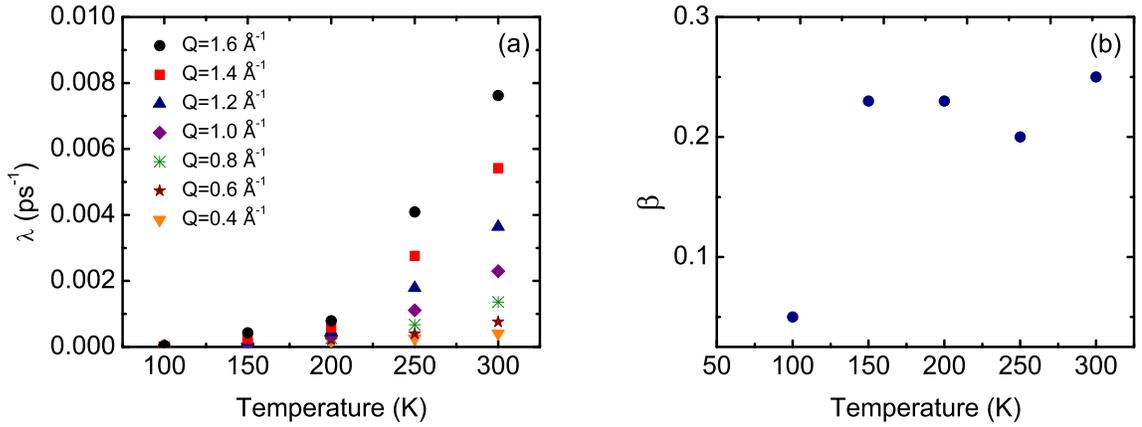


Figure 3.8: Parameters of the model ISF, Eq. (3.1), obtained from the fits of $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$, for $0 < t < 10$ ns shown in Fig. (3.6): (a) the relaxation parameter, λ , and (b) the stretched exponential parameter, β , versus temperatures. The Q values from bottom to top is 0.4, 0.6, 0.8, 1.0, 1.2, 1.4 and 1.6 Å⁻¹.

in Fig. 3.8. Ten times longer simulation allows to determine the fitting parameter λ well, even at temperatures below approximately 170 K. This suggests that $I_{inc}(\mathbf{Q}, t)$ data in

longer time is required to determine the small λ values in the model $I(\mathbf{Q}, t)$. However, we still use small β values at 100 K to have a better fit. The values of β are similar for the two simulation times as shown in Figs. 3.4b and 3.8b.

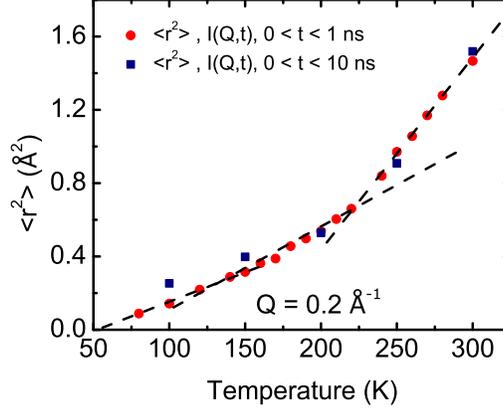


Figure 3.9: The intrinsic MSD $\langle r^2 \rangle$ versus temperature obtained from fits to the $I_{inc}(\mathbf{Q}, t)$ at $Q = 0.2 \text{\AA}^{-1}$ obtained from (1) 100 ns (solid circles) and (2) 1 μs MD simulations (solid squares). The $\langle r^2 \rangle$ is largely independent of the simulation time fitted.

The absolute values of the $\langle r^2 \rangle$ are also consistent with those obtained from the 100 ns simulation except, possibly, at 100 K. Fig. 3.9 compares the values of the $\langle r^2 \rangle$ obtained from the fits to $I_{inc}(\mathbf{Q}, t)$ for times t out to 1 ns and 10 ns. The red solid circles and the blue squares represent the intrinsic MSD $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ in Fig. 3.2 and Fig. 3.6, respectively. The values of the $\langle r^2 \rangle$ obtained from the 1 μs simulation are consistent with those obtained from the 100 ns simulation, except at 100 K. This consistency indicates that new motional process does not observed the simulations between 100 ns and 1 μs . These results also show that (i) $\langle r^2 \rangle$ is a long time ($t \rightarrow \infty$) intrinsic MSD and is independent of the time interval of the data, and (ii) the DT is observed at around 220 K although the $\langle r^2 \rangle$ is independent of simulation time and energy resolution. This suggests that a DT is not simply an artificial result of finite instrument resolution and of time window limitation. In respect of these results,

it can be said that the DT which is observed in the $\langle r^2 \rangle$ at 220 K is an intrinsic property of lysozyme.

3.4 Mean Square Displacement

We compare the intrinsic MSD $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ with the simulated MSD $\Delta^2(t)/2$ and the resolution broadened MSD $\langle r^2 \rangle_R$ to clarify the difference between them. In Sec. 3.4.1, we focus on the temperature and time dependence of the simulated MSD. In Sec. 3.4.2, we compare the $\langle r^2 \rangle$ with the resolution broadened MSD $\langle r^2 \rangle_R$ of lysozyme obtained previous researches.

3.4.1 Simulated Mean Square Displacement

The simulated MSD is one of the important quantities which can be calculated directly from simulations. The simulated MSD is calculated from

$$\Delta^2(t) = \langle [r(t) - r(0)]^2 \rangle \equiv \frac{1}{N} \sum_{i=1}^N \langle [r_i(t) - r_i(0)]^2 \rangle, \quad (3.2)$$

where $r_i(t)$ is the position of nucleus i in the protein at time t . The $\Delta^2(t)$ depends on the time, and increases with increasing time. The correlation between the positions of nuclei, $r_i(t)$ and $r_i(0)$, decreases with time and vanishes in longer times, $t \rightarrow \infty$. In this case, $\langle r_i(t)r_i(0) \rangle$ goes to zero and the expression $\Delta^2(t)$ in Eq. (3.2) reduces to

$$\Delta^2(t \rightarrow \infty) = \langle r^2(\infty) \rangle + \langle r^2(0) \rangle = 2\langle r^2 \rangle_{MD}. \quad (3.3)$$

By using the $r_i(t)$ generated by MD simulation as an input in Eq. (3.2), an $\Delta^2(t)$ for H in proteins can be defined and calculated. We calculated the simulated MSD, $\Delta^2(t)$, of lysozyme for 18 different temperatures out to 1ns for the 100 ns MD simulation, and for 5 different temperatures out to 10 ns for the 1 μ s MD. It is noted that the $\langle r^2 \rangle_{MD}$ in Eq. (3.3) is not the same as the intrinsic $\langle r^2 \rangle$ defined in terms of $I(\mathbf{Q}, t)$ in Eq. (2.11) (see *Appendix D*). As stated in Chap. 2, the largest contribution to the calculated ISF $I_{inc}(\mathbf{Q}, t)$ comes from H nuclei in proteins. Hence, the simulated MSD of lysozyme is only calculated for the non-exchangeable H nuclei in the lysozyme.

The non-exchangeable H does not exchange position with H in the hydration water. In the simulated MSD calculations, we excluded the H in the hydration water and the H in the protein that can exchange positions with H in the hydration water (the exchangeable H) from the sum in Eq. (3.2). The time dependence of the $\Delta^2(t)$ are compared with the intrinsic MSD $\langle r^2 \rangle$ defined in Eq. (2.11). In the model $I(\mathbf{Q}, t)$, we neglect the dynamical diversity of H in proteins. Thus, the $\langle r^2 \rangle$ and $\Delta^2(\infty)/2$ will be the same only if all H are in identical environments in the protein. If there is a dynamical heterogeneity, the $\langle r^2 \rangle$ and $\langle r^2 \rangle_{MD} = \Delta^2(\infty)/2$ will differ.

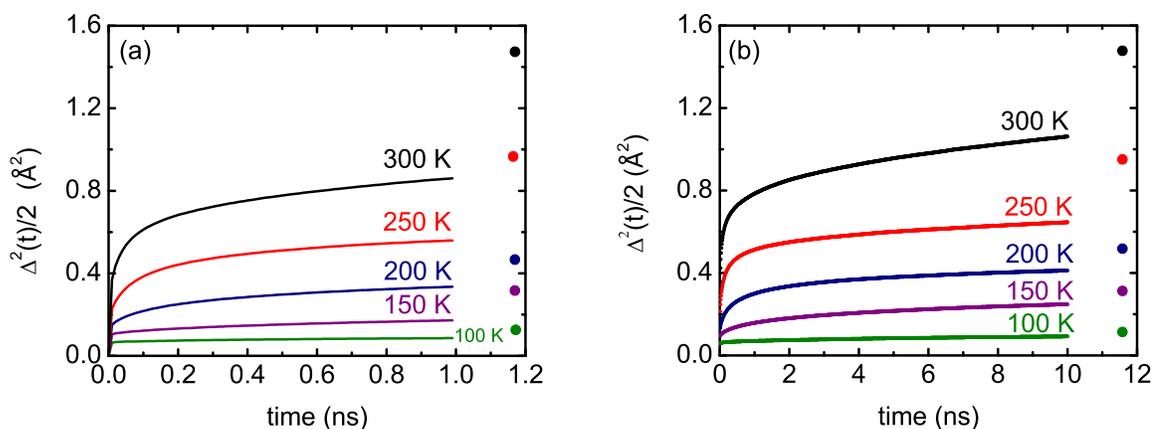


Figure 3.10: The MSD $\Delta^2(t)/2$ defined in Eq. (3.2) of non-exchangeable hydrogen versus time at five different temperatures, 100 K to 300 K (a) for 100 ns MD simulation (b) 1 μ s MD simulation. The dots are the corresponding intrinsic MSD $\langle r^2 \rangle$ at each temperature.

Fig. 3.10 shows $\Delta^2(t)/2$ versus time obtained from the 100 ns and 1 μ s simulations. The $\Delta^2(t)/2$ is calculated out to $t = 1$ ns from 100 ns simulation and out to $t = 10$ ns from 1 μ s simulation. The solid circles represent the intrinsic MSD $\langle r^2 \rangle$ at $Q = 2 \text{ \AA}^{-1}$. The $\langle r^2 \rangle$ from the 1 μ s simulation is slightly larger at 100 K and 150 K (see Fig. 3.9). Fig. 3.10b shows that the $\Delta^2(t)$ appears to have converged after 10 ns, and the $\Delta^2(t = 10 \text{ ns})$ and the $\langle r^2 \rangle$ is quite similar at $T = 100$ K. In contrast, at higher temperatures, the $\Delta^2(t)$ has clearly not reached its terminal value ($t = \infty$) after 10

ns. For example, at $T = 300$ K, the intrinsic $\langle r^2 \rangle$ is approximately 40 – 50 % larger than $\Delta^2(t = 1\text{ns})/2$ and 30 – 40 % larger than $\Delta^2(t = 10\text{ns})/2$. Similarly, the results of previous simulations show that the $\Delta^2(t)$ does not reach to a constant, infinite time value within accessible simulation times of 1 to 100 ns [43, 40, 8]. The reason of this is that the correlation $\langle r(t)r(0) \rangle$ does not have vanished. The infinite time value of a simulated MSD $\langle r^2 \rangle_{MD} = \Delta^2(t \rightarrow \infty)/2$ cannot be calculated within currently accessible simulation times. These results indicate that the $\langle r^2 \rangle$ is the intrinsic long time value of the MSD ($t = \infty$).

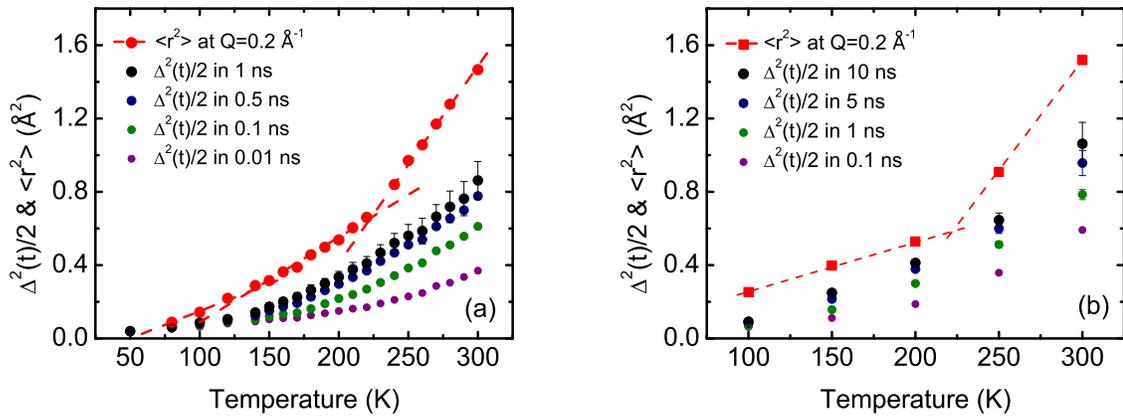


Figure 3.11: Comparison of the intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ and the MSD $\Delta^2(t)/2$ out to times (a) 0.01, 0.1, 0.5 and 1 ns obtained from the 100 ns MD simulation and (b) out to times 0.1, 1, 5 and 10 ns, obtained from the $1 \mu\text{s}$ MD simulation.

Furthermore, we compare the intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ with the $\Delta^2(t)/2$ at different times t obtained from the 100 ns (Fig. 3.11a) and $1 \mu\text{s}$ (Fig. 3.11b) simulations. The error bars on the $\Delta^2(t)/2$ in 1 ns in Fig. 3.11a and in 10 ns Fig. 3.11b correspond to 80% confidence intervals of the $\Delta^2(t)/2$ calculated in 100 different segments in the simulations. The $\Delta^2(t)/2$ lies well below the intrinsic $\langle r^2 \rangle$, especially at high temperature. At 300 K, the $\Delta^2(t)/2$ in 1 ns in Fig. 3.11a is almost half of the intrinsic long-time MSD $\langle r^2 \rangle$, and the $\langle r^2 \rangle$ is 30-40% larger than the $\Delta^2(t)/2$ in

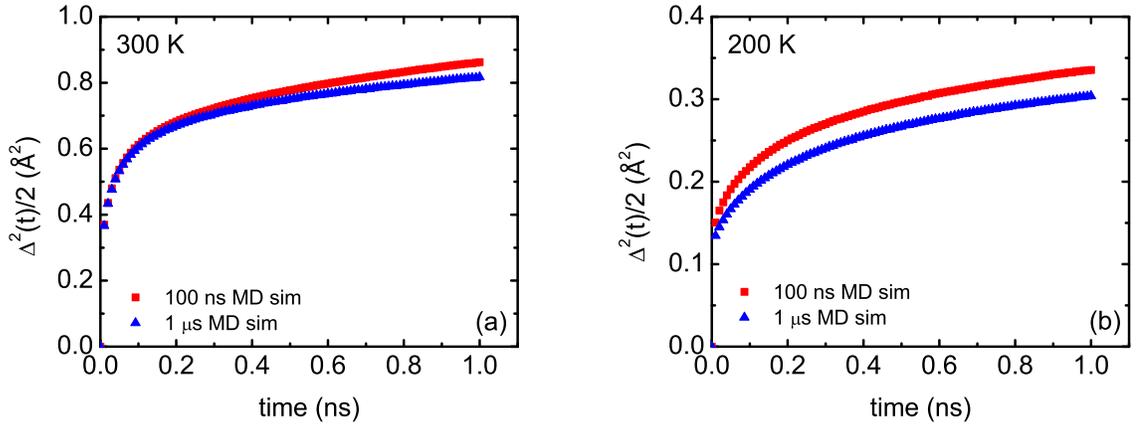


Figure 3.12: The MSD $\Delta^2(t)/2$ of non-exchangeable hydrogen versus time up to 1 ns as calculated from MD simulations of 100 ns (red solid squares) and 1 μs (blue solid triangles) at (a) 300 K and (b) 200 K.

10 ns in Fig. 3.11b. Similar to Fig. 3.10, it is clearly seen that the $\Delta^2(t)$ does not reach its long-time converged value after $t = 10$ ns, specially at higher temperatures in Fig. 3.11. For example, at 300 K the increase in the $\Delta^2(t)/2$ between 1 ns and 10 ns is approximately the same as between 0.1 ns and 1 ns and this suggests that convergence is very slow.

Fig. 3.12 shows the comparison of the $\Delta^2(t)$ over the time range $0 < t < 1$ ns which is calculated from the 100 ns and 1 μs simulations, at 200 K and 300 K. The $\Delta^2(t)$ obtained from data taken out to 1 μs is somewhat smaller than that obtained from the 100 ns simulation, approximately 5-10% small. This suggests that there may be a structural change in the time scale between 100 ns and 1 μs . However, these small differences do not appear to affect $I_{inc}(\mathbf{Q}, t)$ nor the fitted intrinsic $\langle r^2 \rangle$ significantly. The consistency between the two intrinsic $\langle r^2 \rangle$ obtained from the 100 ns and 1 μs simulations supports this result.

3.4.1.1 Center of mass (CM) effect on simulated data

In Sec. 3.4.1, the $\langle r^2 \rangle$, which is obtained from a fit of the model $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$ including all nuclei, is compared with the $\Delta^2(t)/2$, which is calculated for only the non-exchangeable H nuclei. The difference between the $\langle r^2 \rangle$ and $\Delta^2(t)/2$ could be minimal if we fit the model $I(\mathbf{Q}, t)$ to a calculated $I_{inc}(\mathbf{Q}, t)$ which includes only the non-exchangeable H nuclei to obtain the $\langle r^2 \rangle$. We expect that diffusion at low temperatures is less important than that at high temperatures. Hence, we find that the $\Delta^2(t)/2$ appears to have converged after 10 ns and approaches the $\langle r^2 \rangle$ reasonably well, at low temperatures, but it lies below the $\langle r^2 \rangle$ at high temperatures (e.g. 250 K). From these comparisons it would be interesting to evaluate $\Delta^2(t)$ out to longer times to determine whether it converges and to reveal the dynamics contributing. For example, at 300 K, nearly translational diffusion may be possible for some H in the protein that are near the surface or near hydration water. It would be interesting to exclude these H from $\Delta^2(t)$. In this regard, it is also important to exclude the center of mass (CM) motion which becomes important at higher temperature and longer times.

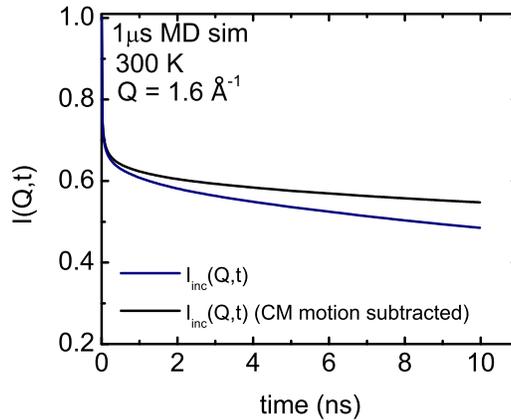


Figure 3.13: The incoherent ISF $I(\mathbf{Q}, t)$ with and without the CM motion subtracted after $t = 10$ ns at 300 K and at $Q = 1.6 \text{ \AA}^{-1}$, obtained from 1 μs MD simulation.

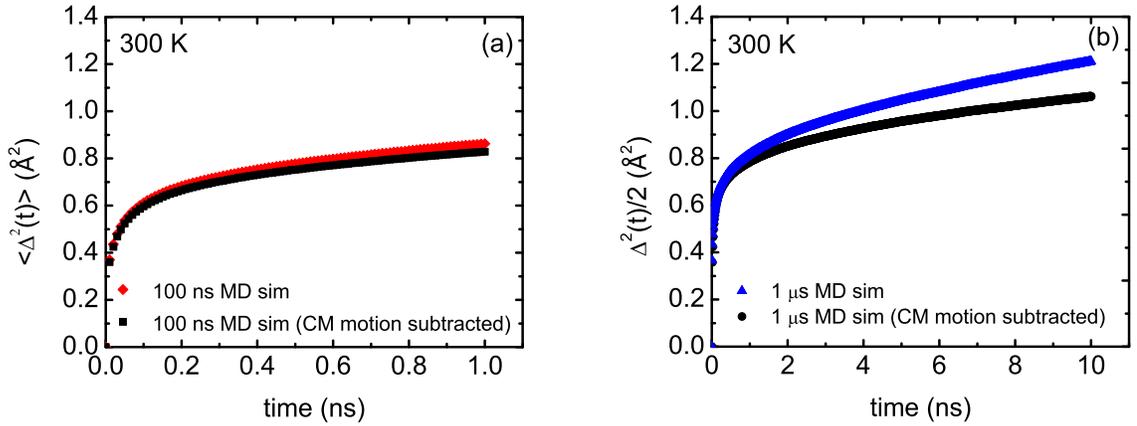


Figure 3.14: The MSD $\Delta^2(t)/2$, with and without the CM motion subtracted after $t = 1$ ns at 300 K, (a) 100 ns MD simulation and (b) 1 μs MD simulation.

Fig. 3.13 shows that the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ at 300 K and $Q = 1.6 \text{ \AA}^{-1}$ (blue line) with and (black line) without CM motion subtracted. In the 100 ns MD simulation, the impact of the CM motion is negligible, so that the CM effect is not subtracted in the incoherent ISF data in Fig. 3.2. However, it is significant for the longer simulation as shown in Fig. 3.13. For the incoherent ISF data in Fig. 3.6, the CM effect is subtracted by superimposing the protein coordinates onto the starting structure before analysis. Similarly, the simulated MSD $\Delta^2(t)$ versus time in Fig. 3.14 shows that the impact on the CM motion is more significant on the results obtained from the 1 μs MD simulation. The difference between the $\Delta^2(t)$ with CM motion (solid triangles) and without CM motion (solid circles) is approximately 5% for the 100 ns MD simulation and 20% for the 1 μs MD simulation.

Fig. 3.15 shows that values of the $\Delta^2(t)$ obtained from the 1 μs simulation with and without the CM motion subtracted. The contribution of the CM motion to the $\Delta^2(t)$ is small. Values of $r_i(t)$ corrected for CM motion were used to calculate $I_{inc}(\mathbf{Q}, t)$.

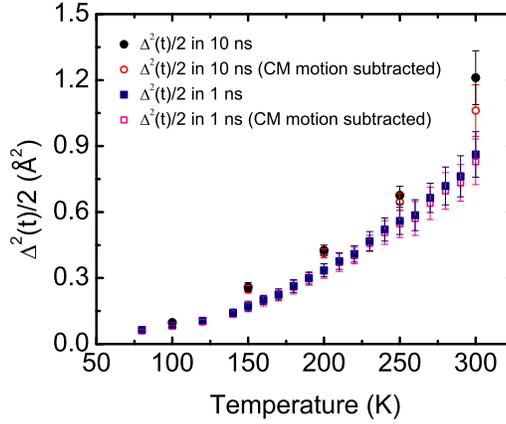


Figure 3.15: The MSD $\Delta^2(t)/2$, with and without the CM motion subtracted after $t = 1$ ns (100 ns MD simulation) and $t = 10$ ns (1 μ s MD simulation).

3.4.2 Resolution Broadened Mean Square Displacement $\langle r^2 \rangle_R$

We calculated the resolution broadened MSD $\langle r^2 \rangle_R$ to compare with the intrinsic MSD $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ in Sec. 3.3. A resolution broadened $\langle r^2 \rangle_R$ defined in Chap. 2 is an MSD that includes the contribution of motions over a finite time only. This finite time, τ_R , corresponds to the resolution width W as $\tau_R = (8 \ln 2)^{1/2} \hbar / W$ for a Gaussian resolution function. The smallest full-width half-maximum (FWHM) readily available today is $W = 1 \mu\text{eV}$ which corresponds to $\tau_R = 1.5$ ns. In order to analyse the impact of a finite resolution width on the MSD, we show the intrinsic MSD $\langle r^2 \rangle$ and the resolution broadened MSD $\langle r^2 \rangle_R$ using the 100 ns MD simulation.

In the MD studies, the observed MSD $\langle r^2 \rangle_{exp}$ is usually compared with the resolution broadened MSD $\langle r^2 \rangle_R$ which is calculated over the same time period which the observed MSD having [96, 87, 106]. This provides a way to have an excellent agreement between a simulated and an observed MSD. Specifically, the resolution broadened MSD, $\langle r^2 \rangle_R$ is obtained from the slope expression in Eq. (1.1) which is used to obtain the observed MSD $\langle r^2 \rangle_{exp}$ in neutron scattering experiments. In the calculation of the $\langle r^2 \rangle_R$, we replacing $S_R^{exp}(\mathbf{Q}, \omega = 0)$ with the calculated resolution broadened DSF

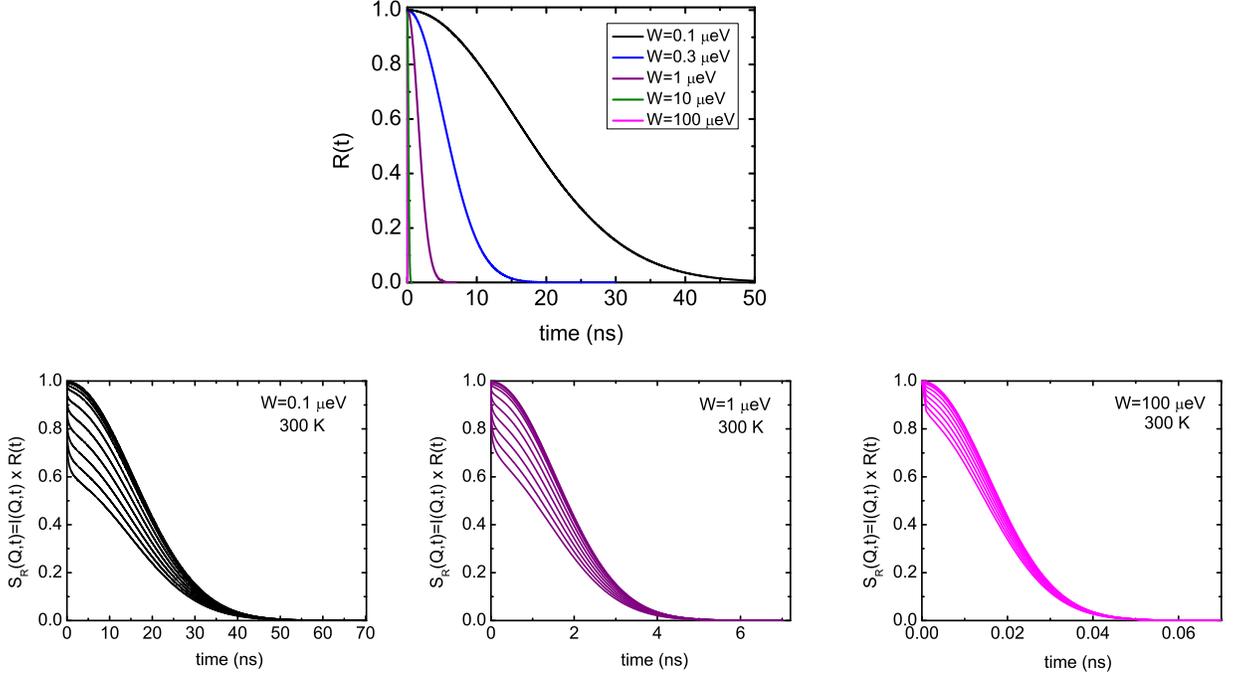


Figure 3.16: (Upper) The Gaussian resolution function $R(t) = \exp(-t^2/2\tau_R^2)$ for $\tau_R = 15, 5, 1.5, 0.15, 0.015$ ns corresponding to $W = 0.1, 0.3, 1, 10, 100 \mu\text{eV}$. (Lower) $S_R(\mathbf{Q}, t) = I(\mathbf{Q}, t) \times R(t)$ is shown in the lower frame in Fig. 3.16 for $W = 0.1, 1, 100 \mu\text{eV}$ at 300 K.

$$S_R(\mathbf{Q}, \omega = 0),$$

$$\langle r^2 \rangle_R = -3 \frac{d \ln S_R(\mathbf{Q}, \omega = 0)}{dQ^2}. \quad (3.4)$$

From Eq. (2.5), the observed, elastic resolution broadened DSF, $S_R(\mathbf{Q}, \omega = 0)$ is

$$S_R(\mathbf{Q}, \omega = 0) = \frac{1}{2\pi} \int dt I(\mathbf{Q}, t) R(t). \quad (3.5)$$

Here, $R(t)$ is Fourier transform of the instrumental resolution function $R(\omega)$ in time. For the Gaussian resolution function $R(\omega)$, $R(t)$ is typically a Gaussian, $R(t) = \exp(-t^2/2\tau_R^2)$ shown in the upper frame in Fig. 3.16. The resolution function cuts off $I(\mathbf{Q}, t)$ after a time $\tau_R = (8 \ln 2)^{1/2} \hbar / W$ for the Gaussian resolution function (see Appendix A). $I(\mathbf{Q}, t)$ is calculated using the fitting parameter in Fig. 3.3 in our model $I(\mathbf{Q}, t)$ in Eq. (3.1). Although $S_R(\mathbf{Q}, t) = I(\mathbf{Q}, t) \times R(t)$ was calculated for five different

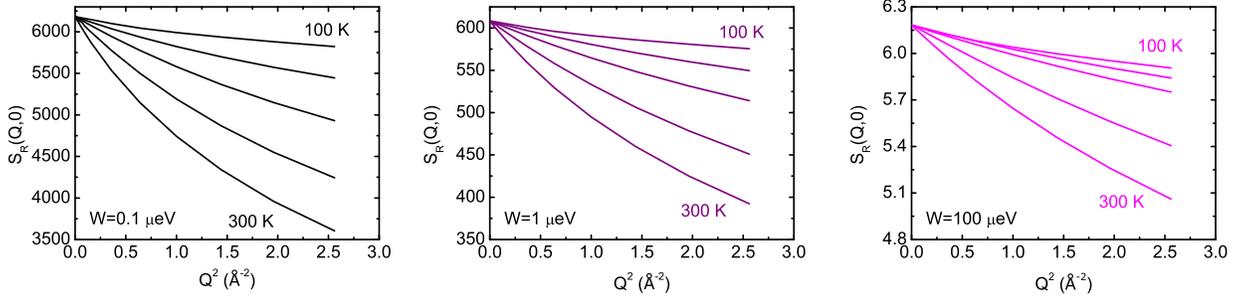


Figure 3.17: $S_R(\mathbf{Q}, \omega = 0)$ for $W = 0.1, 1, 100 \mu\text{eV}$ at 100, 150, 200, 250 and 300 K.

energy resolution widths, only three energy resolution widths are shown in the lower frame in Fig. 3.16 as an example. The energy resolution width from left to right is $W = 0.1, 1, 100 \mu\text{eV}$. Fig. 3.17 shows $S_R(\mathbf{Q}, \omega = 0)$, which is calculated from Eq. (3.5), using the calculated $I(\mathbf{Q}, t)$ and the Gaussian resolution function, for $W = 0.1, 1, 100 \mu\text{eV}$. Here, we obtain $S_R(\mathbf{Q}, \omega = 0)$ by inserting the calculated ISF in Eq. (3.5) and then the $\langle r^2 \rangle_R$ is again obtained using the calculated $S_R(\mathbf{Q}, \omega = 0)$ in Eq. (3.4). In this way an $\langle r^2 \rangle_R$ that has evolved over to a time τ_R is compared with the observed MSD $\langle r^2 \rangle_{exp}$ and the intrinsic MSD $\langle r^2 \rangle$.

The intrinsic $\langle r^2 \rangle$ and the resolution broadened $\langle r^2 \rangle_R$ of lysozyme are compared in Fig. 3.18. The resolution broadened MSD $\langle r^2 \rangle_R$ is calculated for five different resolution width. The resolution width from bottom to top is 100, 10, 1, 0.3 and 0.1 μeV . The $\langle r^2 \rangle_R$ lies below the intrinsic MSD $\langle r^2 \rangle$, even for $W = 0.1 \mu\text{eV}$, corresponding to $\tau_R = 15 \text{ ns}$, a resolution approximately ten times higher than that available today. For a resolution width $W = 1 \mu\text{eV}$ ($\tau_R = 1.5 \text{ ns}$), the $\langle r^2 \rangle$ is found to be approximately twice the $\langle r^2 \rangle_R$. At temperatures lower than 150 K, the difference between the $\langle r^2 \rangle$ and $\langle r^2 \rangle_R$ is negligible. The ratio $\langle r^2 \rangle / \langle r^2 \rangle_R$ is approximately independent of temperature for $T > 150 \text{ K}$. If the relaxation time $\tau = \lambda^{-1}$ of the protein is less than the resolution time τ_R , the significant difference between $\langle r^2 \rangle$ and $\langle r^2 \rangle_R$ is not observed. We expect that $\langle r^2 \rangle_R$ lies below $\langle r^2 \rangle$ if τ_R is less than τ , and $\langle r^2 \rangle_R$ reaches $\langle r^2 \rangle$ if τ_R is longer than τ . The relaxation time of lysozyme corresponding the relaxation parameter λ of

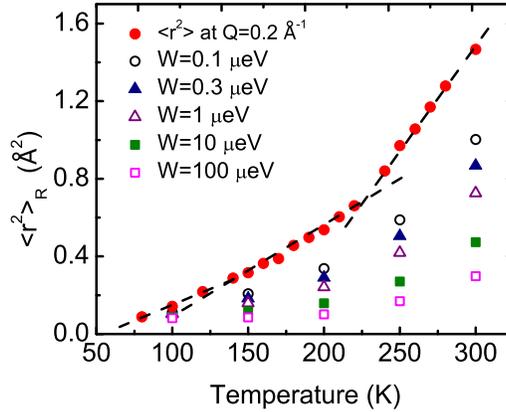


Figure 3.18: The present intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ and resolution broadened MSD $\langle r^2 \rangle_R$ calculated from the same model for the energy resolution widths, W , 0.1, 0.3, 1, 10 and 100 μeV .

lysozyme shown in Figs. 3.3 and 3.7. From these results, we can say $\tau = \lambda^{-1} \geq 1$ ns at low Q and temperatures above 170 K. It is noted that the reason of observing the longer τ at low temperature could be that the system has lower energy at low temperatures. In respect of these results, resolution broadening is important at $W = 1 \text{ \mu eV}$. The functional form of $C(t)$ in the model $I(\mathbf{Q}, t)$ is also important for the sensitivity of the resolution broadening. The sensitivity of the $C(t)$ function will be discussed in Sec. 3.5.

In Fig. 3.19, the intrinsic MSD $\langle r^2 \rangle$ and the resolution broadened MSD $\langle r^2 \rangle_R$ are compared with previous simulated values of $\langle r^2 \rangle_R$ for lysozyme, calculated by Roh et al. [87]. The $\langle r^2 \rangle_R$ at low Q ($Q = 0.2 \text{ \AA}^{-1}$) and $W = 1 \text{ \mu eV}$ is calculated from their simulations of lysozyme which is hydrated to $h = 0.43$. In Fig. 3.19, the solid circles and squares represent the intrinsic MSD $\langle r^2 \rangle$ in Sec. 3.3, the open and solid triangles represent the $\langle r^2 \rangle_R$ at $Q = 0.2 \text{ \AA}^{-1}$ and $W = 1 \text{ \mu eV}$ for $h = 0.40$ and $h = 0.43$, respectively. The resolution broadened MSD $\langle r^2 \rangle_R$ lies below the intrinsic MSD $\langle r^2 \rangle$. The excellent agreement is observed between the resolution broadened MSDs. As expected, the $\langle r^2 \rangle_R$ for $h = 0.43$ is larger than the $\langle r^2 \rangle_R$ for $h=0.4$ due to the sensitivity

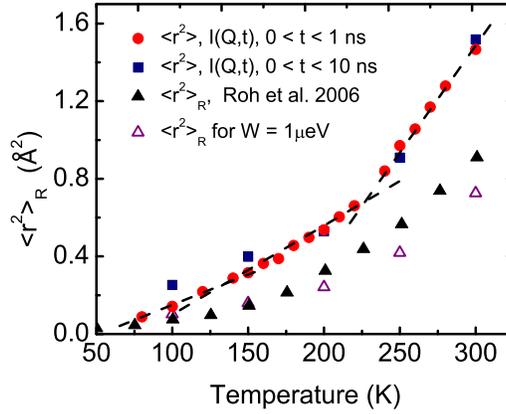


Figure 3.19: The present intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ (obtained from fits to (1) 100 ns (solid squares) and (2) 1 μs MD simulations (solid circles)) and the present resolution broadened MSD $\langle r^2 \rangle_R$ ($W = 1 \mu\text{eV}$) (open triangles) for lysozyme at $h = 0.40$ compared with simulated $\langle r^2 \rangle_R$ at $W = 1 \mu\text{eV}$ by Roh et al. 2006 [87] for lysozyme at $h = 0.43$ (solid triangles).

to hydration level.

In order to clarify the sensitivity to hydration level, we used the observed MSD $\langle r^2 \rangle_{exp}$ data which is taken from Ref. [87] in Fig. 3.20. Similar to Fig. 3.19, we compare the present intrinsic $\langle r^2 \rangle$ with the $\langle r^2 \rangle_{exp}$ observed experimentally in lysozyme using an energy resolution width $W = 1 \mu\text{eV}$ in Fig. 3.20. The observed $\langle r^2 \rangle_{exp}$ of lysozyme at four hydration levels, $h = 0.05, 0.18, 0.3$ and 0.45 , and the present $\langle r^2 \rangle$ are shown in Fig. 3.20. The observed $\langle r^2 \rangle_{exp}$ are very sensitive to the hydration level for $T > T_D$. In Fig. 3.20, the $\langle r^2 \rangle$ for $h = 0.4$ lies above but close to the observed $\langle r^2 \rangle_R$ for $h = 0.45$ and significantly higher than the observed $\langle r^2 \rangle_{exp}$ for lower hydrations, as expected. Comparing the calculated $\langle r^2 \rangle_R$ in Fig. 3.19 with the observed $\langle r^2 \rangle_{exp}$ in Fig. 3.20, we see that the calculated $\langle r^2 \rangle_R$ lies somewhat below but close to the experimental values for similar levels of hydration. Broadly the agreement between the resolution broadened $\langle r^2 \rangle_R$ and observed $\langle r^2 \rangle_{exp}$ is very good, both in terms of the absolute value and in the temperature dependence.

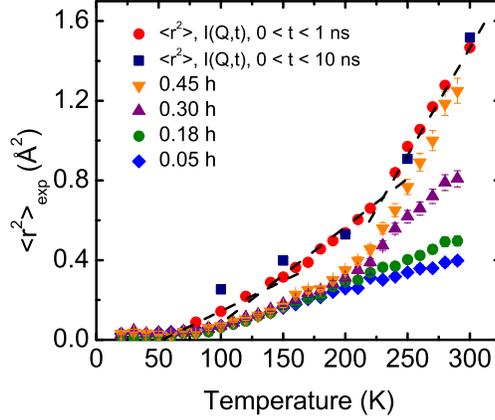


Figure 3.20: The present intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2 \text{ \AA}^{-1}$ for lysozyme ($h = 0.4$), as in Fig. 3.13, compared with the experimental resolution broadened MSD $\langle r^2 \rangle_{exp}$ for $W = 1 \text{ } \mu\text{eV}$ for lysozyme at different hydration levels (h) observed by Roh et al. 2006 [87].

3.4.2.1 Dynamical transition

Similar to Sec. 2.3.2, we discuss the impact of a finite energy resolution width W of employed instrument on the DT temperature T_D . Fig. 3.21 shows the DT temperature on the intrinsic MSD $\langle r^2 \rangle$ and the resolution broadened MSD $\langle r^2 \rangle_R$. The intrinsic MSD $\langle r^2 \rangle$ for lysozyme is obtained from the fit of the model $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$ in Sec. 3.3. The intrinsic MSD shown in Figs. 3.5 and 3.9 displays a clear DT at a transition temperature, T_D , 220 K. It is a long-time, equilibrium MSD that reflects the energy landscape of lysozyme, so that this indicates that the DT is an intrinsic property of the protein rather than being simply a time window effect. Fig. 3.21 shows that the resolution broadening shifts the DT temperature T_D to higher temperatures, similar to previous researches [73, 50, 106, 16].

For a finite resolution width W , motions in the protein in a limited time $\tau_R = (8 \ln 2)^{1/2} \hbar / W$ contribute to the resolution broadened MSD $\langle r^2 \rangle_R$. The intrinsic MSD $\langle r^2 \rangle$ is the resolution broadened MSD $\langle r^2 \rangle_R$ calculated for zero resolution width. For a zero resolution width, even slow, long time motions are contributed to the MSD, so

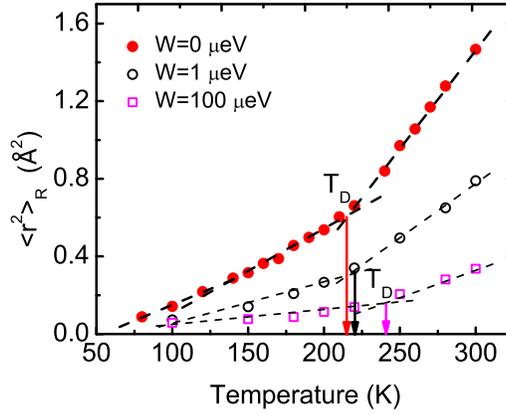


Figure 3.21: The intrinsic MSD ($W = 0$) and resolution broadened MSD $\langle r^2 \rangle_R$ for $W = 1 \mu\text{eV}$ and $W = 100 \mu\text{eV}$ with T_D identified. The MSD are obtained from fits to experiment in Ref. [102] using a model $I(\mathbf{Q}, t)$ that has a simple exponential decay function $C(t) = \exp[-\lambda t]$.

that $\langle r^2 \rangle_R$ for a finite resolution width is always smaller than $\langle r^2 \rangle$ at a given temperature. Slow motions observed longer time than τ_R does not contribute to the resolution broadened MSD $\langle r^2 \rangle_R$. τ_R corresponds to the energy resolution width W , hence the DT is shifted to higher temperatures for higher energy resolution widths. In Fig. 3.21, the impact of the resolution width on the T_D is illustrated by comparing the intrinsic MSD $\langle r^2 \rangle$ with the $\langle r^2 \rangle_R$ for resolution widths $W = 1 \text{ eV}$ and $W = 100 \text{ eV}$ and T_D is explicitly identified. The T_D for the $\langle r^2 \rangle$ is 220 K and shifts to 240 K for the energy resolution width $W = 100 \mu\text{eV}$, corresponding to $\tau_R = 15 \text{ ps}$.

The degree of impact of W on $\langle r^2 \rangle_R$ depends on the rate at which correlations decay in the protein. In the model $I(\mathbf{Q}, t)$, the decay rate depends on the functional form of $C(t)$. In Eq. (3.1), $C(t)$ is chosen as a stretched exponential function which has a long time tail, particularly for small β . This indicates that the reduction of $\langle r^2 \rangle_R$ is significant for small energy resolution width W , corresponding to long τ_R . But the rate of change of $\langle r^2 \rangle_R$ with W is gradual. However, for a simple decay function $C(t)$ in Chap. 3 [102], the reduction of $\langle r^2 \rangle_R$ below $\langle r^2 \rangle$ begins at a larger value of

W and thereafter the reduction increases rapidly with increasing W . The lower the resolution, the larger the apparent T_D . Using this simple model, the increase of the apparent T_D with increasing W can also be readily understood. From Eq. (2.17), the ratio between $\langle r^2 \rangle_R$ and $\langle r^2 \rangle$ becomes

$$\frac{\langle r^2 \rangle_R}{\langle r^2 \rangle} = \left[1 + \frac{W}{I_\infty \lambda}\right]^{-1} \simeq \left[1 - \frac{W}{I_\infty \lambda} + \dots\right]. \quad (3.6)$$

With increasing temperature, the λ increases, as shown in Fig. 3.4 and Fig. 3.8, and the ratio $W/(I_\infty \lambda)$ decreases for a given W . At higher temperatures, the impact of the resolution width is significant, thus $\langle r^2 \rangle_R$ is decreased least by a finite resolution and the larger $\langle r^2 \rangle_R/\langle r^2 \rangle$ is observed.

3.5 Sensitivity of the model to $C(t)$

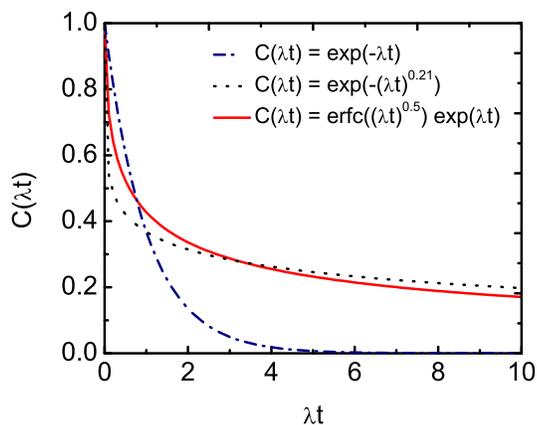


Figure 3.22: Comparison of relaxation functions: a simple exponential (blue dashed dotted line), a stretched exponential with $\beta = 0.5$ (green dashed line), a stretched exponential with $\beta = 0.21$ (black dotted line) and the Mittag-Leffler function (red solid line).

In Sec. 2.3.3, we used a Gaussian $C(t)$ and a Gaussian resolution functions in our model to investigate the sensitivity of our results. Here, we compare four forms of $C(t)$, simple exponential function, stretched exponential function with $\beta = 0.5$

and $\beta = 0.23$, and the Mittag-Leffler function [8], in Fig. 3.22. In Sec. 3.2, $C(t)$ in Eq. (3.1) provides a reasonable fit of the model $I(\mathbf{Q}, t)$ to the calculated $I_{inc}(\mathbf{Q}, t)$ is the stretched exponential function with a small β , i.e. $\beta = 0.23$. As shown in Fig. 3.22, the simple exponential function has a short tail, and the tail of the stretched exponential function becomes longer with decreasing β value. For the simple exponential function used in Chap. 2, the correlations decreases rapidly on a time scale $\tau = \lambda^{-1}$. It was not possible to obtain a good fit to $I_{inc}(\mathbf{Q}, t)$ using a simple exponential function or a stretched exponential function with large β . Similar to the stretched exponential function for a small β , the Mittag-Leffler function [8] has a long range tail reaching out to times a factor of ten beyond $t = \lambda^{-1}$. This indicates that correlations in proteins continue for times well beyond $\tau = \lambda^{-1}$. For this reason, when a stretched exponential function with a small β is used, we expect that $\langle r^2 \rangle$ lies above $\langle r^2 \rangle_R$ even when $\lambda^{-1} \simeq \tau_R$.

3.6 Discussion

We have proposed a fitting method to obtain the intrinsic long-time MSD in proteins from finite time simulations. The intrinsic MSD is uniform and consistent value of MSD. The intrinsic MSD represents the equilibrium MSD as would be predicted by statistical mechanics and the energy landscape, assuming the protein does not go through major structural changes. The intrinsic MSD of lysozyme is obtained from the fit of the ISF model to calculated data from simulations of 100 ns and 1 μ s. The perfect agreement between intrinsic MSDs obtained from simulations of 100 ns and 1 μ s is observed. The intrinsic, long time MSD of lysozyme is found to be approximately twice the MSD that develops after a time of 1.5 ns, as would be observed using neutron instruments with an energy resolution width of $W = 1 \mu$ eV. The intrinsic MSD shows a break in slope with temperature at 220 K as does the finite time MSD. The break at 220 K which is arisen from additional motions at high temperatures is denoted as dynamical transition. These results indicate that the intrinsic MSD is independent of simulation time, and that the dynamical transition is not just a time window effect.

Chapter 4

MOTIONAL DISPLACEMENT IN PROTEINS, THE ORIGIN OF Q DEPENDENT VALUES

As stated in previous chapters, the MSD of nuclei in proteins is extensively investigated using neutron scattering methods. Observed is the elastic scattering ($\omega = 0$) component of the dynamical structure factor (DSF), $S_R^{exp}(\mathbf{Q}, \omega = 0)$, where Q is the wave vector transfer in the scattering. The data is typically analysed assuming that $S(\mathbf{Q}, \omega = 0) = \exp(-\frac{1}{3}Q^2\langle r^2 \rangle)$ where $\langle r^2 \rangle$ is an average MSD of the nuclei in the protein. Since hydrogen (protons) have a large incoherent scattering cross, 10-20 times that of other nuclei in the protein, the average MSD $\langle r^2 \rangle$ is dominated by that of H in the protein (and in its associated hydration water if any). The MSD $\langle r^2 \rangle$ is equal to the the observed MSD $\langle r^2 \rangle_{exp}$ in Chap. 2 and Chap. 3. Typically, the MSD $\langle r^2 \rangle$ is extracted from the observed DSF $S_R^{exp}(\mathbf{Q}, \omega = 0)$ in the usual way as in Eq. (1.1).

The MSD obtained using Eq. (1.1) typically depends on the value of Q at which the derivative is taken. In short, the observed $\langle r^2 \rangle$ is Q dependent. The dependence is unlikely to be physical but rather arises from some over simplification or error in the analysis of the data. Our goal in this chapter is to determine the origin of this Q dependence. It is to propose methods to obtain Q -independent average $\langle r^2 \rangle$ and to use the observed Q dependence to provide information on the distribution of the MSD within the protein.

The upper frame in Fig. 1.3 shows the $\ln S_R^{exp}(\mathbf{Q}, \omega = 0)$ versus Q^2 observed by Daniel et al. in glutamate dehydrogenase [17]. The lines are a guide to the eye through the observed $\ln S_R^{exp}(\mathbf{Q}, \omega = 0)$. Clearly a straight line versus Q^2 is not observed as is required to obtain a Q independent $\langle r^2 \rangle$. The observed $\langle r^2 \rangle$ generally decreases with increasing Q . The lower frame in Fig. 1.3 shows a Q -dependent $\langle r^2 \rangle = \langle x^2 \rangle$

obtained from a simulated value of $S_R(\mathbf{Q}, \omega = 0)$ in lysozyme by Calandrini et al. at ambient temperature [8]. In addition to these examples, Fig. 3.7a shows values of $\langle r^2 \rangle$ obtained by us from fits of a model $I(\mathbf{Q}, t)$ that contains $I(\mathbf{Q}, t = \infty) = \exp[-\frac{1}{3}Q^2\langle r^2 \rangle]$ to simulated $I_{inc}(\mathbf{Q}, t)$ for lysozyme. The fitted values of $\langle r^2 \rangle$ clearly decrease with increasing Q as do observed values, specially at high temperature.

The dependence of $\langle r^2 \rangle$ on Q could arise from three sources. Firstly, the $S(\mathbf{Q}, \omega = 0)$ normalized by $S(\mathbf{Q} = 0, \omega = 0)$ is the infinite time limit of the incoherent intermediate scattering function (ISF), $I(\mathbf{Q}, t) = \langle \exp[-i\mathbf{Q} \cdot [\mathbf{r}(t) - \mathbf{r}(0)]] \rangle$, i.e.

$$\begin{aligned} \frac{S(\mathbf{Q}, \omega = 0)}{S(\mathbf{Q} = 0, \omega = 0)} &= I(\mathbf{Q}, t = \infty) \\ &= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(\infty)) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\ &= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}) \rangle \\ &= \langle \exp[-\langle [\mathbf{Q} \cdot \mathbf{r}]^2 \rangle + \frac{1}{12}(\langle [\mathbf{Q} \cdot \mathbf{r}]^4 \rangle - 3\langle [\mathbf{Q} \cdot \mathbf{r}]^2 \rangle^2) + \dots] \rangle. \end{aligned}$$

The expression $S(\mathbf{Q}, \omega = 0) = \exp[-\frac{1}{3}Q^2\langle r^2 \rangle]$ for the normalized DSF is obtained by neglecting all the higher order cumulants beyond the second order Gaussian cumulant and assuming cubic or spherical symmetry so that $\langle [Q \cdot r]^2 \rangle = \frac{1}{3}Q^2\langle r^2 \rangle$. The Q dependence could arise from making the Gaussian approximation to $I(\mathbf{Q}, t = \infty)$. Secondly, a protein contains over a thousand H in a variety of environments having wide spectrum of MSD. We approximate this diverse

$$I(\mathbf{Q}, t) = \sum_{j=1}^N \langle \exp(-i\mathbf{Q} \cdot (\mathbf{r}_j(t) - \mathbf{r}_j(0))) \rangle \quad (4.1)$$

by a single $I(\mathbf{Q}, t) = \langle \exp[-i\mathbf{Q} \cdot [\mathbf{r}(t) - \mathbf{r}(0)]] \rangle$ when we use Eq. (1.1). That is, we ignore the dynamical diversity of the H in protein and represent all the H by a single scatterer. Thirdly, there could be a genuine dependence on Q . That is a specific Q implies that length scales up to $L = 2\pi/Q$ only can be sampled. If the motional displacement $\langle r^2 \rangle^{1/2}$ is longer than L , the observed $\langle r^2 \rangle^{1/2}$ could limited to L . For example, $Q = 2 \text{ \AA}^{-1}$, $\langle r^2 \rangle$ could limited to lengths of order $\langle r^2 \rangle < \pi^2 \text{ \AA}^2$. From this argument we expect only very large MSD to limited by using a Q of say 2 \AA^{-1} .

In the following sections we investigate the impact of making the Gaussian approximation, of neglecting dynamical diversity and of possible limits to $\langle r^2 \rangle$ at specific Q values. We do this by simulating the protein lysozyme and calculation $I_{inc}(\mathbf{Q}, t)$ arising from all H in lysozyme and then making and not making the above approximations in the analysis of $I_{inc}(\mathbf{Q}, t)$.

4.1 Possible origins of Q -dependent MSD

In Chap. 3, we propose a method to extract the intrinsic, long time MSD $\langle r^2 \rangle$ from the simulated incoherent intermediate scattering function (ISF), $I_{inc}(\mathbf{Q}, t)$. The incoherent ISF $I_{inc}(\mathbf{Q}, t)$ is observed in neutron scattering experiments and calculated from simulations. From Eq. (2.2), the full ISF for only H is written as

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^N \langle e^{-i\mathbf{Q}\cdot\mathbf{r}_i(t)} e^{i\mathbf{Q}\cdot\mathbf{r}_i(0)} \rangle, \quad (4.2)$$

where the sum is over H in the protein. In the classical limit, $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.2) becomes

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^N \langle e^{i\mathbf{Q}\cdot(\mathbf{r}_i(t) - \mathbf{r}_i(0))} \rangle = \frac{1}{N} \sum_{i=1}^N \langle e^{i\mathbf{Q}\cdot\mathbf{\Delta}_i(t)} \rangle, \quad (4.3)$$

where $\mathbf{\Delta}_i(t) = \mathbf{r}_i(t) - \mathbf{r}_i(0)$. In the analysis of neutron scattering experiments, it is usual to represent the ISF in Eq. (2.2), which is summed over all nuclei, by a global $I(\mathbf{Q}, t)$ as shown in Eq. (2.3). Following the same procedure we construct a model of $I(\mathbf{Q}, t)$ in Chap. 2 and modify this model replacing the stretched exponential function $C(t) = \exp(-(\lambda t)^\beta)$ with simple exponential function $C(t) = \exp(-(\lambda t))$ in Chap. 3. In the model $I(\mathbf{Q}, t)$ in Eq. (3.1), we define the intrinsic, long-time MSD as the $\langle r^2 \rangle$ that appears in the ∞ -time limit of $I(\mathbf{Q}, t)$.

As stated in Chap. 3, two assumptions in the model of $I(\mathbf{Q}, t)$ in Eq. (3.1) are made. (1) The cumulant expansion $\langle \exp(i\mathbf{Q}\cdot\mathbf{r}) \rangle$ is

$$\langle \exp(i\mathbf{Q}\cdot\mathbf{r}) \rangle = \exp\left[-\frac{1}{2}\langle [\mathbf{Q}\cdot\mathbf{r}]^2 \rangle + \frac{1}{4!}[\langle [\mathbf{Q}\cdot\mathbf{r}]^4 \rangle - 3\langle [\mathbf{Q}\cdot\mathbf{r}]^2 \rangle^2] + \dots\right]. \quad (4.4)$$

In the model $I(\mathbf{Q}, t)$, we keep only the first term and neglect the higher cumulant terms. This is denoted the Gaussian approximation. (2) $I_{inc}(\mathbf{Q}, t)$ arising from H in a

diverse range of environments can be represented by a global $I(\mathbf{Q}, t)$ in which a single scattering center represents whole protein. This indicates that dynamical diversity is neglected in the model $I(\mathbf{Q}, t)$. The dynamical diversity of H in proteins is denoted heterogeneity. We analyse the data $I_{inc}(\mathbf{Q}, t)$ using a $I(\mathbf{Q}, t)$ which is valid for a single scatterer.

In regards of these assumptions, the Q -dependence of the global MSD $\langle r^2 \rangle$ could arise from three possible reasons. (1) **Gaussian approximation**; In the cumulant expansion, the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.3) is written as

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^N e^{-\frac{1}{2}\langle[\mathbf{Q}\cdot\Delta_i(t)]^2\rangle + \frac{1}{4!}\langle[\mathbf{Q}\cdot\Delta_i(t)]^4\rangle - 3\langle[\mathbf{Q}\cdot\Delta_i(t)]^2\rangle^2 + \dots}. \quad (4.5)$$

where $\Delta_i(t) = \mathbf{r}_i(t) - \mathbf{r}_i(0)$. The model of the ISF in Eq. (3.1) includes only the Gaussian or lowest order (Q^2) term in a cumulant expansion of $I(\mathbf{Q}, \infty)$ (see *Appendix B*),

$$I(\mathbf{Q}, \infty) = e^{-\langle[\mathbf{Q}\cdot\mathbf{r}]^2\rangle + \frac{1}{12}\langle[\mathbf{Q}\cdot\mathbf{r}]^4\rangle - 3\langle[\mathbf{Q}\cdot\mathbf{r}]^2\rangle^2 + \dots}, \quad (4.6)$$

with $\Delta^2(t) = 2\langle r^2 \rangle$ and $\langle[\mathbf{Q}\cdot\mathbf{r}]^2\rangle = \frac{1}{3}Q^2\langle r^2 \rangle$ for a cubic or spherically symmetric system. The whole expression of the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ includes all cumulant terms, whereas the model of $I(\mathbf{Q}, t)$ includes only the Gaussian cumulant term. If the fourth cumulant (Q^4) in Eq. (4.5) is significant and neglected then a Q dependent $\langle r^2 \rangle_{exp}$ would be obtained when we use the $I(\mathbf{Q}, t)$ model includes only the Gaussian term.

The importance of fourth cumulant is explained in *Appendix E*. The magnitude of the fourth cumulant can be characterized by the kurtosis of the motional distribution, $\gamma_\alpha = [\langle r_\alpha^4 \rangle - 3\langle r_\alpha^2 \rangle^2] / \langle r_\alpha^2 \rangle^2$ when Q is chosen along the α axis, $\alpha = x, y, z$. In the cumulant expansion of $\exp(i\mathbf{Q}\cdot(\mathbf{r}_i(t) - \mathbf{r}_i(0)))$, we keep the first two terms

$$\langle \exp(i\mathbf{Q}\cdot\Delta_i(t)) \rangle = \exp\left(-\frac{1}{2}\mu_2 + \frac{1}{4!}\mu_4\right), \quad (4.7)$$

where $\Delta_i(t) = \mathbf{r}_i(t) - \mathbf{r}_i(0)$,

$$\begin{aligned}
\mu_2 &= \mathbf{m}_2, \\
\mu_4 &= \mathbf{m}_4 - 3\mathbf{m}_2^2, \\
\mathbf{m}_2 &= \langle (\mathbf{Q} \cdot \Delta_i(t))^2 \rangle, \\
\mathbf{m}_4 &= \langle (\mathbf{Q} \cdot \Delta_i(t))^4 \rangle.
\end{aligned} \tag{4.8}$$

The kurtosis, that shows the contribution of the fourth cumulant term, is

$$\gamma_\alpha = \frac{\mu_{4\alpha}}{\mu_{2\alpha}^2} = \frac{\mathbf{m}_{4\alpha}}{\mathbf{m}_{2\alpha}^2} - 3, \tag{4.9}$$

where $\mathbf{m}_{2\alpha} = \langle [\mathbf{Q}_\alpha \cdot \Delta_i(t)]^2 \rangle$, $\mathbf{m}_{4\alpha} = \sum_i \langle [\mathbf{Q}_\alpha \cdot \Delta_i(t)]^4 \rangle$ and $\mathbf{Q}_\alpha = Q\alpha$, $\alpha = x, y, z$.

(2) **Neglect of dynamical diversity.** The full $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.3) is the sum over thousands of scattering centers (e.g. 1392 H in lysozyme) which have a spectrum of dynamics and MSDs. In analysis of data or simulation, a model (e.g. $I(\mathbf{Q}, t)$ in Eq. (3.1)) representing a single scattering center is generally fitted to the $I_{inc}(\mathbf{Q}, t)$. The model neglects the dynamical diversity in the protein and this could lead to a Q -dependent MSD. Dynamical diversity has been discussed in Refs. [29, 68, 108].

(3) **The lengths observable may be limited by Q .** The Q -value represent the length-scale of motions in proteins. The low Q values leads to observe the long-distance motion. For example, diffusion up to lengths $\lambda \sim 2\pi/Q$ only can be observed by a technique that has wave vector Q . At $Q \sim 2 \text{ \AA}^{-1}$, large MSDs greater than $\langle r^2 \rangle \gtrsim (2\pi/Q)^2 \sim 9 \text{ \AA}^2$ would not be observable. This could lead to an apparent MSD that decreases with increasing Q .

Our goal is to determine which of these approximations leads to an MSD that depends on Q . To test the Gaussian approximation (GA), we define the Gaussian approximation to $I_{inc}(\mathbf{Q}, t)$:

$$I_{iG}(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^N e^{-\frac{1}{2} \langle [\mathbf{Q} \cdot \Delta_i(t)]^2 \rangle}. \tag{4.10}$$

which includes only the first cumulant, Gaussian, term.

We calculate the ISF data $I_{inc}(\mathbf{Q}, t)$, includes all cumulants, and the $I_{iG}(\mathbf{Q}, t)$, which is the Gaussian approximation to $I_{inc}(\mathbf{Q}, t)$ and does not include the higher cumulants, for H only in the sum over j . Then, we fit the model $I(\mathbf{Q}, t)$ in Eq. (3.1) to $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ to obtain the intrinsic, long-time MSD from these two data set. The impact of the higher cumulants which may contribute significantly to ISF and DSF is investigated by comparing the MSD obtained by fitting the model of $I(\mathbf{Q}, t)$ to the whole expression of the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ and the incoherent ISF including only the first cumulant term, $I_{iG}(\mathbf{Q}, t)$.

As stated above, the dynamical diversity studied by different groups in the literature [29, 68, 108] may make a significant contribution to the ISF $I_{inc}(\mathbf{Q}, t)$. Neglecting the dynamical diversity in the model $I(\mathbf{Q}, t)$ leads to a Q -dependence. To test the dynamical diversity, we evaluated the ISF $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.2) for single protons (single H) in lysozyme. We expect that no dynamical diversity for a single H . Then, we expand our investigation here to single protons having large MSD to test whether $\langle r^2 \rangle$ can be limited by the length $\lambda \sim 2\pi/Q$ observable. The Q -value is the length scale of the motions. Thus, there could be an intrinsic Q dependence in the MSD, i.e. at low Q motions over a longer distance are included which leads to a larger MSD at low Q .

4.2 The impact of the Gaussian approximation for all H in proteins

To investigate the impact of the higher cumulants in the incoherent Full ISF $I_{inc}(\mathbf{Q}, t)$ on the Q -dependence of the MSD $\langle r^2 \rangle$, we calculate the ISF $I_{inc}(\mathbf{Q}, t)$ (includes all cumulants), and $I_{iG}(\mathbf{Q}, t)$ (i.e keeping only the Gaussian term in $I_{inc}(\mathbf{Q}, t)$) for all H in hydrated lysozyme. The $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.3) and $I_{iG}(\mathbf{Q}, t)$ in Eq. (4.10) are calculated by using the 100 ns MD simulation at 300 K. To improve statistic, we divide the trajectory obtained from 100 ns MD simulation to 1000 segment by changing the starting time. Then, the $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ are recalculate averaged over $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ in these segments. The incoherent ISF $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.3) depends on the \mathbf{Q} vector. The $I_{inc}(\mathbf{Q}, t)$ observed in neutron scattering experiments is averaged over \mathbf{Q} vectors which have a same magnitude. In the calculation of $I_{inc}(\mathbf{Q}, t)$,

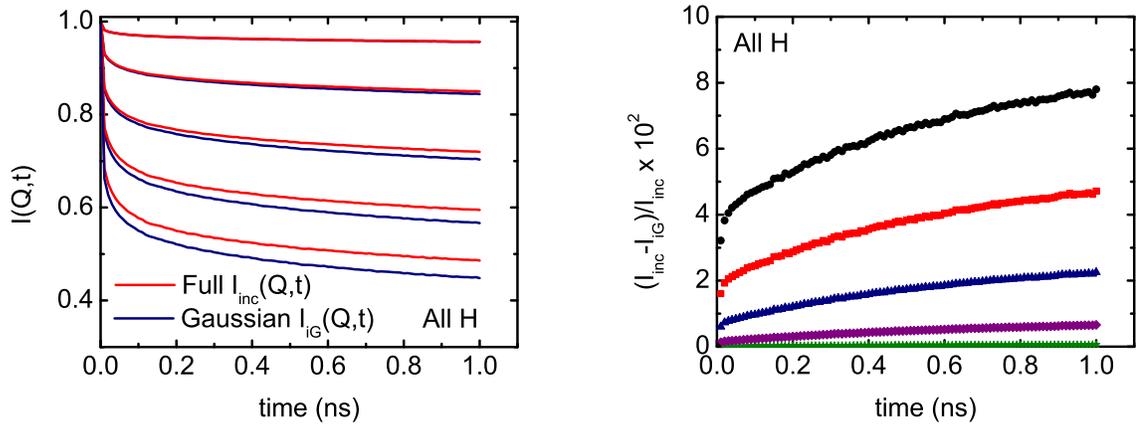


Figure 4.1: LHS : $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ calculated for all H in lysozyme for $0 < t < 1$ ns obtained from a 100 ns MD simulation at 300 K. Comparison of the Full $I_{inc}(\mathbf{Q}, t)$ (red solid lines) containing all cumulants and the Gaussian $I_{iG}(\mathbf{Q}, t)$ (blue solid lines) containing only second order cumulant term. RHS : The percentage of the difference between the Full $I_{inc}(\mathbf{Q}, t)$ and the Gaussian $I_{iG}(\mathbf{Q}, t)$. From top to bottom on LHS and from bottom to top on RHS, Q values are 0.4, 0.8, 1.2, 1.6 and 2 \AA^{-1} .

the average is taken over $I_{inc}(\mathbf{Q}, t)$ for different Q vectors having same magnitude (see *Appendix F*). Then, we fit the model of $I(\mathbf{Q}, t)$ in Eq. (3.1) to $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ to obtain the intrinsic MSD for each.

The comparison of the Full $I_{inc}(\mathbf{Q}, t)$ and Gaussian $I_{iG}(\mathbf{Q}, t)$ for all H at ten different Q values is shown on the LHS in Fig. 4.1. The Q values from top to bottom are 0.4, 0.8, 1.2, 1.6 and 2 \AA^{-1} . The red and blue solid lines represent the $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$, respectively. As shown on the RHS in Fig. 4.1, the difference between $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ is a little, less than 1% at low Q values and increases to 8% at $Q = 2 \text{ \AA}^{-1}$. This difference arises from the contribution of higher cumulants.

Fig. 4.2a and Fig. 4.2b show the fit of the model $I(\mathbf{Q}, t)$ to the $I_{inc}(\mathbf{Q}, t)$ and to $I_{iG}(\mathbf{Q}, t)$ for all H at 300 K, respectively. β is chosen as 0.24 for these fits. The fits are good. The best fit values of $\langle r^2 \rangle$ and λ in the model $I(\mathbf{Q}, t)$ in Eq. (3.1) are shown in Fig. 4.2c and Fig. 4.2d. The blue solid circles represent the $\langle r^2 \rangle$ and λ obtained from

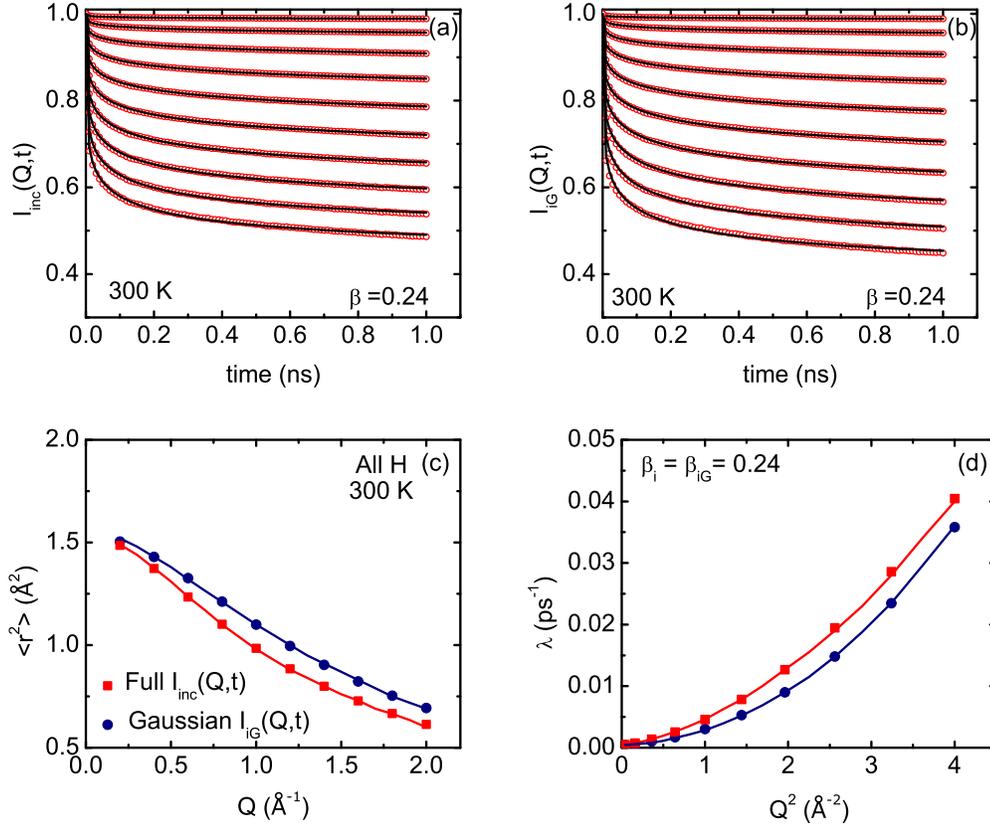


Figure 4.2: (Upper) The fit of $I(\mathbf{Q}, t)$ to (a) $I_{inc}(\mathbf{Q}, t)$ and (b) $I_{iG}(\mathbf{Q}, t)$ for $0 < t < 1$ ns and for all H with $\beta = 0.24$. (Lower) (c) The intrinsic MSD $\langle r^2 \rangle$ and (b) relaxation parameter λ . (Blue solid circles) the parameter obtained from the fit in (a). (Red solid square) for the fit in (b). The solid lines are a guide to the eye.

the fit of the model $I(\mathbf{Q}, t)$ to $I_{iG}(\mathbf{Q}, t)$, and the red solid squares represent the $\langle r^2 \rangle$ and λ , obtained from the fit of the model $I(\mathbf{Q}, t)$ to $I_{inc}(\mathbf{Q}, t)$. The MSD $\langle r^2 \rangle$ obtained from fitting the single scatterer model of $I(\mathbf{Q}, t)$ in Eq. (3.1) to $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ are almost same at low Q value. They differ by approximately 15% at $Q = 2 \text{ \AA}^{-1}$, but have effectively the same Q dependence. Both of MSDs $\langle r^2 \rangle$ are Q -dependence. This suggests that the higher cumulants are not significant. The Q -dependence of $\langle r^2 \rangle$ does not arise from using Gaussian approximation (for lysozyme at least).

4.3 The impact of the Gaussian approximation and dynamical diversity, heterogeneity for an individual H in proteins

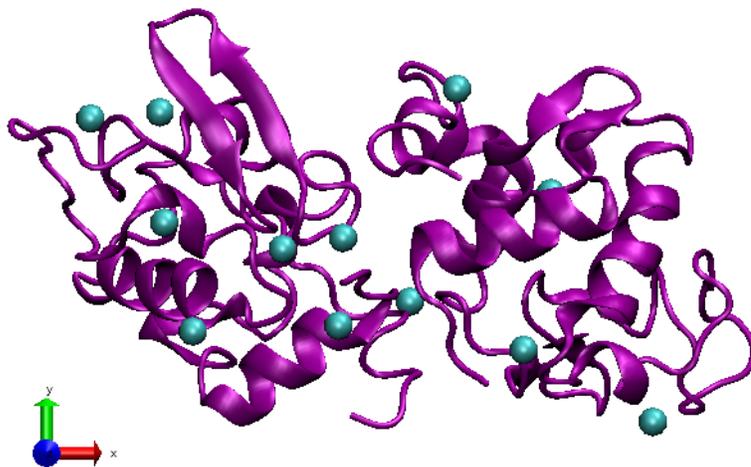


Figure 4.3: Chosen hydrogens in two lysozyme molecules.

In this section, we focus on an individual H in lysozyme. We choose 12 hydrogens from our simulation system including randomly oriented two lysozyme molecule. These 12 hydrogens shown in Fig. 4.3 are non-exchangeable hydrogens in lysozyme. 10 of 12 hydrogens are chosen on sidechain and 2 of them on backbone in lysozyme.

For these single hydrogens in Fig. 4.3, we calculated the ISF $I_{inc}(\mathbf{Q}, t)$ (includes all cumulants), and $I_{iG}(\mathbf{Q}, t)$ (i.e keeping only the Gaussian term in $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.3)). However, we only show the results for two single hydrogens. The comparison of the Full $I_{inc}(\mathbf{Q}, t)$ and the Gaussian $I_{iG}(\mathbf{Q}, t)$ for H in ILE55 and in VAL109 is shown in the upper frame in Fig. 4.4. The Q values from top to bottom are 1, 2, 3 and 4 \AA^{-1} for H in ILE55 and 0.4, 0.8, 1.2 and 1.6 \AA^{-1} for H in VAL109. The red and blue solid lines represent the $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$, respectively. The difference between the Full $I_{inc}(\mathbf{Q}, t)$ and the Gaussian $I_{iG}(\mathbf{Q}, t)$ is shown as percentage in the lower frame in Fig. 4.4. The difference between $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ is around 2%

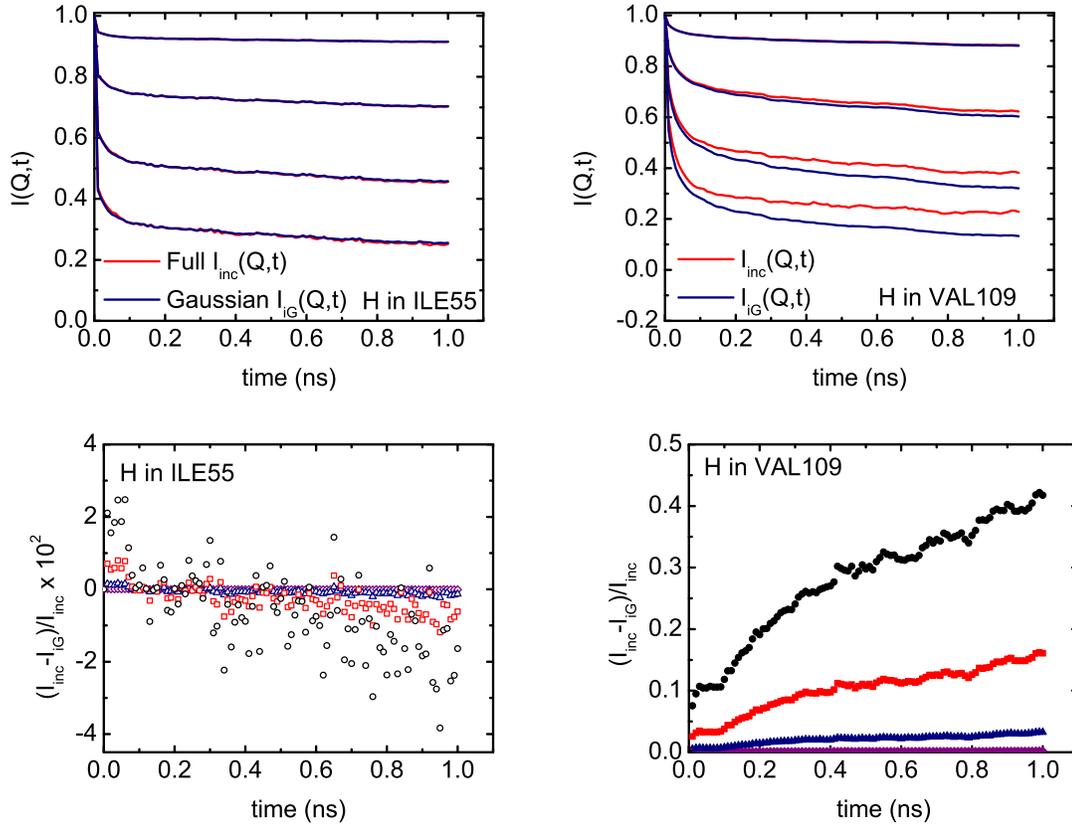


Figure 4.4: (Upper) The comparison of (red solid line) $I_{inc}(\mathbf{Q}, t)$ and (blue solid line) $I_{iG}(\mathbf{Q}, t)$. (Lower) The difference between $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$. H in ILE55 and VAL109.

at $Q = 4 \text{ \AA}^{-1}$ for H in ILE55, however it reaches 40% at $Q = 1.6$ for H in VAL109 due to the contribution of the methyl group rotation.

Fig. 4.5 shows that the best fit values of $\langle r^2 \rangle$ and λ for an individual H in ILE55 and in VAL109, respectively. The blue solid circles represent the $\langle r^2 \rangle$ and λ , obtained from the fit of the model $I(\mathbf{Q}, t)$ to $I_{iG}(\mathbf{Q}, t)$, and the red solid squares represent the $\langle r^2 \rangle$ and λ , obtained from the fit of the model $I(\mathbf{Q}, t)$ to $I_{inc}(\mathbf{Q}, t)$. The fits are good as much as the fits in the upper frame in Fig. 4.2. β value which is used in the fits for H in ILE55 and in VAL109 is 0.148 and 0.29, respectively. The $\langle r^2 \rangle$ for H in ILE55 and in VAL109 remains Q dependent when obtained from a fit to the Gaussian approximation

$I_{iG}(\mathbf{Q}, t)$. They are almost same for H in ILE55 and differ by approximately 40% at $Q = 1.6\text{\AA}^{-1}$ for H in VAL109. They also have effectively the same Q dependence. These results indicate that neglecting higher cumulants in $I_{inc}(\mathbf{Q}, t)$ does not lead to Q independent $\langle r^2 \rangle$. In Fig. 4.5, we see that λ is also Q dependent and increases with increasing Q^2 .

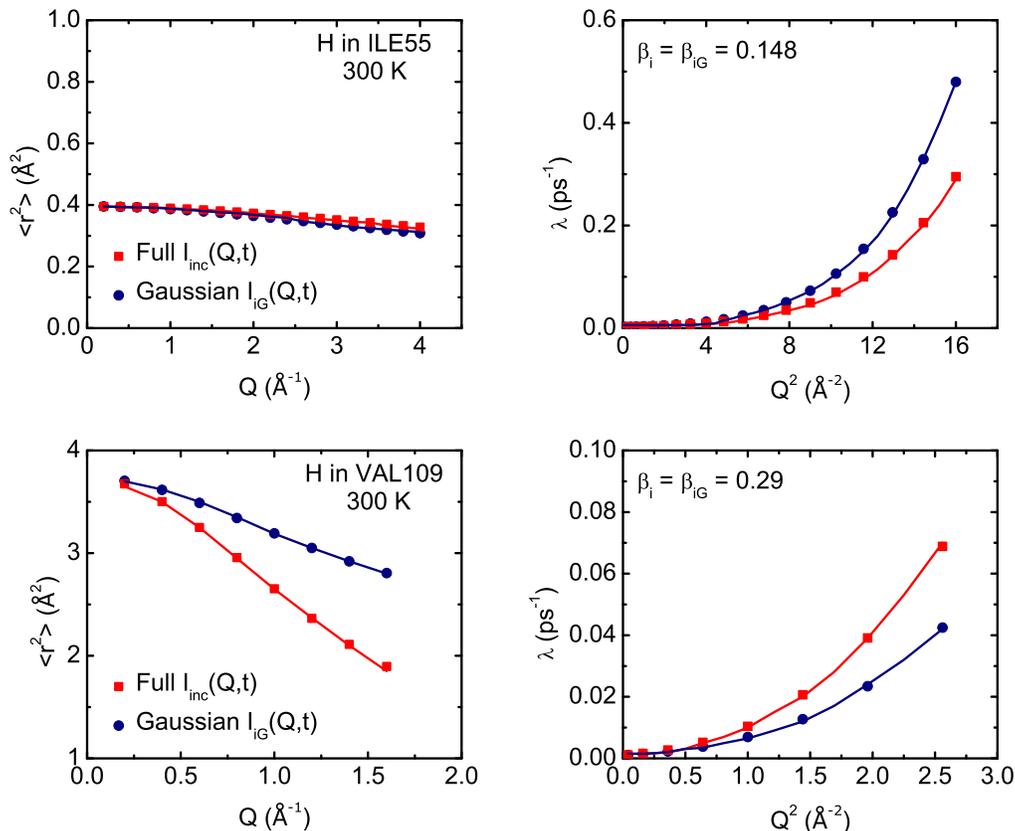


Figure 4.5: (Upper) The parameters, (LHS) $\langle r^2 \rangle$ and (RHS) λ , obtained from fits to the $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ in Fig. 4.4a. The MSD remains Q dependent when obtained from a fit to the Gaussian approximation $I_{iG}(\mathbf{Q}, t)$. The blue solid circles represent the parameters obtained from fits to the $I_{inc}(\mathbf{Q}, t)$. The blue solid squares represent the parameters obtained from fits to the $I_{iG}(\mathbf{Q}, t)$. (Lower) Same as Upper frame for $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$ in Fig. 4.4b.

Fig. 4.6 shows the intrinsic MSD $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ to the Full $I_{inc}(\mathbf{Q}, t)$ for an individual H in five different residue in lysozyme. These

are ALA122, THR40, LYS33, LYS116 and VAL109. For single H that has an average or smaller MSD, the fitted MSD $\langle r^2 \rangle$ is independent of Q , as shown on the LHS in Fig. 4.6. For hydrogens having large MSD (e.g. $\langle r^2 \rangle \sim 5 \text{ \AA}^2$ at $Q \rightarrow 0$) on the RHS in Fig. 4.6, the fitted $\langle r^2 \rangle$ can be somewhat Q dependent. The MSD $\langle r^2 \rangle$ is somewhat smaller at larger Q as if the MSD is limited at large Q . However, the effect is not large. Also since there are so few H that have large MSD, this effect contributes little to the global $I_{inc}(\mathbf{Q}, t)$ of the protein. This indicates that the apparent Q dependence of the MSD $\langle r^2 \rangle$ arises from dynamical diversity. It arises from fitting a model ISF $I(\mathbf{Q}, t)$ (or DSF) representing a single scatterer to data which is a collection of scattering from a large number of dynamically diverse scatters (H) in the protein. When the same model is fitted to $I_{inc}(\mathbf{Q}, t)$ of individual H in the protein (no DD), a Q independent MSD is obtained, specially for hydrogens having an average or smaller MSD.

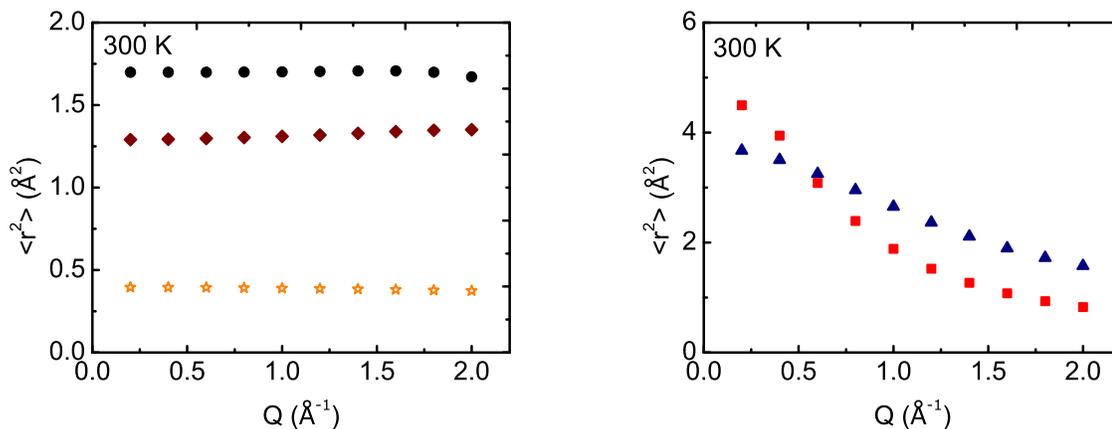


Figure 4.6: (LHS) A Q independent $\langle r^2 \rangle$ is obtained from simulations of a single H (no dynamical diversity). The solid circle is H in ALA122, the solid diamond is H in THR40 and the open star is H in LYS33. RHS : If the $\langle r^2 \rangle$ of a single H is large enough some Q dependence can remain but there are few such H . The solid square is H in LYS116 and the solid triangle is H in VAL109.

Fig. 4.7 shows that the kurtosis γ_α along the α axis, $\alpha = x, y, z$ for four individual H in ILE55, TRP40, ARG68 and VAL109. The kurtosis changes between ± 0.2 for H

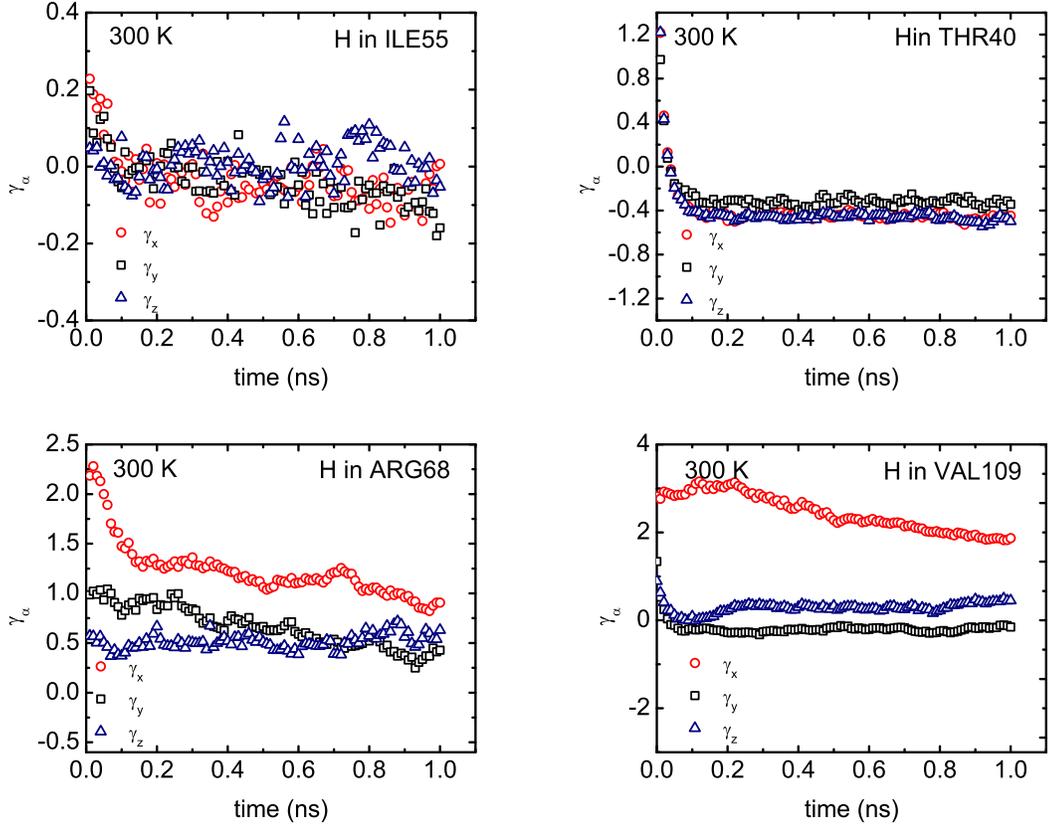


Figure 4.7: The kurtosis term γ along x (red open squares), y (open black squares) and z (blue open triangle) for a single H in lysozyme for $0 < t < 1$ ns obtained from a 100 ns MD simulation at 300 K. H in ILE55, THR40, ARG68 and VAL109.

in ILE 55, but it changes between around 0 and 3 for H in VAL109. The larger kurtosis is observed for a single H which has the larger MSD. This result indicates that the higher order terms are significant for an individual H which has a large MSD.

The $\langle r^2 \rangle$ obtained from simulations of the Full $I_{inc}(\mathbf{Q}, t)$ and the Gaussian approximation $I_{iG}(\mathbf{Q}, t)$ from Eq. (4.9) for an individual H in ALA122 and THR40 is shown on the LHS in Fig. 4.8. The blue solid circles represent the $\langle r^2 \rangle$ and λ , obtained from the fit of the model $I(\mathbf{Q}, t)$ to $I_{iG}(\mathbf{Q}, t)$, and the red solid squares represent the $\langle r^2 \rangle$ and λ , obtained from the fit of the model $I(\mathbf{Q}, t)$ to $I_{inc}(\mathbf{Q}, t)$. For H in ALA122 and THR40, the $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ to $I_{inc}(\mathbf{Q}, t)$ somewhat

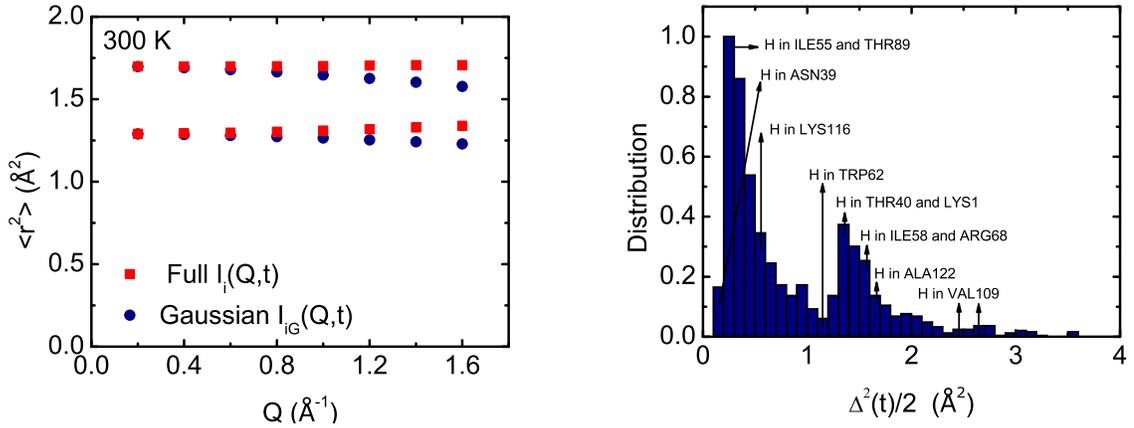


Figure 4.8: LHS : The $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ to the Full $I_{inc}(\mathbf{Q}, t)$ (solid squares) and the Gaussian approximation $I_{iG}(\mathbf{Q}, t)$ (solid circles) for an individual H in ALA122 and THR40. RHS : The distribution of the MSD of individual H in lysozyme calculated using $\Delta^2(t)/2 = \langle (r(t) - r(0))^2 \rangle / 2$ at time $t = 1$ ns at 300 K. The MSD for a single H , $\Delta^2(t)/2 = \langle (r(t) - r(0))^2 \rangle / 2$, out to time 1 ns is calculated from a 100 ns MD simulation.

increases with increasing Q -value since the $\langle r^2 \rangle$ obtained from the fit of the model $I(\mathbf{Q}, t)$ to $I_{iG}(\mathbf{Q}, t)$ decreases with increasing Q -value. They differ slightly at even high Q values. On the RHS in Fig. 4.8, the distribution of the MSD of individual H in lysozyme calculated using $\Delta^2(t)/2 = \langle (r(t) - r(0))^2 \rangle / 2$ at time $t = 1$ ns at 300 K is shown. The distribution is similar to a sum of two Gaussian distribution functions with different centers. It is noted that most of hydrogens have a MSD between 0.2 and 2, and the number of hydrogens having large MSD is few.

4.4 Discussion

We find that the Gaussian approximation to the elastic incoherent DSF,

$$S(\mathbf{Q}, \omega = 0) = I(\mathbf{Q}, t = \infty) = \exp\left(-\frac{1}{3}Q^2\langle r^2 \rangle\right), \quad (4.11)$$

is an accurate approximation out to $Q = 2 - 3 \text{ \AA}^{-1}$ for H in lysozyme. The approximation is tested firstly by calculating the full incoherent function $I_{inc}(\mathbf{Q}, t)$ and the Gaussian

approximation to it, $I_{iG}(\mathbf{Q}, t)$, for all H and showing (1) that the difference between the two is small and (2) that the MSD $\langle r^2 \rangle$ obtained from the two differs insignificantly. It is secondly tested by calculating the leading term beyond the Gaussian, the fourth cumulant, and showing that the fourth cumulant is negligible except for a few isolated H that have very large MSD. The Q dependence of the MSD obtained from experiments does not arise from making the Gaussian approximation to $S(\mathbf{Q}, \omega = 0)$.

We find that the apparent Q dependence of observed MSD arises from neglecting the dynamical diversity of H in a protein in the analysis of data. That is, a model $I(\mathbf{Q}, t)$ or $S(\mathbf{Q}, \omega)$ representing a single scattering center is fitted to data arising from thousands of H in diverse dynamical environments having a wide spectrum of MSD values. It is the use of a single scatterer model to describe a spectrum of scatterers that leads to a Q dependent average MSD. If the same model is fitted to data arising from a single H (any single H), then a Q independent $\langle r^2 \rangle$ is obtained.

For a few exceptional individual H that have very large MSD, the extracted $\langle r^2 \rangle$ remains somewhat Q dependent. In these exceptional cases, the observed length could be limited by Q . There are so few such H that this does not affect the observed $\langle r^2 \rangle$.

The Q dependence can be used to infer and extract some information on dynamical diversity. For example, the Q dependence of the MSD is larger at high temperature than at low temperature. This shows that the dynamical diversity is larger at high temperature. This may be expected since the average MSD is larger indicating that spread of MSDs could be larger. Calculating the histogram of MSDs as a function of temperature would display the temperature dependence of the dynamical diversity. A future project is reproduce the Q dependence of the MSD using models of the distribution of MSDs.

Chapter 5

VIBRATIONAL DYNAMICS OF HYDROGEN IN PROTEINS

As seen in previous chapters the MSD in proteins depends on temperature. At low temperatures, the MSD increases linearly with increasing temperature. The MSD arising from vibration is given as

$$\langle u^2 \rangle = \frac{k_B T}{\phi} \quad (5.1)$$

in the classical limit and in a harmonic approximation [6, 25], where k_B is the Boltzmann constant, T is the temperature and ϕ is a harmonic force constant. At higher temperatures, a significant increase in the slope of the MSD is observed which could also be expressed using Eq. (5.1) using a reduced force constant ϕ at higher temperatures.

In this chapter, our goal is to illustrate that a marked change in the slope of $\langle u^2 \rangle$ with temperature can arise from the vibrational dynamics of a particle in an anharmonic potential well using the self-consistent harmonic (SCH) theory [14, 37, 38, 57]. We consider a single particle in a 1D potential well and investigate the temperature dependence of the MSD. In the harmonic approximation the force constant ϕ is given by the second derivative of the potential $V(u)$ evaluated at the bottom of the well, $\phi = \frac{\partial^2 V}{\partial u^2} |_{u=0}$. In the SCH model, the optimum harmonic force constant ϕ is obtained as the second derivative of the potential averaged over the vibrational distribution of the particle in the well. In a Gaussian potential, as in most potentials, the second derivative of $V(u)$ decreases as u increases. In this way SCH model force decreases as the vibrating particle samples larger values of u in the potential. This reduction in the force constant suggests a gradual increase in the slope of the MSD with temperature. An abrupt change in slope of $\langle u^2 \rangle$ with temperature can also be obtained if a potential

which has hard wall and soft wall components in it. In this model, the effective force constant, ϕ , can be adjusted as decrease with temperature, and this leads to an increase in slope of $\langle u^2 \rangle$ with temperature at T_D .

The SCH theory is applied to investigate the dynamics of a single particle in an anharmonic potential. The aim is to reproduce the change in the slope of MSD $\langle u^2 \rangle$ versus temperature taking account of vibration of a single particle in an anharmonic potential well alone. The present model does not contain any thermally activated, transition rate process and any diffusive motion.

5.1 Dynamical Model

We consider the SCH theory [14, 38, 57, 37] for the case with a single particle in a potential $V(u)$. Note that the SCP theory provides a practical method to determine the effective harmonic force constant ϕ for a single particle in an anharmonic potential. The derivation of the SCH theory is given in *Appendix G*.

The Hamiltonian for the single particle is

$$H = \left(-\frac{1}{2M}\right)\nabla^2 + \frac{1}{2}V(u), \quad (5.2)$$

where u is the displacement from the potential minimum. A harmonic model can be introduced via a harmonic Hamiltonian H_h ,

$$H_h = \left(-\frac{1}{2M}\right)\nabla^2 + \frac{1}{2}u\phi u. \quad (5.3)$$

The usual harmonic force constant is $\phi(u) = (d^2V/du^2)_{u=0}$, the second derivative of potential at the minimum of the well. In the self-consistent harmonic theory, $\phi(u)$ is replaced by the second derivative averaged over the vibrational distribution $\rho(u)$ in the well:

$$\phi = \int_{-\infty}^{\infty} du \rho(u) \frac{d^2V(u)}{du^2}. \quad (5.4)$$

The distribution is a Gaussian distribution as found in a harmonic potential,

$$\rho(u) = [2\pi\langle u^2 \rangle]^{-1/2} \exp\left(-\frac{u^2}{2\langle u^2 \rangle}\right), \quad (5.5)$$

where $\langle u^2 \rangle$ is the usual harmonic expression of the mean square displacement,

$$\langle u^2 \rangle = \frac{\hbar}{2M\omega} \coth\left(\frac{k_B T}{\hbar\omega}\right). \quad (5.6)$$

Here $\omega^2 = \phi/M = \sqrt{\phi/m}$ is the harmonic frequency, M is the particle mass and \hbar is Planck's constant divided by 2π . In the classical limit, $T \geq \theta_E = \hbar\omega/k_B$, Eq. (5.6) reduces to

$$\langle u^2 \rangle = \frac{k_B T}{\phi} \quad (5.7)$$

Eqs. (5.4) to (5.6) constitute the SCH theory. It is implemented by iterating Eqs. (5.4), (5.5) and (5.6) until consistent beginning, for example, with an estimated value of $\langle u^2 \rangle$. The SCH can be derived by summing a class of anharmonic terms as well as by minimizing the free energy or ground state energy at $T = 0$ K.

For asymmetric potentials, the center of the Gaussian distribution function can also change with temperature. Thus, the Gaussian distribution function in Eq. (5.5) is replaced by

$$\rho(u) = \frac{1}{(2\pi\langle u^2 \rangle)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \frac{(u - \Delta)^2}{\langle u^2 \rangle}\right). \quad (5.8)$$

The Δ term in Eq. (5.8) is obtained by minimizing the free energy. In the classical limit the position and momentum are independent variables, so that the kinetic energy is independent of the potential energy. Free energy minimization reduces to minimizing the potential energy,

$$\langle V(u) \rangle = \int_{-\infty}^{\infty} du V(u) \rho(u). \quad (5.9)$$

$\langle V(u) \rangle$ depends on Δ and $\langle u^2 \rangle$. The SCH equations are iterated for given Δ and $\langle V(u) \rangle$ which is minimized with respect to Δ .

To simplify our equations, we normalize parameters to be dimensionless. We use the SCH theory to determine changes in the mean square vibrational amplitude with temperature for different potential models. The temperature dependence of the

mean square vibrational amplitude is well known in a harmonic potential, and hence we start with it. Then, we extend our investigation to different potentials which are similar to potentials in proteins.

5.2 Potential Models

5.2.1 Harmonic Potentials

In order to introduce the model of a particle in 1D potential well which represents an atom in a macromolecule, we begin with a harmonic potential well. At low temperature, $T \lesssim 120$ K, the observed $\langle r^2 \rangle_{exp}$ in biological macromolecules increases linearly with temperature, as in Eq. 5.7. A harmonic potential well is used for determining a low temperature force constant ϕ_L to reproduce the observed straight line for $\langle r^2 \rangle_{exp}$ versus T at low T . The harmonic potential $V(u) = (1/2)\phi_L u^2$ depends on a low temperature force constant ϕ_L and displacement u . In the classical limit, the $\langle u^2 \rangle$ is defined in Eq. 5.7 with $d^2V/du^2 = \phi_L$ where the oscillator (SHO) frequency is $\omega^2 = \omega_L^2 = \phi_L/M$ and a mass is chosen as $M = 20$ Amu which is intended to represent hydrogen bonded in a molecule.

We select arbitrarily the observed MSD $\langle r^2 \rangle_{exp}$ versus T data from literature. A reference point, (u_0^2, T_0) , is chosen from MSD data, to determine the low temperature force constant ϕ_L , where the chosen length u_0^2 corresponds to a convenient temperature T_0 . The MSD versus temperature data for hydrated myoglobin ($h = 0.38$ gr protein/gr water) which is observed by Doster et al. is chosen [25]. A convenient temperature T_0 , arbitrarily choosing $T_0 = 240$ K, and the convenient length $u_0^2 \equiv \langle r^2 \rangle_{exp} = k_B T_0 / \phi_L$ is $u_0^2 = 0.1 \text{ \AA}^2$ at $T_0 = 240$ K. The low temperature force constant for myoglobin is obtained from $(\phi_L/k_B) = T_0/u_0^2$. The MSD for myoglobin at low temperatures is seen as a straight solid line in Fig. 5.1a.

The low temperature force constant is chosen as (ϕ_L/k_B) for harmonic potential. For potential wells $V(u)$ which is not harmonic, the parameters in potential well $V(u)$ is adjusted to obtain a potential function in units of k_B , so that the low temperature force constant is $(d^2V/du^2)_{u=0} = (\phi_L/k_B) = T_0/u_0^2$. This adjustment provides the

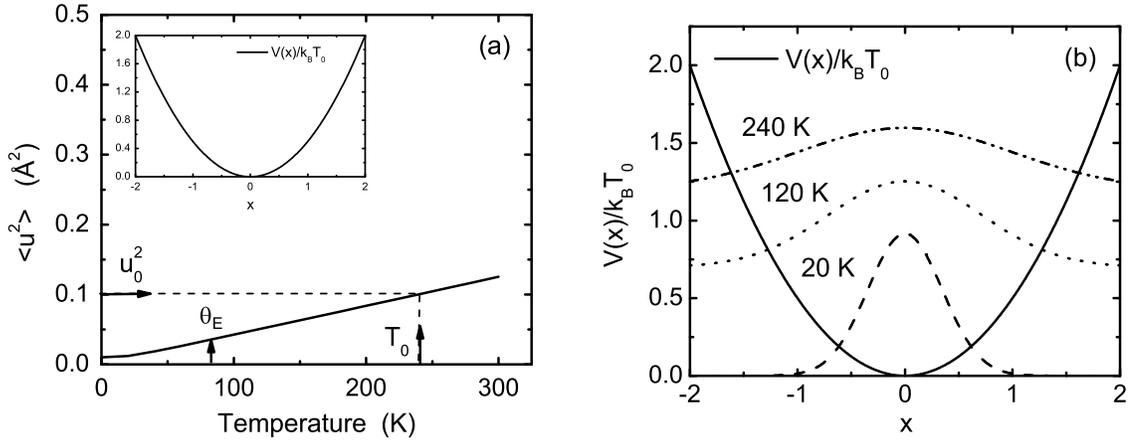


Figure 5.1: (a) The mean square vibrational displacement, (MSD), $\langle u^2 \rangle$, given by (Eq. (5.7)) of an atom in a harmonic well. The slope of $\langle u^2 \rangle$ vs. T in the classical limit, $T \geq \theta_E$, is given by $\langle u^2 \rangle = k_B T / \phi_L$. We choose the low temperature force constant ϕ_L in our model so that it reproduces the observed slope of myoglobin at low temperature, i.e. $r_0^2 = \langle r^2 \rangle_{exp} = k_B T_0 / \phi_L$ where $u_0^2 = 0.1 \text{ \AA}^2$ is the observed MSD at temperature $T_0 = 240 \text{ K}$. (b) The vibrational distribution $\rho(x)$ in harmonic well $v(x) = V(u)/k_B T_0$, $x = u/u_0$ at three temperatures.

same initial slope of $\langle u^2 \rangle$ versus T at low temperatures for any potential. The force constant changes according to the potential, at higher temperatures and the change in the force constant determines the temperature dependence of the $\langle u^2 \rangle$.

To simplify the model, the harmonic potential $V(u) = (1/2)\phi_L u^2$ is written as $V(x) = (1/2)\phi_L x^2 u_0^2$ in the terms of the dimensionless length $x = u/u_0$. The harmonic potential in the form $v(x) = V(x)/k_B T_0$ is shown in the inset of Fig. 5.1a. Using $\phi_L/k_B = T_0/u_0^2$, the potential becomes $v(x) = (1/2)x^2$. Fig. 5.1a shows the MSD $\langle u^2 \rangle$ of a particle in harmonic potential versus T . The MSD is constant at temperatures below the Einstein temperature, $\theta_E = \hbar\omega/k_B$, while it increases linearly with temperature at temperatures above θ_E . Fig. 5.1b shows the harmonic potential and vibrational distribution $\rho(x) = [2\pi\langle x^2 \rangle]^{-1/2} \exp[-x^2/2\langle x^2 \rangle]$ at three temperatures, 20, 120 and 240 K, where $\langle x^2 \rangle = \langle u^2 \rangle / u_0^2$. Since $d^2v(x)/dx^2 = 1$ for all x for the

harmonic potential, the SCH force constant ϕ given by Eq. 5.4 remains $\phi = \phi_L$ at all T independent of the width of $\rho(x)$. For a harmonic potential, the SCH model is the same as the harmonic approximation.

5.2.2 Gaussian Potentials

The Gaussian function is a good prediction to reproduce the temperature behaviour of the MSD $\langle u^2 \rangle$. The inset of Fig. 5.2a shows the Gaussian potential,

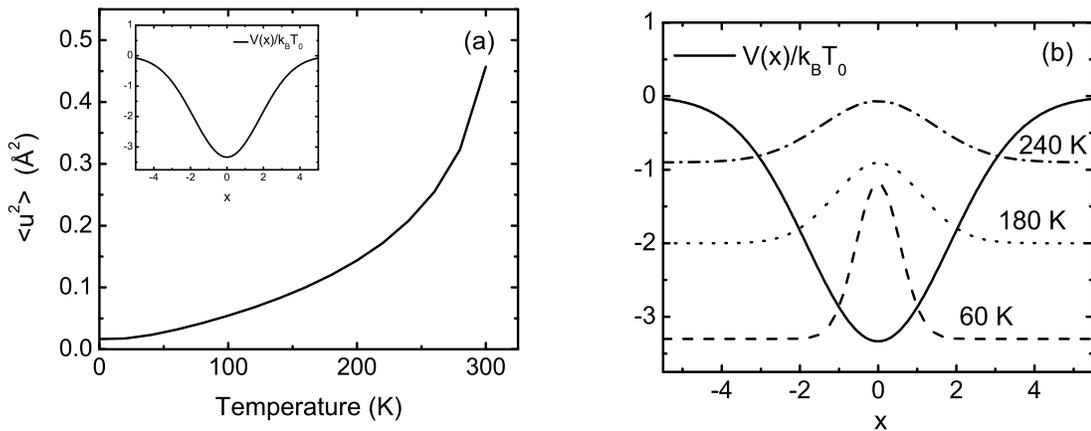


Figure 5.2: (a) The MSD, $\langle u^2 \rangle$, of a particle in a Gaussian well, $V(x)$, where $x = u/u_0$. The slope of $\langle u^2 \rangle$ with temperature shows a marked increase at a $T \sim 250$ K simulating a dynamical transition at $T_D \sim 250$ K. (b) The vibrational distribution $\rho(x)$ in the Gaussian well at three temperatures.

$$V(u) = -Ae^{-\alpha u^2}, \quad (5.10)$$

in the form $v(x) = V(x)/k_B T_0$ with the dimensionless length $x = u/u_0$ by adjusting parameters A and α . In the Gaussian potential, d^2V/du^2 decreases with increasing u and the force constant ϕ in Eq. (5.5), decreases with increasing temperature as the distribution function, $\rho(u)$, broadens. The reduce in the force constant suggests that the slope of $\langle u^2 \rangle$ versus T ($\langle u^2 \rangle = k_B T/\phi$) increases with increasing temperature as shown in Fig. 5.2a. The change in the MSD depends on the chosen parameters in

the Gaussian potential. Fig. 5.2b shows the distribution function at three different temperatures, 60, 180 and 240 K. The $\rho(x)$ begins with sample regions of $V(x)$ where d^2V/dx^2 is quite small (even negative) at a temperature around 250 K. The distribution function becomes broader with increasing temperature. Thus the SCH force constant ϕ decreases rapidly with increasing T at temperatures larger than 250 K, so that the rapid increase in the slope of the MSD $\langle u^2 \rangle = k_B T / \phi$ is observed.

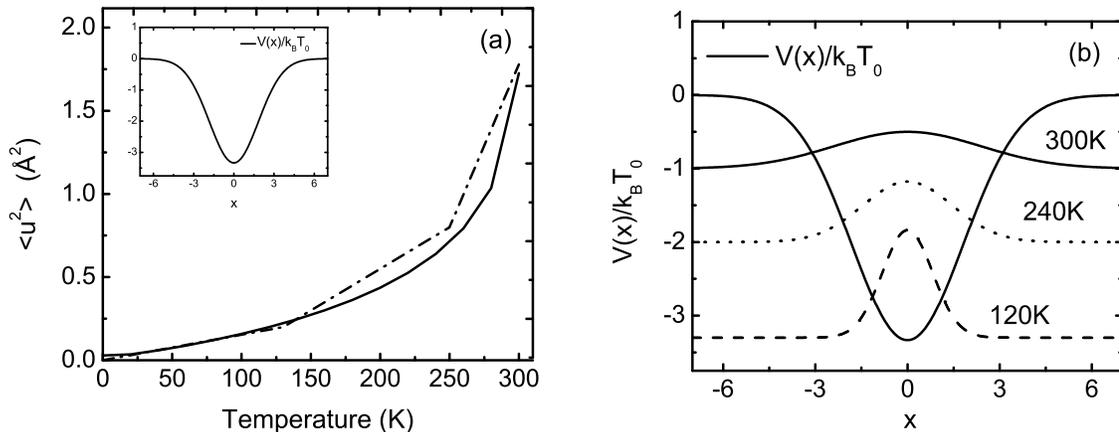


Figure 5.3: (a) The calculated MSD $\langle u^2 \rangle$ for a Gaussian potential well (solid line). The low temperature force constant ϕ_L of the Gaussian is set to reproduce the low temperature slope of $\langle u^2 \rangle$ vs. T in purple membrane (PM). The calculated $\langle u^2 \rangle$ at higher temperature for the Gaussian $V(u)$ reproduces the observed $\langle u^2 \rangle$ for PM (dashed-dotted line) [106] well at higher temperature ($T \gtrsim 200$ K). The vibrational distribution $\rho(x)$ is shown in part (b).

In the Gaussian potential, the $\langle u^2 \rangle$ changes as depending on the second derivative of $V(u)$ that changes with u . The significant change in the $\langle u^2 \rangle$ is observed at temperature, T_D , when the second derivative of $V(u)$ changes rapidly at energies $V(u) \sim k_B T_D$. For example, we choose the MSD versus temperature data for purple membrane (PM) which is observed by Wood et al. [106]. The parameters are adjusted to obtain the low temperature force constant ϕ_L . In Fig. 5.3a, solid line represents the $\langle u^2 \rangle$ versus T of a particle in the Gaussian potential in the inset of Fig. 5.3a and

the dashed line represents the observed MSD $\langle u^2 \rangle$ of PM. The reduce in the force constant leads to obtain the broaden distribution function shown in Fig. 5.3b. As shown Fig. 5.3a, the observed $\langle u^2 \rangle$ versus T in PM is reproduced well at low and at high temperatures using a Gaussian potential.

5.2.3 Symmetric Potentials

We consider a single particle in cases with different symmetric potentials to investigate the vibrational dynamics of a particle and observe how well a marked change in the slope of MSD $\langle u^2 \rangle$ versus temperature is reproduced. As shown in the inset of Fig. 5.4a, the symmetric potential is chosen as sum of two harmonic potentials which have different force constants. Similar to the harmonic potential, the low temperature force constant is determined using $(\phi_L/k_B) = T_0/u_0^2$, where $u_0^2 = 0.1 \text{ \AA}^2$ and $T_0 = 240 \text{ K}$ is chosen from the observed MSD data for hydrated myoglobin [25]. The MSD for a particle in the potential is shown in the inset of Fig. 5.4a. The short dashed line in Fig. 5.4a indicates the slope of $\langle u^2 \rangle$ at low temperature which is set by the large force constant ϕ_L . At high temperatures the slope of $\langle u^2 \rangle$ is set by a much smaller force constant ϕ_H that characterizes the shallow harmonic well at large values of x . The slope of MSD changes at temperature $T \sim 150 \text{ K}$. The long dashed line in Fig. 5.4a indicates the steeper slope of $\langle u^2 \rangle$ at higher temperatures. This temperature depends on the energy at which the $V(x)$ crosses over from a large force constant ϕ_L to the smaller one, ϕ_H that is determined by adjusting the parameters in potential. Fig. 5.4a shows the distribution function of the particle in the potential in the inset of Fig. 5.4a. At higher temperatures, the distribution function $\rho(x)$ becomes more broaden, but it still continues to sample the steep well potential near $x = 0$. Therefore, the slope will always be gradual in the transition from the low (L) to high (H) temperatures.

Fig. 5.5a shows the MSD $\langle u^2 \rangle$ for a particle in a potential which is again the sum of two harmonic components, but a barrier separates the two components as depicted in the inset of Fig. 5.5a. Modifying the potential in the inset in Fig. 5.4a with a barrier between two harmonic potential does not lead to a significant change in the MSD $\langle u^2 \rangle$.

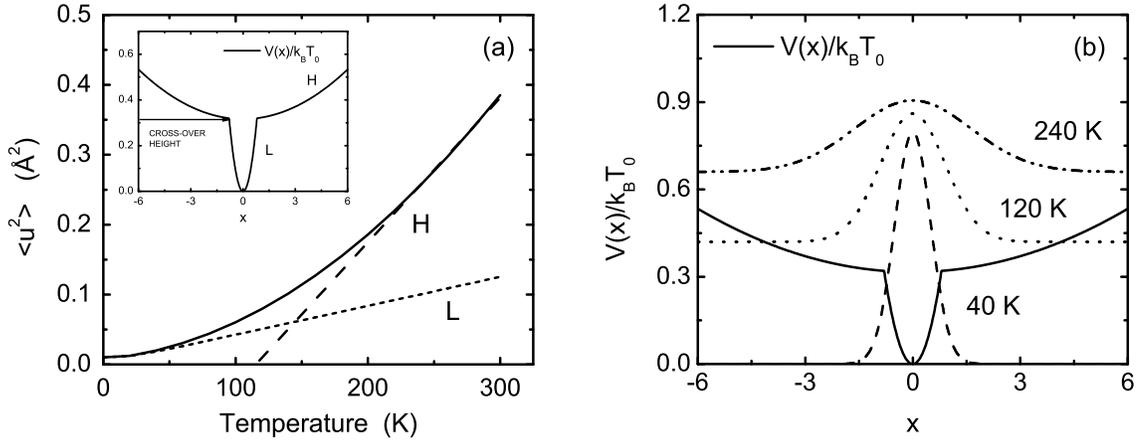


Figure 5.4: (a) The $\langle u^2 \rangle$ in a potential well composed of two harmonic components characterized by force constants ϕ_L and ϕ_H . The slope crosses over gradually from a low (L) temperature to high (H) temperature value. (b) The $\rho(x)$ for a potential composed of two harmonic components.

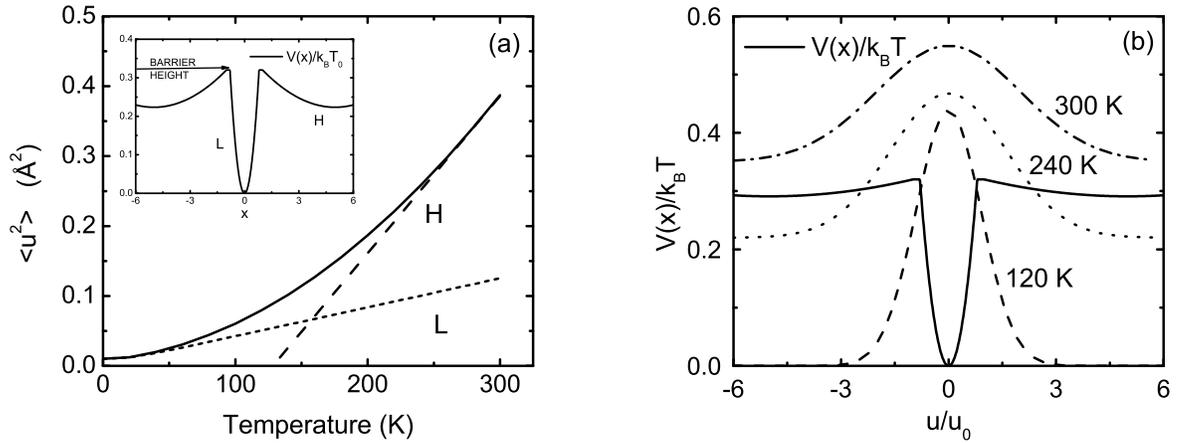


Figure 5.5: (a) The $\langle u^2 \rangle$ for an atom in a potential well composed of two harmonic components with a barrier separating the two components (see inset). (b) The $\rho(x)$ for a potential composed of two harmonic components with a barrier.

The MSDs $\langle u^2 \rangle$ in Fig. 5.4a and Fig. 5.5a have the same basic character. This suggests that a barrier and displacing the minimum of the shallow potentials to finite of x make a small difference in $\langle u^2 \rangle$. As shown in Fig. 5.5b, the distribution function samples the steep well potential near $x = 0$ at even 300 K. For the potentials which are symmetric about $x = 0$, the MSD and the distribution function are very similar.

In Sec.5.2.3, we employed the SCH model to two potentials which are symmetric around $x = 0$. This allows to reproduce the change in the slope of $\langle u^2 \rangle$ versus T at low temperature, however the significant change at higher temperatures is not observed because the center of the distribution function around $x = 0$ does not change with temperature. Using these potential models with SCH theory is not useful to reproduce the temperature dependence of the MSD $\langle u^2 \rangle$ which is observed in macromolecules.

5.2.4 Asymmetric Potentials

We analyse the dynamics of a single particle in potentials which are asymmetric with respect to $x = 0$. Note that the distribution function $\rho(x)$ is not centered around $x = 0$ in an asymmetric potential, and this causes a significant change in the MSD at higher temperatures. To accommodate the change in the center of $\rho(x)$, we use the distribution function defined in Eq. 5.8. As done before, the SCH $\langle u^2 \rangle$ and ϕ are obtained from Eq. (5.6) and Eq. (5.4), respectively. The Δ is obtained by minimizing the potential energy $\langle V(x) \rangle$ as a function of Δ . Since we are in the classical limit, minimization of the free energy reduces to minimization of the potential energy.

A double well potential which consists of two harmonic potential wells separated by a barrier is a good model to investigate the impact of the change in the center of distribution function on the MSD. The inset of Fig. 5.6a shows the double well potential which have two harmonic potential wells with the same force constant ϕ_L . The force constant of wells is determined from the expression for the low temperature force constant, $(\phi_L/k_B) = T_0/u_0^2$, where $u_0^2 = 0.1 \text{ \AA}^2$ and $T_0 = 240 \text{ K}$. Fig. 5.6b shows the $\rho(x)$ versus $x = u/u_0$ at three different temperatures, 60, 120 and 240 K. The center of $\rho(x)$ is around $x = 0$ at 60 K and this indicates that the particle in the double well

potential is confined to the well centered at $x = 0$ at low temperatures. At higher temperatures, the center of the $\rho(x)$ moves to the right as temperature increases until it reaches the midpoint between two harmonic wells. As shown in Fig. 5.6b, the $\rho(x)$ is centered symmetrically between the two wells at $T = 240$ K. In Fig. 5.6a, the change in the slope of the MSD for a particle in the double well potential does not reproduce the marked increase in slope of $\langle u^2 \rangle$ with increasing temperature well.

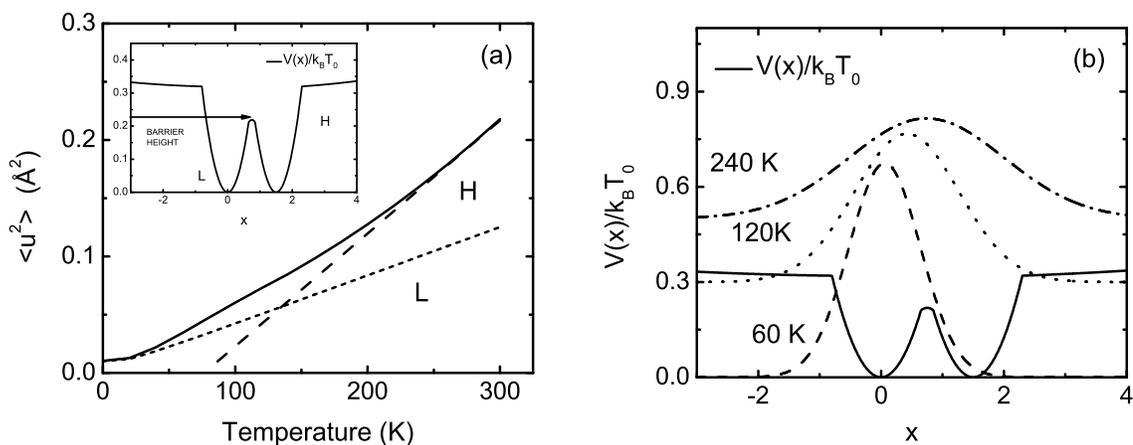


Figure 5.6: (a) The MSD, $\langle u^2 \rangle$, for a double well potential. The height of the barrier (at $x_B = 0.75$) between the two wells (see inset) is relatively small, $V(x_B)/k_B T_0 \simeq 0.2$. (b) The $\rho(x)$ for a particle in a double well potential. The center of $\rho(x)$, Δ moves from one well ($\Delta = 0$) at $T = 0$ K to the midpoint between the two wells ($\Delta = x_B$) at a relatively low temperature. The slope at high temperature is set by the harmonic well force constant (H) at larger x .

We introduce another asymmetric model which is intended to simulate temperature dependence of the motion of a particle near a surface or a wall as shown in Fig. 5.7. This asymmetric potential model consists a harmonic potential well, which has a wall on LHS and a small barrier on the RHS before the second harmonic potential well, centered at $x = 0$. The wall on the LHS formed as continuing to high energy ($\sim 2k_B T_0$) has the large low temperature force constant ϕ_L (near $x = 0$). The height of the barrier

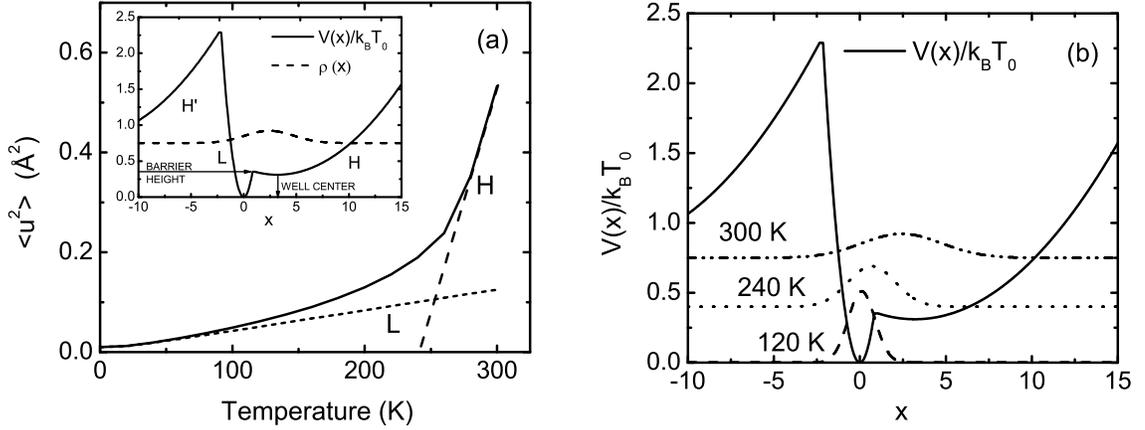


Figure 5.7: (a) The MSD $\langle u^2 \rangle$ for a well that has a hard wall component. The $\langle u^2 \rangle$ vs. T shows a large increase in slope at $T \simeq 250$ K, an increase arising from the center of $\rho(x)$ moving rapidly away from the wall as T increases as shown in part (b).

on the RHS is chosen as low. The force constant of the harmonic potential well and of the wall on the LHS are determined using the expression for low temperature force constant, $(\phi_L/k_B) = T_0/u_0^2$, and the reference point ($u_0^2 = 0.1 \text{ \AA}^2$ and $T_0 = 240$ K) taken from the MSD data for hydrated myoglobin [25]. The force constant of the well on the RHS is chosen to be smaller than the ϕ_L . Fig. 5.7b shows the distribution function which determines the motion of the particle in the asymmetric potential in the inset of Fig. 5.7a with temperature. The center of the distribution function is around $x = 0$ at 120 K, but it moves to the right with increasing temperature after 120 K until the center of the distribution function reaches to the minimum of the well centered around $x = 3$. Thus, the particle moves to the right away from the wall and crosses the barrier on the RHS when it has enough energy at higher temperatures. This allows that the MSD $\langle u^2 \rangle$ of the particle in the asymmetric potential model increases very rapidly at higher temperatures as shown in Fig. 5.7a since the particle has enough energy to cross the barrier on the RHS. At higher temperatures, the slope of $\langle u^2 \rangle$ is very steep. The steep slope is associated with the particle moving rapidly away from the wall as T is

increased i.e. $\Delta(T)$ increasing rapidly with T at higher temperatures. The potential shown in Fig. 5.7 is a simple but very flexible potential. It can be use to reproduce a large or small change in slope of $\langle u^2 \rangle$ versus T with the change in slope taking place at any temperature. It could also describe the eventual instability of a protein even at higher temperatures.

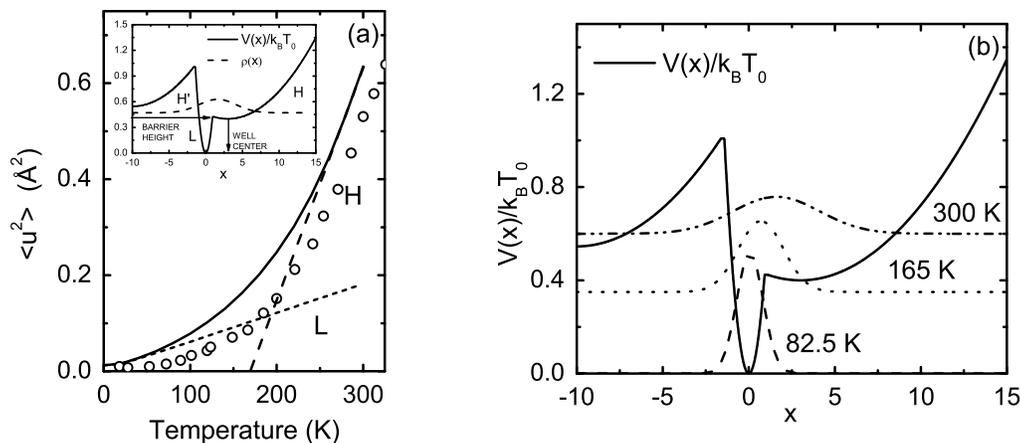


Figure 5.8: (a)The MSD $\langle u^2 \rangle$ for a well that has a hard wall component. The well parameters (ϕ_L , ϕ_H , barrier height and well center) are adjusted to reproduce the $\langle u^2 \rangle$ vs. T observed for myoglobin by Doster et al. at low and high temperature. The observed values of Doster et al. [25] as presented in Ref.[6]. The calculated $\langle u^2 \rangle$ clearly reproduce the observed $\langle u^2 \rangle$ for myoglobin well.(b)The distribution function $\rho(x)$ versus $x = u/u_0$.

As an example, we choose the $\langle u^2 \rangle$ observed by Doster et al. [25] for hydrated myoglobin, presented in Ref. [6]. To reproduce the $\langle u^2 \rangle$ for hydrated myoglobin, we adjust the parameters (barrier height, well center and ϕ_H) that determine the asymmetric potential. Similar to Fig. 5.7, the particle moves away from the wall at higher temperatures which is indicated in the change in the center of the distribution function $\rho(x)$ shown in Fig. 5.8b. The rate of the change in the slope of $\langle u^2 \rangle$ depends on the rate of the change of $\Delta(T)$ with T rather than the force constant ϕ_H describing the potential well on the RHS away wall. In this sense the physical origin of the slope

of $\langle u^2 \rangle$ at higher temperatures is quite different from that obtained for the symmetric Gaussian potential shown in Fig. 5.2.

The results in Sec. 5.2.2 and 5.2.4 show the temperature dependence of the $\langle u^2 \rangle$ reproduced by using the Gaussian potential or an anharmonic potential with hard wall. In a Gaussian potential, the second derivation of $V(u)$ changes with u , thus we observe the change in the slope of $\langle u^2 \rangle$ versus temperature. If the $d^2V(u)/du^2$ changes rapidly at energies $V(r) \sim k_B T_D$, the significant change in $\langle u^2 \rangle$ is observed at a dynamical transition temperature T_D . For an anharmonic potential with wall, the changes in the slope of $\langle u^2 \rangle$ depend on the motion of the particle in respect to the wall. If the particle moves away from the wall at temperature $T \gtrsim T_D$, the large slope is observed for $T \geq T_D$. In a Gaussian potential and an anharmonic potential with a wall, the parameters are adjusted to reproduce the observed $\langle u^2 \rangle$ of myoglobin and purple membrane.

5.3 Suggested Potential Models in Literature

5.3.1 Two State Model

Two state potential model is introduced by Frauenfelder et al, Keller and Debrunner, Doster et al. [35, 53, 25]. We investigated the dynamics of a single particle in a two state potential model which is asymmetric potential around $x = 0$. The distribution function $\rho(x)$ is not centered around $x = 0$ in an asymmetric potential. A two state model consists two harmonic potential wells which are separated by a barrier and have different force constants. Similar to the previous sections, this potential model is used for reproducing the observed MSD for myoglobin [25], by adjusting the parameter in the potential.

The inset of Fig. 5.9a shows the two state model that we used. The force constant of the well centered $x = 0$ is determined from the expression for the low temperature force constant, $(\phi_L/k_B) = T_0/u_0^2$, where $u_0^2 = 0.1 \text{ \AA}^2$ and $T_0 = 240 \text{ K}$. The force constant of the harmonic well on RHS is chosen to be smaller than the force constant of the harmonic well on LHS. Fig. 5.9b shows the $\rho(x)$ versus $x = u/u_0$ at

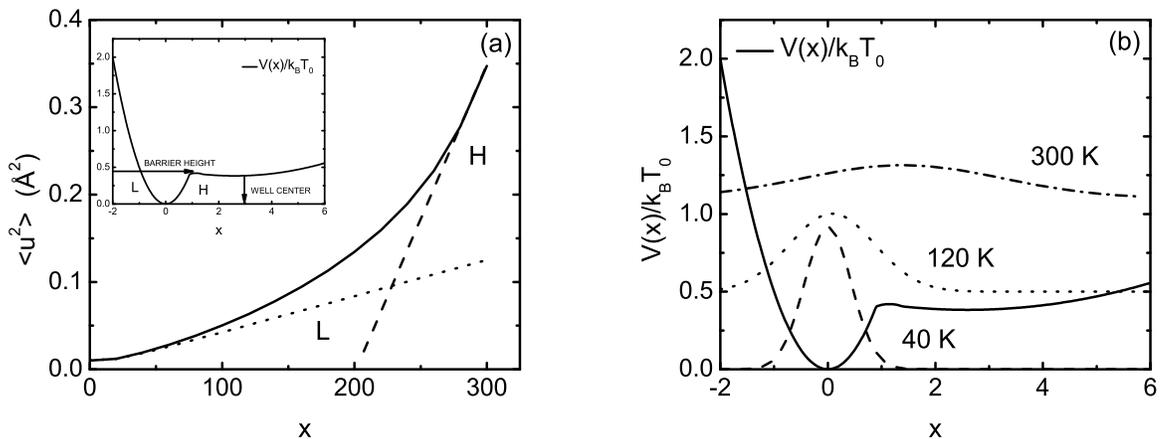


Figure 5.9: (a) The MSD, $\langle u^2 \rangle$, for a particle in the two state potential well. (b) The $\rho(x)$ for a particle in the two state potential.

three different temperatures, 40, 120 and 300 K. The center of $\rho(x)$ is around $x = 0$ at 40 K and this indicates that the particle in the two state potential model is confined to the well centered at $x = 0$ at low temperatures. At higher temperatures, the center of the $\rho(x)$ moves to the right as temperature increases. Fig. 5.9a shows that this potential coupled with the SCH dynamics does not lead to a $\langle u^2 \rangle$ that agrees well with experiment, essentially because the thermally activated transitions central to this model are not included.

Our goal is to illustrate that a change in the slope of the MSD $\langle u^2 \rangle$ versus temperature can be obtained within vibration of a mass in an anharmonic potential. It does not imply that thermally activated processes and diffusion is not important in the long time dynamics. The various dynamical processes may contribute to the dynamics of proteins as reviewed in the paper of Doster et al. [24].

5.3.2 Potential for H

The potential energy for H in interfacial water is obtained from the distribution of hydrogens measured by George Reiter [85]. The potential for hydrogen bond has a

double minima and it is symmetric. The potential is in unit of meV and distance in Å. The idea is to compare the experimentally obtained potential with the model potential that we have used. The theoretical model of potential of hydrogen is described by $V(u)$. Fig. 5.10a shows that the experimental potential and the theoretical potential model exactly match. The potential model is

$$V(u) = \begin{cases} \frac{1}{2}k(u+z)^2 + D & \text{if } -\infty < u \leq -z \\ A \exp(-\frac{u^2}{2B^2}) + C & \text{if } -z \leq u \leq z \\ \frac{1}{2}k(u-z)^2 + D & \text{if } z \leq u < \infty \end{cases}$$

Same as in previous potential models, a reference point is chosen as $u_0^2 = 0.1 \text{ \AA}^2$ and $T_0 = 240 \text{ K}$. Using $x = u/u_0$ and dividing potential to $1/k_B T_0$, the potential is rescaled

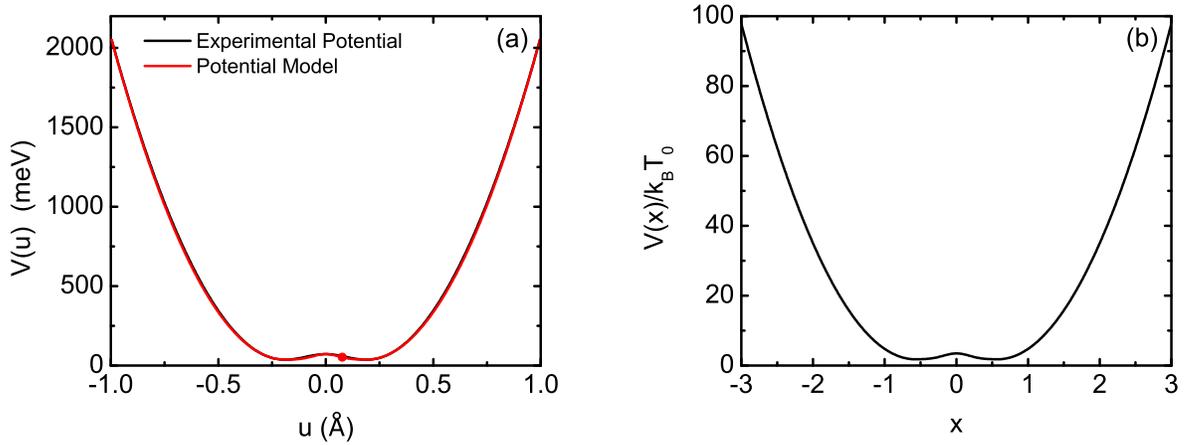


Figure 5.10: (a) Comparison of the potential of H obtained from numerical data and modelled potential function. (b) The unitless potential $V(x)/k_B T_0$.

$$v(x) = \frac{V(x)}{k_B T_0} = \begin{cases} \frac{1}{2}\lambda(x+z')^2 + D' & \text{if } -\infty < x \leq z' \\ A' \exp(-\frac{x^2}{2B'^2}) + C' & \text{if } -z' \leq x \leq z' \\ \frac{1}{2}\lambda(x-z')^2 + D' & \text{if } z' \leq x < \infty \end{cases}$$

where $z' = z/u_0$, $B' = B/u_0$, $\lambda = \frac{k u_0^2}{k_B T_0} = 32.62$, $D' = \frac{D}{k_B T_0} = 1.79$, $A' = \frac{A}{k_B T_0} = 1.69$ and $C' = \frac{C}{k_B T_0} = 1.78$. As shown in Fig. 5.10b, the magnitude of the potential is too

large to compare with the mean square displacement of the experimental potential of hydrogen bond and the mean square displacement of our model, therefore λ is set to 1 and other parameters are adjusted. Fig. 5.11a shows the $\langle u^2 \rangle$ versus T of a particle in the adjusted potential of H in Fig. 5.11b. The vibrational distribution $\rho(x)$ versus $x = u/u_0$ at three different temperatures, 50, 150 and 300 K is shown in in Fig. 5.11b. The temperature dependence of MSD $\langle u^2 \rangle$ is not well produced at low and at high temperatures.

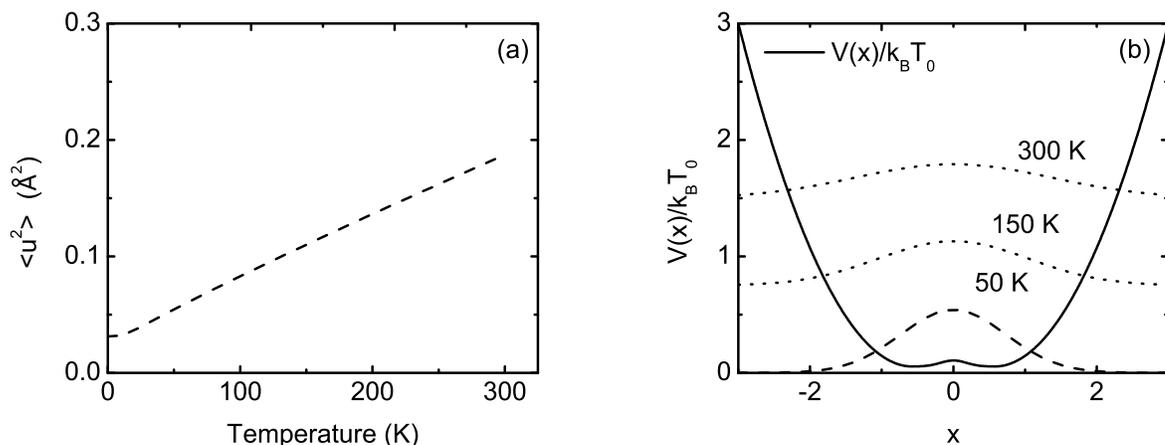


Figure 5.11: (a) The MSD, $\langle u^2 \rangle$, for a particle in the potential on RHS. (b) The vibrational distribution $\rho(x)$.

5.4 Potentials in Proteins

Our aim is to compare the potential models mentioned in Section 5.2 with the potentials seen by H in proteins. H in proteins is bounded in a wide spectrum of sites and solvent water. The H potential varies depending on its separation from its neighbors and the bond angle of H . The model potentials represent an average over these widely varying potentials and angles of H bonds.

The dynamics of H in an H -bond is governed by the H -bond energy. The potential energy of H bounded in a single H bond is discussed in Ref. [72]. The single

H bond consists of a donor atom which donates some charges to the base H^+ (proton) nuclei. The donor atom has a (-) charge, the H^+ still a (+) charge, and the two terms describe a polar molecule. This $D^- - H^+$ polar molecule interacts with another polar molecule, $A^- - AB^+$ covalent bounded pair where A^- and AB^+ are called as an acceptor and acceptor base, respectively. The H potential energy is the potential energy of the H in a single dimer bond: $D^- - H^+ \cdots A^- - AB^+$ as a function of distance δ_{HA} between the H^+ and the acceptor A^- . The δ_{HA} distance denoted by the dotted line.

The H -bond in a single formamide dimer is a simple example to explain the H potential energy. The formamide dimer is, for example, an important component of side chain-side chain bonds. The formamide dimer consists of a tightly bound $N^- - H^+$ pair in which hydrogen is bounded to a second pair $O^- = C^+$, an acceptor-acceptor base pair. In the formamide dimer ($N - H \cdots O = C$), the H -bond is indicated by the dotted line between the H^+ (the donor H^+) and the acceptor O^- in the acceptor-acceptor base pair. Energy as a function δ_{HA} and angles is presented in Ref. [72]. If we assume that this energy is equal to the potential energy of H (i.e. rigidly bound to N), then the H potential energy is clearly anharmonic and asymmetric, and it has a hard core. The potential energy has an anharmonic shape similar to the potential models which we used, especially the potential model having a hard wall component (see Sec.5.2.4).

Fig. 5.12 shows the formamide H -bond energy versus separation r between the H and the acceptor (O), as calculated by Morozov and Kortemme [72] by using the density functional theory (DFT)(see Fig. 4 of Ref.[21]). The H -bond energy is divided by $k_B T_0 = 0.476$ kcal/mole = 1.99 kJ/mole where $T_0 = 240$ K and the separation r by $u_0 = 0.33$ Å ($x = r/u_0$) so that the H -bond energy is in the same unit as the present model potentials. The DFT energy is very similar to the CHARMM27 H bond potential [64]. In the CHARMM27 potential, the dependence of the hydrogen bond potential on separation is represented by a standard 12-6 potential Lennard-Jones potential [51] plus an electrostatic [64] term or a dipole-dipole term [95]. In

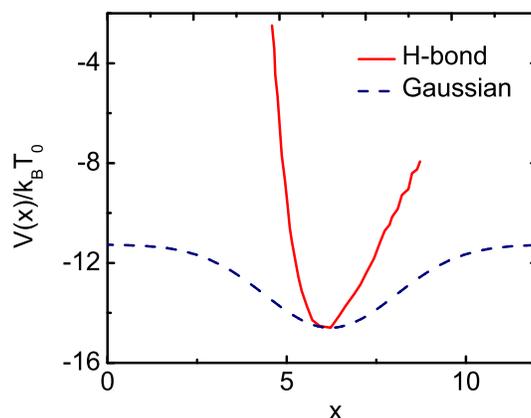


Figure 5.12: H -bond potential $V(r)$ in a formamide dimer versus H^+ to acceptor (O^-) separation, r , from Fig. 4 of Morozov and Kortemme [72]. The potential is shown in units $x = r/u_0$ where $u_0 = 0.33 \text{ \AA}$ and $V(x)/k_B T_0$ where $T_0 = 240 \text{ K}$ and $k_B T_0 = 0.476 \text{ kcal/mole}$. Also shown is the present Gaussian model potential in these units.

Fig. 5.12 the H -bond energy is compared with the present model Gaussian potential. The Gaussian well minimum is moved to coincide with that of the H -bond potential.

Fig. 5.12 shows that the H -bond potential has clear anharmonic character as does the model present Gaussian potential. In addition the H -bond potential is asymmetric around the minimum energy. Both the anharmonic and asymmetric characters will lead to an increase in slope of $\langle u^2 \rangle$ at higher temperatures, as found here in Fig. 5.3 for the model Gaussian potential and in Fig. 5.7 for the model hard walled potential. As anticipated, the H -bond potential is much narrower and stronger than the model Gaussian. High frequency, small amplitude motion is expected for H in an H -bond whereas the Gaussian represents an average for H over all sites.

Hydrogen is also attached to the backbone of proteins. In this position, the H rides on the backbone and its dynamics is determined by the lower frequency, larger amplitude dynamics of the backbone. Specifically, H is a component of amino acids which form the backbone and are attached to the backbone. To describe the dynamics

of amino acids in the backbone, a simplified potential between amino acids is often introduced [2, 5]. Each amino acid is approximated by a single (heavy) atom or pseudo-atom. The effective or coarse-grained (CG) potential between the pseudo-atoms is obtained using several methods one of which is simulating the amino acids in proteins using molecular dynamics (MD) and determining the effective interaction between them which represents their properties. In the MD simulation a potential between all the atoms in the amino acids is typically used, denoted an all-atom potential. The CG potential is subsequently used in calculation of the longer time scale dynamics of the backbone or of larger polypeptides in the protein. The dynamics of H in these amino acids is largely governed by these CG potentials.

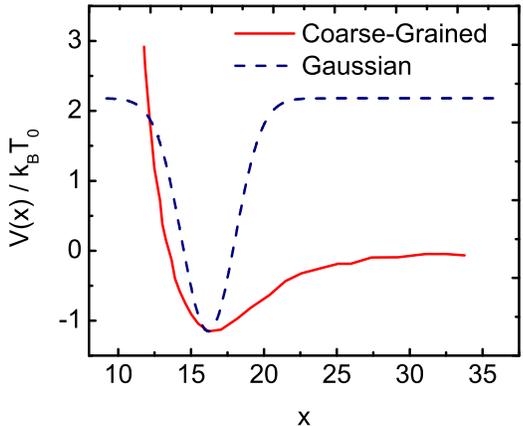


Figure 5.13: Coarse-Grained potential $V(r)$ between two TRP3 amino acids in a protein from Fig. 3 of Basdevant et al. [2]. The potential is in units of $x = r/r_0$ and $V(x)/k_B T_0$ as in Fig. 5.12 and compared with the present Gaussian model potential.

Specifically, Basdevant et. al. [2] and Ha-Duong [39] represent the CG potential by a repulsive ($1/r^6$) term plus a Gaussian attractive term. If the amino acid is charged there is also an electrostatic term. A representative CG potential between two TRP3 amino acids [2] is shown in Fig. 5.13 and compared with our model Gaussian potential. The CG potential is again shown with separation divided by $u_0 = 0.33\text{\AA}$ and energy

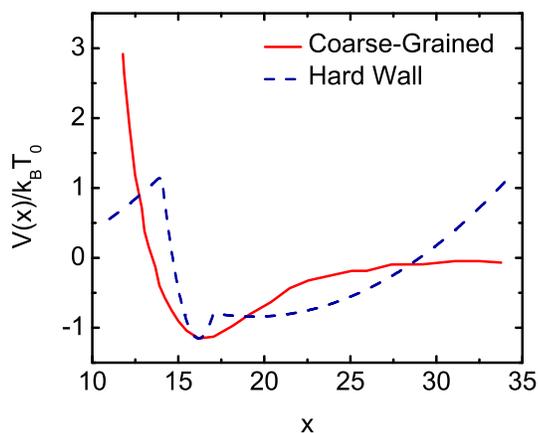


Figure 5.14: Coarse-Grained potential shown in Fig. 5.13 compared with the present Hard Wall model potential.

divided by $k_B T_0$ to coincide with the units of our model potentials. The minimum of our Gaussian model potential has been shifted to coincide with that of the CG potential. From Fig. 5.13 we see that the CG potential is asymmetric and anharmonic so a change in the slope of $\langle u^2 \rangle$ with increasing temperature can be expected. Indeed the shape of the CG potential is similar to that of the the present hard wall potential as shown in Fig. 5.14. The present hard wall model leads to a significant change of slope of $\langle u^2 \rangle$ with increasing temperature as shown in Fig. 5.8 and a similar change can be expected for the CG potential. From Fig. 5.14 we see that the CG potential is broader and weaker (on the right side) than our model Gaussian. Thus lower frequency, larger amplitude motions than average are expected for H riding on amino acids.

In sum, hydrogen in H -bonds and in amino acids see effective potentials that are anharmonic and similar in character to the model potentials used here. The H -bond potential is stiffer and the CG potential between amino acids is softer than the present model potentials which represent an average over a spectrum of potentials seen by H in proteins.

5.5 Discussion

Our aim is to illustrate the change in the slope of MSD $\langle u^2 \rangle$ versus temperature at T_D within a simple model of vibration, a single particle in an anharmonic potential. Using Self-Consistent-Harmonic theory, we investigated dynamics of a particle in different potential models. The SCH vibrational dynamics is also applied the “two-state” model potential introduced by Frauenfelder et al, Keller and Debrunner and Doster et al.[25, 53]. For the “two-state” model potential, a change in slope of the MSD $\langle u^2 \rangle$ versus temperature is not illustrated, because the thermally activated transitions central to this model are not included. It is not stated that thermally activated processes and diffusion are not important in the long time dynamics of proteins. We successfully reproduced a break in $\langle u^2 \rangle$ with temperature for myoglobin and purple membrane within vibration alone.

Chapter 6

CONCLUSION

In this thesis we have investigated and proposed models to describe the mean square motional displacements (MSD) of hydrogen (H) in proteins as observed in neutron scattering experiments. We have developed methods to obtain values of the MSD, $\langle r^2 \rangle$, that are intrinsic to the protein and independent of the instrument settings used in the experiment. Using the same model and method we can obtain the intrinsic, long-time MSD from finite time simulations of a protein.

For example, in Chap. 2 we presented a method in which the intrinsic $\langle r^2 \rangle$ can be extracted from data measured by instruments that have different energy resolution widths W , one that is independent of the instrument resolution. The intrinsic long-time MSD is defined as an equilibrium MSD which is measured by using an instrument that has zero energy resolution width $W = 0$. The method includes the fit of the model of the resolution broadened dynamical structure factor (DSF), $S_R(\mathbf{Q}, \omega)$, to the resolution broadened data. The model contains the intrinsic MSD $\langle r^2 \rangle$, the instrument resolution width W and a parameter describing the motional processes that contribute to the MSD. By fitting the model to data taken on instruments that have different energy resolutions, we show that an intrinsic MSD $\langle r^2 \rangle$ can be obtained, one that is independent of the resolution width W of the instrument.

An expression for the resolution broadened MSD $\langle r^2 \rangle_R$ which is equal to the observed MSD $\langle r^2 \rangle_{exp}$ is also obtained. By fitting the model of $\langle r^2 \rangle_R$ to an observed MSD $\langle r^2 \rangle_{exp}$, we show that the intrinsic MSD $\langle r^2 \rangle$ can also be obtained from the observed MSD data if the observed DSF data $S_R^{exp}(\mathbf{Q}, \omega = 0)$ is not available. The model of $\langle r^2 \rangle_R$ in Eq. (2.17) is useful to clarify the relation between the observed MSD $\langle r^2 \rangle_{exp}$ and the intrinsic MSD $\langle r^2 \rangle$. From the expression for $\langle r^2 \rangle_R$, it is clearly seen

that the observed $\langle r^2 \rangle_R$ lies below the intrinsic MSD $\langle r^2 \rangle$ for a finite resolution width W .

The intrinsic MSD $\langle r^2 \rangle$ obtained by fitting the model $S_R(\mathbf{Q}, \omega = 0)$ to data is found to depend on temperature. There is a marked increase in the intrinsic MSD at a dynamical transition (DT) temperature T_D as is observed. The DT temperature T_D is different for different proteins. The energy resolution width W of the instrument also has an impact on the DT temperature T_D . The intrinsic T_D which is observed in the intrinsic MSD $\langle r^2 \rangle$ is found to be less than the resolution broadened T_D . The impact of resolution broadening on the DT is to shift the DT temperature T_D to higher temperatures.

We apply our method to existing data in the literature and successfully obtain the intrinsic MSD of proteins. The intrinsic MSD in proteins shows a clear DT at a temperature $T_D = 220$ K. Since the intrinsic MSD is a long-time, equilibrium MSD, this indicates that the DT is an intrinsic property of the protein rather than being simply a time window effect.

The model of $S_R(\mathbf{Q}, \omega = 0)$ is very simple and has several limitations, such as using a simple representation $C(t) = \exp(-\lambda t)$ to describe the relaxation of correlations in the protein. This $C(t) = \exp(-\lambda t)$ represents only a single diffusion process. We used this simple exponential function to describe the motional processes because the experimental data is not significantly precise to distinguish between a simple and more sophisticated model. However, the model could be improved by using a stretched exponential $C(t) = \exp(-(\lambda t)^\beta)$ which represents several diffusion mechanisms contributing to $C(t)$.

In Chap. 3, we extended our research to extracting the intrinsic long-time MSD $\langle r^2 \rangle$ of H in proteins from finite time simulations. In this case the simulated data is more precise, in contrast experimental data. Thus, our fits to simulations are sensitive to the model of $C(t)$. In this case, the model of the incoherent ISF $I(\mathbf{Q}, t)$ constructed in Chap. 2 is modified by replacing a simple exponential $C(t)$ by a stretched exponential function. By fitting the model $I(\mathbf{Q}, t)$ to the simulated data, we obtained the intrinsic,

long-time MSD in proteins from finite time simulations. The intrinsic MSD represents the equilibrium MSD as would be predicted by statistical mechanics and the energy landscape, assuming the protein does not go through major structural changes.

In the application of this model, we calculated the incoherent ISF from simulations of 10 ns and $1\mu\text{s}$, then applied the method to extract the intrinsic long-time MSD $\langle r^2 \rangle$ of lysozyme from the simulated data. The intrinsic $\langle r^2 \rangle$ obtained from data out to 1 ns and to 10 ns is found to be the same. The $\langle r^2 \rangle$ is compared with the $\langle r^2 \rangle_R$ that has developed over a limited time $\tau_R \sim \hbar/W$. The $\langle r^2 \rangle$ is found to be independent of simulation time, and approximately twice the resolution broadened MSD $\langle r^2 \rangle_R$ that develops after a time of 1.5 ns. The $\langle r^2 \rangle_R$ developed in 1.5 ns corresponds to an observed MSD $\langle r^2 \rangle_{exp}$ measured using neutron instruments that have an energy resolution width of $W = 1 \mu\text{eV}$. We observe a dynamical transition in the intrinsic MSD as in the finite time MSD. The DT temperature T_D and the difference between the intrinsic MSD $\langle r^2 \rangle$ and the resolution broadened MSD $\langle r^2 \rangle_R$ are sensitive to the energy resolution width W . In addition to this, the ratio of the intrinsic to finite time MSD is sensitive to the $C(t)$ function in the model which describes the motions in the protein. Although we use a stretched exponential function for $C(t)$ in the model of $I(\mathbf{Q}, t)$, our model does not contain ballistic motion and vibrational motion. In future applications, more sophisticated expressions for $C(t)$ that combine vibrational motion at short times and diffusion at longer times could be used in the model of $I(\mathbf{Q}, t)$.

One shortcoming of measurements in neutron scattering is that the observed MSD depends on the wave vector (Q) of the neutron data used to obtain the MSD. Similar to the Q -dependence of the observed MSD $\langle r^2 \rangle_{exp}$, the intrinsic MSD $\langle r^2 \rangle$ is found to be dependent of Q . This Q dependence in MSD is often attributed to use of the Gaussian approximation made to the scattering function in the analysis of the data. In Chap. 4 we analysed the origin of the Q -dependence of MSD. Possible origins of a Q -dependent MSD are the Gaussian approximation or the dynamical diversity of hydrogen in the protein. In the most current methods of data analysis, it is common to use the Gaussian approximation (GA) in the model and to develop a global model

to represent whole protein by a single scattering center. In the GA, we keep the lowest order (Q^2) term and neglect the higher order terms. In this case, if the higher order terms are significant, a Q -dependent MSD could be obtained in the data analysis.

To test the impact of the GA, we have calculated the incoherent ISF, both the full ISF $I_{inc}(\mathbf{Q}, t)$ and ISF in the GA $I_{iG}(\mathbf{Q}, t)$ from simulations of lysozyme. Then the difference between the $\langle r^2 \rangle$ is investigated by fitting the model $I(\mathbf{Q}, t)$ to $I_{inc}(\mathbf{Q}, t)$ and $I_{iG}(\mathbf{Q}, t)$. We have found that the MSD extracted from the full ISF and the Gaussian approximation is the same at low Q -values specially and still Q dependent in both cases. That indicates that the Gaussian approximation is valid. Also, higher cumulants, the terms beyond the Gaussian approximation, (e.g. terms in Q^4) are not significant. The other possible origin for the Q -dependent MSD is the neglecting the dynamical diversity in the model which is used to analyse the data. To test this, we calculated the incoherent ISF $I_{inc}(\mathbf{Q}, t)$ for an individual hydrogen, because there is not any dynamical diversity for an individual H. Then, we fit the model $I(\mathbf{Q}, t)$ to $I_{inc}(\mathbf{Q}, t)$ of a single hydrogen. We have found that the apparent Q dependence of the MSD arises from the “dynamical diversity” of the H in lysozyme. Specifically, if the ISF of an individual H in the protein is calculated (no diversity), then the MSD is independent of Q . The Q dependence arises from ignoring the dynamical diversity in the data analysis.

Finally, in Chap. 5 we have investigated a vibrational model of the temperature dependence of the MSD. Our purpose is simply to illustrate that a break in $\langle u^2 \rangle$ with temperature is possible within vibration alone. A change in the slope of the MSD $\langle u^2 \rangle$ versus temperature is illustrated from vibration of a mass in an anharmonic potential. We have successfully reproduced the MSD versus temperature data for myoglobin and purple membrane by using a Gaussian potential and an anharmonic potential contains hard and soft walls.

BIBLIOGRAPHY

- [1] P. J. Artymiuk, C. C. F. Blake, D. W. Rice, and K. S. Wilson. The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme of 6Å resolution. *Acta. Cryst.*, B38:778, 1982.
- [2] N. Basdevant, D. Borgis, and T. Ha-Duong. A coarse-grained protein-protein potential derived from an all-atom force field. 111:9390, 2007.
- [3] T. Becker, J. A. Hayward, J. L. Finney, R. M. Daniel, and J. C. Smith. Neutron frequency windows and the protein dynamical transition. *Biophys. Journ.*, 87:1436, 2004.
- [4] T. Becker and J. C. Smith. Energy resolution and dynamical heterogeneity effects on elastic incoherent neutron scattering from molecular system. *Phys. Rev. E*, 67:021904, 2003.
- [5] M. R. Betancourt and S. J. Omovie. Pairwise energies for polypeptide coarse-grained models derived from atomic force field. *J. Chem. Phys.*, 130:195103, 2009.
- [6] D. J. Bicout and G. Zaccai. Protein flexibility from the dynamical transition. *Biophys. Journ.*, 80:1115, 2001.
- [7] R. Biehl, B. Hoffmann, M. Monkenbusch, P. Falus, S. Préost, R. Merkel, and D. Richter. Direct observation of correlated interdomain motion in alcohol dehydrogenase. *Phys. Rev. Lett.*, 101:138102, 2008.
- [8] V. Calandrini, V. Hamon, K. Hinsén, P. Calligari, M.-C. Bellissent-Funel, and G. R. Kneller. Relaxation dynamics of lysozyme in solution under pressure: combining molecular dynamics simulations and quasielastic neutron scattering. *Chem. Phys.*, 345:289, 2008.
- [9] V. Calandrini and G. R. Kneller. Influence of pressure on the slow and fast fractional relaxation dynamics in lysozyme: A simulation study. *J. Chem. Phys.*, 128:065102, 2008.
- [10] G. Caliskan, R. M. Briber, D. Thirumalai, V. Garcia-Sakai, S. A. Woodson, and A. P. Sokolov. Dynamic transition in tRNA is solvent induced. *J. Am. Chem. Soc.*, 128:32, 2006.

- [11] G. Caliskan, A. Kisliuk, A. Tsai, C. Soles, and A. P. Sokolov. Protein dynamics in viscous solvents. *J. Chem. Phys.*, 118:4230, 2003.
- [12] S.-H. Chen, M. Lagi, X.-Q. Chu, Y. Zhang, C. Kim, A. Faraone, E. Fratini, and P. Baglioni. Dynamics of a globular protein and its hydration water studied by neutron scattering and MD simulations. *Spectroscopy*, 24:1, 2010.
- [13] S.-H. Chen, L. Liu, E. Fratini, P. Baglioni, A. Faraone, and E. Mamontov. Observation of fragile-to-strong dynamics crossover in protein hydration water. *Proc. Nat. Acad. Sci. U.S.A.*, 103:9012, 2006.
- [14] P. F. Choquard. *“The anharmonic crystal”*. W. A. Benjamin, New York, 1967.
- [15] X.-Q. Chu, E. Fratini, P. Baglioni, A. Faraone, and S.-H. Chen. Observation of a dynamic crossover in RNA hydration water which triggers a dynamic transition in the biopolymer. *Phys. Rev. E*, 77:011908, 2008.
- [16] R. M. Daniel, R. V. Dunn, J. L. Finney, and J. C. Smith. The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.*, 32:69, 2003.
- [17] R. M. Daniel, J. L. Finney, V. Reat, R. Dunn, M. Ferrand, and J. C. Smith. Enzyme dynamics and activity: time-scale dependence of dynamical transitions in glutamate dehydrogenase solution. *Biophys. Journ.*, 77:2184, 1999.
- [18] R. M. Daniel, J. C. Smith, M. Ferrand, S. Héry, R. Dunn, and J. L. Finney. Enzyme activity below the dynamical transition at 220K. *Biophys. Journ.*, 75:2504, 1998.
- [19] P. Debye. Zur Theorie der spezifischen Wärme. *Ann. Phys.*, 39:789, 1912.
- [20] S. Dellerue, A. Petrescu, J. C. Smith, S. Longeville, and M. C. Bellissent-Funel. Collective dynamics of a photosynthetic protein probed by neutron spin-echo spectroscopy and molecular dynamics simulation. *Physica B*, 276:514, 2000.
- [21] T. E. Dirama, G. A. Carri, and A. P. Sokolov. Coupling between lysozyme and glycerol dynamics: microscopic insight from molecular-dynamics simulations. *J. Chem. Phys.*, 122:244910, 2005.
- [22] T. E. Dirama, J. E. Curtis, G. A. Carri, and A. P. Sokolov. Coupling between lysozyme and trehalose dynamics: microscopic insight from molecular-dynamics simulations. *J. Chem. Phys.*, 124:034901, 2006.
- [23] W. Doster. The protein-solvent glass transition. *Biochim. Biophys. Acta*, 1804:3, 2010.
- [24] W. Doster, S. Busch, A.M. Gaspar, M.-S. Appavou, J. Wuttke, and H. Scheer. Dynamical transition of protein-hydration water. *Phys. Rev. Lett.*, 104:098101, 2010.

- [25] W. Doster, S. Cusack, and W. Petry. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature (London)*, 337:754, 1989.
- [26] W. Doster, S. Cusack, and W. Petry. Dynamical instability of liquidlike motions in a globular protein observed by inelastic neutron scattering. *Phys. Rev. Lett.*, 65:1080, 1990.
- [27] R. V. Dunn, V. Reat, J. Finney, M. Ferrand, J. C. Smith, and R. M. Daniel. Enzyme activity and dynamics: xylanase activity in the absence of fast anharmonic dynamics. *Biochem. J.*, 346:355, 2000.
- [28] R. Elber and M. Karplus. Multiple conformational states of proteins : A molecular dynamics analysis of myoglobin. *Science*, 235:318, 1987.
- [29] N. Engler, A. Ostermann, N. Niimura, and F. G. Parak. Hydrogen Atoms in Proteins: Positions and Dynamics. *Proc. Nat. Acad. Sci. U.S.A.*, 100:10243, 2003.
- [30] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577, 1995.
- [31] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak. Slaving: Solvent fluctuation dominate protein dynamics and functions. *Proc. Nat. Acad. Sci. U.S.A.*, 99:16047, 2002.
- [32] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and R.D. Young. Bulk-solvent and hydration-shell fluctuations, similar to α - and β -fluctuations in glasses, control protein motions and fluctuations. *Proc. Nat. Acad. Sci. U.S.A.*, 101:14408, 2004.
- [33] M. Ferrand, A. J. Dianoux, W. Petry, and G. Zaccai. Thermal motions and function of bacteriorhodopsin in purple membranes: Effects of temperature and hydration studied by neutron scattering . *Proc. Nat. Acad. Sci. U.S.A.*, 90:9668, 1993.
- [34] J. Fitter, R. E. Lechner, and N. A. Dencher. Picosecond molecular motions in Bacteriorhodopsin from neutron scattering. *Biophys. Journ.*, 73:2126, 1997.
- [35] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598, 1991.
- [36] F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik, and G. Zaccai. Protein dynamics studied by neutron scattering. *Quart. Rev. Biophys.*, 35:327, 2002.
- [37] N. S. Gillis, N. R. Werthamer, and T. R. Koehler. Properties of crystalline argon and neon in the self-consistent phonon approximation. *Phys. Rev.*, 165:951, 1968.

- [38] H. R. Glyde and M. L. Klein. Anharmonic effects and lattice dynamics of insulators. *Critical Reviews in Solid State Sciences*, 2:181, 1971.
- [39] T. Ha-Duong. Protein Backbone Dynamics Simulations Using Coarse-Grained Bonded Potentials and Simplified Hydrogen Bonds. *J. Chem. Theory Comput.*, 6:761, 2010.
- [40] V. Hamon, P. Calligari, K. Hinsén, and G. R. Kneller. Simulation studies of structural changes and relaxation processes in lysozyme under pressure. *J. Non-Cryst. Solids*, 352:4417, 2006.
- [41] J. A. Hayward, R. M. Daniel, J. L. Finney, and J. C. Smith. Use of computer simulation in the interpretation of elastic neutron scattering in complex molecular system: a small protein in various environments. *Chem. Phys.*, 292:389, 2003.
- [42] J. A. Hayward, J. L. Finney, R. M. Daniel, and J. C. Smith. Molecular dynamics decomposition of temperature-dependent elastic neutron scattering by a protein solution. *Biophys. Journ.*, 85:679, 2003.
- [43] J. A. Hayward and J. C. Smith. Temperature dependence of protein dynamics: computer simulation analysis of neutron scattering properties. *Biophys. Journ.*, 82:1216, 2002.
- [44] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463, 1997.
- [45] B. Hess, C. Kutzner, D. V. Spoel, and E. Lindahl. GROMACS 4: Algorithm for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435, 2008.
- [46] L. Hong, X. Cheng, D. C. Glass, , and J. C. Smith. Surface hydration amplifies single-well protein atom diffusion propagating into the macromolecular core. *Phys. Rev. Lett.*, 108:238102, 2012.
- [47] L. Hong, N. Smolin, B. Lindner, A. P. Sokolov, and J. C. Smith. Three classes of motion in the dynamic neutron-scattering susceptibility of a globular protein. *Phys. Rev. Lett.*, 107:148102, 2011.
- [48] W. G. Hoover. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695, 1985.
- [49] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, 120:9665, 2004.
- [50] M. Jasnin, L. van Eijck, M. M. Koza, J. Peters, C. Laguri, H. Lortat-Jacob, and G. Zaccai. Dynamics of heparan sulfate explored by neutron scattering. *Phys. Chem. Chem. Phys.*, 12:3360, 2010.

- [51] J. E. Jones. On the determination of molecular fields. II. From the equation of state of a gas. *Proc. Roy. Soc. A*, 106:463, 1924.
- [52] W. L. Jorgensen and J. Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657, 1988.
- [53] H. Keller and P. G. Debrunner. Evidence of conformational and diffusional mean square displacements in frozen aqueous solution of oxymyoglobin. *Phys. Rev. Lett.*, 45:68, 1980.
- [54] S. Khodadadi, S. Pawlus, J. H. Roh, V. G. Sakai, E. Mamontov, and A. P. Sokolov. The origin of the dynamic transition in proteins. *J. Chem. Phys.*, 128:195106, 2008.
- [55] G. R. Kneller and K. Hinsén. Quantitative Model for the Heterogeneity of Atomic Position Fluctuations in Proteins: A Simulation Study. *J. Chem. Phys.*, 131:045104, 2009.
- [56] G. R. Kneller and J. C. Smith. Liquid-like-side-chain dynamics in myoglobin. *J. Mol. Biol.*, 242:181, 1994.
- [57] T. R. Koehler. Theory of the self-consistent harmonic approximation with application to solid Neon. *Phys. Rev. Lett.*, 17:89, 1966a.
- [58] P. Kumar, Z. Yan, L. Xu, M. G. Mazza, S. V. Buldyrev, S. H. Chen, S.-H. Sastry, and H. E. Stanley. Glass Transition in Biomolecules and the Liquid-Liquid Critical Point of Water. *Phys. Rev. Lett.*, 97:177802, 2006.
- [59] M. Lagi, X. Chu, C. Kim, F. Mallamace, P. Baglioni, and S. H. Chen. The low-temperature dynamic crossover phenomenon in protein hydration water: simulation vs experiments. *J. Phys. Chem. B*, 112:1571, 2008.
- [60] A. Lamy, M. Souaille, and J. C. Smith. Simulation evidence for experimentally detectable low-temperature vibrational inhomogeneity in a globular protein. *Biopolymers*, 39:471, 1996.
- [61] A. L. Lee and A. J. Wand. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature (London)*, 411:501, 2001.
- [62] U. Lehnert, V. Reat, M. Weik, G. Zaccai, and C. Pfister. Thermal motion in bacteriorhodopsin at different hydration levels studied by neutron scattering: correlation with kinetics and light-induced conformational changes. *Biophys. Journ.*, 75:1945, 1998.
- [63] A. Lerbret, F. Affouard, P. Bordat, A. Hedoux, Y. Guinet, and M. Descamps. Molecular dynamics simulations of lysozyme in water/sugar solutions. *Chem. Phys.*, 345:267, 2008.

- [64] A. D. MacKerrell, D. D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J. Phys. Chem. B*, 102:3586, 1998.
- [65] S. Magazù, G. Maisano, F. Migliardo, and C. Mondelli. α , α -Trehalose/Water Solutions. VII:Elastic Incoherent Neutron Scattering Study on Fragility. *J. Phys. Chem. B*, 108:13580, 2004.
- [66] S. Magazù, F. Migliardo, and A. Benedetto. Mean square displacement from elastic incoherent neutron scattering evaluated by spectrometers working with different energy resolution on dry and hydrated (H_2O and D_2O) lysozyme . *J. Phys. Chem. B.*, 114:9268, 2010.
- [67] A. A. Maradudin, E. W. Montroll, and G. H Weiss. “*Theory of Lattice Dynamics in the Harmonic Approximation*”. Academic Press, 1963.
- [68] L. Meinhold, D. Clement, M. Tehei, R. Daniel, J. L. Finney, and J. C. Smith. Protein Dynamics and Stability: The Distribution of Atomic Fluctuations in Thermophilic and Mesophilic Dihydrofolate Reductase Derived Using Elastic Incoherent Neutron Scattering. *Biophys. Journ.*, 94:4812, 2008.
- [69] L. Meinhold and J. C. Smith. Pressure-dependent transition in protein dynamics at about 4 kbar revealed by molecular dynamics simulation. *Phys. Rev. E*, 72:061908, 2005.
- [70] L. Meinhold, J. C. Smith, A. Kitao, and A. H. Zewail. Picosecond fluctuating protein energy landscape mapped by pressure-temperature molecular dynamics simulation. *Proc. Nat. Acad. Sci. U.S.A.*, 104:17261, 2007.
- [71] Y. Miao, Z. Yi, D. C. Glass, L. Hong, M. Tyagi, J. Baudry, N. Jain, and J. C. Smith. Temperature-dependent dynamical transitions of different classes of amino acid residue in a globular protein. *J. Am. Chem. Soc.*, 134:19576, 2012.
- [72] A. V. Morozov and T. Kortemme. Potential functions for hydrogen bonds in protein structure prediction and design. *Adv. Protein Chem.*, 72:1, 2005.
- [73] H. Nakagawa, H. Kamikubo, and M. Kataoka. Effect of conformational states on protein dynamical transition. *Biochim. Biophys. Acta*, 1804:27, 2010.
- [74] H. Nakagawa, A. Tokuhisa, H. Kamikubo, Y. Joti, A. Kitao, and M. Kataoka. Dynamical heterogeneity of protein dynamics studies by elastic incoherent neutron scattering and molecular simulations. *Mater. Sci. Eng. A*, 442:356, 2006.

- [75] A. Oleinikova, N. Smolin, I. Brovchenko, A. Geiger, and R. Winter. Formation of spanning water networks on protein surfaces via 2D percolation transition. *J. Phys. Chem. B*, 109:1988, 2005.
- [76] A. Ostermann, R. Waschipky, F. G. Parak, and G. U. Nienhaus. Ligand binding and conformational motions in myoglobin. *Nature (London)*, 404:205, 2000.
- [77] A. Paciaroni, S. Cinelli, and G. Onori. Effect of the environment of the protein dynamical transition: a neutron scattering study. *Biophys. Journ.*, 83:1157, 2002.
- [78] A. Paciaroni, E. Cornicchi, A. De Francesco, M. Marconi, and G. Onori. Conditioning action of the environment on the protein dynamics studied through elastic neutron scattering. *Eur. Biophys. J.*, 35:591, 2006.
- [79] F. Parak and E. W. Knapp. A consistent picture of protein dynamics. *Proc. Nat. Acad. Sci. U.S.A.*, 81:7088, 1984.
- [80] F. Parak, E. W. Knapp, and D. Kucheida. Protein dynamics: Mössbauer spectroscopy on deoxymyoglobin crystal. *J. Mol. Biol.*, 161:177, 1982.
- [81] F. G. Parak. Physical aspects of protein dynamics. *Rep. Prog. Phys.*, 66:103, 2003.
- [82] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. App. Phys.*, 52:7182, 1981.
- [83] J. Perez, J. M. Zanotti, and D. Durand. Evolution of the internal dynamics of two globular proteins from dry powder to solution. *Biophys. Journ.*, 77:454, 1999.
- [84] J. Pieper, A. Buchsteiner, N. A. Dencher, R. E. Lechner, and T. Hauß. Transient protein softening during the working cycle of a molecular machine. *Phys. Rev. Lett.*, 100:228103, 2008.
- [85] G. Reiter, C. Burnham, D. Homouz, P. M. Platzman, J. Mayers, T. Abdul-Redah, A. P. Moravsky, J. C. Li, C.-K. Loong, and A. I. Kolesnikov. Anomalous Behavior of Proton Zero Point Motion in Water Confined in Carbon Nanotubes. *Phys. Rev. Lett.*, 97:247801, 2006.
- [86] T. Róg, K. Murzyn, and G. R. Kneller. Software news and updates, nMoldyn: A program package for a neutron scattering oriented analysis of molecular dynamics simulations. *J. Comp. Chem.*, 24:657, 2003.
- [87] J. H. Roh, J. E. Curtis, S. Azzam, V. N. Novikov, I. Peral, Z. Chowdhuri, R. B. Gregory, and A. P. Sokolov. Influence of hydration on the dynamics of lysozyme. *Biophys. Journ.*, 91:2573, 2006.

- [88] J. H. Roh, V. N. Novikov, R. B. Gregory, J. E. Curtis, Z. Chowdhuri, and A. P. Sokolov. Onsets of anharmonicity in protein dynamics. *Phys. Rev. Lett.*, 95:038101, 2005.
- [89] J. A. Rupley. Protein hydration and function . *Adv. Protein Chem.*, 41:37, 1991.
- [90] J. Smith, K. Kuczera, and M. Karplus. Dynamics of myoglobin: comparison of simulation results with neutron scattering spectra . *Proc. Nat. Acad. Sci. U.S.A.*, 87:1601, 1990.
- [91] J. C. Smith. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Quart. Rev. Biophys.*, 24:227, 1991.
- [92] C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature (London)*, 420:102, 2002.
- [93] A. P. Sokolov, H. Grimm, and R. Kahn. Glassy dynamics in DNA: Ruled by water of hydration? *J. Chem. Phys.*, 110:7053, 1999.
- [94] A. P. Sokolov, H. Grimm, A. Kisliuk, and A. J. Dianoux. Slow relaxation process in DNA. *J. Biol. Phys.*, 27:313, 2001.
- [95] C-L. Sun and C-S. Wang. Estimation on the intramolecular hydrogen-bonding energies in proteins and peptides by the analytic potential energy function. *J. Mol. Struct.:THEOCHEM*, 956:38, 2010.
- [96] M. Tarek, G. J. Martyna, and D. J. Tobias. Amplitudes and frequencies of protein dynamics: analysis of discrepancies between neutron scattering and molecular dynamics simulations. *J. Am. Chem. Soc.*, 122:10450, 2000.
- [97] M. Tarek and D. J. Tobias. The dynamics of protein hydration water: A quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. *Biophys. Journ.*, 79:3244, 2000.
- [98] M. Tarek and D. J. Tobias. Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys. Rev. Lett.*, 88(13):138101, 2002.
- [99] A. L. Tournier and J. C. Smith. Principle components of the protein dynamical transition. *Phys. Rev. Lett.*, 91:208106, 2003.
- [100] A. M. Tsai, D. A. Neumann, and L. N Bell. Molecular dynamics of solid-state lysozyme as affected by glycerol and water: a neutron scattering study. *Biophys. Journ.*, 79:2728, 2000.
- [101] D. Vural and H. R. Glyde. Vibrational dynamics of hydrogen in proteins. *Phys. Rev. E*, 83:031922, 2011.

- [102] D. Vural and H. R. Glyde. Intrinsic mean-square displacements in proteins. *Phys. Rev. E*, 86:011926, 2012.
- [103] D. Vural, L. Hong, J. C. Smith, and H. R. Glyde. Long-time mean-square displacements in proteins. *Phys. Rev. E*, 88:052706, 2013.
- [104] I. Waller. Über eine verallgemeinerte Streuungsformel. *Z. Physik*, 51:213, 1928.
- [105] K. Wood, C. Caronna, P. Fouquet, W. Haussler, F. Natali, J. Ollivier, A. Orecchini, M. Plazanet, and G. Zaccai. A benchmark for protein dynamics: Ribonuclease A measured by neutron scattering in a large wavevector-energy transfer range. *Chem. Phys.*, 345:305, 2008.
- [106] K. Wood, S. Grudinin, B. Kessler, M. Weik, M. Johnson, G. R. Kneller, D. Oesterhelt, and G. Zaccai. Dynamical heterogeneity of specific amino acids in bacteriorhodopsin. *J. Mol. Biol.*, 380:581, 2008.
- [107] K. Wood, U. Lehnert, B. Kessler, G. Zaccai, and D. Oesterhelt. Hydration dependence of active core fluctuation in bacteriorhodopsin. *Biophys. Journ.*, 95:194, 2008.
- [108] Z. Yi, Y. Miao, J. Baudry, N. Jain, and J. C. Smith. Derivation of Mean-Square Displacements for Protein Dynamics from Elastic Incoherent Neutron Scattering. *J. Phys. Chem. B.*, 116:5028, 2012.
- [109] G. Zaccai. How soft is a protein? a protein dynamics force constant measured by neutron scattering. *Science*, 288:1604, 2000.
- [110] J. M. Zanotti, M. C. Bellissent-Funel, and J. Parello. Dynamics of a globular protein studies by neutron scattering and solid-state NMR. *Physica B*, 234-236:228–230, 1997.
- [111] L. Zhu, D. Frenkel, and P. G. Bolhuis. Role of Fluctuations in Ligand Binding Cooperativity of Membrane Receptors. *Phys. Rev. Lett.*, 106:168103, 2011.

Appendix A

THE OBSERVATION TIME τ_R IN THE INSTRUMENTS IN NEUTRON SCATTERING

The resolution broadened DSF $S_R(\mathbf{Q}, \omega)$ is

$$\begin{aligned} S_R(\mathbf{Q}, \omega) &= \int_{-\infty}^{\infty} d\omega' S(\mathbf{Q}, \omega') R(\omega - \omega') \\ &= \int_{-\infty}^{\infty} dt \exp(i\omega t) I(\mathbf{Q}, t) R(\omega - \omega') \end{aligned} \quad (\text{A.1})$$

where $R(\omega)$ is the instrument energy resolution function. Commonly observed is the elastic component of the resolution broadened DSF,

$$S_R(\mathbf{Q}, \omega = 0) = \int_{-\infty}^{\infty} dt I(\mathbf{Q}, t) R(t). \quad (\text{A.2})$$

The convolution in time coordinates in $S_R(\mathbf{Q}, \omega)$ is

$$S_R(\mathbf{Q}, t) = I(\mathbf{Q}, t) R(t). \quad (\text{A.3})$$

For example, the energy resolution function could be chosen as a Gaussian function,

$$R(\omega) = [2\pi\sigma^2]^{-1/2} \exp\left(-\frac{1}{2} \frac{\omega^2}{\sigma^2}\right) \quad (\text{A.4})$$

that is normalized to unity,

$$\int d\omega R(\omega) = 1. \quad (\text{A.5})$$

From the shape of a Gaussian, the FWHM of the Gaussian is related to σ by

$$W_\omega \equiv FWHM = (8 \ln 2)^{1/2} \sigma, \quad (\text{A.6})$$

where σ is equal to $1/\tau_R$. In units of energy $\epsilon = \hbar\omega$, the FWHM in energy is

$$\begin{aligned} W &\equiv \hbar W_\omega = (8 \ln 2)^{1/2} \hbar \sigma = (8 \ln 2)^{1/2} \hbar / \tau_R \\ \tau_R &= \frac{(8 \ln 2)^{1/2} \hbar}{W}. \end{aligned} \quad (\text{A.7})$$

where $\hbar = \frac{h}{2\pi} = \frac{4.135}{2\pi} \mu eV.ns$. The $R(t) = \exp(-t^2/2\tau_R^2)$, which is a Fourier transform of the $R(\omega)$ in Eq. (A.4), cut off the $S_R(\mathbf{Q}, t)$ at time τ_R ,

$$S_R(\mathbf{Q}, t) = I(\mathbf{Q}, t) \exp(-t^2/2\tau_R^2). \quad (\text{A.8})$$

Appendix B

INTERMEDIATE SCATTERING FUNCTION AT INFINITE TIME LIMIT

The purpose of this appendix is to derive the expression

$$I_\infty(\mathbf{Q}) = \exp\left(-\frac{1}{3}Q^2\langle r^2 \rangle\right), \quad (\text{B.1})$$

to identify the approximations made to obtain it and to clarify the meaning of $\langle r^2 \rangle$. $I_\infty(\mathbf{Q})$ is the $t \rightarrow \infty$ limit of the intermediate incoherent DSF, $I(\mathbf{Q}, t)$ defined in Eq. (2.3). $I(\mathbf{Q}, \infty)$ is independent of time t . The Fourier transform of $I_\infty(\mathbf{Q})$ is therefore purely elastic (zero except at $\omega = 0$) and given by $S_{el}(\mathbf{Q}, \omega) = I_\infty(\mathbf{Q})\delta(\omega)$. This shows that $I_\infty(\mathbf{Q})$ arises from structure in the protein since translationally invariant systems such as gases and liquids that have no structure have no elastic scattering. Thus the motions that contribute to $I_\infty(\mathbf{Q})$ and $\langle r^2 \rangle$ are therefore vibrations, hindered rotations, restricted diffusion and all motions that in some way are restricted by or reflect a structure. $I_\infty(\mathbf{Q})$ is exactly the Debye-Waller factor that appears in the scattering of X-rays or neutrons from crystals [19, 104, 67]. In crystalline solids, $I_\infty(\mathbf{Q})$ is the reduction in intensity of a Bragg peak arising from atomic vibration in the solid. For H in proteins, $I_\infty(\mathbf{Q})$ is the reduction in intensity of the incoherent elastic scattering at any \mathbf{Q} value arising from all restricted motions of the H .

To obtain $I_\infty(\mathbf{Q})$, we begin with $I_{inc}(\mathbf{Q}, t)$ given by Eq. (2.2). The first approximation is to reduce Eq. (2.2) to Eq. (2.3). If Eq. (2.2) were retained then terms beyond the Gaussian in Eq. (B.1) will be obtained [68, 55, 108]. Following the arguments below Eq. (2.10), the $t \rightarrow \infty$ limit of Eq. (2.3) is,

$$\begin{aligned} I_\infty(\mathbf{Q}) &= I(\mathbf{Q}, t = \infty) = \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(\infty)) \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\ &= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle. \end{aligned} \quad (\text{B.2})$$

To arrive at the final line of Eq. (B.2) we have assumed: (1) that the position, $\mathbf{r}(\infty)$ at $t \rightarrow \infty$ of each H is uncorrelated with its position $\mathbf{r}(0)$ at $t = 0$ so the expectation values of $\mathbf{r}(\infty)$ and $\mathbf{r}(0)$ are independent and (2) that the protein properties are independent of the time when they are observed so that $\langle r^n(\infty) \rangle = \langle r^n(0) \rangle = \langle r^n \rangle$. $I_\infty(\mathbf{Q})$ is the product of two identical exponentials, one containing i and the other $-i$.

To proceed we make a cumulant expansion of each expectation value in $I_\infty(\mathbf{Q})$,

$$\langle \exp(ix) \rangle = \exp\left[\sum_{n=1}^{\infty} \frac{(i)^n}{n!} \mu_n\right] \quad (\text{B.3})$$

where the μ_n are cumulants and $x = \pm(\mathbf{Q} \cdot \mathbf{r})$. The cumulants μ_n are a combination of moments,

$$\begin{aligned} \mu_1 &= \langle x \rangle \\ \mu_2 &= \langle x^2 \rangle - \langle x \rangle^2 \\ \mu_3 &= \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3 \\ \mu_4 &= \langle x^4 \rangle - 3\langle x^2 \rangle^2 - 4\langle x^3 \rangle \langle x \rangle + 12\langle x^2 \rangle \langle x \rangle^2 - 6\langle x \rangle^4. \end{aligned} \quad (\text{B.4})$$

Because of the $\pm i$ in the exponentials in Eq. (B.2), the odd cumulants cancel. In $I_\infty(\mathbf{Q})$ only even cumulants appear. Thus,

$$\begin{aligned} I_\infty(\mathbf{Q}) &= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}) \rangle = \exp\left[2 \sum_{n \text{ even}} \frac{(i)^n}{n!} \mu_n\right] \\ &= \exp\left[-\mu_2 + \frac{1}{12}\mu_4 - \frac{2}{6!}\mu_6 + \dots\right] \end{aligned} \quad (\text{B.5})$$

We also take $\langle \mathbf{r} \rangle = 0$. Thus up to fourth order,

$$I_\infty(\mathbf{Q}) = \exp\left[-\langle [\mathbf{Q} \cdot \mathbf{r}]^2 \rangle + \frac{1}{12}[\langle [\mathbf{Q} \cdot \mathbf{r}]^4 \rangle - 3\langle [\mathbf{Q} \cdot \mathbf{r}]^2 \rangle^2]\right] \quad (\text{B.6})$$

On neglecting the fourth order cumulant and assuming cubic symmetry so that $\langle [\mathbf{Q} \cdot \mathbf{r}]^2 \rangle = Q^2 \langle r_z^2 \rangle = \frac{1}{3}Q^2 \langle r^2 \rangle$, where the z axis is chosen parallel to \mathbf{Q} , we obtain $I_\infty(\mathbf{Q})$ in Eq. (B.1). The fourth order cumulant will be small if \mathbf{Q} is small or if the motional distribution of the H atom is approximately Gaussian.

Eq. (B.1) defines the intrinsic $\langle r^2 \rangle$. This $\langle r^2 \rangle$ is the same as that defined in simulations as the $t \rightarrow \infty$ limit of $\frac{1}{2} \langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle$ since $\langle \mathbf{r}(t) \cdot \mathbf{r}(0) \rangle = 0$ at $t \rightarrow \infty$ and $\langle r^2(\infty) \rangle = \langle r^2(0) \rangle = \langle r^2 \rangle$. Explicitly, the intrinsic value of $\langle r^2 \rangle$ is defined as the value that appears in the Debye-Waller factor $I_\infty(\mathbf{Q})$.

Appendix C

THE MODEL OF $S_R^N(\mathbf{Q}, \omega = 0)$ FOR GAUSSIAN CORRELATION AND RESOLUTION FUNCTIONS

For the Gaussian $C(t) = \exp(-\lambda^2 t^2/2)$ and the resolution function $R(\omega) = \frac{1}{\sqrt{2\pi W^2}} e^{-\omega^2/2W^2}$, the incoherent ISF becomes

$$\begin{aligned} I(\mathbf{Q}, t) &= I(\mathbf{Q}, t = \infty) + I'(\mathbf{Q}, t) \\ I(\mathbf{Q}, t) &= I_\infty + (1 - I_\infty) \exp(-\lambda^2 t^2/2). \end{aligned} \quad (\text{C.1})$$

where $I'(\mathbf{Q}, t) = (1 - I(\mathbf{Q}, t = \infty))C(t)$. From Eq. 2.15, the elastic resolution broadened DSF becomes

$$\begin{aligned} S_R(\mathbf{Q}, \omega = 0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt I(\mathbf{Q}, t) \exp(-\lambda^2 t^2/2) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt I(\mathbf{Q}, \infty) \exp(-W^2 t^2/2) \\ &+ \frac{1}{2\pi} \int_{-\infty}^{\infty} dt (1 - I(\mathbf{Q}, \infty)) \exp(-\lambda^2 t^2/2) \exp(-W^2 t^2/2) \\ &= I_\infty \sqrt{\frac{1}{2\pi W}} + (1 - I_\infty) \sqrt{\frac{1}{2\pi(W^2 + \lambda^2)}} \end{aligned} \quad (\text{C.2})$$

The expression in Eq. (C.2) is divided to $S_R(\mathbf{Q} = 0, \omega = 0)$ at $\mathbf{Q} = 0$ to obtain the the expression for the “normalized” $S_R(\mathbf{Q}, \omega = 0)$. From Eq. (C.2), the $S_R(\mathbf{Q} = 0, \omega = 0) = (2\pi W)^{-1/2}$ and “normalized” $S_R(\mathbf{Q}, \omega = 0)$ is

$$S_R^N(\mathbf{Q}, \omega = 0) = I_\infty + (1 - I_\infty) \frac{W}{\sqrt{W^2 + \lambda^2}}, \quad (\text{C.3})$$

The resolution broadened MSD $\langle r^2 \rangle_R$ using the expression in Eq. (1.1) and our model $S_R^N(\mathbf{Q}, \omega = 0)$ of Eq. (C.3),

$$\begin{aligned} \langle r^2 \rangle_{slope} &= -3 \frac{d \log S_R^N(\mathbf{Q}, \omega = 0)}{dQ^2} \Big|_{Q=Q_{exp}} \\ &= \langle r^2 \rangle \left[1 + \frac{W}{I_\infty (\sqrt{W^2 + \lambda^2} - W)} \right]^{-1}. \end{aligned} \quad (\text{C.4})$$

Appendix D

COMPARISON OF INTRINSIC MSD AND SIMULATED MSD

In neutron scattering experiments, the ISF $I_{inc}(\mathbf{Q}, t)$ is

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^N \langle b_i \exp(-i\mathbf{Q} \cdot \mathbf{r}_i(t)) \exp(i\mathbf{Q} \cdot \mathbf{r}_i(0)) \rangle, \quad (\text{D.1})$$

where $r_i(t)$ is the position of nucleus i . The incoherent ISF is sum over all nuclei in protein. The $I_{inc}(\mathbf{Q}, t)$ in Eq. (D.1) is commonly represented by a global $I(\mathbf{Q}, t)$. In the global $I(\mathbf{Q}, t)$, we neglect the heterogeneity (which will be discuss in Chap. 5) and a single scatterer center represents whole protein. The global $I(\mathbf{Q}, t)$ is

$$I(\mathbf{Q}, t) = \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(t)) \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle. \quad (\text{D.2})$$

The $t = \infty$ limit value of the incoherent ISF defines the intrinsic MSD $\langle r^2 \rangle$;

$$I(\mathbf{Q}, t = \infty) = \exp(-\frac{1}{3}Q^2 \langle r^2 \rangle). \quad (\text{D.3})$$

In MD simulations, the simulated MSD is calculated from

$$\Delta^2(t) = \langle [r(t) - r(0)]^2 \rangle \equiv \frac{1}{N} \sum_{i=1}^N \langle [r_i(t) - r_i(0)]^2 \rangle, \quad (\text{D.4})$$

where $r_i(t)$ is the position of nucleus i which is obtained from simulation. After long times, $\Delta^2(t \rightarrow \infty) = 2\langle r^2 \rangle_{MD}$.

The MSDs $\langle r^2 \rangle$ defined in Eq. (D.3) and $\langle r^2 \rangle_{MD}$ defined in Eq. (D.4) are not exactly same. Because $\langle r^2 \rangle$ is obtained from $I_{inc}(\mathbf{Q}, t)$ data which is average over ISF data of each nuclei, since $\langle r^2 \rangle_{MD}$ is calculated as an average over MSD of each nucleus. The average over the nuclei in the protein made when the full $I_{inc}(\mathbf{Q}, t)$ is represented by a global $I(\mathbf{Q}, t)$ is not the same as the average over the nuclei made in Eq. (D.4).

Appendix E

THE IMPORTANCE OF FOURTH CUMULANT TERM

The intermediate scattering function $I(\mathbf{Q}, t)$ at infinite time limit ($t = \infty$) is,

$$I(\mathbf{Q}, t) = I_\infty = \exp[-\langle[\mathbf{Q} \cdot \mathbf{r}]^2\rangle + \frac{1}{12}[\langle[\mathbf{Q} \cdot \mathbf{r}]^4\rangle - 3\langle[\mathbf{Q} \cdot \mathbf{r}]^2\rangle^2] + \dots] \quad (\text{E.1})$$

For Q parallel to direction α

$$\begin{aligned} \langle[\mathbf{Q} \cdot \mathbf{r}]^2\rangle &= Q^2\langle r_\alpha^2\rangle \\ \langle[\mathbf{Q} \cdot \mathbf{r}]^4\rangle &= Q^4\langle r_\alpha^4\rangle \end{aligned}$$

$$\begin{aligned} I_\infty &= \exp[-Q^2\langle r_\alpha^2\rangle + \frac{1}{12}Q^4[\langle r_\alpha^4\rangle - 3\langle r_\alpha^2\rangle^2]] \\ &= \exp[-Q^2\langle r_\alpha^2\rangle[1 - \frac{1}{12}Q^2\gamma_\alpha\langle r_\alpha^2\rangle]] \end{aligned} \quad (\text{E.2})$$

where $\gamma_\alpha = \frac{\langle r_\alpha^4\rangle - 3\langle r_\alpha^2\rangle^2}{\langle r_\alpha^2\rangle^2}$.

The 4th cumulant is significant is the kurtosis γ is large enough that $\frac{1}{12}Q^2\gamma_\alpha\langle r_\alpha^2\rangle$ is comparable to 1. Note $\langle r_\alpha^2\rangle \approx \frac{1}{3}\langle r^2\rangle$ e.g. for example $\gamma_\alpha \sim 1$, $\langle r^2\rangle \sim 2 \text{ \AA}^2$, $Q^2 \sim 4 \text{ \AA}^{-2} \implies \frac{1}{12}Q^2\gamma_\alpha\langle r^2\rangle \sim \frac{4}{12} \sim \frac{1}{3}$.

Thus, for large $\langle r^2\rangle \sim 3 \text{ \AA}^2$ and $\gamma \sim 1 - 2$, we expect the 4th cumulant to be significant at $Q \sim 2 \text{ \AA}^{-1}$.

Appendix F

CALCULATING THE FULL ISF AND GAUSSIAN ISF

The incoherent ISF $I_{inc}(\mathbf{Q}, t)$ in Eq. (4.1) depends on the \mathbf{Q} vector. The $I_{inc}(\mathbf{Q}, t)$ observed in neutron scattering experiments is averaged over \mathbf{Q} vectors which have a same magnitude. In the calculation of $I_{inc}(\mathbf{Q}, t)$, the Q vector is chosen as $\mathbf{Q} = Q\mathbf{x}$, $\mathbf{Q} = Q\mathbf{y}$ and $\mathbf{Q} = Q\mathbf{z}$, then we take an average over these three calculations:

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{3}[I_{inc}(Q\mathbf{x}, t) + I_{inc}(Q\mathbf{y}, t) + I_{inc}(Q\mathbf{z}, t)]. \quad (\text{F.1})$$

where $I_{inc}(Q\mathbf{x}, t)$, $I_{inc}(Q\mathbf{y}, t)$ and $I_{inc}(Q\mathbf{z}, t)$ are calculated from Eq. (4.1). Fig. (F.1) shows that the $I_{inc}(\mathbf{Q}, t)$ data calculated from Eq. (F.1) for all H is agreed well with the $I_{inc}(\mathbf{Q}, t)$ data calculated by SASSENA.

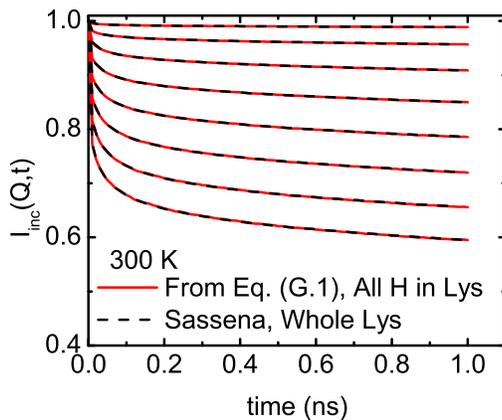


Figure F.1: The comparison of the ISF, $I_{inc}(\mathbf{Q}, t)$, for all H in lysozyme (red solid lines) calculated from Eq. (F.1) and $I_{inc}(\mathbf{Q}, t)$, for whole lysozyme (black dashed lines) calculated by using SASSENA.

Similar to the calculation of $I_{inc}(\mathbf{Q}, t)$, $I_{iG}(\mathbf{Q}, t)$ is calculated for $\mathbf{Q} = Q\mathbf{x}$, $\mathbf{Q} = Q\mathbf{y}$ and $\mathbf{Q} = Q\mathbf{z}$, then averaged over them:

$$I_{iG}(\mathbf{Q}, t) = \frac{1}{3}[I_{iG}(Q\mathbf{x}, t) + I_{iG}(Q\mathbf{y}, t) + I_{iG}(Q\mathbf{z}, t)]. \quad (\text{F.2})$$

where $I_{iG}(Q\mathbf{x}, t)$, $I_{iG}(Q\mathbf{y}, t)$ and $I_{iG}(Q\mathbf{z}, t)$ are calculated from Eq. (4.9).

Appendix G
THE SELF CONSISTENT HARMONIC THEORY

In this section we derive the self consistent harmonic theory for a single particle of mass M in an arbitrary potential, $V(u)$ used in the present article. The particle is described by the Hamiltonian

$$\hat{H} = \hat{K} + V(u) \quad (\text{G.1})$$

where $\hat{K} = -(\hbar^2/2M)d^2/du^2$ is the kinetic energy operator. We introduce a model harmonic Hamiltonian

$$\hat{H}_h = \hat{K} + \frac{1}{2}\phi u^2 \quad (\text{G.2})$$

and corresponding harmonic density matrix

$$\rho_h = \frac{e^{-\beta H_h}}{\text{Tr}(e^{-\beta H_h})} \quad (\text{G.3})$$

where $\beta = (k_B T)^{-1}$. Expectation values evaluated using ρ_h are

$$\langle \hat{O} \rangle_h = \text{Tr}\{\rho_h \hat{O}\} \quad (\text{G.4})$$

where \hat{O} is any operator. The corresponding model harmonic Helmholtz free energy is

$$\begin{aligned} F_h &= \langle \hat{H}_h \rangle_h - TS_h \\ &= \langle \hat{H}_h \rangle + (k_B T) \text{Tr}\{\rho_h \log \rho_h\} \\ &= k_B T \log 2 \sinh \left(\frac{1}{2} \beta \hbar \omega \right) \end{aligned} \quad (\text{G.5})$$

where $\omega = (\frac{\phi}{M})^{\frac{1}{2}}$ is the model simple harmonic oscillator frequency. In terms of \hat{H}_h the particle \hat{H} is

$$\hat{H} = \hat{H}_h + V(u) - \frac{1}{2}\phi u^2. \quad (\text{G.6})$$

The basic concept is to use ρ_h as a trial or model density matrix with which to evaluate the particle free energy. The trial particle free energy is

$$\begin{aligned} F_{trial} &= \langle H \rangle_h + (k_B T) \text{Tr}\{\rho_h \log \rho_h\} \\ &= F_h + \langle V(u) \rangle_h - \frac{1}{2}\phi \langle u^2 \rangle_h. \end{aligned} \quad (\text{G.7})$$

The F_{trial} is an upper bound to the exact Helmholtz free energy, the Gibbs-Bogolibov variational principle. We consider F_{trial} as a functional of the model harmonic force constant ϕ and the MSD $\langle u^2 \rangle_h$. We minimize the F_{trial} with respect to ϕ and $\langle u^2 \rangle_h$ independently holding the other variable constant to find the optimum ϕ and $\langle u^2 \rangle_h$. This variation gives

$$\frac{\delta F_{trial}}{\delta \langle u^2 \rangle_h} = \frac{1}{2} \langle \frac{d^2 V(u)}{du^2} \rangle_h - \frac{1}{2}\phi = 0 \quad (\text{G.8})$$

$$\frac{\delta F_{trial}}{\delta \phi} = \frac{1}{2} \left(\frac{\hbar}{2M\omega} \right) \coth \left(\frac{1}{2}\beta\hbar\omega \right) - \frac{1}{2}\langle u^2 \rangle_h = 0 \quad (\text{G.9})$$

The first term in Eq. G.8 is obtained by making a Taylor's expansion of $\langle V(u) \rangle_h$ in Eq. G.7 about $V(0)$, $V(u) = e^{u(\frac{d}{du})}V(0)$, and a cumulant expression of the exponential $\langle e^{u(\frac{d}{du})} \rangle_h$. For, a harmonic system with Gaussian distributions, $\langle u \rangle_h = 0$ and all cumulants beyond the second cumulant vanish so that

$$\begin{aligned} \langle V(u) \rangle_h &= \langle e^{u(\frac{d}{du})} \rangle_h V(0) \\ &= e^{\frac{1}{2}\langle u^2 \rangle_h (\frac{d^2}{du^2})} V(0). \end{aligned} \quad (\text{G.10})$$

Differentiation with respect to $\langle u^2 \rangle_h$ then yields the first term in Eq. G.8 and Eq. G.8 leads immediately to Eq. G.1 with $\langle \rangle_h$ expressed as a average in configuration space. The average in configuration space can be obtained by Fourier transforming $V(u)$ and again using a cumulant expansion,

$$\begin{aligned}
\langle \nabla^2 V(u) \rangle_h &= \nabla^2 \int dq V(q) \langle e^{iqu} \rangle_h \\
&= \nabla^2 \int dq V(q) e^{-\frac{1}{2}q^2 \langle u^2 \rangle_h} \\
&= \int du e^{-\frac{1}{2}u^2 / \langle u^2 \rangle_h} \nabla^2 V(u)
\end{aligned} \tag{G.11}$$

where d/du is denoted by ∇ . The first term in Eq. G.9 is obtained by differentiating Eq. G.5 for F_h with respect to ω using $\omega^2 = \phi/M$. Eq. G.9 leads immediately to Eq. 5.6. The ϕ represents the optimum harmonic force constant representing a particle in an anharmonic well $V(u)$ in which the particle has a MSD $\langle u^2 \rangle$.