



**Thermal energy transport and signal transduction of
biomolecular machines: Molecular dynamics
simulation study of proteins**

(生体分子機械の熱エネルギー輸送と情報伝達:
タンパク質の分子動力学シミュレーション)

Tingting WANG

Supervisor: Assoc. Prof. Dr. Takahisa YAMATO

DISSERTATION

Submitted in Partial Fulfilment of the Requirements for
the Degree of Philosophy Doctor in Physics

Computational Biophysics Laboratory (B-Lab)
Graduate School of Science
September 2023

This page intentionally left blank.

Acknowledgement

I would like to take this opportunity to express my deepest gratitude and appreciation to all those who have supported and guided me throughout my doctoral journey.

First and foremost, I am immensely grateful to my supervisor, Takahisa Yamato, for his unwavering support, invaluable guidance, and immense knowledge. For these three years, he provided the most excellent scientific research environment, the best platform and resources, so that I can devote myself to scientific research and academic exchange. Without his dedication, I am not able to contribute myself to this area. I am truly fortunate to have had the opportunity to work under his mentorship.

I am also thankful to the lab members: Yuko Okamoto, Akihiro Kimura, Wataru Sugiyama, and fellow researchers: Valérie Belle, Alessio Bonucci, and Véronique Receveur-Bréchet, who have provided a stimulating academic environment and valuable discussions. Their support, encouragement, and intellectual exchanges have been invaluable throughout the course of my research and will be the spiritual wealth of my life.

My heartfelt thanks also go to my mom, Xinhua Huang, and my love, Zhongyu Guo, for their unconditional love, encouragement, and unwavering belief in my abilities. Their constant support and understanding have always been a constant source of motivation, and I am grateful for their presence in my life.

Lastly, a motto of Wang Yangming is for myself, "*Knowledge as Action*", to remind myself: "*Keep the enthusiasm for knowledge and research, put the ideas into practice and make the contributions*".

My story with Nagoya University from the winter of 2020 to the summer of 2023 will be one of the most precious and memorable experiences of my whole life. Thanks again for all of you.

This page intentionally left blank.

Contents

1	Introduction	1
1.1	Proteins	1
1.2	Vibrational energy transport in proteins	2
1.3	Linear response theory	4
1.4	Molecular dynamics simulations	5
1.5	Current calculation for proteins	6
1.6	Machine learning methods	6
1.7	Knowledge gap	7
1.7.1	The anisotropic and non-uniform flow of thermal energy	7
1.7.2	Determinates of protein energy transport	8
1.7.3	Factors affecting heat transfer through nonbonded contacts	9
1.8	Organization of the thesis	9
2	Theory, Materials and Methods	11
2.1	Thermal conductivity	11
2.2	Derivation of inter-atomic heat current	12
2.3	Linear-homopolymer-like model	13
2.4	Cross-correlation correction on linear-homopolymer-like model	15
2.5	Inter-residue heat current analysis	19
2.6	Inter-residue energy flow analysis	20
2.7	Parameterization of the Fe^{3+} metal cofactor	20
2.8	Equilibrium molecular dynamics simulations	22
2.8.1	Systems setup	23
2.8.2	Minimization	24
2.8.3	Equilibration	24
2.8.4	<i>NPT</i> sampling and <i>NVE</i> MD simulations	24
2.9	Definition and classification of nonbonded contacts	25

2.9.1	Definition	25
2.9.2	Classification	25
2.10	Hydrogen bond occurrence probability	26
2.11	Random forest regression method	26
2.12	Computational modeling for <i>Thalassosira pseudonana</i> CP12 homodimer	27
2.12.1	FASTA sequence of CP12	27
2.12.2	AlphaFold2 prediction	29
2.12.3	Harmonic restrained MD simulations	29
2.12.4	Restrained-ensemble MD simulations	30
3	Site-selective heat current analysis with linear-homopolymer-like model	33
3.1	The overall thermal conductivity	33
3.2	Local thermal transport	34
3.3	Cross-correlation correction	35
3.4	Residue-type dependence	38
3.5	Local density dependence	39
3.6	Thermal diffusivity and temperature relaxation time	41
3.7	Water model	42
4	Heat and energy transfer through nonbonded contacts	43
4.1	Energy transfer and heat transfer	43
4.2	Interaction type dependence	44
4.3	Peptide bonds? or nonbonded contacts?	45
4.4	Thermal energy transport through hydrogen bonds	46
4.5	Important features of thermal transport through nonbonded contacts	50
5	Energy transport and its function in sensory domain of <i>BjFixL</i>	53
5.1	Effects of ligand binding and dimerization	55
5.2	Vibrational energy transfer pathways and residue interaction network	58
5.3	Reorganization of vibrational energy transfer pathways upon ligand binding	60
6	Structural and dynamical characterization of CP12 protein	63
6.1	AlphaFold2 model	63
6.2	Molecular dynamic simulations	65
7	Conclusions and future Directions	69
7.1	Conclusions	69
7.2	Future directions	71

Contributions	73
Published articles	73
Conference talks	73
Bibliography	75

List of Figures

1.1	The overall growth of protein structural data. Data were obtained from the RCSB Protein Data Bank (https://www.rcsb.org), which serves as a valuable repository for storing and sharing these protein structures.	1
1.2	20 amino acids in proteins.	2
1.3	Milestone studies on vibrational energy transport properties in proteins.	3
1.4	Workflow diagram of CURP program for energy/heat currents calculations, taken from https://curp.jp	6
2.1	Thermal transport through backbone model. Local heat currents, $h_{\alpha,\alpha}$ and $h_{\alpha,\alpha+1}$, occur within each residue α and between each pair of residues α and $\alpha + 1$	14
2.2	Short-range cross-correlation approximation. Cross-correlations are considered only for nearby pairs (dashed line).	16
2.3	Allocation of cross-correlation term.	17
2.4	Molecules used to obtain the tuned force field parameters for met-FixLH dimer system. (a) Molecule used in Gaussian optimization calculations to obtain the Fe^{3+} metal associated bond and angle parameters. (b) Molecule used used in Gaussian RESP partial charge calculations to obtain the partial charges for the metal site.	21
2.5	Molecules used to obtain the tuned force field parameters for met-FixLH-imd dimer system. (a) The molecule used in Gaussian optimization calculations to obtain the Fe^{3+} metal associated bond and angle parameters. (b) The molecule used in Gaussian RESP partial charge calculations to obtain the partial charges for the metal site.	21
2.6	Computational procedure for HP36 protein system.	22
2.7	Computational procedure for <i>Bj</i> FixLH protein systems.	23
2.8	(A) Refinement scheme for the model obtained from AlphaFold2. (B) Starting model for the harmonic restrained MD simulations. C) Distances (blue) and distance distributions (black) between spin label pairs of the starting model for reMD simulation and DEER experiment data, respectively.	28

2.9	Time evolution of the CA-CA distances. We consider five residue pairs, S39-S39, S46-S46, S56-S56, S83-S83, and C150-C150 in the dimer. For each pair, the CA-CA distance are plotted as a function of time during harmonic restrained MD simulation using Amber.	30
2.10	The fixation between CYR1 spin label and its attached residues. The all-atom CYR1 spin label has a main chain like amino acid. The reMD simulation adds and fixes the spin label model by overlapping the main chain of spin label attached residue and the main chain of spin labels with a harmonic force constant of 10 kcal/(mol·Å ²).	31
3.1	The thermal conductivity of HP36. The autocorrelation function of the heat current of the entire molecule of HP36 was calculated and ensemble-averaged, and plotted as a function of time (blue). The inset shows the short-time region (0 - 0.2 ps). The heat current autocorrelation function was integrated over time using the trapezoidal rule, yielding a quantity denoted as λ as a function of integration time (shown in orange), and a running mean of λ was computed with a window size of 100 fs and is illustrated in red. The volume of HP36 was set to 4845.447 Å ³ . ⁵⁹	33
3.2	The contribution factors without cross-correlation corrections were calculated and plotted for both intra-residue ($c_{\alpha,\alpha}$, shown in red) and inter-residue ($c_{\alpha,\alpha+1}$, shown in green) interactions, as described in equations (Equation 2.15) and (Equation 2.16), respectively. The data points, such as $c_{1,2}$, represent the contribution factor between residue numbers $\alpha = 1$ and $\alpha = 2$. The horizontal axis displays the residue number, α , along with the corresponding residue names represented in the one-letter code. The grey-shaded regions correspond to the α -helical regions.	34
3.3	The cross-correlation contribution factor between nearest neighbor residues, $c_{\alpha,\alpha+1}^{\text{cross}}$, defined as $\xi_{\alpha,\alpha+1}/\Lambda$, was plotted as a function of the residue number. Each data point represents the value of $c_{\alpha,\alpha+1}^{\text{cross}}$ between residue α and $\alpha + 1$. For example, the data point $c_{1,2}^{\text{cross}}$ corresponds to the cross-correlation between residues 1 and 2.	35
3.4	Contribution factors with cross-correlation correction. See the caption to Figure 3.2.	35
3.5	Volume dependence of \tilde{c} , a residue-wise contribution factor after cross-correlation correction, for charged (green), polar (orange) and hydrophobic (blue) residues, respectively.	39
3.6	A scatter plot and linear correlation analysis were performed to examine the relationship between the intra-residue thermal conductivity and the local density, represented by the mass density (ρ), calculated as the ratio of residue mass to residue volume (unit: g/cm ³).	39
3.7	Comparison of the residue volumes in this study with those in water (blue) and the protein interior (green) ¹²⁵ . The second row of the x-axis represents the amino acid names using the one-letter format, with the residues buried into the protein interior highlighted in red. The third row indicates the residue types, with hydrophobic residues shown in blue, charged residues in green, and polar residues in orange.	40
4.1	Correlation of $\Lambda_{\alpha,\beta}$ and $G_{\alpha,\beta}$	43

4.2	Frequency histogram of local thermal conductivity in HP36. It shows the occurrence of $\lambda_{\alpha,\beta}$ for different types of residue pairs (α, β) in nonbonded native contacts, together with the residue-wise thermal conductivity and those between adjacent residue pairs along the chain. For comparison, the value of the thermal conductivity of the entire HP36 molecule is indicated by the red dashed line, as well as that of bulk water by the black dashed line. ¹³⁷	44
4.3	Nonbonded contacts that can compete with polypeptide chain.	45
4.4	Correlation between the inter-residue thermal conductivity through hydrogen bonding and their hydrogen bond occurrence probability (P_{HB}) during MD simulations. (a): all residue pairs in contact with hydrogen bonds; (b) residue pairs ($\alpha, \alpha + 4$) with α -helical hydrogen bonding.	46
4.5	Distribution of average contact distance, the shortest interatomic distance between residue pairs in nonbonded native contacts. (a) all residue pairs in nonbonded native contacts; (b) residue pairs ($\alpha, \alpha+4$) in contact with α -helical hydrogen bonds.	49
4.6	(a) Cross plot of $\lambda_{\alpha,\beta}$ and predicted $\lambda_{\alpha,\beta}$ of nonbonded contacts ($\langle d_c \rangle < 6$) using random forest regression model; (b) VIP score plot of variable importance of variables. $\langle d_c^2 \rangle$: contact (shortest) distance of residue α and β ; P_{HB} : hydrogen bonds occurrence probability; $\langle \delta d_c^2 \rangle$: the variance in contacts distance; $V_{\alpha\beta}$: summation of the volumes of residue pair in contacts; RT_α (RT_β): residue type; IT: interaction type.	50
4.7	(a) Cross plot of $\lambda_{\alpha,\beta}$ between those obtained from the MD simulations and those predicted using random forest model for nonbonded native contacts ($\langle d_c \rangle < 4$) using random forest regression model; (b) VIP score plot of variable importance. $\langle d_c^2 \rangle$: contact (minimum interatomic) distance between residue α and β ; P_{HB} : hydrogen bonds occurrence probability; $\langle \delta d_c^2 \rangle$: the variance of contact distance; $V_{\alpha\beta}$: summation volume of contacts; RT_α (RT_β): residue type; IT: interaction type.	51
4.8	Correlation map. (a) Pairwise Pearson correlation coefficients, r ; (b) p -values.	52
5.1	Crystal structure of met- <i>Bj</i> FixLH dimer. The role of the dimer interface and the coiled-coil linker region in the $\text{J}\alpha$ helix is unclear.	55
5.2	Heatmap of G of (1) met-FixLH dimer protein and (2) met-FixLH-imd dimer protein. The unit of measure on the color bar on the right is $\text{kcal}^2/(\text{mol}^2 \text{ fs})$. Residue numbers and secondary structure names are labeled on the axes. Secondary structures α -helixes (green) and β -sheets (orange) are distinguished by different colors.	56
5.3	(a) Linear regression of G between chain A and chain B in met-FixLH dimer. (b) Linear regression of G between met-FixLH dimer and deoxy-FixLH monomer. G of deoxy-FixLH is from the previous study of our group ²⁹ . (c) Relationship of G between chain A and chain B in met-FixLH-imd dimer.	57
5.4	Energy exchange network of (a) met-FixLH dimer protein and (b) met-FixLH-imd dimer protein.	57

5.5	The strong interactions, represented by dash lines, are defined by the residue pairs with G values $> G_{\text{met}} + 2\sigma_{\text{met}} = 6.58 \times 10^{-3}$ [kcal ² /(mol ² fs)], where G_{met} ($= 2.42 \times 10^{-4}$) and σ_{met} ($= 3.17 \times 10^{-3}$) [kcal ² /(mol ² fs)] are the mean and the standard deviation of the G -values of all residue pairs, respectively. (A) met-FixL dimer, (B) met-FixLH-imd dimer. The interactions involved with the heme and the dimer interface are, shown in red and distinguished from the other interactions, shown in blue. Panels A/B, R1, R2, R3, show important amino acid residues in the regions around the heme ligand and on the dimer interface; close to the N-terminus; and close to the C-terminus, respectively. It should be noted that the connections between the imidazole ligand and residue HIS200, ARG206, ARG220 and LEU236 are not shown in Panel B1.	59
5.6	Molecular dynamic simulations of met- <i>Bj</i> FixLH without force restraints on N-terminal of A' α helix. (A) The crystal structure model of MD simulations. (B) The time evolution of distance between carbon atom (CA) of residue ILE128s.	60
5.7	Reorganization of the network of vibrational energy transfer pathways. Strengthened (Weakened) paths upon ligand binding with $r\Delta G > (<) 0.5$ are indicated by dotted lines in red (blue). The residue name and residue number (colored in purple and black) represent the residues from chain A and chain B, respectively.	61
6.1	Top 5 homodimer models for WT, S46C and S56C mutants are shown with each residue being colored according to its per-residue confidence score (pLDDT): Blue (high), cyan (high medium), yellow (low medium) and red (low).	64
6.2	AlphaFold2 model of wild-type <i>Thalassosira pseudonana</i> CP12 dimer. One monomer is shown in cartoon representation using a color scheme based on confidence measure (blue: high, cyan: high medium, yellow: low medium, red: low), and the other monomer presents a translucent surface format. Two AWD_VEEL motifs and two cysteine residues (C142 and C150) are shown with sticks and spheres, respectively.	65
6.3	(A) Comparison of SAXS curves between AlphaFold2 model and experiment; (B) Comparison of SAXS curves between initial structure of restrained-ensemble molecular dynamics (reMD) simulation and experimental data; In both (A) and (B), an insert figure shows the SAXS curve at low q region ($0 - 0.1 \text{ \AA}^{-1}$) and gives the R_g for both experiment and predicted model. (C) Initial structure of CP12 with 25 copies of R1 spin labels for the production run of reMD simulations; (D) Comparison of experimental and calculated distance distribution between the spin labels on 5 the residue pairs. The SAXS discrepancy between simulated and experimental data, χ^2 , for both (A) and (B), are marked in blue text.	66
6.4	Probability distribution of radius of gyration (R_g). We used of 20, 000 sampled structures from reMD simulations.	67

List of Tables

2.1	Quantum calculation results of met-FixLH and met-FixLH-imd. Hartree Energy (unit: eV) comparison among different spin states of met-FixLH and met-FixLH-imd.	22
3.1	The second nearest cross-correlation. $\xi_{\alpha,\alpha+2}$ was calculated in a similar manner as $\xi_{\alpha,\alpha+1}$ using Equation 2.15, and contribution factor, $e_{\alpha,\alpha+2}^{\text{cross}}$, was calculated as $\xi_{\alpha,\alpha+2}/\Lambda$. The values of ξ are shown in the unit of $(\text{\AA}\cdot\text{kcal/mol})^2/\text{fs}$	36
3.2	Local thermal transport property after the cross-correlation correction. $\tilde{\Lambda}_{\alpha,\alpha}$ ($\alpha=1, 2, \dots, 36$) and $\tilde{\Lambda}_{\alpha,\alpha+1}$ ($\alpha=1, 2, \dots, 35$) are shown in unit of $(\text{\AA}\cdot\text{kcal})^2/\text{fs}$ (Equation 2.12), while the value of Λ was 9.79 in the same unit (Equation 2.13).	37
3.3	Pearson correlation coefficient (r), p -value and slope of linear relationship between \tilde{c} and volume in Figure 3.5. Generally, a value of $ r > 0.3$, or a p -value < 0.05 is often considered to be statistically significant, indicating that the pair of variables under consideration are correlated.	40
4.1	Hydrogen bond occurrence probability (P_{HB}) and averaged shortest distance ($\langle d_c \rangle$) between residue α and $\alpha+4$. MC : mainchain, SC: sidechain.	47
4.2	Inter-residue thermal conductivity and hydrogen bond formation capacity P_{HB} between residue pair α and $\alpha+4$. Also the contact distance (shortest interatomic distance between the residue pair) is shown for each pair.	47

Chapter 1

Introduction

1.1 Proteins

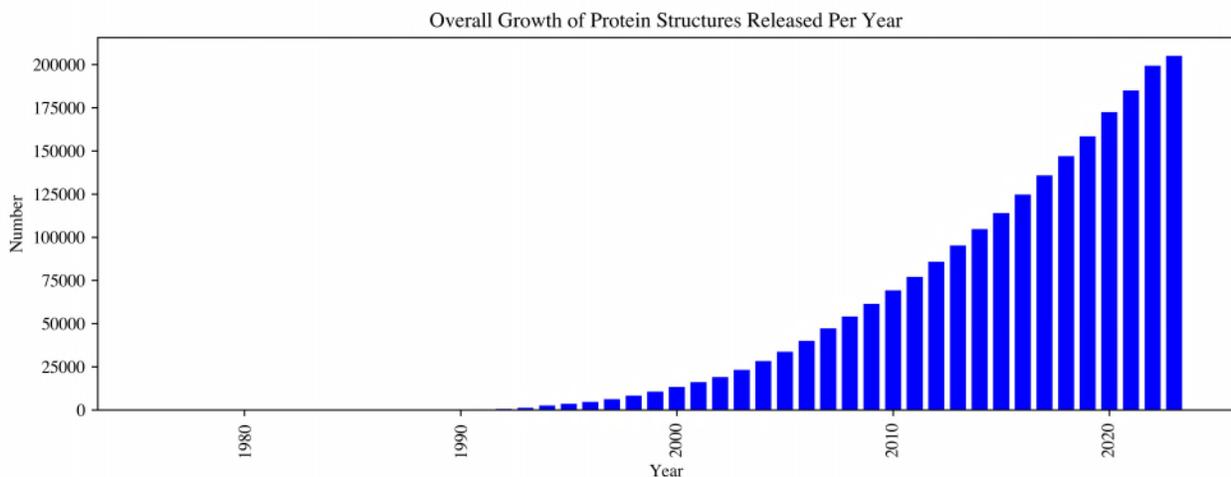
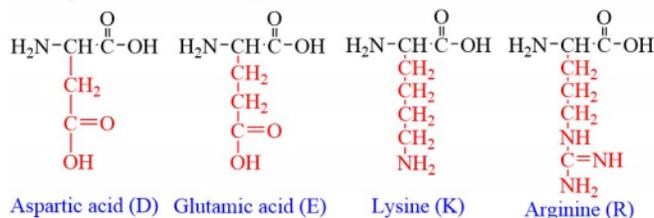


Figure 1.1: The overall growth of protein structural data. Data were obtained from the RCSB Protein Data Bank (<https://www.rcsb.org>), which serves as a valuable repository for storing and sharing these protein structures.

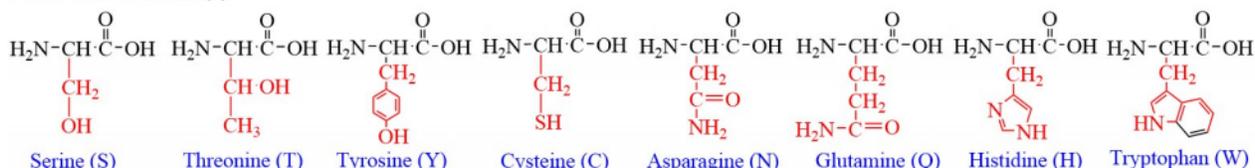
Although proteins and protein-based materials, such as meat, egg, silk and leather, have been common materials we use for millennia, they were recognized as a distinct category of materials only after Antoine Fourcroy and others in the 18th century found that these materials coagulate or flocculate under high temperature or acid conditions. In 1958, the first X-ray crystallographic structure of myoglobin, a globular protein molecule, was published by John Kendrew, hereafter, a new era in protein structural biology began.¹ Since then, protein structure determination has evolved beyond X-ray crystallography. With advancements in techniques such as nuclear magnetic resonance (NMR), high-resolution electron microscopy, and computational methods, an increasing number of protein structures (Figure 1.1) are being characterized. These efforts are driven by the goal of achieving atomic-level resolution and a deeper understanding of

protein structure, property and function.

Charged amino acids (4)



Polar amino acids (8)



Hydrophobic amino acids (8)

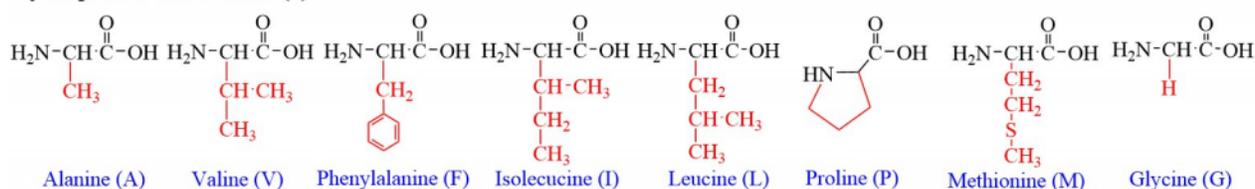


Figure 1.2: 20 amino acids in proteins.

Most natural proteins are linear polymers with a sequence constructed by 20 different amino acids (Figure 1.2). The distinguishing feature among them is the unique side chain attached to the central carbon atom (C) that connects the hydrogen atom, amino group (NH_2), and carboxyl group (COOH). They are usually classified into three different groups based on the chemical properties of their side chains. Apart from intrinsically disordered proteins, globular proteins fold into specific structures, normally known as “native conformations”, characterized by four levels of hierarchies: (1) primary structure, which refers to the amino acid sequence; (2) secondary structure, including α -helices, β -sheets, and turns/loops, which are stabilized by hydrogen bonds; (3) tertiary structure, representing the overall three-dimensional structure of a protein; and (4) quaternary structure, formed by multiple protein subunits.

1.2 Vibrational energy transport in proteins

Protein and protein-based materials are considered as nonmetallic solids. The thermophysical properties of some common protein materials have been measured and recorded in a series of data called *Thermophysical properties of matter* collected by the TPRC Data Series in the 1970s. In this data series, numerous data on the thermal conductivity, specific heat, and thermal diffusivity of protein materials are recorded and reported in volume 2, volume 5, and volume 10, respectively.

Vibrational energy transfer in proteins plays a significant role in conformational changes, energy dissipation, protein folding and unfolding, bimolecular interactions, enzymatic reactions, and protein dynamics. The current methods

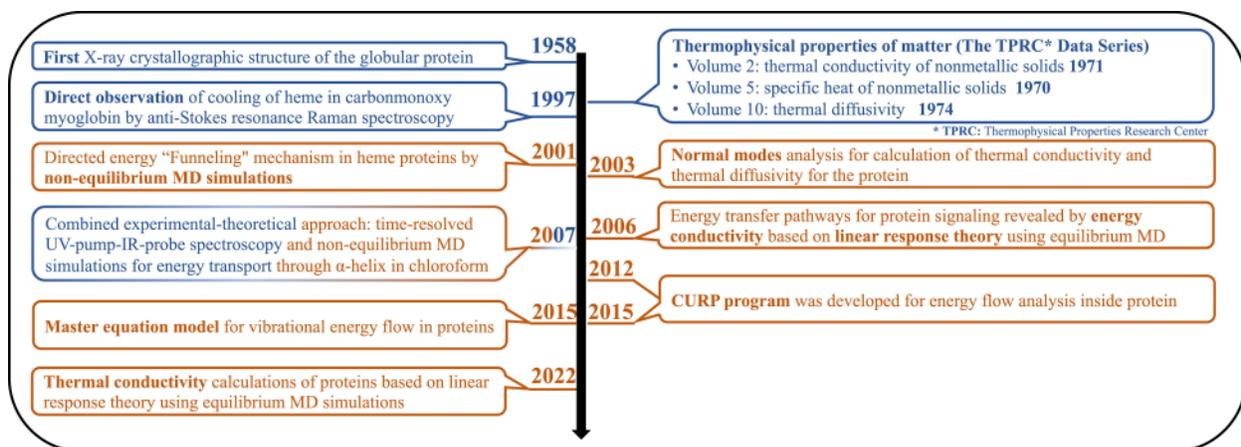


Figure 1.3: Milestone studies on vibrational energy transport properties in proteins.

of investigating vibrational thermal energy transport in proteins include experimental and computational approaches. Milestone studies on vibrational energy transport properties in proteins are shown in Figure 1.3. From a macroscopic point of view, experimental methods involve measuring the thermal conductivity or diffusivity of protein samples using techniques such as laser flash analysis, 3ω method, or Raman spectroscopy.² From a microscopic point of view, time-resolved experimentation approaches, such as time-resolved ultraviolet resonance Raman spectroscopy and UV-pump-IR-probe transients spectroscopy, have been used to investigate the vibrational energy transport mechanism in peptides and proteins.³⁻⁶ These methods enabled direct observation of the redistribution of excess energy deposited into either natural group of protein, like heme, or unnatural amino acid sites connected to short peptides and proteins.

In 1997, Mizutani Yasuhisa and his colleagues utilized time-resolved ultraviolet resonance Raman spectroscopy to directly observe the energy flow in heme proteins, taking advantage of the efficient photothermal conversion properties of the heme group.⁷ The redistribution of excess energy can be probed by employing anti-Stokes ultraviolet resonance Raman spectroscopy on a tryptophan residue as a probe, enabling the mapping of energy flow with high spatial resolution. In 2007, they observed a double exponential decay in the cooling process of the heme group in myoglobin, with time constants of 1.9 ± 0.6 ps and 16 ± 9 ps for vibrational population decay, and 3.0 ± 1.0 ps and 25 ± 14 ps for temperature relaxation. Further investigations revealed that the primary pathway for energy transfer from the heme group to the protein is not mediated by the covalent bond between the heme and the proximal or the protein backbone, but instead occurs through atomic contacts between the heme and specific residues.^{3,8-10} In 2007, Gerhard Stock and his coworkers investigated the energy transfer through a 3_{10} -helix in chloroform solvent, whose N-terminal is attached with a heater of α -aminoisobutyric acid, with a combined experimental-theoretical approach: UV-pump-IR-probe transients spectra and non-equilibrium molecular dynamics (MD) simulations. Non-equilibrium MD simulations were also employed to simulate the dissipation of excitation energy from an unnatural amino acid (acting as a built-in heater), allowing the mapping of vibrational energy transport pathways originating from the heater. It demonstrated that 70% of the energy absorbed by the helix quickly equilibrates within a timeframe of less than a picosecond to the solvent and subsequently spreads along the helix through a diffusive-like process. The non-equilibrium MD simulations mimic the

laser excitation of the molecules by the non-equilibrium phase-space initial condition for the solute and the solvent atoms. The fitted experimental and simulated thermal diffusivity values along the peptide chain with 1D diffusion process were $10 \text{ \AA}^2 \text{ ps}^{-1}$ and $2 \text{ \AA}^2 \text{ ps}^{-1}$, respectively.⁵ Subsequently, the non-equilibrium MD simulations techniques were employed on various proteins, such as villin headpiece subdomain (HP36)^{11,12}, photoswitchable PDZ domain (PDZ2S)¹³, heptahelical transmembrane proteins¹⁴ and so on.⁶ Numerous results are yielded by these studies including scaling rules for vibrational energy transport in proteins through the backbone and nonbonded native contacts, competition of vibrational energy transport between bonds and contacts, and the role of energy transport on the allosteric mechanism in functional proteins. Besides, John E. Straub and his coworkers have employed non-equilibrium classical MD simulations to study the spatially anisotropic “funneling” mechanism for the kinetic energy relaxation of the excited heme group in solvated myoglobin by directly depositing excess kinetic energy on heme. They demonstrated that the kinetic energy decay of heme in native myoglobin and His93Gly mutant myoglobin were a single exponential with a time constant of 5.9 ps and 8.8 ps, respectively, while the kinetic energy relaxation of heme in cytochrome *c* was a biphasic exponential decay process with relaxation time constants of 1.5 ps and 10.1 ps for fast and slow processes, respectively.¹⁵⁻¹⁷

In addition to the non-equilibrium MD simulations mentioned above,^{12,18-22} computational and theoretical approaches²³⁻²⁷, including equilibrium MD simulations, normal mode analysis, i.e. the master equation model, and anharmonic network models, have also been employed to simulate the flow of vibrational energy in proteins.

One approach involves the utilization of a transport-coefficient-like quantity derived from the autocorrelation function of energy flow, which providing a measure of the amount of energy transferred per unit of time.²⁸ This methodology has been employed to construct a network of amino acid residues that mediates the propagation of intramolecular vibrational energy. This model (energy exchange network model) was employed to investigate the allosteric signal transduction mechanisms and energy transport pathways using equilibrium MD simulations.²⁹⁻³¹ Another approach, the master equation model, has been used to predict energy transport pathways in proteins.^{12,32,33} These predictions were compared to those obtained from all-atom non-equilibrium MD simulations. Theoretical studies were also conducted on scaling rules between vibrational energy transport coefficient and either contact distance or the variance of contacts distance for amino acid residue pairs in nonbonded native contacts in proteins.^{12,34} It is worth noting that vibrational transport occurs not only within the protein molecule but also across the protein-solvent interface and within the solvent. Theoretical analyses have been conducted to investigate these aspects^{15,35}.

1.3 Linear response theory

The process of transferring heat and energy with a temperature gradient is typically thought of as a non-equilibrium system. However, by considering carriers as particles, the fluctuation-dissipation transport theory, pioneered by Green and Kubo,³⁶⁻⁴⁰ provides general expressions for transport coefficients. These expressions are applicable at all times and densities and are derived from correlation or autocorrelation functions computed from an equilibrium system. The key idea is based on the linear response theory, which offers a framework for illustrating the relationships between perturbations that are applied to a system and their responses. As long as the state of the system stays close enough

to the equilibrium, the equilibrium fluctuations can dictate the non-equilibrium response. The system's response to a small perturbation is linearly related to the magnitude of the perturbation. The analysis of the dynamic characteristics and transport phenomena of molecules and materials can be performed by using linear response theory in the context of molecular dynamics simulations and statistical mechanics. A general expression of transport coefficient γ can be represented in terms of the integral of time correlation functions as shown in Equation 1.1.^{41,42}

$$\gamma = \int_0^{\infty} \langle \dot{\mathbf{A}}(t) \cdot \dot{\mathbf{A}}(0) \rangle dt \quad (1.1)$$

1.4 Molecular dynamics simulations

Molecular dynamics (MD) simulations⁴³ are computational techniques that are widely used to study the behavior and dynamics of molecules at the atomic level. Furthermore, it can be used to explore the heat, energy and signal flow in proteins based on the linear response theory. The MD simulations are especially suitable for studies of such transport phenomena in proteins as it allows for the *direct* measurement of energy/heat flows with high spatiotemporal resolution. Not only the biophysical properties of proteins, MD simulations are also useful for structural characterization and refinement of highly flexible proteins. We explored such possibility by integrating advanced MD simulation techniques with the experimental small-angle X-ray scattering (SAXS) and electron paramagnetic resonance/double electron–electron resonance (EPR/DEER) data.^{44–47}

In MD simulations, the laws of classical mechanics are used to numerically solve the equations of motion for molecular systems. By specifying the initial positions and velocities of the particles, the system's time evolution can be simulated. This allows us to observe and analyze various physical and chemical phenomena that occur on the atomic scale. We need force field functions to accurately capture the behavior of protein systems using MD simulations. In general, the force field functions consist of potential energy terms representing bonded interactions for bonds, angles, dihedral angles, and nonbonded interactions of van der Waals forces and electrostatic interactions (Equation 1.2).

$$\begin{aligned} E_{total} &= E_{bonded} + E_{nonbonded} \\ E_{bonded} &= \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi[1 + \cos(n\chi - \sigma)] \\ E_{nonbonded} &= \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon R_{ij}} \right] \end{aligned} \quad (1.2)$$

In Equation 1.2, K_b , K_θ and K_χ are force constants for bonds, angles and dihedrals, respectively; b and b_0 are force constant the bond length and the equilibrium bond length; θ and θ_0 are the angle and the equilibrium angle, respectively; χ is the dihedral value, n is the periodicity, and σ is the phase. $E_{nonbonded}$ contains two parts. The first term is also known as Lennard-Jones (LJ) potential, describing the attractive dispersion and repulsive interactions. The second term corresponds to Coulomb's law and is used to describe the electrostatic interactions. q_i and q_j are the charges

of atoms i and j . R_{ij} is the interatomic distance and ϵ is the dielectric constant.

Force field parameters are typically derived from experimental data and quantum mechanical calculations. They aim to strike a balance between accuracy and computational efficiency, allowing simulations of biologically relevant timescales and system sizes. Different force fields for different systems have been developed, each with its own set of assumptions, approximations, and parameterization strategies. Commonly used force fields in MD simulations of biomolecular systems include CHARMM⁴⁸, AMBER⁴⁹, GROMOS⁵⁰, and OPLS-AA⁵¹. These force fields have been extensively tested and parameterized for a wide range of biomolecules, including proteins, nucleic acids, lipids, and carbohydrates. The associated molecular dynamics software packages and servers have also undergone substantial development based on these force-fields, including CHARMM (<http://www.charmm.org>),⁵² Amber (<http://amber.scripps.edu>),⁵³ GROMACS (<http://www.gromacs.org>),⁵⁴ NAMD (<http://www.ks.uiuc.edu/Research/namd>),⁵⁵ and GENESIS (<https://www.r-ccs.riken.jp/labs/cbrt>)⁵⁶.

1.5 Current calculation for proteins

CURP (CURrent calculation for Proteins, <https://curp.jp>) package is a computational tool designed by YAMATO group to calculate the energy/heat currents and atomic stress tensors in proteins.⁵⁷⁻⁵⁹ It operates by utilizing atomic coordinates and velocity trajectories obtained from molecular dynamics (MD) simulations. In this thesis, the CURP program is used to perform the energy flow, energy transport coefficient, heat current, thermal conductivity calculations. A workflow diagram of CURP calculations is shown in Figure 1.4.

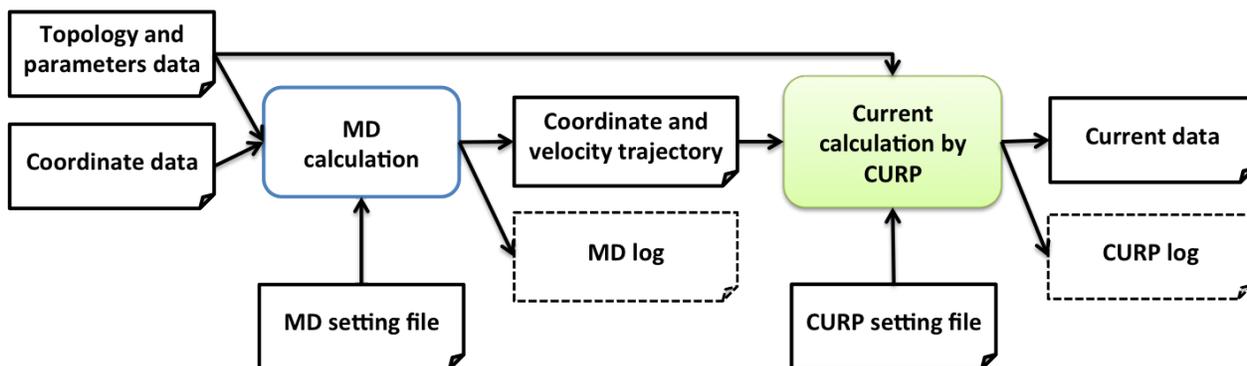


Figure 1.4: Workflow diagram of CURP program for energy/heat currents calculations, taken from <https://curp.jp>.

1.6 Machine learning methods

Machine learning methods play a crucial role in protein science by enabling the analysis, prediction, and understanding of various aspects related to proteins, including structure, function, interactions, and dynamics.⁶⁰⁻⁶³ These methods leverage computational algorithms and statistical models to extract meaningful information from protein data and address complex challenges in the field.

One of the key applications of machine learning in protein science is protein structure prediction. Predicting the three-dimensional structure of a protein from its amino acid sequence is a fundamental problem in biology. Machine learning algorithms, such as deep learning-based approaches, have shown remarkable success in predicting protein structures with high accuracy.^{64–66} These methods utilize large training datasets, incorporating known protein structures, to learn patterns and relationships between sequences and structures, enabling the generation of accurate structure predictions for previously unknown proteins.

On the other hand, machine learning methods play a significant role in regression tasks in protein science, where the goal is to predict continuous or quantitative properties of proteins based on various input features. Random forest regression method among them can make accurate predictions by building multiple decision trees on randomly sampled subsets of the training data and then averaging the predictions of all the trees to obtain the final prediction. It can handle both numerical and categorical features, automatically handles missing values and outliers, and is resistant to overfitting. The estimate of feature importance can be useful for feature selection. As described in Chapter 1.7.3, the thermal transfer through nonbonded contacts is influenced by multiple factors related to various static and dynamical properties of proteins, interior of which we observe complex and heterogeneous environments. To get a better understanding of the thermal transfer through nonbonded contacts, a non-linear regression model is developed between inter-residue thermal conductivity values and static and dynamical properties of proteins by using the machine learning based approach of random forest methods.

1.7 Knowledge gap

1.7.1 The anisotropic and non-uniform flow of thermal energy

Thermodynamic and other environmental factors, such as temperature^{67–72} as well as protein-solvent interactions,^{73–78} govern the structure, dynamics, and function of proteins. Although thermal transport is one of the fundamental biophysical characteristics of proteins, it is still unclear how heat flow interacts with the molecular mechanism of protein activity. The structure of a folded protein exhibits significant heterogeneity, resulting in anisotropic and non-uniform thermal energy flow during conformational fluctuations. To gain insights into the site-specific thermal transport properties of proteins, the development of a theoretical method would be valuable. A possibility is to introduce a concept of “residue-wise thermal conductivity”, which allows us to analyze protein thermal transport at the level of individual amino acid residues. However, several challenges arise when implementing such an approach. Firstly, the small spatial scale occupied by each amino acid residue within the protein matrix introduces boundary effects that may significantly influence the residue-wise thermal conductivity. These effects cannot be neglected and must be considered in the analysis. Additionally, the local conformational fluctuations within a residue may not be independent of those occurring in the surrounding amino acid residues. This interdependence poses further complexities in accurately characterizing the site-specific thermal transport properties. To overcome these problems, we have developed a theoretical model, *linear-homopolymer-like model*, which incorporates the influence of cross

correlations effectively into the site-selective heat current analysis based on the autocorrelation function formalism. With this model, we aim to enhance our understanding of how thermal energy is transmitted and distributed at the residue level within protein interior.

1.7.2 Determinates of protein energy transport

Proteins are complex biomolecules composed of amino acids that are connected by peptide bonds. These molecules exhibit a hierarchical organization, including primary (linear sequence of amino acids), secondary (local patterns such as α helices or β strands), tertiary (overall folding of the protein), and quaternary (association of multiple protein subunits) structures. Within this heterogeneous and intricate system, residues participate in various interactions, including bonded interactions and nonbonded interactions. Bonded interactions such as peptide and disulfide bonds in proteins keep the molecular topology, i.e., the pattern of connectivity of atoms, while nonbonded interactions such as ionic bonds (e.g. salt bridges between oppositely charged residues), and hydrogen bonds (e.g. interactions between backbone atoms or polar side chains) are responsible for maintaining and stabilizing the secondary, tertiary and quaternary structures of proteins. Van der Waals interactions in proteins are a type of weak nonbonded interaction that occur between atoms that are close to each other but not covalently bonded. They mainly contribute to the stabilization of tertiary and quaternary structures of proteins by optimizing the surface contact between residues and subunits. Other types of nonbonded interactions in proteins include electrostatic interactions (such as those between charged residues or between dipoles), π - π stacking (such as those between aromatic residues), etc, also play important roles in protein folding, stability, function, and interactions.

During thermal fluctuation and vibrational energy relaxation of proteins, transfer of kinetic or potential energy occurs from one part of the protein to another or between protein and the surrounding environment. Different types of interactions can affect the rate and efficiency of energy transfer in proteins. For example, the peptide bond not only keeps the primary structure intact, but also provides major pathways for energy transfer. Nonbonded native contacts, found in the folded protein, stabilize the tertiary structure and play important roles in protein functions. For instance, we expect that a protein molecule, as a biomolecular machine, should experience recognition of native contacts especially at the regulation site and the active site, in response to external stimulus, such as ligand binding at light illumination. Accordingly, we also expect possible alteration of energy transfer pathways induced by the stimulus.

Using non-equilibrium MD simulations, the significance of different types of nonbonded contacts and the competition with the backbone have been studied in several proteins. A computational study showed the dominating energy transfer efficiency of the backbone with a much higher energy diffusion constant of $1.25 \text{ nm}^2 \text{ ps}^{-1}$ in HP36 than through nonbonded contacts with much smaller diffusion constants (1.1×10^{-4} for polar contacts and 3.6×10^{-7} for nonpolar contacts).¹² In a joint experimental and computational study, the hydrogen bonds are found to be the dominant energy transfer pathway in β -hairpin fold.⁶ On the contrary, a series of experimental findings on the heme proteins clearly demonstrated that the primary pathway for energy transfer from the heme group to the protein does not involve the covalent bond between the heme and the protein (Trp residue). Instead, the predominant route involves atomic

interactions between the heme group and the Trp residue.^{3,4,9} To quantitatively access the competition of different pathways (backbone and nonbonded contacts) and different types of interactions, it would be helpful to develop a method capable of measuring the heat transfer ability between residue and residue.

1.7.3 Factors affecting heat transfer through nonbonded contacts

For structural property, David M. Leitner's group found that the correlations of energy transfer rate are linearly correlated with the mean-square distance for charged contacts between charged residue pairs in myoglobin.³⁴ For dynamics, they also discovered that the energy transfer rate has a good correlation with the variance in the length of the contact for polar and hydrophobic contacts.^{12,30,31,34,79-82} The hydrogen bonds have been found to facilitate the thermal transport in proteins and protein-based materials. For instance, the thermal conductivity of π -helices, which have a stronger hydrogen bond strength, is two folds and three folds higher than that of α - and 3_{10} - helices, respectively.⁸³ Hydrogen bonding has been identified as the main contributor to the increase in the thermal conductivity of spider silk protein.⁸⁴ It has been reported that the thermal conductivity of spider silk protein is 1 - 2 orders of magnitude higher than that of globular proteins. Furthermore, hydrogen bonding has been shown to facilitate dominant vibrational energy transfer pathways in β -hairpin structures, competing with energy transport along the backbone.⁶ The scaling rules for rates of vibrational energy transfer in proteins have been extensively studied in various proteins, including HP36^{12,80}, myoglobin³⁴, deoxy-HbI/oxy-HbI^{30,79,81}, A2A adenosine receptor⁸¹, and GPCR proteins^{31,82}, by David M. Leitner and his colleagues. These scaling rules have been validated and found to work particularly well for hydrogen bonding contacts, especially those within a distance of less than 2.8 Å. However, given the importance of van der Waals interactions, especially in the hydrophobic core of water-soluble proteins, it would be helpful to consider a different approach as well. Due to the complexity of protein systems, the thermal energy transport through contacts should be affected by a number of both static and dynamic properties, including their structure, compositions, dynamics, and interaction types. Thus, the relationship between heat and energy transport efficiency through contacts and these protein properties is expected to be non-linear, rather than linear.

1.8 Organization of the thesis

In this dissertation, we developed a theoretical framework for analyzing the local thermal transport properties based on the autocorrelation function formalism and applied it to two protein systems to study the heat, energy, and signaling flow. In addition, due to the complexity of protein systems, advanced machine learning-based methods were utilized to identify the contributing factors of thermal transport in protein and the structural characterization of intrinsically disordered regions.

Chapter 1 provides an introduction, including an overview of current studies and methods related to vibrational energy transport in proteins and protein solutions. It, then, introduces the concepts of linear response theory and molecular dynamics simulation techniques. It highlights the knowledge gap in understanding vibrational energy transport in proteins and protein solutions and emphasizes the significance of applying machine learning algorithms to

address the protein-related problems.

Chapter 2 details the methodology used to investigate thermal transport properties of proteins: the calculations process of the overall thermal conductivity of proteins and the formalism for the site-selective heat current analysis. It also introduces a novel integrated approach that combines machine learning algorithms, MD simulation techniques, and SAXS and EPR/DEER experiments for the structural characterization of intrinsically disordered protein complexes.

Chapter 3 presents a further theoretical development for the analysis of local thermal transport properties using the autocorrelation function formalism with a model, linear-homopolymer-like model, applied to a small α -helical protein known as the villin headpiece subdomain (HP36) with a special emphasis on the role of heat current along the polypeptide chain. Equilibrium molecular dynamics simulations were employed to study the thermal behavior of HP36 within this framework. This chapter provides detailed insights into the local thermal transport phenomena occurring in HP36 and establishes a foundation for further research in this area.

Chapter 4 focuses on the roles of nonbonded contacts in HP36 proteins. The concept of inter-residue thermal conductivity was introduced as a measure of heat transfer between pairs of residues. The competition of different types of interactions heat transfer is discussed. In addition, the machine learning approach of random forest regression was used to explore the non-linear relationship between thermal conductivity and static/dynamic properties of proteins.

Chapter 5 investigates the role of energy transport in signaling mechanism of protein. Two forms of oxygen sensor proteins exhibit totally different activities, inactive and active forms, despite minimal structural differences. The signaling pathways upon ligand binding and allosteric effects were identified by using the energy transport network model.

In Chapter 6, machine learning methods were integrated with molecular dynamics simulations and experimental results to characterize the conformational ensemble of a new type of chloroplast protein (CP12) with intrinsically disordered regions.

Finally, Chapter 7 provides a comprehensive summary of the conclusions drawn from this thesis. It highlights the key findings and insights gained from the study of protein thermal transport properties and the application of machine learning algorithms. Additionally, this chapter explores future research directions and suggests areas for further investigation.

Chapter 2

Theory, Materials and Methods

2.1 Thermal conductivity

According to the Green–Kubo relations, the steady-state transport coefficient can be obtained by dividing the space–time integral of the flux–flux equilibrium correlation function by $k_B T$. A very detailed derivation process can be found in APPENDIX B: *Derivation of Green–Kubo Relation* of book *Heat transfer physics* by Massoud Kaviani.⁴² Like other transport processes, e.g. viscosity, the thermal conductivity of the material can be derived from Green–Kubo relation and expressed in terms of the time autocorrelation function (ACF) of the heat current vector, \mathbf{h} , based on equilibrium fluctuation,

$$\lambda = \frac{1}{3Vk_B T^2} \int_0^\infty \langle \mathbf{h}(t) \cdot \mathbf{h}(0) \rangle dt, \quad (2.1)$$

where λ is the thermal conductivity, V is the volume, k_B is the Boltzmann constant, T is the absolute temperature, the angle brackets of $\langle \mathbf{h}(t) \cdot \mathbf{h}(0) \rangle$ denote ensemble average.³⁶

Autocorrelation is a measure of how well a dynamic system’s signature aligns with a time-shifted version of itself, representing the cross-correlation of a signal with its own shifted counterpart. In the case of thermal conductivity, the duration of autocorrelation is related to the material’s ability to transfer heat and energy. Materials with high thermal conductivity exhibit long-lasting correlation, indicating slow dissipation of fluctuations from equilibrium. In contrast, materials with low thermal conductivity display short-lived correlation. One important point is that the Green-Kubo approach allows for the calculation of transport properties from an equilibrium system, providing valuable insights even in non-equilibrium scenarios.

2.2 Derivation of inter-atomic heat current

In this section, we have obtained an atomistic formulation for the heat current term of thermal conductivity expression (Equation 2.1) in a molecular system. The atomistic representation of the instantaneous heat current, denoted as \mathbf{h} , is derived by taking the time derivative of the energy distribution,⁵⁹

$$\mathbf{h} \equiv \frac{d}{dt} \sum_{i=1}^N (E_i \mathbf{r}_i) = \sum_{i=1}^N \left(E_i \frac{d\mathbf{r}_i}{dt} + \mathbf{r}_i \frac{dE_i}{dt} \right) \quad (2.2)$$

, where E_i , \mathbf{r}_i are per atom energy and the position vector of atom i , and N is the total number of atoms.

The on the right-hand side of Equation 2.2 can be divided into two terms: the convective term and the virial term. In the context of this equation, it is commonly recognized that the convective term plays a significant role in gaseous systems, but its contribution is minimal in solids. On the other hand, the virial term is predominant in solids and biomolecular materials, such as proteins⁸⁵.

The total energy (E) of a protein system, shown in Equation 2.3, in classical molecular mechanics is calculated by summing the kinetic energy contributions from each atom, which is determined by its mass (m_i) and momentum (\mathbf{p}_i), and the potential energy term ($V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$) that depends on the positions of all atoms ($\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$).

$$E = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N), \quad (2.3)$$

Alternatively, the potential energy term ($V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$) of a protein system in classical molecular mechanics can also be expressed as a function of the inter-atomic distances ($r_{ij} = |\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j|$) between all pairs of atoms (i, j). In this representation, the force acting on atom i is obtained by taking the partial derivative of V with respect to the position of atom i .

$$\mathbf{F}_i = - \sum_{(k,j)} \frac{\partial V}{\partial r_{kj}} \nabla_i(r_{kj}) = \sum_{j(\neq i)} - \frac{\partial V}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} = \sum_{j(\neq i)} \mathbf{F}_{ij} \quad (2.4)$$

, where $\nabla_i = (\partial/\partial\{\mathbf{r}_i\}_x, \partial/\partial\{\mathbf{r}_i\}_y, \partial/\partial\{\mathbf{r}_i\}_z)$.⁵⁷

The time derivatives of the potential energy (V) and the total energy (E) can be expressed as:

$$\frac{dV}{dt} = \sum_{(i,j)} \frac{\partial V}{\partial r_{ij}} \frac{dr_{ij}}{dt} = -\frac{1}{2} \sum_i^N \sum_j^N \mathbf{F}_{ij} \cdot (\mathbf{v}_i - \mathbf{v}_j) \quad (2.5)$$

and

$$\frac{dE}{dt} = \sum_i^N \mathbf{v}_i \cdot \mathbf{F}_i + \frac{dV}{dt} = \sum_i^N \sum_j^N \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) = \sum_i \frac{dE_i}{dt} \quad (2.6)$$

respectively, where E_i denotes the per atom energy of atom i . While it is possible to calculate the time derivative of E_i using the formula provided, obtaining the explicit form of E_i itself is not feasible in practice.

As a result, the total heat current can be obtained through the following derivation process:

$$\begin{aligned} \mathbf{h} &= \sum_i^N \mathbf{r}_i \frac{dE_i}{dt} = \sum_i^N \sum_j^N \mathbf{r}_i \left\{ \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \right\} \\ &= \frac{1}{2} \left[\sum_i^N \sum_j^N \mathbf{r}_i \left\{ \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \right\} + \sum_j^N \sum_i^N \mathbf{r}_j \left\{ \frac{1}{2} \mathbf{F}_{ji} \cdot (\mathbf{v}_j + \mathbf{v}_i) \right\} \right] \\ &= \frac{1}{2} \sum_i^N \sum_j^N (\mathbf{r}_i - \mathbf{r}_j) \left\{ \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \right\} \\ &= \sum_{(i,j)} (\mathbf{r}_i - \mathbf{r}_j) \left\{ \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \right\} \equiv \sum_{(i,j)} \mathbf{h}_{ij} \end{aligned} \quad (2.7)$$

, where $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$.

The time derivative of E_i thus can be expressed as

$$\frac{dE_i}{dt} = \sum_{j(\neq i)}^N \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j), \quad (2.8)$$

, where \mathbf{F}_{ij} is the force acting on atom i due to atom j .

Thus, we obtained atomistic expression for the heat current as

$$\mathbf{h} = \sum_{i=1}^N \mathbf{r}_i \frac{dE_i}{dt} = \sum_i^N \mathbf{r}_i \sum_j^N \left\{ \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \right\}, \quad (2.9)$$

with the inter-atomic heat current between atoms i and j , \mathbf{h}_{ij} , denoted as

$$\mathbf{h}_{ij} = (\mathbf{r}_i - \mathbf{r}_j) \left\{ \frac{1}{2} \mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \right\}. \quad (2.10)$$

2.3 Linear-homopolymer-like model

The inter-atomic heat current, as described in Equation 2.10, provides insight into the essential biophysical nature of a protein molecule, i.e. its structure ($\mathbf{r}_i - \mathbf{r}_j$), dynamics ($\mathbf{v}_i + \mathbf{v}_j$) and interactions (\mathbf{F}_{ij}). We expect that the local thermal transport of a highly non-uniform molecule such as proteins can be effectively illustrated by analyzing the ACF

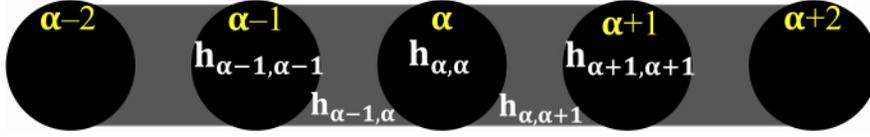


Figure 2.1: Thermal transport through backbone model. Local heat currents, $h_{\alpha,\alpha}$ and $h_{\alpha,\alpha+1}$, occur within each residue α and between each pair of residues α and $\alpha + 1$.

of the local heat current within individual residues and between pairs of residues. It is well-known that amino acids serve as the fundamental building blocks of proteins. Therefore, it would be advantageous if we could investigate the local thermal transport characteristics on a residue-by-residue basis. The spatial scale of each residue is considerably smaller than the macroscopic level, and the local fluctuations within each residue can be influenced by the surrounding protein environment. As a result, we need a special care when discussing the local thermal transport property using the ACF formalism.

To address this issue, we have devised a two-stage strategy. In the first stage, we have introduced a simple model (Figure 2.1) that focuses on the thermal transport along the polypeptide chain. This simple model allows us to analyze and understand the local thermal transport phenomena more effectively. In the second stage (Chapter 2.4), we have carefully examined the theoretical consistency between the ACF formalism based on the local heat current and the ACF formalism based on the overall heat current. As a result, we have derived correction terms that ensure the accuracy of the local heat current analysis.

In our previous study,⁵⁹ we found that the thermal transport between the entire mainchain and each sidechain exhibited negligible magnitude and showed no significant positional dependence. Additionally, our observations revealed that the thermal transport in a protein molecule predominantly takes place either along the polypeptide chain or within individual sidechains. To better understand the thermal transport properties at the residue level, we introduced a model comprising N nodes representing the residues, as depicted in Figure 2.1. Here, N denotes the total number of residues in the protein. Within this model, we assumed the occurrence of heat currents within each residue as well as between adjacent residue pairs along the sequence, while long-range heat currents between residue pairs in nonbonded contacts were neglected. Then, the total heat current can be represented as a summation of the partial heat currents:

$$\mathbf{h} = \sum_{\alpha=1}^N \mathbf{h}_{\alpha,\alpha} + \sum_{\alpha=1}^{N-1} \mathbf{h}_{\alpha,\alpha+1} \quad (2.11)$$

Here, $h_{\alpha,\alpha}$ represents the intra-residue heat current within residue α , which is the sum of heat currents between all pairs of atoms i and j within a single residue ($\sum_i^{n_\alpha} \sum_j^{n_\alpha} h_{ij}$). Similarly, $h_{\alpha,\alpha+1}$ represents the inter-residue heat current between a pair of adjacent residues α and $\alpha + 1$, calculated as the sum of heat currents between all atoms i in residue α and all atoms j in residue $\alpha + 1$ ($\sum_i^{n_\alpha} \sum_j^{n_{\alpha+1}} h_{ij}$).

In order to characterize the residuewise thermal transport property, we introduced several quantities: $\Lambda_{\alpha,\alpha}$, $\Lambda_{\alpha,\alpha+1}$, and Λ . These quantities are defined as the ACFs of the partial heat currents within a residue ($\Lambda_{\alpha,\alpha}$), between adjacent

residues ($\Lambda_{\alpha,\alpha+1}$), and for the total heat current (Λ), respectively,

$$\Lambda_{\alpha,\alpha} = \int \langle \mathbf{h}_{\alpha,\alpha}(t) \cdot \mathbf{h}_{\alpha,\alpha}(0) \rangle dt \quad (2.12)$$

$$\Lambda_{\alpha,\alpha+1} = \int \langle \mathbf{h}_{\alpha,\alpha+1}(t) \cdot \mathbf{h}_{\alpha,\alpha+1}(0) \rangle dt$$

$$\Lambda = \int \langle \mathbf{h}(t) \cdot \mathbf{h}(0) \rangle dt \quad (2.13)$$

The contribution of the partial heat current, $\mathbf{h}_{\alpha,\beta}$, to the overall thermal transport property of the molecule can be expressed using a contribution factor, denoted as c ,

$$c_{\alpha,\beta} \equiv \frac{\Lambda_{\alpha,\beta}}{\Lambda} \quad (2.14)$$

2.4 Cross-correlation correction on linear-homopolymer-like model

To illustrate why we need to take into account the cross-correlation in the linear-homopolymer-like model, we use a hypothetical dipeptide composed of residues A and B as an example. The thermal conductivity, denoted as λ , of the molecule can be determined by calculating the heat current (\mathbf{h}) autocorrelation function, $\langle \mathbf{h}(0) \cdot \mathbf{h}(t) \rangle$. Based on Equation 2.10, \mathbf{h} can be expressed as the sum of three components: \mathbf{h}_A , \mathbf{h}_B , and \mathbf{h}_{AB} . The first two components represent the heat currents occurring within residues A and B, respectively, while the third component, \mathbf{h}_{AB} , corresponds to the heat current between residues A and B.

Considering the non-uniform heat flow in the dipeptide, we need to calculate the local thermal conductivities, denoted as λ_A , λ_B , and λ_{AB} , independently for each of the three parts. This can be achieved by calculating the heat current autocorrelation functions $\langle \mathbf{h}_A(0) \cdot \mathbf{h}_A(t) \rangle$, $\langle \mathbf{h}_B(0) \cdot \mathbf{h}_B(t) \rangle$, and $\langle \mathbf{h}_{AB}(0) \cdot \mathbf{h}_{AB}(t) \rangle$. In an ideal case where \mathbf{h}_A , \mathbf{h}_B , and \mathbf{h}_{AB} fluctuate independently, we can neglect the cross-correlation terms between them. Thus, we have $\langle \mathbf{h}(0) \cdot \mathbf{h}(t) \rangle = \langle \mathbf{h}_A(0) \cdot \mathbf{h}_A(t) \rangle + \langle \mathbf{h}_B(0) \cdot \mathbf{h}_B(t) \rangle + \langle \mathbf{h}_{AB}(0) \cdot \mathbf{h}_{AB}(t) \rangle$. Finally, we can obtain the value of λ by combining the calculated values of λ_A , λ_B , and λ_{AB} .

In reality, however, strong interactions exist between residues A and B through the peptide bonding, leading to non-negligible cross-correlations. These cross-correlations are represented by $\langle \mathbf{h}_A(0) \cdot \mathbf{h}_B(t) \rangle$, $\langle \mathbf{h}_A(0) \cdot \mathbf{h}_{AB}(t) \rangle$, and $\langle \mathbf{h}_B(0) \cdot \mathbf{h}_{AB}(t) \rangle$. As a result, the value of λ cannot be accurately reconstructed using only λ_A , λ_B , and λ_{AB} . In this analysis, it is essential to consider the influence of these cross-correlations in order to obtain effective local thermal conductivities that accurately characterize the non-uniform thermal transport property of proteins. Thus, by incorporating the cross-correlations, we aim to derive the mathematical expression of the local thermal conductivities that are consistent with the overall thermal conductivity of the entire molecule.

The analysis of the local thermal transport property using Equation 2.12 leads to a problem: when dividing

the overall heat current into the summation of partial heat currents (Equation 2.11), Equation 2.13 includes both autocorrelation and cross-correlation terms, while the cross-correlation terms, $\langle \mathbf{h}_{\alpha,\beta}(0) \cdot \mathbf{h}_{\alpha',\beta'}(t) \rangle$ with $(\alpha, \beta) \neq (\alpha', \beta')$, are missing in Equation 2.12. As a result, we anticipate that the summation of all contribution factors, $\sum_{\alpha=1}^N c_{\alpha,\alpha} + \sum_{\alpha=1}^{N-1} c_{\alpha,\alpha+1}$, does not equal unity. If this summation is greater (or less) than 1, it implies that the average intensity of the local thermal transport is overestimated (or underestimated) compared to its actual value.

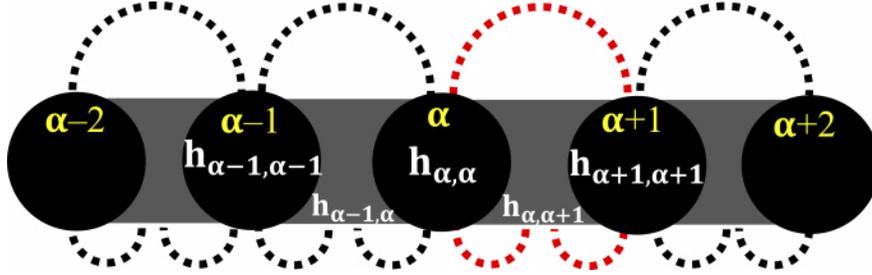


Figure 2.2: Short-range cross-correlation approximation. Cross-correlations are considered only for nearby pairs (dashed line).

To address this issue, we explored the role of cross-correlations in this section. We made the assumption that the local fluctuation of thermal transfer at a particular site in a protein is not influenced by that at another distant site. Based on this assumption, we introduced a short-range cross-correlation approximation, shown in Figure 2.2. In order to evaluate these cross-correlations, it is useful to define a “dimer” as a pair of adjacent residues, α and $\alpha + 1$ ($\alpha = 1, \dots, N - 1$), and introduce the concept of intra-dimer heat current, denoted as $\mathbf{h}_{\{\alpha,\alpha+1\},\{\alpha,\alpha+1\}}$. The cross-correlation between residues α and $\alpha + 1$, denoted as ξ , can then be expressed as shown by the red dashed line in Figure 2.2,

$$\xi_{\alpha,\alpha+1} = \Lambda_{\{\alpha,\alpha+1\},\{\alpha,\alpha+1\}} - (\Lambda_{\alpha,\alpha} + \Lambda_{\alpha,\alpha+1} + \Lambda_{\alpha+1,\alpha+1}) \quad (2.15)$$

$$\Lambda_{\{\alpha,\alpha+1\},\{\alpha,\alpha+1\}} = \int \langle \mathbf{h}_{\{\alpha,\alpha+1\},\{\alpha,\alpha+1\}}(t) \cdot \mathbf{h}_{\{\alpha,\alpha+1\},\{\alpha,\alpha+1\}}(0) \rangle dt \quad (2.16)$$

Next, we derived correction terms for $\Lambda_{\alpha,\alpha}$ and $\Lambda_{\alpha,\alpha+1}$ in order to obtain their corrected counterparts, $\tilde{\Lambda}_{\alpha,\alpha}$ and $\tilde{\Lambda}_{\alpha,\alpha+1}$. The goal was to ensure that the summation of these corrected terms, $\tilde{\Lambda}_{\alpha,\alpha}$ ($\alpha = 1, \dots, N$) and $\tilde{\Lambda}_{\alpha,\alpha+1}$ ($\alpha = 1, \dots, N - 1$), approaches the value of Λ . In other words, we aimed to achieve the condition:

$$\Lambda \simeq \sum_{\alpha=1}^N \tilde{\Lambda}_{\alpha,\alpha} + \sum_{\alpha=1}^{N-1} \tilde{\Lambda}_{\alpha,\alpha+1}, \quad (2.17)$$

Based on the short-range cross-correlation assumption, we assumed that the influence of $\xi_{\alpha,\alpha+1}$ is limited to $\Lambda_{\alpha-1,\alpha}$, $\Lambda_{\alpha,\alpha}$, $\Lambda_{\alpha,\alpha+1}$, $\Lambda_{\alpha+1,\alpha+1}$, and $\Lambda_{\alpha+1,\alpha+2}$ (see Figure 2.3), we derived a model where the contribution of $\xi_{\alpha,\alpha+1}$ to $(\tilde{\Lambda}_{\alpha-1,\alpha}$ and $\tilde{\Lambda}_{\alpha+1,\alpha+2})$, $(\tilde{\Lambda}_{\alpha,\alpha}$ and $\tilde{\Lambda}_{\alpha+1,\alpha+1})$, and $(\tilde{\Lambda}_{\alpha,\alpha+1})$ is represented by $u\xi_{\alpha,\alpha+1}$, $v\xi_{\alpha,\alpha+1}$, and $w\xi_{\alpha,\alpha+1}$,

$$\mathbf{T} = \frac{w}{2}\mathbf{I} + v\mathbf{T}_1 + u\mathbf{T}_2, \quad (2.22)$$

, where \mathbf{I} is the $N \times N$ identity matrix.

The values of u, v, w were derived as follows. First, the total weight of $\xi_{\alpha, \alpha+1}$ in Equation 2.17 should be one, as shown in Figure 2.3. Therefore, we have the following equation:

$$2u + 2v + w = 1 \quad (2.23)$$

Next, we expect that the influence of $\xi_{\alpha, \alpha+1}$ on $\Lambda_{\alpha, \alpha+1}$ can be greater than that of either $\xi_{\alpha-1, \alpha}$ or $\xi_{\alpha+1, \alpha+2}$, and we introduce an arbitrary assumption:

$$2u = w \quad (2.24)$$

Also, we assume that the total weight of $\xi_{\alpha-1, \alpha}, \xi_{\alpha, \alpha+1}, \xi_{\alpha+1, \alpha+2}$ in $\tilde{\Lambda}_{\alpha, \alpha+1}$ is equal to the total weight of $\xi_{\alpha, \alpha+1}, \xi_{\alpha+1, \alpha+2}$ in $\tilde{\Lambda}_{\alpha+1, \alpha+1}$ for $\alpha = 2, \dots, N-2$, then we can express the relation as follows:

$$2u + w = 2v \quad (2.25)$$

From Equation 2.23, Equation 2.24, and Equation 2.25, we obtain $u = \frac{1}{8}, v = w = \frac{1}{4}$. Accordingly, Equation 2.18 and Equation 2.19 become $\tilde{\Lambda}_{\alpha, \alpha} = \Lambda_{\alpha, \alpha} + \frac{1}{4}(\xi_{\alpha-1, \alpha} + \xi_{\alpha, \alpha+1})$, and $\tilde{\Lambda}_{\alpha, \alpha+1} = \Lambda_{\alpha, \alpha+1} + \frac{1}{4}\xi_{\alpha, \alpha+1} + \frac{1}{8}(\xi_{\alpha-1, \alpha} + \xi_{\alpha+1, \alpha+2})$, respectively.

For the terminal residues, we assume that

$$\tilde{\Lambda}_{1,1} = \Lambda_{1,1} + v'\xi_{1,2} \quad (2.26)$$

$$\tilde{\Lambda}_{1,2} = \Lambda_{1,2} + w'\xi_{1,2} + u\xi_{2,3} \quad (2.27)$$

$$\tilde{\Lambda}_{N-1,N} = \Lambda_{N-1,N} + u\xi_{N-2,N-1} + w'\xi_{N-1,N} \quad (2.28)$$

$$\tilde{\Lambda}_{N,N} = \Lambda_{N,N} + v'\xi_{N-1,N} \quad (2.29)$$

Then, it is necessary for the total weight of $\xi_{1,2}(\xi_{N-1,N})$ in $\tilde{\Lambda}_{1,1}, \tilde{\Lambda}_{1,2}, \tilde{\Lambda}_{2,2}, \tilde{\Lambda}_{2,3}$ ($\tilde{\Lambda}_{N,N}, \tilde{\Lambda}_{N-1,N}, \tilde{\Lambda}_{N-1,N-1}, \tilde{\Lambda}_{N-2,N-1}$) to be equal to one. In other words, we have the following equation:

$$v' + w' + v + u = 1. \quad (2.30)$$

Additionally, we assume that the weight v' is greater than the weight u' , and we introduce the following relation:

$$w' = 2v'. \quad (2.31)$$

As a result, we obtain the values of v' and w' as $\frac{5}{24}$ and $\frac{5}{12}$ respectively, based on equations Equation 2.30 and Equation 2.31.

In summary, the values of u, v, w in Equation 2.22 are $\frac{1}{8}, \frac{1}{4}, \frac{1}{4}$, respectively. To account for the terminal effects, we separately evaluated $\tilde{\Lambda}_{1,1}, \tilde{\Lambda}_{1,2}, \tilde{\Lambda}_{N-1,1}$, and $\tilde{\Lambda}_{N,N}$ using Equation 2.18 and Equation 2.19 with values of v' and w' as $\frac{5}{24}$ and $\frac{5}{12}$ respectively, instead of using $v = w = \frac{1}{4}$.

We defined the contribution factor after cross-correlation correction as follows,

$$\tilde{c}_{\alpha,\beta} \equiv \frac{\tilde{\Lambda}_{\alpha,\beta}}{\Lambda} \quad (2.32)$$

All of the heat currents and their autocorrelation function calculations were performed using by our CURP program, version 1.3⁵⁹ based on 50 trajectories of *NVE* molecular dynamics simulations, each of which started from the last step of different sampling trajectory in isothermal-isobaric *NPT* ensemble at $T = 300$ K and $P = 0.987$ atm.

2.5 Inter-residue heat current analysis

Considering the molecular functional importance of nonbonded native contacts in a folded protein, it would be helpful to analyze the thermal energy transfer through such native contacts.

Based on the inter-atomic heat current between atoms i and j in a molecule, h_{ij} , the heat current between residue α and β can be expressed as:

$$h_{\alpha,\beta} = \sum_{i \in \alpha} \sum_{j \in \beta} h_{ij} \quad (2.33)$$

By using the following quantity:

$$\Lambda_{\alpha,\beta} = \int \langle h_{\alpha,\beta}(t) \cdot h_{\alpha,\beta}(0) \rangle dt \quad (2.34)$$

the inter-residue thermal conductivity, $\lambda_{\alpha,\beta}$, can be expressed as,

$$\lambda_{\alpha,\beta} = \frac{1}{3(V_\alpha + V_\beta)k_B T^2} \int \langle \mathbf{h}_{\alpha,\beta}(t) \cdot \mathbf{h}_{\alpha,\beta}(0) \rangle dt = \frac{\Lambda_{\alpha,\beta}}{3(V_\alpha + V_\beta)k_B T^2} \quad (2.35)$$

where V_α (V_β) is the volume of residue α (β), k_B is the Boltzmann constant, T is the absolute temperature.

2.6 Inter-residue energy flow analysis

The atomistic representation of the instantaneous energy flow (J) between two atoms in a molecule, i and j , can be expressed by:

$$J_{ij} = \frac{1}{2}(\mathbf{v}_i \cdot \mathbf{F}_{ij} - \mathbf{v}_j \cdot \mathbf{F}_{ji}) = \frac{1}{2}\mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j) \quad (2.36)$$

, where J_{ij} is the inter-atom energy flow between atom i and atom j .

Then, the energy flow between residue α and β can be expressed as:

$$J_{\alpha,\beta} = \sum_{i \in \alpha} \sum_{j \in \beta} J_{ij} \quad (2.37)$$

Here, we introduced a transport coefficient-like quantity, *energy conductivity*, denoted as $G_{\alpha,\beta}$, to measure the amount of energy transferred per unit of time between two residues through native contacts.

$$G_{\alpha,\beta} = \int \langle J_{\alpha,\beta}(t) J_{\alpha,\beta}(0) \rangle dt \quad (2.38)$$

From Equation ?? and Equation 2.35, we can see that the relationship between atomistic heat current and energy flux is:

$$\mathbf{h}_{ij} = (\mathbf{r}_i - \mathbf{r}_j) J_{ij} \quad (2.39)$$

In this study, the integration time of the autocorrelation function for both heat current and energy flux was set as 60 ps. All calculations of $\lambda_{\alpha,\beta}$ and $G_{\alpha,\beta}$ between each pair of residues in native contact using the *CURrent calculations in Proteins* (CURP) program of version 1.3 developed by our lab.⁵⁹

2.7 Parameterization of the Fe³⁺ metal cofactor

To investigate the interplay between allosteric signal transduction and the thermal energy transfer in the oxygen sensor domain of *BjFixL* proteins. We conducted the MD simulation. The force-field parameters of the non-standard part

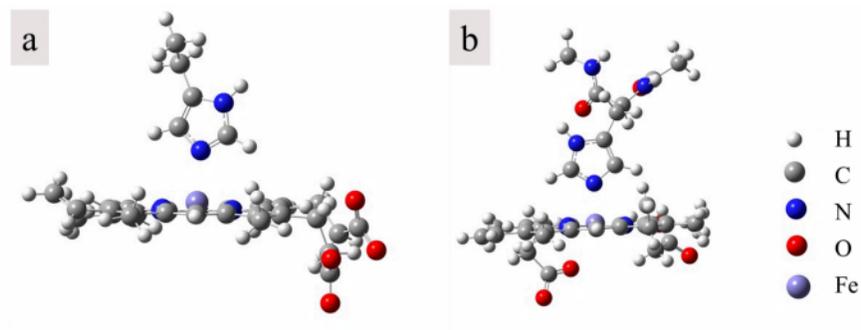


Figure 2.4: Molecules used to obtain the tuned force field parameters for met-FixLH dimer system. (a) Molecule used in Gaussian optimization calculations to obtain the Fe^{3+} metal associated bond and angle parameters. (b) Molecule used in Gaussian RESP partial charge calculations to obtain the partial charges for the metal site.

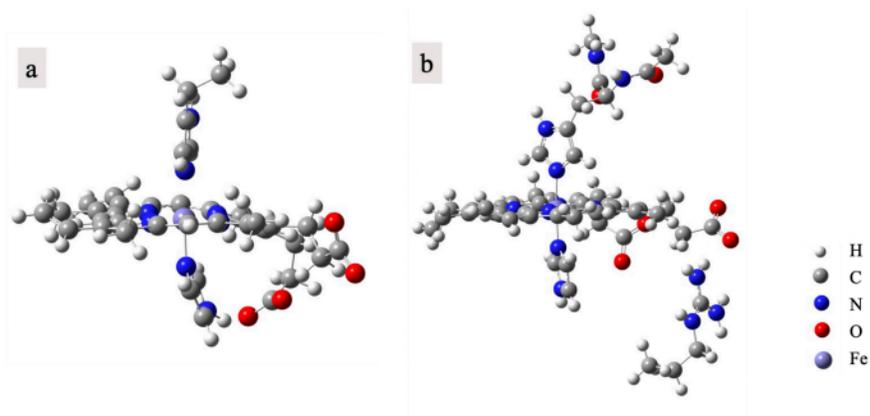


Figure 2.5: Molecules used to obtain the tuned force field parameters for met-FixLH-imd dimer system. (a) The molecule used in Gaussian optimization calculations to obtain the Fe^{3+} metal associated bond and angle parameters. (b) The molecule used in Gaussian RESP partial charge calculations to obtain the partial charges for the metal site.

of dimeric FixLHs were generated using MCPB.py module^{86,87} of AmberTools 15. A small model and a large model of two dimeric FixLH proteins generated from MCPB.py (Figure 2.4: met-FixLH, Figure 2.5: met-FixLH-imd) were used to obtain the Fe³⁺ metal associated bond and angle parameters and generate the atomic partial charges for the metal site, respectively. The small model consists of heme, the coordinated sidechains of the proximal H200 with/without ligand. In the large model, the coordinated sidechain has its free N- and C- termini capped with N-methyl (NME) and acetyl groups (ACE) to mimic the protein backbone. Optimization and force constant calculations for the small model and the Merz-Kollman RESP charge calculations for the large model were performed using Gaussian 16 package with the B3LYP/6-31G(d) level of theory^{88,89} and the ultrafine grid. Fe³⁺ with spin state $S=5/2$ and $S=1/2$ ⁹⁰ were used in Gaussian calculations for met-FixLH and met-FixLH-imd protein, respectively (Table 2.1).

Table 2.1: Quantum calculation results of met-FixLH and met-FixLH-imd. Hartree Energy (unit: eV) comparison among different spin states of met-FixLH and met-FixLH-imd.

	Spin = 2	Spin = 4	Spin = 6
Met-Heme	-3402.1853060	-3402.1959457	-3402.1976629
Imdazole-Heme	-3627.8693349		-3627.8189808

2.8 Equilibrium molecular dynamics simulations

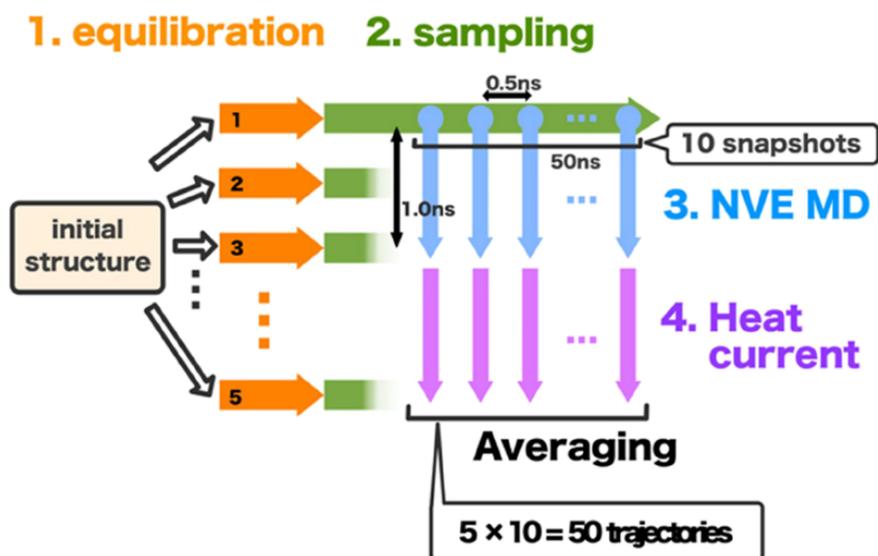


Figure 2.6: Computational procedure for HP36 protein system.

The procedures of MD simulations for heat current and energy flow calculations on HP36 (Figure 2.6) in Chapter 3 and Chapter 4, and on two *Bj*FixLH proteins (Figure 2.7) in Chapter 5 were described in this chapter. All molecular simulations in Chapter 3, Chapter 4, and Chapter 5 were performed by using Amber 19 package.⁹¹

2.8.1 Systems setup

(1) HP36 system

The structure of villin headpiece protein (HP36) was obtained from the protein data bank and its PDB code is 1VII, which was determined by the nuclear magnetic resonance (NMR) technique.⁹² To reproduce the vibrational energy transfer accurately along the polypeptide chain, we used a small time step, $\Delta t = 0.5$ fs, for MD simulations for generating time series of the heat current and energy flow. Thus, we keep the system size as small as possible in this study. The HP36 protein molecule was placed in an octahedral solvent box containing 2329 water molecules of the TIP3P model,^{93,94} using the LEaP program of AmberTools 19. All charged residues were considered in their standard protonation state at pH = 7.0. Amber ff19SB force field functions⁹⁵ were used for the protein atoms. Four chloride and two sodium ions were added to neutralize the simulation box. For efficient long-range electrostatic interaction calculations, the particle mesh Ewald (PME) method was used under periodic boundary conditions.

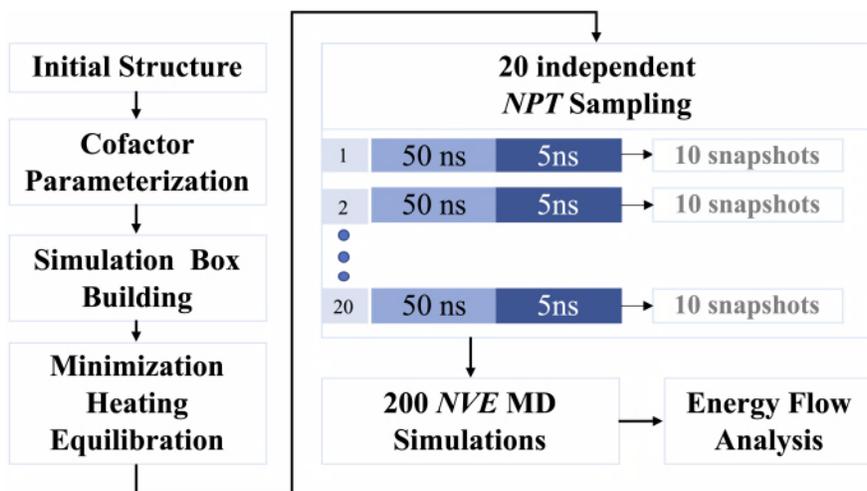


Figure 2.7: Computational procedure for *BjFixLH* protein systems.

(2) *BjFixL* systems

We constructed two models of ferric FixLH dimer, (a) the ligand-free state (met-FixLH) and (b) the imidazole-bound state (met-FixLH-imd), based on the X-ray crystallographic structure of FixL dimer. It is known that the deletion of the transmembrane region from *RmFixL*, which corresponds to the PAS-A domain of *BjFixL*, gives rise to no significant defect in the signaling functions.⁹⁶ Besides, the A' α and J α helices are involved in the formation of the dimer interface, and considering the potential influence of the *BjFixLH*s dimer interface on the signaling process. Thereby, we considered 142 residues from E128 to L269 in the subsequent MD simulations.

For both protein structures, the N- and C- termini were capped with uncharged ACE (acetyl) and NME (N-methylamine) groups, respectively. For both systems, we modeled the histidines H235 and H259 (H144, H150, H162, H200, H214) as neutral forms with each epsilon (delta) nitrogen protonated. All other residues were considered in their standard protonation state at pH = 7.0. The systems were solvated with the TIP3P⁹⁴ water molecules in a periodic cubic

box and sodium ions were used to neutralize the systems, then, additional Na and Cl ions were added to achieve a salt concentration of ~ 0.15 M. Each solvated system contains about $\sim 60,000$ atoms.

The Amber ff14SB force field⁹⁷ was used to model the standard residues of the proteins except for H200. Quantum calculations were performed to tune the force field to model the heme, ligands, and H200 and calculation details were described in Chapter 2.7. The long-range electrostatic interactions were treated with the particle mesh Ewald method⁹⁸ and nonbonded particle-particle interactions⁹⁹ were considered using a 9.0 \AA cutoff. The time step for all MD simulations was set as 2.0 fs and SHAKE was used to constrain hydrogens for production run MD simulations. To prevent the dissociation of α helices, a harmonic restraint with a spring constant of $10 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$ was applied to the bond between two CA atoms of residue I128 of chain A and chain B.

2.8.2 Minimization

The periodic boundary condition was applied to all simulation systems. Nonbonded particle-particle interactions⁴⁹ were taken into account using a distance cutoff of 9 \AA , and long-range electrostatic interactions were handled using the particle mesh Ewald (PME) method¹⁰⁰.

The energy minimization of simulation systems contains three steps, (1) fixing all the heavy atoms and only relaxing hydrogen atoms, (2) optimizing the side-chain atoms with positional restraints of $99.9 \text{ kcal}/(\text{mol} \text{ \AA}^2)$ on the backbone atoms (N, C α , C, O), (3) the positions of the main chain atoms were optimized with positional restraints of $2.0 \text{ kcal}/(\text{mol} \text{ \AA}^2)$. After that, a Maxwell-Boltzmann distribution of initial atomic velocities was generated at 0.1 K and each system was heated from 0.1 to 300 K for 50 ps by performing constant temperature, constant volume (*NVT*) MD simulations with the relaxation time of 0.1 ps and a positional restraint of $2.0 \text{ kcal}/(\text{mol} \text{ \AA}^2)$ on the backbone atoms. With the same positional restraints, a 50-ps *NVT* MD simulation was performed at $T = 300 \text{ K}$, followed by a 200 ps Langevin MD simulation at $T = 300 \text{ K}$ with a collision frequency of 2.0 ps^{-1} with positional restraints switched-off.

2.8.3 Equilibration

We started five (twenty) independent MD simulations from the minimum-energy conformation of HP36 (each of the *BjFixL*) system, with different Maxwell-Boltzmann velocity distributions at $T = 0.1 \text{ K}$, while keeping positional restraints imposed on the mainchain atoms. Subsequently, the system temperature was gradually increased from 0.1 to 300 K . We, then, continued constant temperature, constant volume (*NVT*) MD simulation at $T = 300 \text{ K}$ for 200 ps without positional restraints. The thermal equilibrium of the system was achieved by using the constant temperature, constant pressure (*NPT*) MD simulation at $T = 300 \text{ K}$, $P = 1 \text{ atm}$ for another 700 ps .

2.8.4 *NPT* sampling and *NVE* MD simulations

Next, for the HP36 system, we continued 5 independent *NPT* MD simulations for 56 ns , including a 50-ns run with a time step, Δt , of 2 fs with SHAKE constraints imposed on the bonds involving hydrogen atoms, 1-ns run with a time step of 0.5 fs with SHAKE constraints switched off only for the protein part. We, then, continued another 5-ns run

with the same condition for conformational sampling. During the last 5-ns trajectories of the five independent *NPT* MD simulations, we saved snapshots with atomic coordinates and velocities every 500 ps. We thus obtained 50 different initial conditions, each of which was used for the subsequent constant volume, constant energy (*NVE*) MD simulation for 1 ns with a time step of 0.5 fs. During these *NVE* MD simulations, we saved atomic velocities every 0.5 fs, while atomic coordinates were saved every 1 fs.

For *BjFixL* systems, 20 independent *NPT* MD simulations were performed for 55 ns at conditions of $T = 300$ K and $P = 0.978$ atm, to conduct the conformational samplings. From the last 5-ns trajectory of each *NPT* simulation, 10 snapshots with their atomic coordinates and velocity information were saved every 0.5 ns. The atomic coordinates and velocities of *NVE* MD simulations were saved every 10 fs for further calculations of energy flow and its autocorrelation function of residue pairs in native contacts.

2.9 Definition and classification of nonbonded contacts

2.9.1 Definition

To describe the intramolecular interaction networks in proteins, the protein contact networks (PCNs)^{101–103} have been introduced, aiming to unravel the relationship between protein structure and function. In PCNs, two residues whose alpha-carbon atoms (CA) are within 4 - 8 Å of each other are considered to have contacts and interactions. In this study, we defined nonbonded contacts based on a geometric criterion where residue pairs with the nearest atom distance of 6 Å or less at least in each *NVE* MD simulation trajectory were considered to have a significant role in heat and energy transport, as the thermal transport becomes negligible for pairs separated farther than this threshold. A total of 319 residue pairs in HP36 were selected using the *pickup-respairs* module of CURP 1.3.

2.9.2 Classification

The nonbonded contacts among residues are categorized into five types based on the geometric structure of HP36 and MD simulation trajectories: hydrogen bond contacts, π stacking contacts between aromatic side chains, electrostatic contacts for both residues that are charged or polar, and hydrophobic contacts for either of them being hydrophobic residue. Hydrogen bond contacts are identified from MD simulation trajectories by the *cpptraj* module of AmberTools 19, with a geometric criterion, i.e., the distance between atom X and atom Y from a pair of shorter than 3.0 Å and the angle X-H...Y or Y-H...X within 145-180°. Here, X(Y) \in residue A (B), and X(Y) is either O or N atom in protein. On the other hand, π stacking contacts are identified using the RING3.0 server (<https://ring.biocomputingup.it>) based on the NMR structure of HP36.

2.10 Hydrogen bond occurrence probability

The hydrogen bonds are found to facilitate the thermal transport in materials made of α -helices.^{21,83} To quantitatively evaluate the average strength of hydrogen bonds during thermal fluctuations, we assessed their integrity during the simulation process by calculating the lifetime of detected hydrogen bond with the quantity, hydrogen bond occurrence probability (P_{HB}), as defined below. The *cpptraj* module in AMBER was utilized to probe hydrogen bonds between residues and their lifetime in the simulation. The definition of hydrogen bonds and screening criterion are the same as those in Chapter 2.9. The hydrogen-bonding-occurrence probability between a pair of residues during a certain period of time can be calculated by the following formula:

$$P_{HB} = \frac{\sum_{i=1}^N n_i}{N} \quad (2.40)$$

where n_i is the number of hydrogen bonds formed between the pair, and N is the number of snapshots in the MD trajectory.

2.11 Random forest regression method

The random forest regression model was applied to predict the values of inter-residue thermal conductivity, $\lambda_{\alpha,\beta}$, for residue pairs (α, β), by using sklearn's *RandomForestRegressor* function in Python. The $\lambda_{\alpha,\beta}$ values were considered as the target variable and seven properties, each of which represents either static or dynamical feature of a protein, were considered as explanatory variables: inverse of contact (shortest) distance ($\langle d_c \rangle$), P_{HB} , the inverse of variance in the contact distance ($\langle \delta d_c^2 \rangle$), the summation of the volumes of contact residue pair ($V_{\alpha\beta}$), residue type (RT_α and RT_β), and interaction type (IT). The whole dataset was split into two groups for model fitting and evaluation: training set (80%) and testing set (20%). The mean absolute error (MAE) loss was calculated for the split quality measurement. The bootstrapping method was used to randomly sample the subsets of the training dataset to build the model. The random forest regression model (estimator) performance is affected by the choice of its hyperparameters, such as the number of decision trees and the maximum number of splits for each decision tree. To improve the accuracy of predictions and avoid underfit or overfit problems, the *RandomizedSearchCV* function in the scikit-learn library was used to tune the hyperparameters and to obtain the best estimator. R^2 and root-mean-squared error (RMSE) for the training data set and for the testing data set were calculated to evaluate the model performance. Because the explanatory variables were considered contain categorical data, such as, *interaction type* and *residue type*, we applied the *LabelEncoder* function of Python to give a unique integer to each categorical explanatory variable for regression analysis. The feature importance for all explanatory variables was evaluated using permutation feature importance rather than impurity-based feature importance to avoid misleading for high cardinality features.

For constructing random forest models, we used two types of datasets with $\langle d_c \rangle < 6 \text{ \AA}$ (*dataset L*) and that within $\langle d_c \rangle < 4 \text{ \AA}$ (*dataset S*). The distance criterion of 6 \AA is used as a default value by the CURP program so that the selected

residue pairs contain not only strongly interacting ones, but also the loosely interact to each other. Alternatively, the distance criterion of 4 Å is often used to extract weakly interacting residue pairs in constructing the network graph of nonbonded native contacts.¹⁰¹

2.12 Computational modeling for *Thalassosira pseudonana* CP12 homodimer

Although structure modeling of globular protein with its well-defined native conformation has become practical by using deep-learning algorithms based techniques, there still remains various obstacles to characterize the conformational ensemble of the intrinsically disordered protein (IDP) or the protein with highly flexible disordered regions (IDRs). To address this issue, we developed a three-stage modeling technique and applied it to *Thalassosira pseudonana* chloroplast protein (CP12) with IDRs. First, AlphaFold-Multimer was used to obtain the CP12 dimer's initial guess structures.^{66,104} Then, the harmonic restrained all-atom molecular dynamics (MD) simulations were used to roughly refine those structures so that they became consistent with the experimental results (SAXS¹⁰⁵ and EPR/DEER). Finally, restrained-ensemble molecular dynamics (reMD) simulations^{106–109} were employed to further refine those structures, ensuring that the distance distributions of all the spin pairs introduced in the reMD simulation became consistent with those of the experimental EPR/DEER data are consistent.

2.12.1 FASTA sequence of CP12

From the complete amino acid sequence of wild type of CP12, we used the segment of 163 residues (shown in blue) served as the input (WT) for AlphaFold2 modeling as follows:

```
MKIFLASLIGSCAAFAPAPFGKSPTALFGRVDTSAAIEAALDASKKFGSTSSEARVLWDIVEEMDASDNSVAS
KAPIVDSEYEAKVKLSLQMLTKTKAELDQVKALADDLKGVKLASPSVGSSAPDDSVMEALAAARAATEEFGQ
SSPQARLAWETVEEIAASPVDIRAPLDEECLIELIEGCEALEKFQAALGSR (WT)
```

In addition, to consider the influence of mutants in the EPR/DEER experiment, two mutant sequences are considered to predict the CP12's dimer. Their sequences are as follows:

```
AAIEAALDASKKFGSTSSEARVLWDIVEEMDASDNSVASKAPIVDCEYEAKVKLSLQMLTKTKAELDQVK
ALADDLKGVKLASPSVGSSAPDDSVMEALAAARAATEEFGQSSPQARLAWETVEEIAASPVDIRAPLDEE
SLIELIEGSEALEKFQAALGSR (S46C)
```

```
AAIEAALDASKKFGSTSSEARVLWDIVEEMDASDNSVASKAPIVDSEYEAKVKSLCQMLTKTKAELDQVK
ALADDLKGVKLASPSVGSSAPDDSVMEALAAARAATEEFGQSSPQARLAWETVEEIAASPVDIRAPLDEE
SLIELIEGSEALEKFQAALGSR (S56C)
```

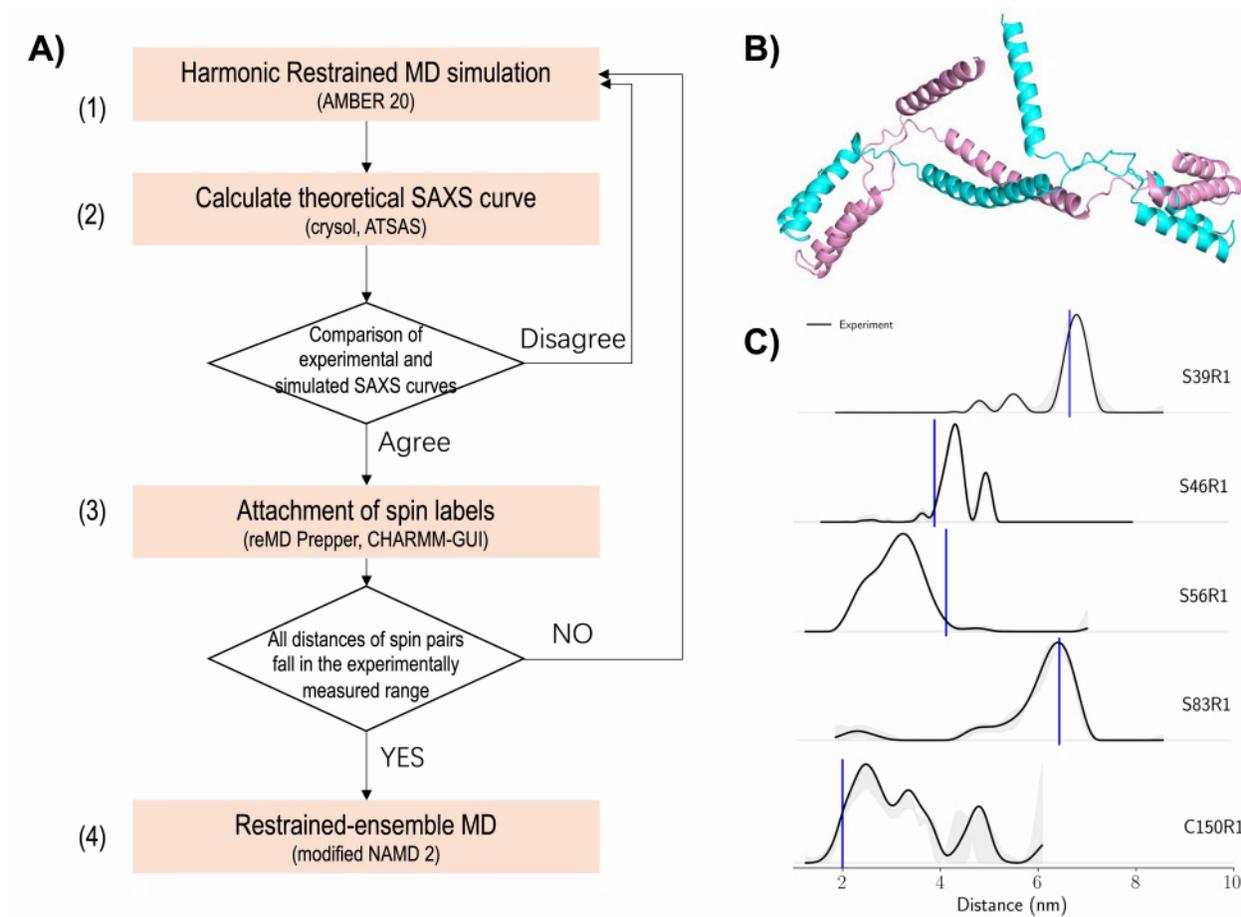


Figure 2.8: (A) Refinement scheme for the model obtained from AlphaFold2. (B) Starting model for the harmonic restrained MD simulations. (C) Distances (blue) and distance distributions (black) between spin label pairs of the starting model for reMD simulation and DEER experiment data, respectively.

2.12.2 AlphaFold2 prediction

Using AlphaFold v2.1.1-Multimer (AF2) and the default databases,^{64,66,104} the structure of the CP12 homodimer was predicted. From NCBI¹⁰⁵, the wild-type (WT) amino acid sequence was retrieved. For multiple sequence alignments (MSAs) lookup and structural template matching, an incomplete FASTA sequence of 163 residues (the blue portion of the sequence in Chapter 2.12.1) was used in the AF2 input file. Each of the homodimeric structure of WT, S46C, and S56C mutants was predicted and the top five ranked models was generated for each. To evaluate the reliability of the predictions residuewise, the predicted local distance difference test score (pLDDT, on a scale from 0 to 100, where 100 denotes the most confident) was used to assess the per-residue confidence for the model.

2.12.3 Harmonic restrained MD simulations

In the AF2 model (Figure 2.8 B), we observed two well-structured C-terminal helices were overfolded with the coiled-coil region in the AF2 model (Figure 6.2), despite their high flexibility by the EPR/DEER experiment (C150 in Figure 2.8 C). Therefore, these two C-terminal helices were manually shifted away from the dimer's coiled-coil portion (Figure 2.8 B).

All-atom harmonic restrained MD simulations without spin labels were followed by restrained-ensemble MD (reMD) simulations with spin labels introduced in order to refine the model and generate a conformational ensemble that is realistic and fits both the experimental SAXS curve¹⁰⁵ and the spin-spin distance distributions obtained by EPR/DEER experiments. A logic diagram for the simulation procedure is displayed in Figure 2.8 A. The reMD simulation¹⁰⁶ technique for the model with all-atom spin labels is suitable to exploit protein's structure in their native environment based on multiple distance histograms information obtained from EPR/DEER spectroscopy due to the high flexibility of R1 spin labels with 5 dihedral angles. A care must be taken for constructing an initial model for reMD simulations: a spin pair distance that is unlikely in the distance distribution obtained by the EPR/DEER experiment could result in an unexpectedly large disturbance on the simulation system, leading to unsuccessful conformational sampling. Therefore, before each run of the reMD simulation, a screening on the initial model based on the distance information between all the spin pairs was required to avoid such a violation.

The calculation process is detailed below:

- (1) Based on the modified AF2 model, we conducted harmonic restricted MD simulations using the Amber 20 package.¹¹⁰ The Amber ff19SB force field⁹⁵ was used for the protein, which was immersed into a cubic periodic box filled with water solvent molecules modeled by the OPC model¹¹¹ with the LEaP program of AmberTools20. At pH = 7.0, all charged residues were taken into consideration in their standard protonation condition. To neutralize the simulation box, we added 28 sodium ions, and the total number of atoms become 178352. With a distance cutoff of 9 Å, nonbonded particle-particle interactions¹¹² were taken into account, and the particle mesh Ewald (PME) method¹¹³ was used to treat long-range electrostatic interactions. After the simulation system was minimized, heated, 10 rounds of MD simulations were conducted, with harmonic restraint of a spring force of 30 kcal/(mol·Å²) applied to the CA-CA lengths across the dimer for the corresponding spin-labeled residues S39,

S46, S56, S83, and C150. Each round, MD simulations was performed for 100ps with a time step of 2 fs at $T = 300$ K and $P = 0.978$ atm. Every 1 ps, the atomic coordinates were stored. The initial/target distances for each pair of CA atoms were, respectively, 73/68 Å for S39, 57/43 Å for S46, 26/32 Å for S56, 51/64 Å for S83, and 12/25 Å for C150. In the first 2 ps of each cycle, the target distances were readily achieved (Figure 2.9).

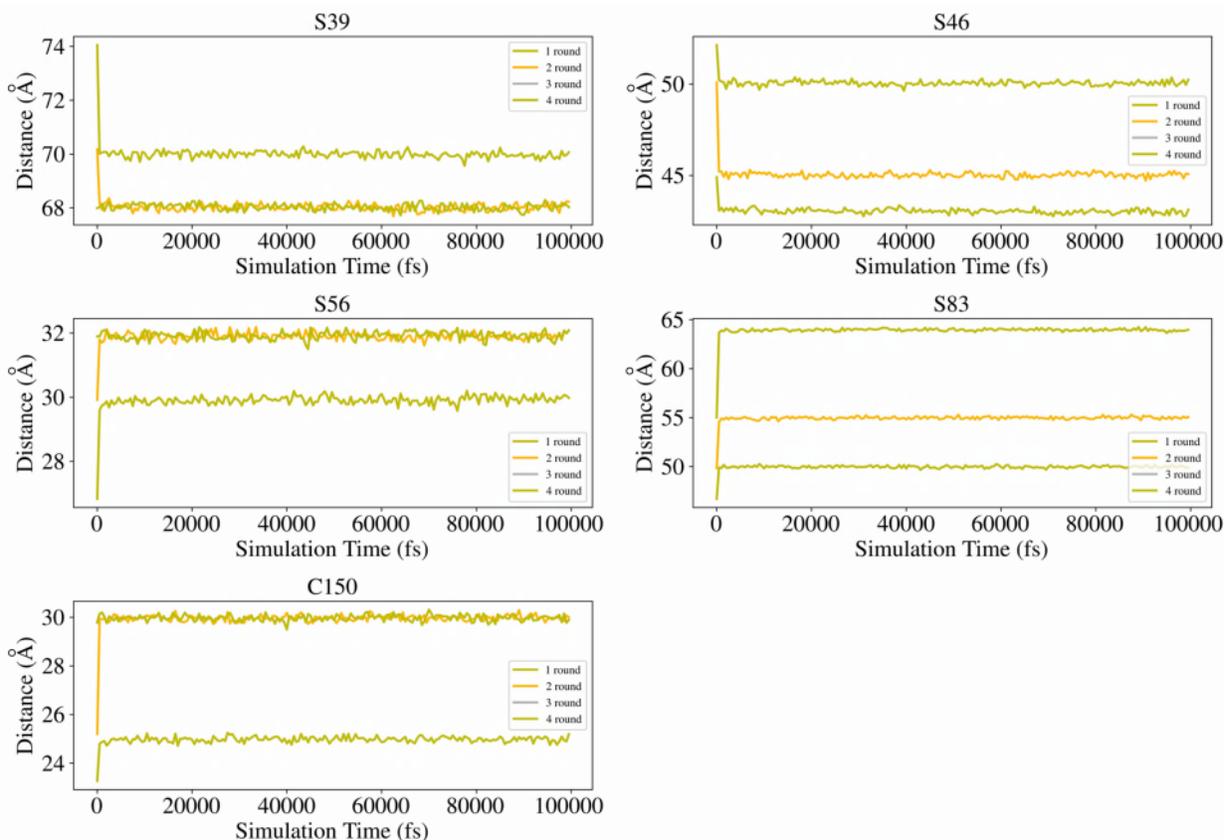


Figure 2.9: Time evolution of the CA-CA distances. We consider five residue pairs, S39-S39, S46-S46, S56-S56, S83-S83, and C150-C150 in the dimer. For each pair, the CA-CA distance are plotted as a function of time during harmonic restrained MD simulation using Amber.

- (2) Using the reMD Prepper module of CHARMM-GUI¹⁰⁹, we attached the all-atom spin label CYR1¹¹⁴ to the residues S39, S46, S56, S83, and C150, respectively. After removing the waters from the simulation model, we used CRY SOL¹¹⁵ of ATSAS-3.0.4-2¹¹⁶ to simulate SAXS curve and compared it with the experiment¹⁰⁵. If the simulated SAXS curve and the experimentally obtained curve were in good agreement, we then measured the initial distance of each spin pair.

2.12.4 Restrained-ensemble MD simulations

- (3) We then performed the reMD simulations using a modified version of NAMD 2^{107,109} with an all-atom CHARMM36m protein force field¹¹⁷, if all the spin pair distances were within the experimentally reported range. In order to save computational resources, we attached 25 copies of all-atom CYR1 spin labels with the same initial



Figure 2.10: The fixation between CYR1 spin label and its attached residues. The all-atom CYR1 spin label has a main chain like amino acid. The reMD simulation adds and fixes the spin label model by overlapping the main chain of spin label attached residue and the main chain of spin labels with a harmonic force constant of $10 \text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$.

coordinates to the corresponding residues S39, S46, S56, S83, and C150 using reMD Prepper in a vacuum. The N, C α , C, and O atoms of each spin label were fixed to the corresponding atom positions in the labeled residues (Figure 2.10) throughout the entire reMD simulations using a force constant of $10 \text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$. The force field of all-atom CYR1 spin label¹¹⁸ was obtained using CHARMM-GUI. We allowed the spatial overlap between these 25 copies by ignoring their mutual interactions. Five independent all-atom reMD simulations, each with five different random number seeds were conducted at 303.15 K using Langevin dynamics and a damping coefficient of 5 ps^{-1} . Prior to each reMD production run, we carried out minimization and equilibration, keeping the positions of the backbone atoms with harmonic restraints of $2 \text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$ imposed on them, while no restraints were imposed on the sidechain atoms. For further improvement of the models, we turned off the harmonic restrictions imposed on the backbone atoms and continued the reMD simulations for 2 ns with a 0.5-fs time step. The particle mesh Ewald (PME) method¹¹⁹ was used to evaluate the long-range electrostatic interactions, and the nonbonded interactions were truncated at a distance cutoff of 10 \AA . During the reMD simulations a force constant of $100 \text{ kcal}/(\text{mol}\cdot\text{\AA}^2)$ and a bin width of 0.025 nm , the conformational ensemble of the system was generated in such a way that the distance distributions of each spin label pair were restrained to those of the experimental distance distribution histograms. In each reMD production run, the atomic coordinates were saved every 1 ps. Since we have 25 copies of each spin label, a total of 625 distances were calculated for each pair of spin labels from a single snapshot of their trajectories, and a total of 1, 250 000 data points were generated for each pair of spin labels from a single reMD production run. Trajectory analysis and protein visualization were carried out using PyMOL¹²⁰ and VMD¹²¹, respectively.

Chapter 3

Site-selective heat current analysis with linear-homopolymer-like model

3.1 The overall thermal conductivity

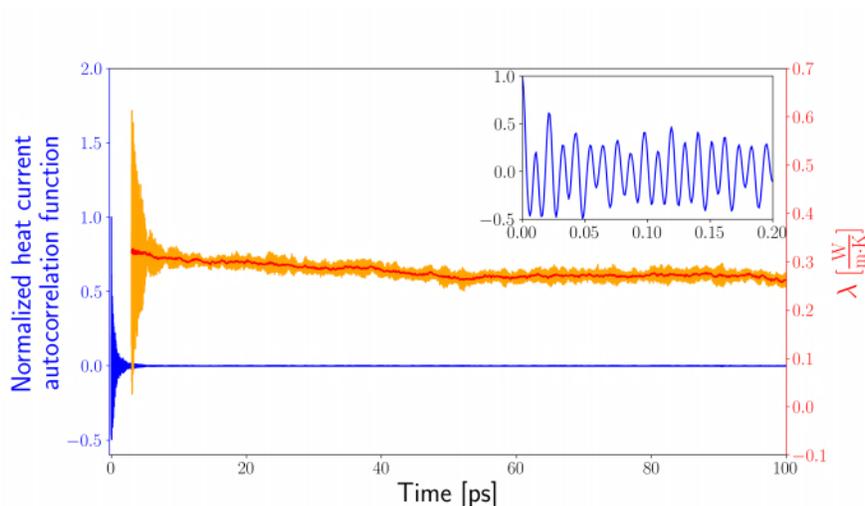


Figure 3.1: The thermal conductivity of HP36. The autocorrelation function of the heat current of the entire molecule of HP36 was calculated and ensemble-averaged, and plotted as a function of time (blue). The inset shows the short-time region (0 - 0.2 ps). The heat current autocorrelation function was integrated over time using the trapezoidal rule, yielding a quantity denoted as λ as a function of integration time (shown in orange), and a running mean of λ was computed with a window size of 100 fs and is illustrated in red. The volume of HP36 was set to 4845.447 \AA^3 .⁵⁹

The overall thermal conductivity, denoted as λ , of the HP36 protein was determined using Equation 2.1 in this study. To calculate the total volume of HP36 and individual atom volumes, we utilized the VLDP (Voronoi Laguerre Delaunay Protein) method through a web server¹²². After obtaining the ensemble-averaged heat current autocorrelation function, we performed time integration using the trapezoidal rule, as depicted in Figure 3.1. The value of λ converged

at approximately $t = 60$ ps, resulting in a thermal conductivity of 0.26 ± 0.01 W/(m·K). It is worth noting that the calculated value of λ slightly differs from the previous calculation,⁵⁹ primarily due to the utilization of a “flexible” TIP3P water model in the previous study. The dependence of the water model on thermal conductivity will be discussed further in Chapter 3.7. For subsequent analyses, we set the upper limit of the time integration of the local heat current autocorrelation function to 60 ps.

3.2 Local thermal transport

The contribution factors for intra-residue and inter-residue heat currents were presented in Figure 3.2. In general, the intra-residue contribution factors were found to be higher than the inter-residue factors (c), except for certain pairs of terminal residues (Met1-Leu2, Leu35-Phe36 and Gly12-Met13). The average intra-residue contribution factor (see Equation 2.14) was 0.029, which was nearly twice as large as the average inter-residue contribution factor (0.016). This suggests that intra-residue thermal transport makes a dominant contribution to the overall heat current, which is consistent with the results of the master equation analysis.¹² In addition to these two dominating thermal energy transport pathways, the contributions from sidechain-sidechain heat currents across nonbonded native contacts were relatively small.⁵⁹

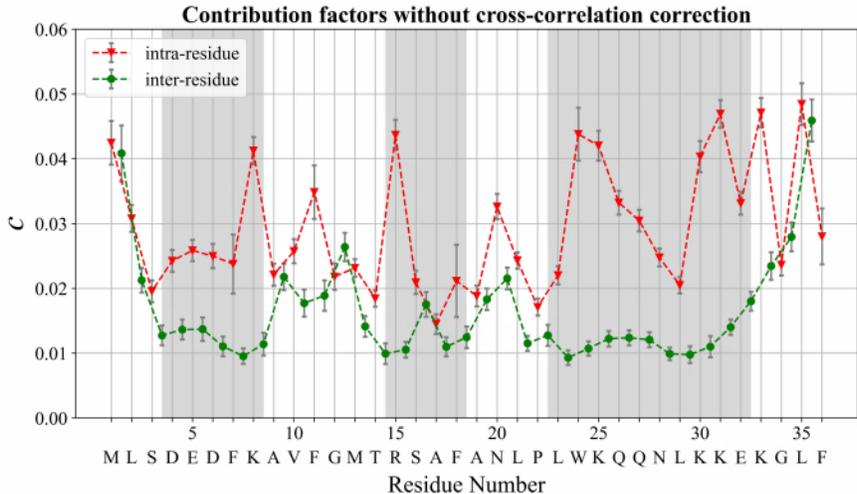


Figure 3.2: The contribution factors without cross-correlation corrections were calculated and plotted for both intra-residue ($c_{\alpha,\alpha}$, shown in red) and inter-residue ($c_{\alpha,\alpha+1}$, shown in green) interactions, as described in equations (Equation 2.15) and (Equation 2.16), respectively. The data points, such as $c_{1,2}$, represent the contribution factor between residue numbers $\alpha = 1$ and $\alpha = 2$. The horizontal axis displays the residue number, α , along with the corresponding residue names represented in the one-letter code. The grey-shaded regions correspond to the α -helical regions.

The sum of the intra-residue contribution factors, $\sum_{\alpha=1}^{36} c_{\alpha,\alpha}$, was 1.06, and the sum of the inter-residue contribution factors, $\sum_{\alpha=1}^{35} c_{\alpha,\alpha+1}$, was 0.57. The deviation of $1.06 + 0.57 = 1.63$ from 1 suggests the presence of non-negligible cross-correlation effects among different partial heat currents. In Chapter 3.3, we will delve into a more detailed analysis of these cross-correlation effects.

3.3 Cross-correlation correction

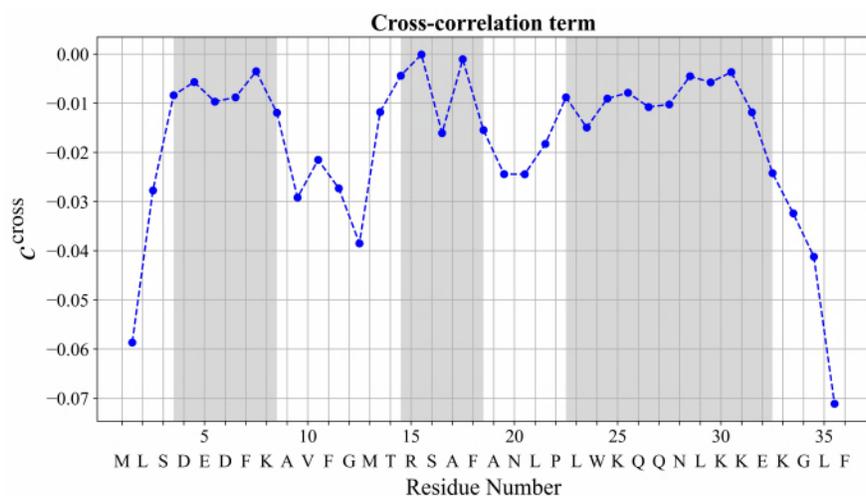


Figure 3.3: The cross-correlation contribution factor between nearest neighbor residues, $c_{\alpha, \alpha+1}^{\text{cross}}$, defined as $\xi_{\alpha, \alpha+1}/\Lambda$, was plotted as a function of the residue number. Each data point represents the value of $c_{\alpha, \alpha+1}^{\text{cross}}$ between residue α and $\alpha + 1$. For example, the data point $c_{1,2}^{\text{cross}}$ corresponds to the cross-correlation between residues 1 and 2.

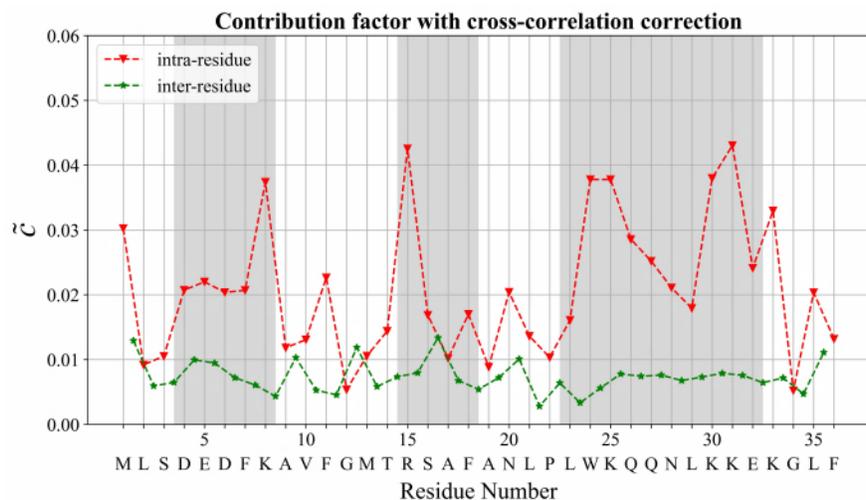


Figure 3.4: Contribution factors with cross-correlation correction. See the caption to Figure 3.2.

In this study, we assumed that the cross-correlation effect is short range along the polypeptide sequence (Chapter 2.4). Interestingly, the cross-correlation effect exhibits secondary structure dependence (Figure 3.3): the α -helical regions are less affected by the cross-correlation. The reason for this observation is currently unclear and will be investigated further in future studies.

All of the contribution factors decreased after the cross-correlation correction (Figure 3.4), yielding a similar pattern to Figure 3.2. The total intra-residue contribution was 0.75, approximately three times larger than that of the inter-residue contribution of 0.26. To validate the assumption of short-range cross-correlation, we also calculated the contributing factors for the second nearest cross-correlation between residue pairs of α and $\alpha+2$ ($\xi_{\alpha, \alpha+2}$), ($\alpha = 1, 2, \dots$,

34), using the same method described in Chapter 2.4. The total contribution due to the second nearest cross-correlation was 0.03 (Table 3.1), while that for the nearest cross-correlation was -0.62 (Table 3.2), indicating that the second nearest cross-correlation has a minimal impact on the overall heat current.

To validate the linear-homopolymer-like model, we compared Λ and $\tilde{\Lambda}$. The former was derived from the exact heat current of the entire molecule, while the latter was based on the linear-homopolymer-like model. The calculation results showed that $\tilde{\Lambda}$ overestimated Λ only by 0.9%, indicating that the linear-homopolymer-like model successfully represents the thermal transport property of the entire molecule.

Table 3.1: The second nearest cross-correlation. $\xi_{\alpha,\alpha+2}$ was calculated in a similar manner as $\xi_{\alpha,\alpha+1}$ using Equation 2.15, and contribution factor, $c_{\alpha,\alpha+2}^{\text{cross}}$, was calculated as $\xi_{\alpha,\alpha+2}/\Lambda$. The values of ξ are shown in the unit of $(\text{\AA}\cdot\text{kcal/mol})^2/\text{fs}$.

Residue (α)	Residue ($\alpha + 2$)	$\xi_{\alpha,\alpha+2}$	$c_{\alpha,\alpha+2}^{\text{cross}}$
ARG15	ALA17	0.012	0.001
ALA17	ALA19	-0.010	-0.001
PHE7	ALA9	0.027	0.003
MET13	ARG15	-0.033	-0.003
PHE18	ASN20	-0.007	-0.001
GLN26	ASN28	0.032	0.003
LEU2	ASP4	0.036	0.004
ASP4	ASP6	-0.007	-0.001
TRP24	GLN26	0.027	0.003
LYS25	GLN27	0.045	0.005
LYS30	GLU32	0.010	0.001
SER3	GLU5	0.036	0.004
VAL10	GLY12	0.029	0.003
GLU32	GLY34	0.038	0.004
ALA19	LEU21	0.024	0.002
LEU21	LEU23	-0.027	-0.003
GLN27	LEU29	0.019	0.002
LYS33	LEU35	0.086	0.009
LEU23	LYS25	0.063	0.006
ASN28	LYS30	-0.396	-0.040
LEU29	LYS31	0.028	0.003
LYS31	LYS33	0.047	0.005
ASP6	LYS8	0.000	0.000
PHE11	MET13	0.051	0.005
ALA9	PHE11	0.042	0.004

Residue (α)	Residue ($\alpha + 2$)	$\xi_{\alpha,\alpha+2}$	$c_{\alpha,\alpha+2}^{\text{cross}}$
SER16	PHE18	0.015	0.001
GLY34	PHE36	0.034	0.003
GLU5	PHE7	-0.028	-0.003
ASN20	PRO22	0.026	0.003
THR14	SER16	-0.024	-0.002
MET1	SER3	0.042	0.004
GLY12	THR14	0.006	0.001
PRO22	TRP24	0.014	0.001
LYS8	VAL10	0.022	0.002

Table 3.2: Local thermal transport property after the cross-correlation correction. $\tilde{\Lambda}_{\alpha,\alpha}$ ($\alpha=1, 2, \dots, 36$) and $\tilde{\Lambda}_{\alpha,\alpha+1}$ ($\alpha=1, 2, \dots, 35$) are shown in unit of $(\text{\AA} \cdot \text{kcal})^2/\text{fs}$ (Equation 2.12), while the value of Λ was 9.79 in the same unit (Equation 2.13).

Residue (α)	$\tilde{\Lambda}_{\alpha,\alpha}$	$\tilde{\Lambda}_{\alpha,\alpha+1}$
1	2.96×10^{-1}	1.27×10^{-1}
2	9.00×10^{-2}	5.81×10^{-2}
3	1.03×10^{-1}	6.34×10^{-2}
4	2.03×10^{-1}	9.76×10^{-2}
5	2.15×10^{-1}	9.27×10^{-2}
6	2.00×10^{-1}	7.05×10^{-2}
7	2.03×10^{-1}	5.94×10^{-2}
8	3.66×10^{-1}	4.24×10^{-2}
9	1.16×10^{-1}	1.01×10^{-1}
10	1.28×10^{-1}	5.16×10^{-2}
11	2.22×10^{-1}	4.46×10^{-2}
12	5.24×10^{-2}	1.16×10^{-1}
13	1.04×10^{-1}	5.71×10^{-2}
14	1.41×10^{-1}	7.19×10^{-2}
15	4.17×10^{-1}	7.80×10^{-2}
16	1.65×10^{-1}	1.31×10^{-1}
17	9.99×10^{-2}	6.63×10^{-2}
18	1.67×10^{-1}	5.29×10^{-2}
19	8.66×10^{-2}	7.07×10^{-2}
20	2.00×10^{-1}	9.87×10^{-2}

Residue (α)	$\tilde{\Lambda}_{\alpha,\alpha}$	$\tilde{\Lambda}_{\alpha,\alpha+1}$
21	1.33×10^{-1}	2.73×10^{-2}
22	1.01×10^{-1}	6.27×10^{-2}
23	1.57×10^{-1}	3.26×10^{-2}
24	3.70×10^{-1}	5.49×10^{-2}
25	3.70×10^{-1}	7.62×10^{-2}
26	2.80×10^{-1}	7.27×10^{-2}
27	2.47×10^{-1}	7.45×10^{-2}
28	2.06×10^{-1}	6.62×10^{-2}
29	1.76×10^{-1}	7.15×10^{-2}
30	3.72×10^{-1}	7.73×10^{-2}
31	4.22×10^{-1}	7.41×10^{-2}
32	2.36×10^{-1}	6.30×10^{-2}
33	3.23×10^{-1}	7.02×10^{-2}
34	5.11×10^{-2}	4.60×10^{-2}
35	1.99×10^{-1}	1.09×10^{-1}
36	1.29×10^{-1}	
total	7.35	2.53

3.4 Residue-type dependence

By applying the site-selective heat current analysis based on the linear-homopolymer-like model, we were able to evaluate the residue-by-residue local thermal conductivity. Figure 3.5 shows the dependence of the intra-residue contribution factors on the residue type. The residue volume, V_α , was calculated as the sum of the atomic volumes of the constituent atoms. We obtained the average residue volume using five representative structures of HP36 in thermal equilibrium at $T = 300$ K using the VLDP web server.¹²²

We observed that the values of the corrected contribution factors, \tilde{c} , are proportional to the residue volume. Additionally, the slope of the regression line depends on the residue type, indicating that the residue-wise thermal conductivity is sensitive to the residue type. We define the residue-wise thermal conductivity, λ_α , of residue α as $\tilde{\Lambda}_{\alpha,\alpha}/(3V_\alpha k_B T)$, and we see that λ_α decreases in the order of charged, polar, and hydrophobic residues, in consistent with the previous report that the thermal diffusion along an α -helix composed entirely of polar residues is faster than that of its non-polar residue counterpart.²¹

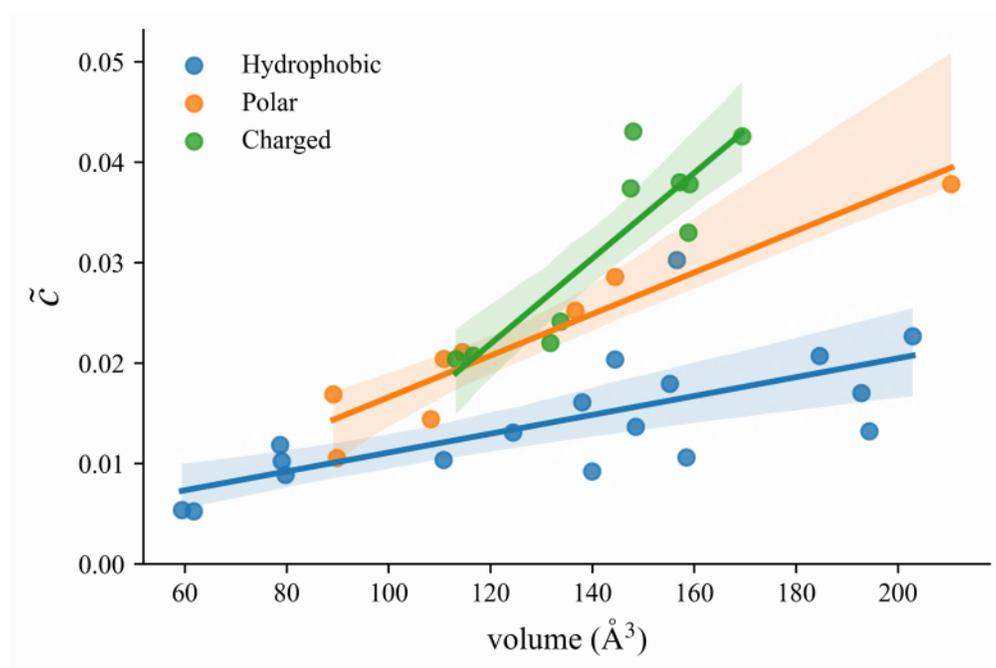


Figure 3.5: Volume dependence of \tilde{c} , a residue-wise contribution factor after cross-correlation correction, for charged (green), polar (orange) and hydrophobic (blue) residues, respectively.

3.5 Local density dependence

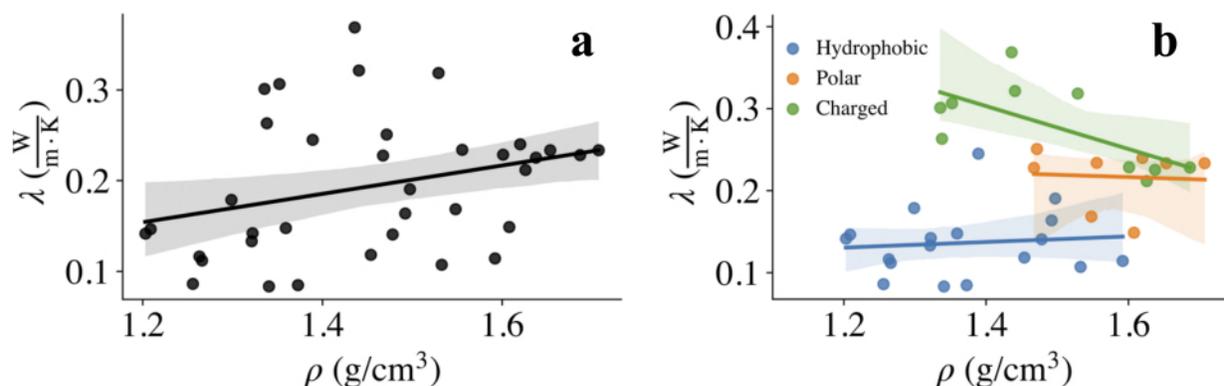


Figure 3.6: A scatter plot and linear correlation analysis were performed to examine the relationship between the intra-residue thermal conductivity and the local density, represented by the mass density (ρ), calculated as the ratio of residue mass to residue volume (unit: g/cm^3).

The density of a material is widely recognized as an important determinant of its thermophysical properties.³ The product of material density (ρ) and its specific heat capacity (c_p), denoted as ρc_p , is commonly used as a measure of the thermal energy storage capacity of a material.¹²³ A higher value of ρc_p indicates a greater capacity of the material to store thermal energy. In macroscopic homogeneous materials, it is generally observed that higher densities are associated with higher thermal conductivities. This relationship is evident in various materials, such as high-density polyethylene having a higher thermal conductivity ($0.44 \text{ W}/(\text{m}\cdot\text{K})$) compared to low-density polyethylene ($0.3 \text{ W}/(\text{m}\cdot\text{K})$).

However, in the case of heterogeneous materials such as proteins, the local density can vary significantly from one site to another. This is due to the different packing of amino acid residues, resulting in a wider distribution of residue densities. The relationship between local density and thermal conductivity in proteins is shown in Figure 3.6, a scatter plot of the residue-wise thermal conductivity versus local density of each residue. The local mass density values range from 1.20 to 1.70 g/cm³, shown in Figure 3.6, which is somewhat broader compared to those determined experimentally and theoretically (1.33 to 1.44 g/cm³).¹²⁴ This wider range of residue densities could be mainly attributed to the residue-wise volume variation between different environments, i.e., buried or exposed. The buried residues, which are surrounded by other residues, tend to have a higher density than the exposed residues, with greater solvent accessibility, leading to a broader distribution of residue densities within the protein molecule.^{125,126}

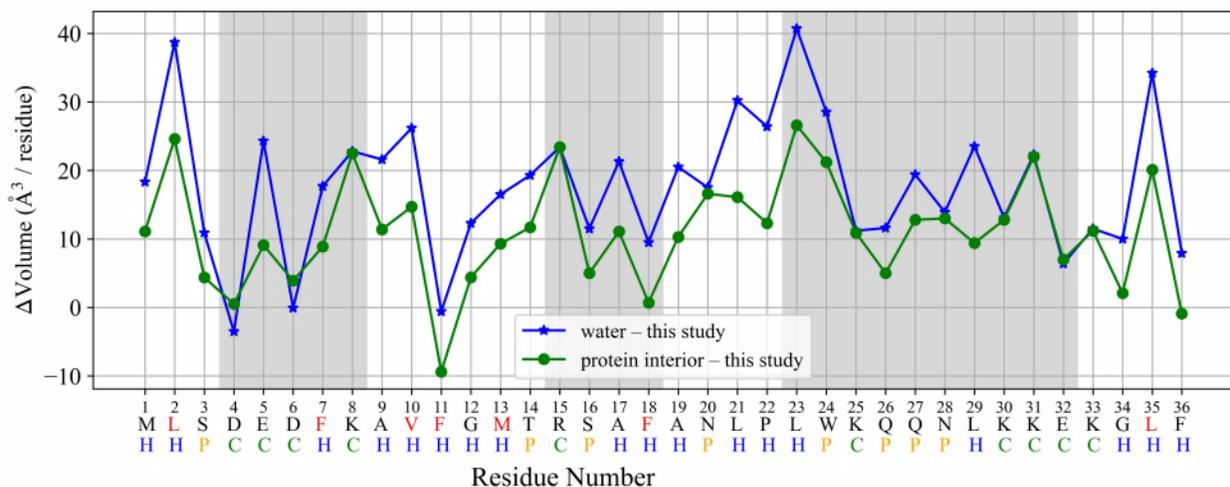


Figure 3.7: Comparison of the residue volumes in this study with those in water (blue) and the protein interior (green)¹²⁵. The second row of the x-axis represents the amino acid names using the one-letter format, with the residues buried into the protein interior highlighted in red. The third row indicates the residue types, with hydrophobic residues shown in blue, charged residues in green, and polar residues in orange.

Table 3.3: Pearson correlation coefficient (r), p -value and slope of linear relationship between \bar{c} and volume in Figure 3.5. Generally, a value of $|r| > 0.3$, or a p -value < 0.05 is often considered to be statistically significant, indicating that the pair of variables under consideration are correlated.

Residue type	r	p -value	slope
overall	0.56	0.0004	1.55×10^{-4}
charged	0.88	0.0008	4.3×10^{-4}
polar	0.95	0.0004	2.1×10^{-4}
hydrophobic	0.68	0.0019	9.4×10^{-5}

The linear regression analysis of the residue-wise thermal conductivity with cross-correlation correction shows a weak density dependence (Figure 3.6 a). The data points exhibit a broad distribution, and the correlation between thermal conductivity and density is not statistically significant, as indicated by a Pearson correlation coefficient of 0.3

and a p -value of 0.078. Pearson correlation analysis for all data in Figure 3.6 a and sub data points classified by residue types in Figure 3.6 b are shown in Table 3.3.

In Figure 3.6 b, the linear regression lines for hydrophobic and polar residues have almost constant thermal conductivity values, while charged residues show a negative proportionality, although the reason for this observation is unclear. One possible explanation is that charged residues tend to have larger volumes compared to other residue types.¹²⁵ Additionally, as shown in Figure 3.6 b, the thermal conductivity of polar residues is approximately twice that of hydrophobic residues, in line with the previous observations of faster heat diffusion in α -helices consisting of polar residues than their non-polar-residue counterparts.

For the HP36 protein, the subtotal contribution factors (\tilde{c}) for charged, polar, and hydrophobic residues were 0.32, 0.17, and 0.23, respectively. Although HP36 contains smaller proportions of charged residues (10/36) than that of hydrophobic residues (18/36), the charged residues play more important roles in the thermal transport due to the relatively larger contribution to the overall heat current.

3.6 Thermal diffusivity and temperature relaxation time

In the field of molecular biophysics, vibrational energy relaxation in proteins has attracted more attention than the thermal transport properties of proteins. A pioneering study by Mizutani and Kitagawa⁷ investigated the excess energy dissipation in myoglobin and observed double exponential decay in the population of the ν_4 mode of the heme after flash photolysis. They estimated the temperature relaxation times of the vibrational mode to be 3.0 and 25 ps, respectively, assuming a Boltzmann distribution for the vibrational mode. This highlights the importance of picosecond dynamics in the temperature relaxation of proteins, which involves energy redistribution processes.

It is possible to estimate two parameters, namely the thermal diffusivity and temperature relaxation time, using the following equation^{22,123,127}:

$$\lambda = \frac{\rho c_p R^2}{\tau} = \rho c_p \alpha \quad (3.1)$$

In this equation, λ represents the thermal conductivity, τ is the relaxation time, α is the thermal diffusivity, ρ is the protein density, c_p is the heat capacity, and R is the effective radius of the protein, which can be estimated from the protein volume, $V = \frac{4\pi R^3}{3}$.

The estimated values of τ fall within the range of 8.4 to 13.9 ps, corresponding to a temperature relaxation rate ranging from 0.072 to 0.119 ps⁻¹. It should be noted that the reported values of vibrational energy relaxation times range from a few picoseconds to several tens of picoseconds.^{7,15,16,26,127,128} The estimated thermal diffusivity is in the range of 7.9 to 13.2 Å² ps⁻¹, in good agreement with values obtained from nonequilibrium molecular dynamics simulations.^{5,129} The density of HP36 was calculated as $\rho = 1.57$ g/cm³, slightly larger than the widely used value of 1.35 g/cm³.¹²⁴ The heat capacity was assumed to be in the range of 0.30 to 0.50 cal/g deg.¹³⁰

3.7 Water model

In our previous study on the thermal conductivity of the HP36 protein⁵⁹, we utilized the Amber ff14SB force field for the protein atoms and the TIP3P water model with the SHAKE constraints turned off, although this usage of the TIP3P model is unconventional. It is known that simulations using the standard rigid TIP3P water model can exhibit anomalous diffusion.^{131,132} To investigate the influence of the water models on protein thermal conductivity, we employed the Amber ff19SB force field in combination with the TIP3P water model, again with the SHAKE constraints turned off, but only for the protein in this particular study. As mentioned earlier, the calculated thermal conductivity (λ) of the entire protein was found to be $0.26 \pm 0.01 \text{ W}/(\text{m}\cdot\text{K})$, which is close to the value of $0.3 \pm 0.01 \text{ W}/(\text{m}\cdot\text{K})$ obtained in our previous study.⁵⁹ This suggests that solvent properties may influence protein thermal conductivity through protein-solvent interfaces.

As discussed earlier in this chapter, protein-solvent interactions play a crucial role in determining the structure, dynamics, and function of proteins. Notably, Straub, Leitner, and their colleagues have conducted significant studies on the energy transport across protein-water interfaces. Sagnella et al. observed spatially directed “funneling” of kinetic energy from the heme group to the surrounding solvent for the dissipation of excess energy in myoglobin following flash photolysis.¹⁵ Agbo, Xu, Zhang, et al. investigated the thermal conductance between cytochrome c and water and demonstrated that the protein-water interface does not pose a greater Kapitza resistance to heat flow compared to the protein itself. Interestingly, thermal conductance at protein-solvent interfaces varies for different types and shapes of proteins, ranging from 100 to 330 $\text{MWK}^{-1}\text{m}^{-2}$.^{22,35,127,133,134} Hamzi et al. studied the dependence of thermal conductance at protein-water interfaces on the types of amino acid residues, and they observed that hydrophobic and aromatic amino acids tend to exhibit lower interfacial thermal conductance.¹³⁵ Other types of interfaces have also been investigated. For instance, ultrafast energy dissipation from peptide helices to chloroform solvents on the timescale of 0.5 ps was observed through a collaboration between experimental and theoretical approaches.⁵ To gain insights into how different water models could affect thermal boundary conductance, it is useful to consider the vibrational density of states of both the protein and water.^{134,136} In contrast to our previous study,⁵⁹ the present study using the rigid TIP3P water model shows that some high-frequency vibrational modes of the solvent are absent, resulting in a decrease in thermal conductance at the protein-water interface. Consequently, it is plausible that the vibrational energy distribution, especially for surface amino acid residues near the protein-water interface, may be affected. Moreover, the local heat capacities of such amino acid residues might also be influenced, leading to changes in the overall protein thermal conductivity. Systematic studies of the influence of solvent models on the thermal transport properties of proteins and protein-solvent interface should be further investigated to validate these hypotheses.

Chapter 4

Heat and energy transfer through nonbonded contacts

4.1 Energy transfer and heat transfer

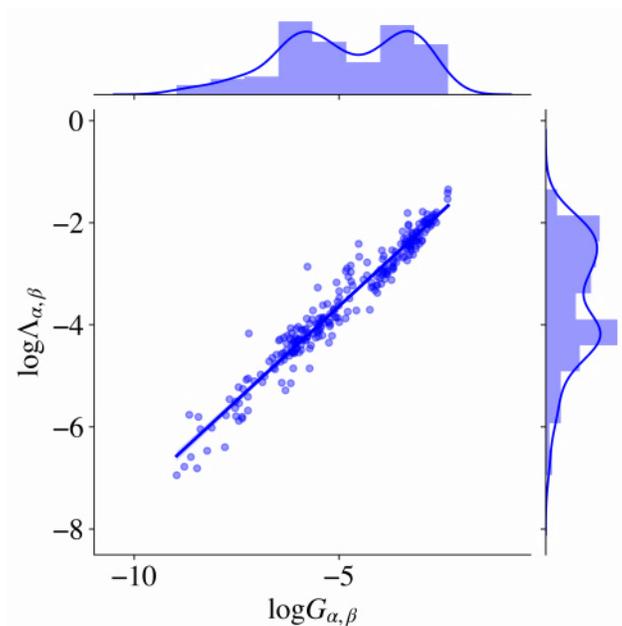


Figure 4.1: Correlation of $\Lambda_{\alpha,\beta}$ and $G_{\alpha,\beta}$.

In the previous studies, we explored the energy transport network in proteins based on the time-integrated ACF of the *inter-residue energy flow* ($G_{\alpha,\beta}$).^{23,28,29,58,80} Considering the similarity between the interatomic current of energy (Equation 2.36) and heat (Equation 2.39), we are allowed to take a similar approach. To compare the time-integrated ACF of inter-residue current of heat with that of energy, we showed a scatter plot between $\Lambda_{\alpha,\beta}$ and $G_{\alpha,\beta}$ in Figure 4.1 and

carried out a linear regression analysis. As expected, the Pearson correlation analysis showed a statistically significant correlation between $\Lambda_{\alpha,\beta}$ and $G_{\alpha,\beta}$, with $r = 0.97$, $**p \ll 0.01$ (8.0×10^{-173}). The estimated linear regression model is as follows: $\log \Lambda_{\alpha,\beta} = 0.74 \times \log G_{\alpha,\beta} + 0.06$ with $R^2 = 0.95$. The high similarity between $\Lambda_{\alpha,\beta}$ and $G_{\alpha,\beta}$ implies that the network patterns of energy transport and heat transport should also be similar to each other.

4.2 Interaction type dependence

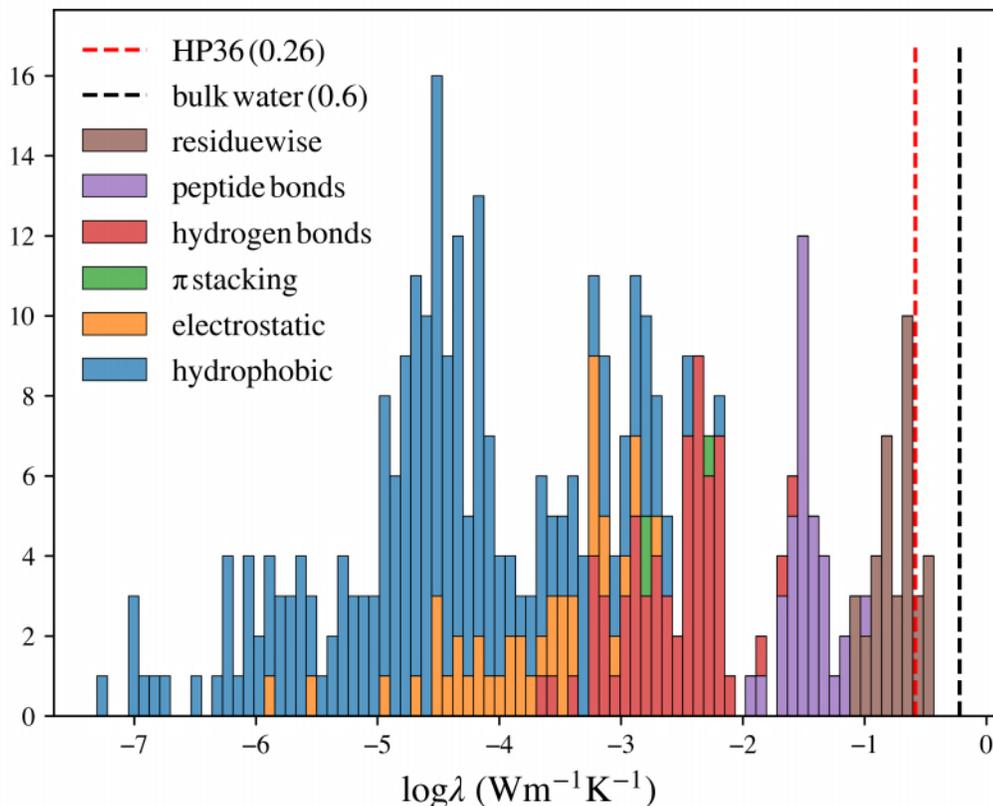


Figure 4.2: Frequency histogram of local thermal conductivity in HP36. It shows the occurrence of $\lambda_{\alpha,\beta}$ for different types of residue pairs (α, β) in nonbonded native contacts, together with the residue-wise thermal conductivity and those between adjacent residue pairs along the chain. For comparison, the value of the thermal conductivity of the entire HP36 molecule is indicated by the red dashed line, as well as that of bulk water by the black dashed line.¹³⁷

Figure 4.2 shows the frequency histogram of the calculated local thermal conductivities in HP36. The histogram reveals that the hydrogen-bond contacts have the greatest values of $\lambda_{\alpha,\beta}$ among all the nonbonded native contacts, with a peak at around 2.5×10^{-2} W/(m·K), followed by the electrostatic contacts with $\lambda_{\alpha,\beta}$ values ranging from 10^{-5} to 10^{-3} W/(m·K) with having broader and less sharply peaked distributions than that of the hydrogen-bond contacts. In the hydrophobic core of HP36, there are three π stacking contacts: Phe7-Phe11 ($\lambda_{7,11} = 4.9 \times 10^{-3}$), Phe7-Phe18 ($\lambda_{7,11} = 1.5 \times 10^{-3}$), and Phe11-Phe18 ($\lambda_{11,18} = 1.5 \times 10^{-3}$), with $\lambda_{\alpha,\beta}$ values comparable with hydrogen-bond contacts. Hydrophobic contacts, constitute the majority of the nonbonded native contacts with a broad distribution ranging from 10^{-7} to 10^{-2} W/(m·K). Their $\lambda_{\alpha,\beta}$ values are relatively smaller, with peaks at around 5.9×10^{-4}

$W/(m \cdot K)$. In summary, the frequency histogram of the calculated local thermal conductivities for nonbonded native contacts exhibits a strikingly broad distribution, where different types of nonbonded contacts have significantly different contributions to thermal transport in the protein, with hydrogen bonds playing the dominant role.

4.3 Peptide bonds? or nonbonded contacts?

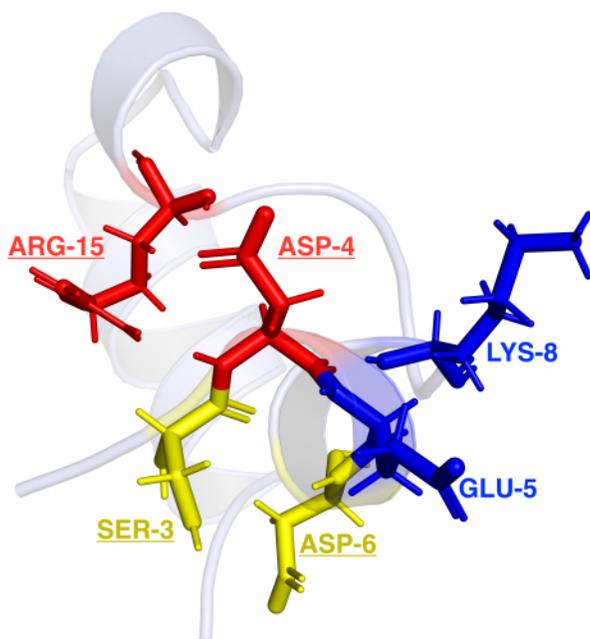


Figure 4.3: Nonbonded contacts that can compete with polypeptide chain.

There is an ongoing debate on the dominant pathway of thermal energy transport in proteins.^{3,6,12} To address the issue, we carried out site-selective heat current analysis (Figure 4.2). As a result, the distribution of local thermal transport coefficients demonstrated substantial heterogeneity in protein thermal transport at microscopic scale: The residue-wise λ values are ranging from about 0.08 to 0.3 $W/(m \cdot K)$, while the inter-residue thermal conductivity between adjacent residue pairs along the chain ranges from about 0.01 to 0.08 $W/(m \cdot K)$, indicating that the importance of the polypeptide chain as a major pathway of thermal transport in protein. It should be noted that, however, that there are three exceptions $\lambda_{3,6} = 2.4 \times 10^{-2}$, $\lambda_{4,15} = 1.3 \times 10^{-2}$, and $\lambda_{5,8} = 2.1 \times 10^{-2}$, shown in Figure 4.3, where their λ values exceed $10^{-2} W/(m \cdot K)$ being comparable to the local thermal transport coefficient for the pathways along the polypeptide chain. It is worthy mentioning that the residue pairs, Ser3-Asp6 and Asp4-Arg15, have also been identified as having a relatively larger local energy diffusivities in the theoretical study by non-equilibrium MD simulations and master equation model,¹² indicating their possible roles to as “shortcuts” on the thermal transport network of the protein.

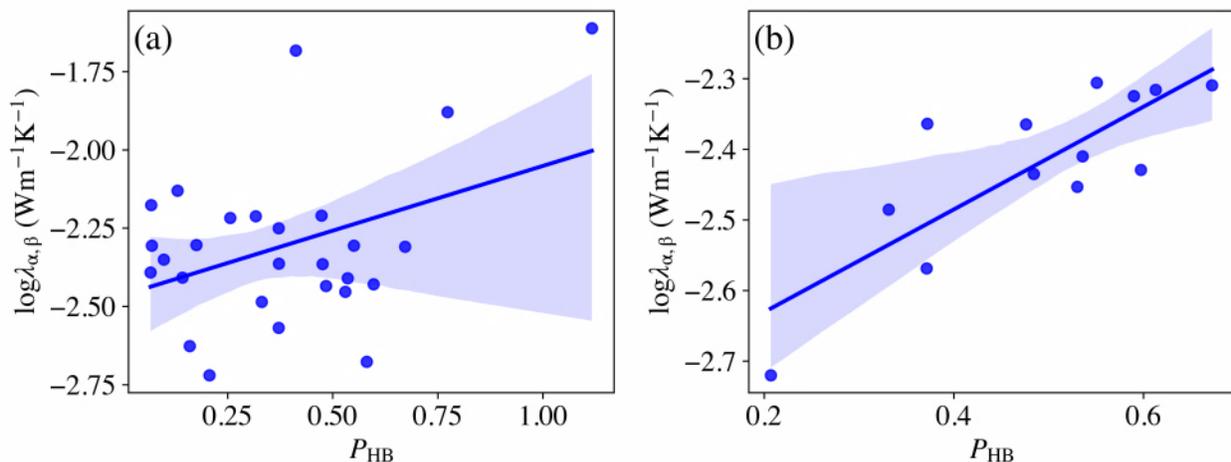


Figure 4.4: Correlation between the inter-residue thermal conductivity through hydrogen bonding and their hydrogen bond occurrence probability (P_{HB}) during MD simulations. (a): all residue pairs in contact with hydrogen bonds; (b) residue pairs ($\alpha, \alpha + 4$) with α -helical hydrogen bonding.

4.4 Thermal energy transport through hydrogen bonds

The values of inter-residue thermal conductivity for residue pairs in contacts with hydrogen bonds are generally greater than those of the other types of nonbonded contacts (Figure 4.2), in line with the previous studies in the literature.^{6,21} The linear regression analysis with a scatter plot between the values of inter-residue thermal conductivity ($\lambda_{\alpha,\beta}$) between hydrogen bonded residue pairs and their hydrogen bond occurrence probability (P_{HB}) during the MD simulations (Figure 4.4a) yielded a Pearson correlation coefficient of $r = 0.51$ (> 0.3) and a $**p$ -value of 0.007 ($\ll 0.05$), indicating a positive correlation between them.

In an α -helical protein, two types of hydrogen bonds are recognized: The first type, known as central hydrogen bonds stabilize the helical structure with hydrogen bonds between the carbonyl oxygen (O) of residue α and the amino group (N-H) of residue $\alpha+4$. The second type occurs either as a sidechain-sidechain or sidechain-mainchain hydrogen bond. Miño et al. reported that the heat diffusion along the α -helical polypeptide chain containing both types of hydrogen bonds is two times faster than those containing only the first type of hydrogen bonds, highlighting the important role of the second type of hydrogen bonds in proteins.²¹ To investigate the role of the first type of hydrogen bond in the heat transport in HP36, we selected the values of inter-residue thermal conductivity for 32 residue pairs between residue α and $\alpha+4$ ($n=1, 2, \dots, 32$). Figure 4.4b. The NMR structure of HP36 (PDB: 1VII) contains three alpha-helices, H1 (residues 4-8), H2 (residues 15-18), and H3 (residues 23-32). We also observed transient formation of hydrogen bonds between ($\alpha, \alpha + 4$) residue pairs during MD simulations, in addition to those found in the NMR structure. Thereby, 19 members of these hydrogen bonds are shown in Table 4.1, among which 13 of the α -helical hydrogen bonds were selected for the linear regression analysis (Figure 4.4b). In the previous study,²¹ the heat diffusion along the α -helical chain with both types of hydrogen bonds is twice faster than those with only the first type of hydrogen bonds, indicating the important role of the second type of hydrogen bonds in proteins. To investigate the role of the first type of hydrogen bond in heat transport in HP36, we found that almost all hydrogen bonds between pairs of residue ($\alpha, \alpha+4$) were of the

first type with only one exception between the mainchain of 16SER and the sidechain of 20ASN.

Table 4.1: Hydrogen bond occurrence probability (P_{HB}) and averaged shortest distance ($\langle d_c \rangle$) between residue α and $\alpha+4$. MC : mainchain, SC: sidechain.

acceptor residue number	acceptor atom	donorH residue number	donorH atom	donor atom	P_{HB}	acceptor type	donor type	averaged distance
3SER	O	7PHE	H	N	0.60	MC	MC	2.05
4ASP	O	8LYS	H	N	0.37	MC	MC	2.19
5GLU	O	9ALA	H	N	0.21	MC	MC	2.32
7PHE	O	11PHE	H	N	0.67	MC	MC	1.92
15ARG	O	19ALA	H	N	0.53	MC	MC	2.04
22PRO	O	26GLN	H	N	0.33	MC	MC	2.11
23LEU	O	27GLN	H	N	0.55	MC	MC	1.99
24TRP	O	28ASN	H	N	0.48	MC	MC	2.03
25LYS	O	29LEU	H	N	0.54	MC	MC	2.00
26GLN	O	30LYS	H	N	0.48	MC	MC	2.06
27GLN	O	31LYS	H	N	0.59	MC	MC	2.00
28ASN	O	32GLU	H	N	0.61	MC	MC	1.96
29LEU	O	33LYS	H	N	0.37	MC	MC	2.12

We conducted further analysis of the properties of the interaction between residue pairs ($\alpha, \alpha + 4$) (Table 4.2). The values of inter-residue thermal conductivity for the pairs with no hydrogen bond are significantly smaller than those with hydrogen bonds. Two special residues pairs (2LEU-6ASP and 21LEU-25LYS) did not form hydrogen bonds but had relatively larger values. This may be because their interactions are stronger than those contacts with no hydrogen bonds, which are all hydrophobic residues with weak Van der Waals interactions.

Table 4.2: Inter-residue thermal conductivity and hydrogen bond formation capacity P_{HB} between residue pair α and $\alpha+4$. Also the contact distance (shortest interatomic distance between the residue pair) is shown for each pair.

Residue Number	Residue Name	Residue Number	Residue Name	P_{HB}	Interaction Type	$\lambda_{\alpha, \alpha+4}$	Contacts Distance (Å)
1	MET	5	GLU	0	hydrophobic	1.99×10^{-4}	8.33
2	LEU	6	ASP	0	charged	3.29×10^{-3}	3.08
3	SER	7	PHE	0.6	hydrogen bonds	3.72×10^{-3}	2.05
4	ASP	8	LYS	0.37	hydrogen bonds	4.33×10^{-3}	2.19

Residue Number	Residue Name	Residue Number	Residue Name	P_{HB}	Interaction Type	$\lambda_{\alpha, \alpha+4}$	Contacts Distance (Å)
5	GLU	9	ALA	0.21	hydrogen bonds	1.90×10^{-3}	2.32
6	ASP	10	VAL	0.47	hydrogen bonds	6.16×10^{-3}	2.10
7	PHE	11	PHE	0.67	pi stacking	4.91×10^{-3}	1.92
8	LYS	12	GLY	0.13	hydrogen bonds	7.40×10^{-3}	2.34
9	ALA	13	MET	0	hydrophobic	4.29×10^{-5}	5.07
10	VAL	14	THR	0	hydrophobic	8.14×10^{-7}	7.83
11	PHE	15	ARG	0	hydrophobic	4.31×10^{-5}	6.08
13	MET	17	ALA	0	hydrophobic	2.42×10^{-3}	2.83
14	THR	18	PHE	0.58	hydrogen bonds	2.10×10^{-3}	1.98
15	ARG	19	ALA	0.53	hydrogen bonds	3.52×10^{-3}	2.04
16	SER	20	ASN	0.18	hydrogen bonds	4.06×10^{-3}	2.64
17	ALA	21	LEU	0	hydrophobic	5.66×10^{-4}	2.85
18	PHE	22	PRO	0	hydrophobic	1.65×10^{-5}	5.70
19	ALA	23	LEU	0	hydrophobic	1.95×10^{-5}	6.67
21	LEU	25	LYS	0	charged	2.23×10^{-3}	2.56
22	PRO	26	GLN	0.33	hydrogen bonds	3.27×10^{-3}	2.11
23	LEU	27	GLN	0.55	hydrogen bonds	4.95×10^{-3}	1.99
24	TRP	28	ASN	0.48	hydrogen bonds	3.67×10^{-3}	2.03
25	LYS	29	LEU	0.54	hydrogen bonds	3.89×10^{-3}	2.00
26	GLN	30	LYS	0.48	hydrogen bonds	4.32×10^{-3}	2.06
27	GLN	31	LYS	0.59	hydrogen bonds	4.74×10^{-3}	2.00

Residue Number	Residue Name	Residue Number	Residue Name	P_{HB}	Interaction Type	$\lambda_{\alpha, \alpha+4}$	Contacts Distance (\AA)
28	ASN	32	GLU	0.61	hydrogen bonds	4.83×10^{-3}	1.96
29	LEU	33	LYS	0.37	hydrogen bonds	2.70×10^{-3}	2.12
30	LYS	34	GLY	0.05	hydrogen bonds	2.23×10^{-3}	2.62
31	LYS	35	LEU	0	hydrophobic	2.02×10^{-4}	3.9
32	GLU	36	PHE	0	hydrophobic	3.46×10^{-4}	5.72

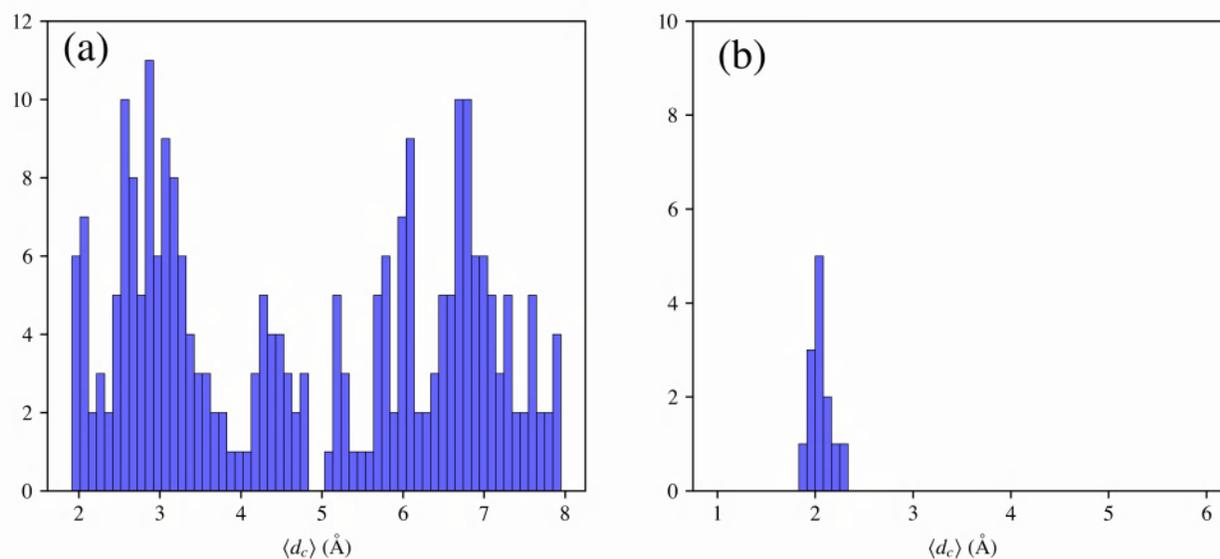


Figure 4.5: Distribution of average contact distance, the shortest interatomic distance between residue pairs in nonbonded native contacts. (a) all residue pairs in nonbonded native contacts; (b) residue pairs $(\alpha, \alpha+4)$ in contact with α -helical hydrogen bonds.

A smaller data set consisting of the inter-residue thermal conductivity and P_{HB} for residue pairs with α -helical hydrogen bonds was used to fit a linear regression relationship, as shown in Figure 4.4b. The Pearson correlation coefficient and $**p$ -value are 0.80 (> 0.3) and 0.001 ($\ll 0.05$), respectively, indicating a stronger correlation than that for the all hydrogen bonded residue pairs. The contact distance is supported to be a factor that can affect thermal transport properties.³⁴ In addition, the histogram of all contact distance and contact distance for residue pairs of $\alpha, \alpha+4$ are shown in Figure 4.5. We can see that the contact distance of all residue pairs has a broad range ($\sim 2 - 8 \text{\AA}$) in Figure 4.5a. It is worth mentioning that the contact distance exhibits a broad distribution (2 - 4 \AA) for the dataset used in Figure 4.4a, whereas the range of those for the smaller dataset used in Figure 4.4b is limited within $2.1 \pm 0.2 \text{\AA}$ (Figure 4.5b).

4.5 Important features of thermal transport through nonbonded contacts

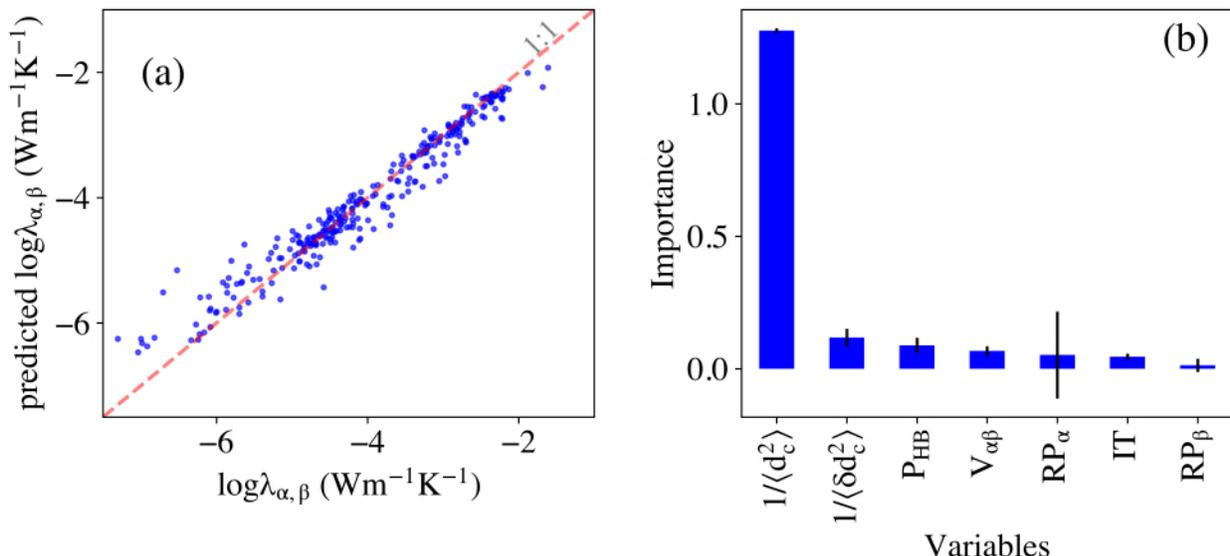


Figure 4.6: (a) Cross plot of $\lambda_{\alpha,\beta}$ and predicted $\lambda_{\alpha,\beta}$ of nonbonded contacts ($\langle d_c \rangle < 6$) using random forest regression model; (b) VIP score plot of variable importance of variables. $\langle d_c^2 \rangle$: contact (shortest) distance of residue α and β ; P_{HB} : hydrogen bonds occurrence probability; $\langle \delta d_c^2 \rangle$: the variance in contacts distance; $V_{\alpha\beta}$: summation of the volumes of residue pair in contacts; RT_α (RT_β): residue type; IT : interaction type.

To get a better understanding of the role of these features in the thermal transport in protein, we applied the random forest regression model to predict the $\log \lambda_{\alpha,\beta}$ values using seven features as predictors. The comparison of calculated and predicted $\log \lambda_{\alpha,\beta}$ together with the feature importances are shown in Figure 4.7 (dataset S) and Figure 4.6 (dataset L). The model of *dataset L* resulted in r^2 -values of 0.95 and 0.89 for training set and testing set, and the corresponding RMSE were 0.24 and 0.4, respectively. Only the contact distance made the considerable contributions to the $\log \lambda_{\alpha,\beta}$ among all the feature variables. The model of *dataset S* resulted in r^2 -values of 0.89 and 0.75 for training set and testing set, and the corresponding RMSE were 0.10 and 0.18, respectively. The top three most important features decreased in the order of $1/\langle d_c^2 \rangle > 1/\langle \delta d_c^2 \rangle > P_{\text{HB}}$. After decreasing the dataset size, we recognized a slight decrease in the prediction performance (R^2) for the training datasets, whereas a bigger decrease for testing set. The difference in feature importances between the two datasets indicates that the contact distance plays a dominant role in determining the value of $\log \lambda_{\alpha,\beta}$ for a wide range of contacts. On the other hand, for short-distance contacts ($\langle d_c \rangle < 4 \text{ \AA}$), both the average squared deviation ($\langle \delta d_c^2 \rangle$) and the hydrogen bonding occurrence probability (P_{HB}) become increasingly important, in consistent with the scaling rule with the harmonic oscillator model for the hydrogen bonding contacts.¹² As a test, we used a much smaller dataset ($\langle d_c \rangle < 2.8 \text{ \AA}$) for the analysis, leading to a very poor prediction performance due to the limitation of the data points. For the smaller threshold for the contact distance, the proportion of hydrogen bonding contacts increase. Consequently, the importance of both $\langle \delta d_c^2 \rangle$ and P_{HB} becomes more pronounced. Moreover, a pairwise correlation analysis of all features and $\lambda_{\alpha,\beta}$ values was performed and their Pearson correlation coefficients were shown in Figure 4.8. We found that the correlation coefficient (r) values between $1/\langle d_c^2 \rangle$, P_{HB} , and $1/\langle \delta d_c^2 \rangle$ were

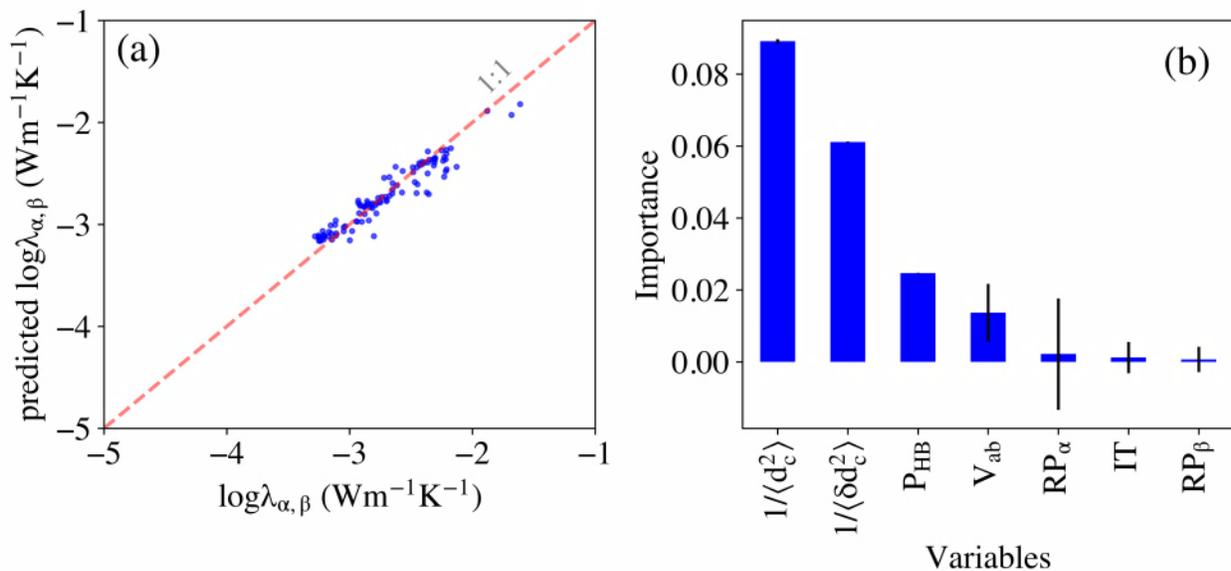


Figure 4.7: (a) Cross plot of $\lambda_{\alpha,\beta}$ between those obtained from the MD simulations and those predicted using random forest model for nonbonded native contacts ($\langle d_c \rangle < 4$) using random forest regression model; (b) VIP score plot of variable importance. $\langle d_c^2 \rangle$: contact (minimum interatomic) distance between residue α and β ; P_{HB} : hydrogen bonds occurrence probability; $\langle \delta d_c^2 \rangle$: the variance of contact distance; $V_{\alpha\beta}$: summation volume of contacts; RT_α (RT_β): residue type; IT: interaction type.

all greater than 0.3 and the corresponding $**p$ values are all below 0.01, indicating a statistically significant correlation between them.

Although this study has examined the significance of feature variables that capture the static and dynamic properties of proteins in thermal transport through nonbonded contacts, it is important to note that the analysis was conducted on a limited number of pairs. Therefore, a more comprehensive model is required to ensure its validity and applicability to a broader range of nonbonded contacts in other proteins.

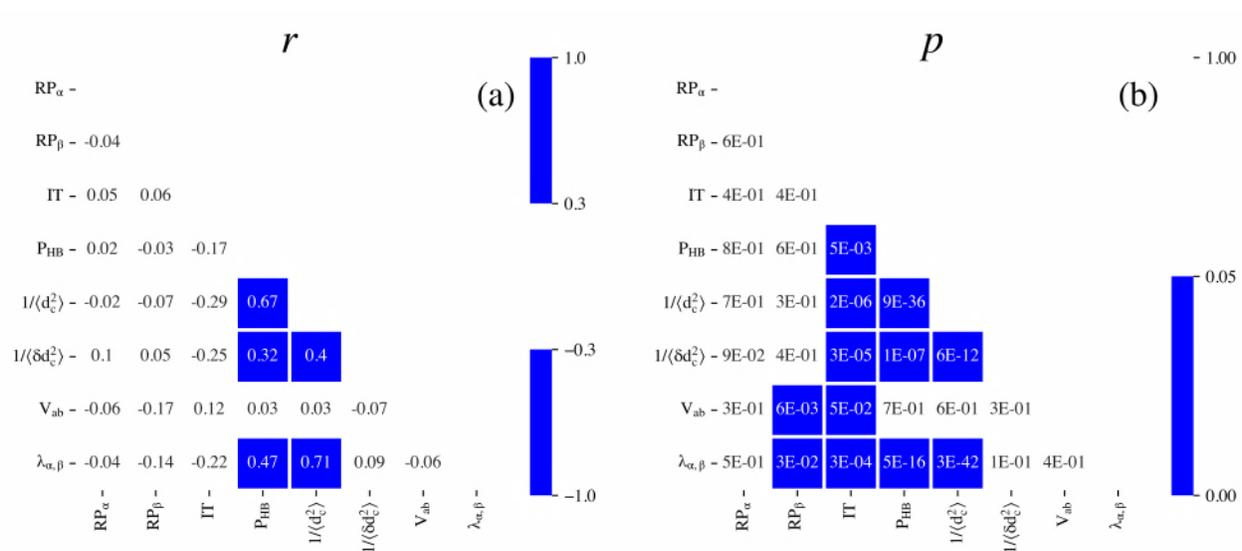


Figure 4.8: Correlation map. (a) Pairwise Pearson correlation coefficients, r ; (b) p -values.

Chapter 5

Energy transport and its function in sensory domain of *BjFixL*

Bradyrhizobium japonicum lives symbiotically with leguminous plants in root nodules, where they fix nitrogen gas from the atmosphere into nitrogenous fertilizer for the plants under hypoxic conditions. Its FixL/FixJ two-component system (TCS) is responsible for sensing diatomic gas molecules (such as O₂, CO, and NO) and regulates the expression of the genes related to nitrogen fixation.¹³⁸ The FixL protein of *Bradyrhizobium japonicum* (*BjFixL*) is comprised of two distinct domains: a tandem Per-Arnt-Sim (**PAS**) domain consisting of PAS-A and PAS-B (**BjFixLH**); and a histidine kinase (**HK**) domain controlling autophosphorylation and phosphotransferase. *BjFixLH* senses the gaseous diatomic molecules through a heme *b* cofactor, whose one side of the axial ligand coordinated with the side chain of amino acid HIS200 of *BjFixLH* protein and the other axial position coordinated with/without gas diatomic molecules.^{139–144} When the heme iron is at its unliganded ferrous/ferric metal state, its conjugated HK domain will be activated and undergo autophosphorylation and phosphotransferase, while binding of strong-field ligands (O₂, CO, NO, cyanide, imidazole, etc.) will inactivate HK domain and impede the expression of nitrogen fixation.^{143,145,146}

Several hypotheses on the signaling mechanism of the PAS-B domain upon ligand binding have been proposed: *spin-state*,^{147–149} *ligand-induced conformational changes*,^{96,141,150–152} *redox potential of heme*,¹⁴⁹ and *affinity (pKa)*.¹⁵³ For the *spin-state hypothesis*, it has been reported that all the high-spin forms (*BjFixL*, met-*BjFixL* and F⁻-*BjFixL*) are active while all the low-spin forms (O₂-*BjFixL*, NO-*BjFixL*, CO-*BjFixL*, imidazole-*BjFixL* and CN⁻-*BjFixL*) are inactive.^{147,154} However, the degree of inhibition strongly depends on the type of ligand, i.e. not fully inhibited (CO and NO) and fully inhibited (oxygen, imidazole and cyano),¹⁵⁴ indicating the limitation of the *spin-state hypothesis* for the comprehensive explanation of the signaling mechanism. Currently, the *ligand-induced conformational changes* mechanism is generally accepted to switch kinase activity on/off by long-range effect of ligand binding.¹⁵⁵ Based on the crystal structural analysis, the averaged positional differences with met-*BjFixL* in the FG loop are measured as 0.1, 0.9, 1.4, 1.7 Å in NO, CN⁻, imidazole, and oxygen binding forms of *BjFixL*s, respectively.¹⁵⁶ A distal ARG220 has a special

selectivity on oxygen with an orientation change breaking its hydrogen bond with heme propionate side chain 7 to form a new one with oxygen to stabilize the oxygen molecule.¹⁵⁷⁻¹⁵⁹ ARG206 has a common positive influence on different ligands binding during signal transduction through interactions with the heme edge.^{146,156,160} ARG208, ILE209 and ILE210 of *Rhizobium meliloti* (*RmFixL*) corresponding to residue 214-216 of the FG loop in *BjFixL*, regulate the activity of the HK domain through their interactions with O₂.^{161,162} Crystallographic data analysis and mutagenesis experiments have suggested that the rearrangement of the hydrogen bonding and salt bridges between the heme propionates and the FG loop upon ligands binding might be responsible for FixL allosteric transition.

Unlike other membrane-embedded histidine kinases,^{96,141,163} *BjFixL* is a water-soluble sensor and forms a homodimer in the cytoplasm.^{161,164,164,165} The sensory domain of *BjFixL*s forms inner and outer surfaces after folding, where the inner surface binds to a heme *b* cofactor and the outer surface makes contacts with the opposing monomer and the flanking J α helix to form a homodimer.¹⁶⁶ The dimer interface in FixLH homodimer is continuous and highly connected, comprising of $\alpha'\alpha$, A α , B β helices, G β , H β , I β sheets and the J α helix.¹⁵⁰ Mutation experiments^{157-162,167,168} pointed out that multiple residues surrounding the heme core and on the dimer interface may play a key role in the signal transmission event according to the decrease of activity in the HK domain at different extents. The functional roles of the dimer interface in other heme proteins have been investigated for signal transduction and modulation,^{150,169} facilitating the cooperative binding of oxygens in Hbl¹⁷⁰ or controlling the enzymatic activity.¹⁷¹ The importance of the potential dimer interface was highlighted by studies of the monomeric form of *BjFixL*Hs.²⁹ As pointed out by a previous study, there are no direct interactions between the PAS-B domain and HK domain in *BjFixL* but a coiled-coil linker (J α helix) connecting them and the reduction of HK domain activity in met-*BjFixL* by residues mutagenesis of J α helix suggested the J α helix could be a potential signal transmission gate from PAS-B domain to the HK domain.¹⁶⁵ However, the precise roles of the dimer interface and the coiled-coil linker region between the PAS-B domain and HK domain have not been detailed yet, due to the lack of structural information at the atomic level for the *BjFixL* dimer so far (Figure 5.1). There remain questions about the role of the dimer interface and coiled-coil linker J α helix in the signal transduction mechanism from the heme-bound PAS domain to the HK domain.

The dynamics and changes in quaternary structures also play roles in the signaling mechanism of the PAS sensor domain^{150,172} Considering the ubiquitous thermal vibrations and conformational fluctuations in biomolecules under physiological conditions^{4,10} ceaseless energy exchange among amino acid residues occurs through interaction between them. It has been demonstrated the analysis of energy flow across nonbonded native contacts in proteins provides a useful tool for characterizing the network of residue-residue interactions.^{28-30,34,58,79,81,170,173-175} To illustrate such transport property of local energy flow, we quantified a measure of the efficiency of local energy transport, denoted hereafter as G , in terms of the autocorrelation function of energy flow. The concept of G was introduced as an analogy with the diffusion constant, D , which is expressed in terms of the velocity autocorrelation function. Importantly, this method can illustrate both static and dynamic effects on biomolecular functions. In fact, recent studies^{28-30,34,58,79-81,170,174} have demonstrated that G serves as a good measure of vibrational energy transfer rate across native contacts. In addition, we observed a scaling relationship between the rate and the inversed value of the variance in the distance between hydrogen-bonded contacts, while the rate is found to scale as a power law in the distance between charged groups.³⁴

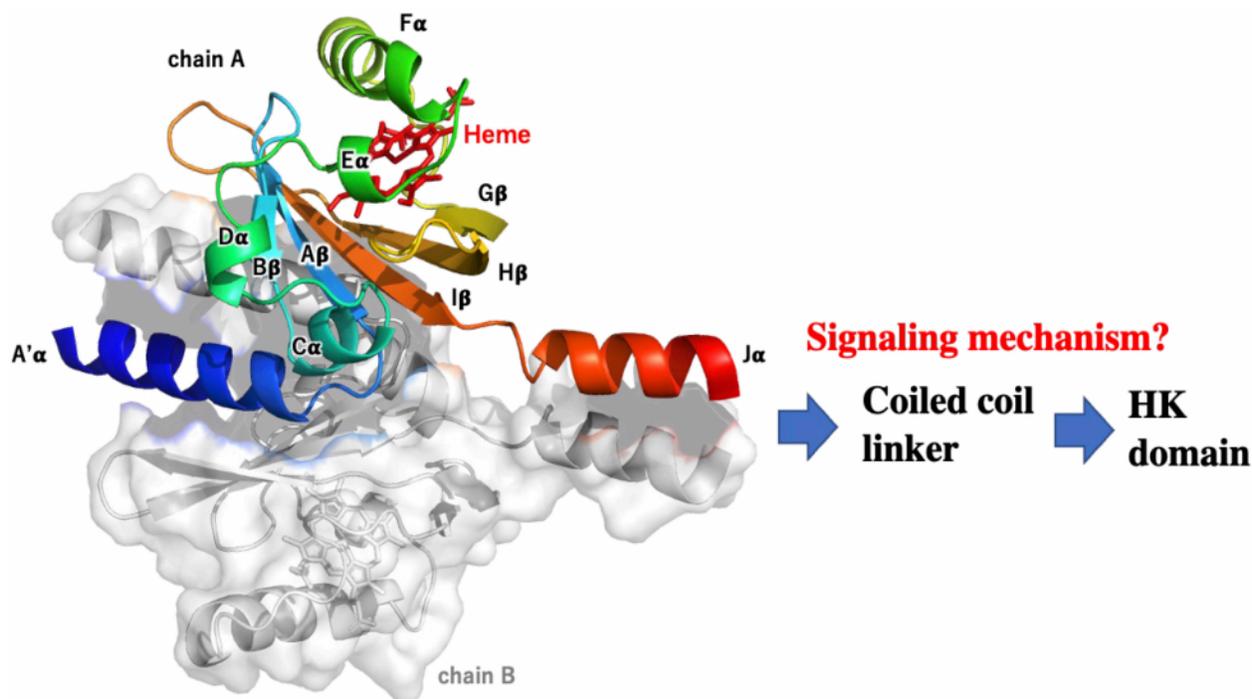


Figure 5.1: Crystal structure of met-*Bj*FixLH dimer. The role of the dimer interface and the coiled-coil linker region in the $J\alpha$ helix is unclear.

In this study, we analyzed the energy flow in the active form (met-*Bj*FixL) and the inactive form (met-*Bj*FixL-IMD) based on the equilibrium molecular dynamics (MD) simulations using the X-ray crystallographic structure of the *Bj*FixL dimer. The pattern of energy flow was illustrated with a network graph based on the local energy flow coefficient, G , for both forms of *Bj*FixL, and the effect of ligand binding was investigated with special attention to the dimer interface. Before we analyze the interaction network of native contacts between amino acid residues, we performed the calculations in three stages: (1) MD simulations to obtain the equilibrium conformational ensemble of each state of dimeric *Bj*FixL protein, (2) calculations of the time series of the energy flow between native contacts in the protein, and (3) the mobility of energy flow was quantified for each contact. We used our original computer program, CURP (CURrent calculation for Proteins, <https://curp.jp>),^{57,59,173} for the second and third stages. Finally, our results are discussed in the context of experimental studies in the literatures.

5.1 Effects of ligand binding and dimerization

Using the CURP program we conducted energy flow analysis of met-FixLH (met-FixLH-imd) model, considering 3485 (4069) residue pairs in native contacts whose nearest interatomic distances are less than 6 Å (Figure 5.2). Overall, the heatmaps of G , which appear in the lower left (chain A) and upper right (chain B), are similar to each other. To more quantitatively evaluate the similarity of properties between the monomers, we conducted linear regression analysis and examined the correlation of G -values between the corresponding residue pairs in chain A and B (Figure 5.3). As a

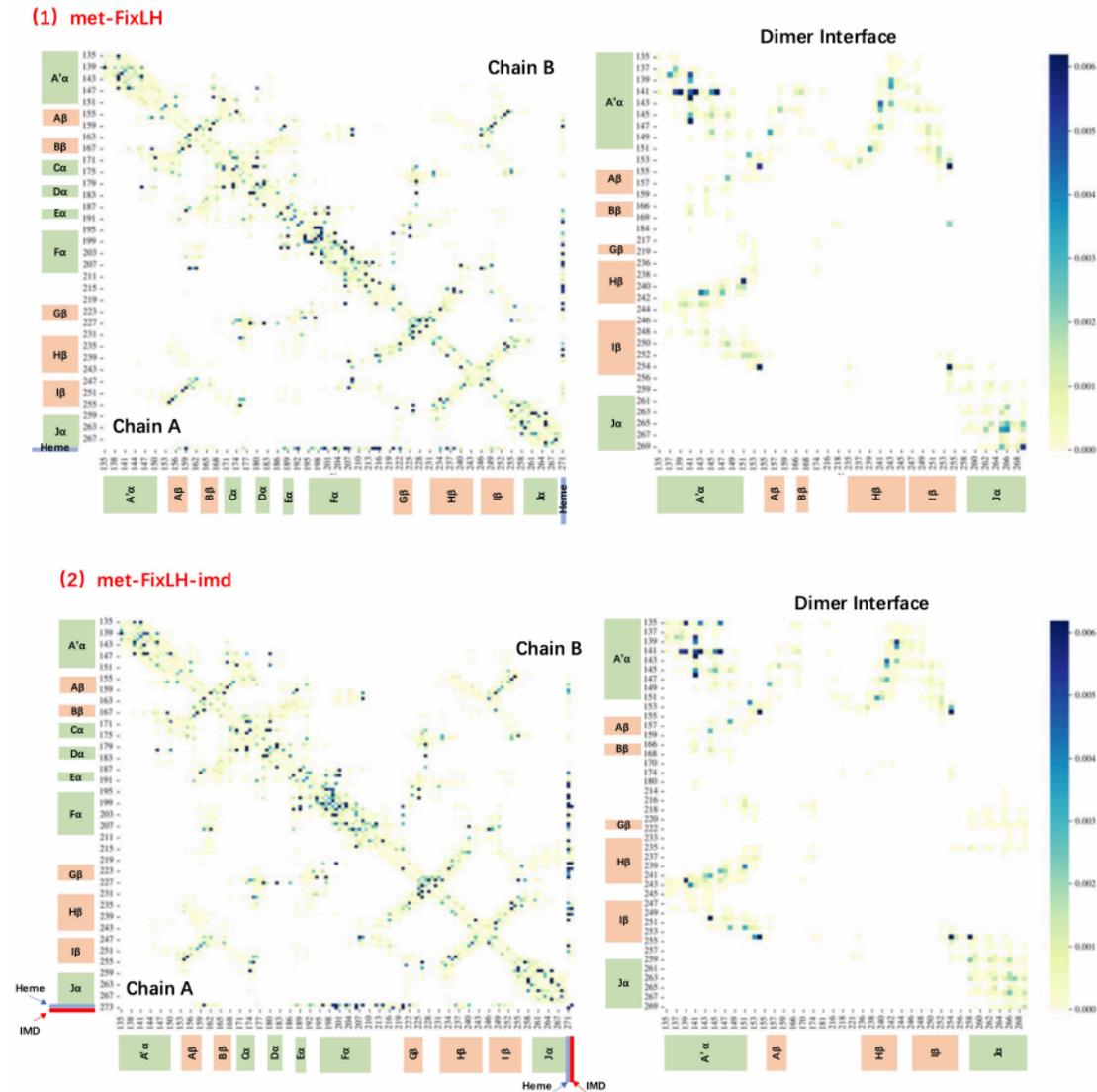


Figure 5.2: Heatmap of G of (1) *met-FixLH* dimer protein and (2) *met-FixLH-imd* dimer protein. The unit of measure on the color bar on the right is $\text{kcal}^2/(\text{mol}^2 \text{ fs})$. Residue numbers and secondary structure names are labeled on the axes. Secondary structures α -helices (green) and β -sheets (orange) are distinguished by different colors.

result, the correlation coefficients are 0.9998 ($R^2 = 0.997$) and 0.9823 ($R^2 = 0.984$) for met-FixLH and met-FixLH-imd, respectively. This observation indicates that chain A and B of met-FixLH-imd behave slightly more independently than met-FixLH, in line with the ligand-induced asymmetry found in the other histidine sensor kinase family.¹⁷⁶ In the previous study, we investigated the effect of ligand binding on the residue interactions within a monomeric unit of PAS-B domain of FixL based on the MD simulation.²⁹ Regarding ligand binding to FixL, it should be noted there is no cooperativity observed between the monomers.³⁰ The correlation coefficient of G -values between the chain A of met-FixLH (this study) and the deoxy FixLH²⁹ is 0.984 ($R^2 = 0.994$), which is slightly smaller than that for the met-FixLH dimer, indicating that the interactions across the dimer interface enhance the synchronization between the monomers, although the essential feature of the sensory domain is determined by the monomeric structure itself.

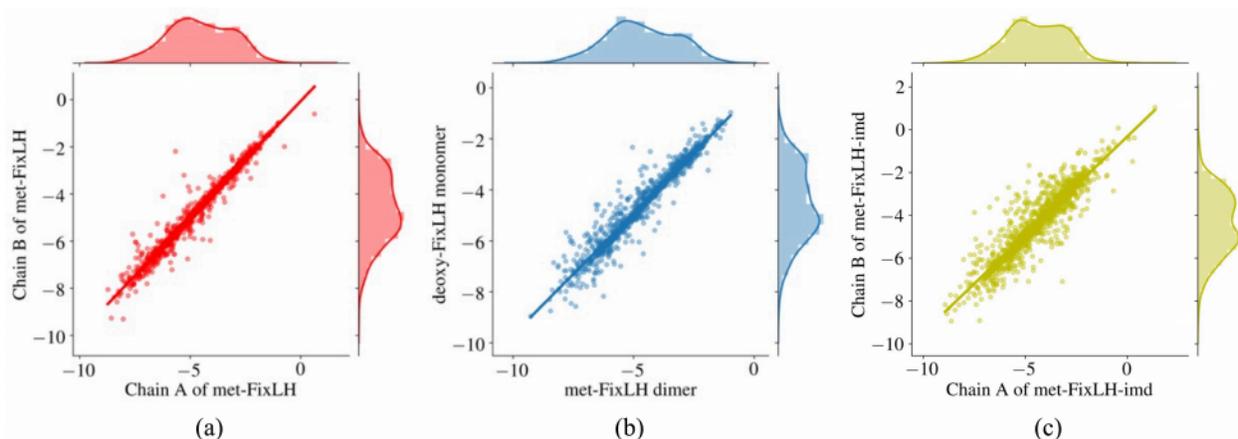


Figure 5.3: (a) Linear regression of G between chain A and chain B in met-FixLH dimer. (b) Linear regression of G between met-FixLH dimer and deoxy-FixLH monomer. G of deoxy-FixLH is from the previous study of our group²⁹. (c) Relationship of G between chain A and chain B in met-FixLH-imd dimer.

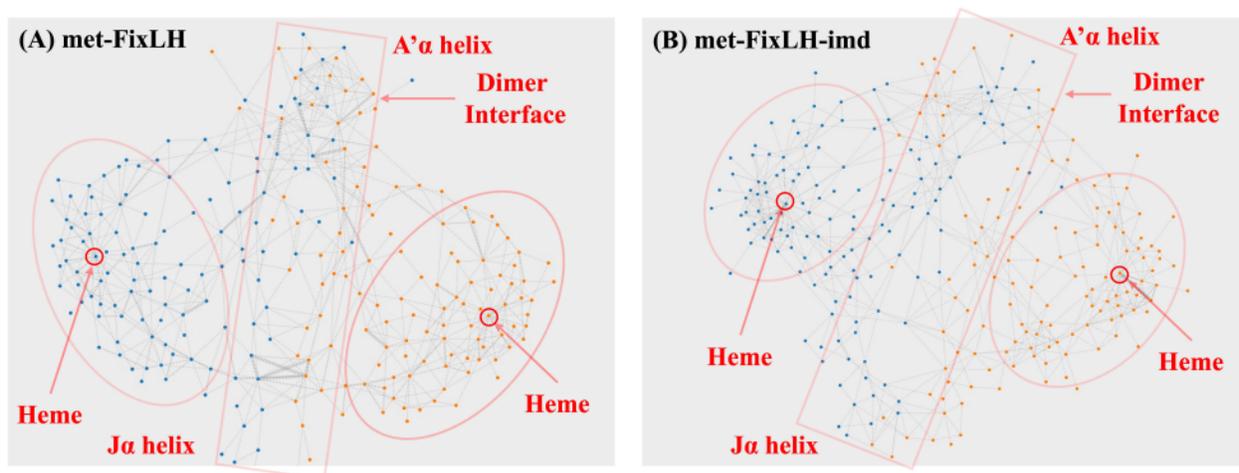


Figure 5.4: Energy exchange network of (a) met-FixLH dimer protein and (b) met-FixLH-imd dimer protein.

The network pattern of G , illustrated in Figure 5.4, shows three important regions where busy traffic of energy flow occurs: two heme-cores and the dimer interface. This indicates that the heme core efficiently serves as the origin of

the signal transduction to the downstream functional domain through the dense network of vibrational energy transfer. Similarly, intensive energy exchange across the dimer interface indicates stable connections between monomers, which is consistent with its low dimer-monomer dissociation equilibrium constant and the continuousness feature of FixLH¹⁵⁰. Also, the strong interactions on the dimer interface may imply its potential role in signal transduction as a bridge between the heme core and the HK domain of FixL domain.

5.2 Vibrational energy transfer pathways and residue interaction network

Important regions with busy energy traffic in met-FixLH and met-FixLH-imd are detailed in Figure 5.5 A(1) and B(1), respectively. The strong interactions are mainly observed on the dimer interface and around the heme core, where the heme (and heme ligand) is wrapped by the shell of EF loop, F α , FG loop, G β , and H β .

Around the heme core, we found busy traffic of vibrational energy transfer in two regions: residue groups in direct contact with heme, and those having no direct contact with heme. **Region (1)**, is consisted of MET192 (EF loop), HIS200(F α), TYR203(F α), ARG206(F α), HIS214(FG loop), ARG220(G β), VAL-222(G β), and ILE238(H β) (Figure 5.5 A(R1)) and ASP196(F α), ARG199(F α), HIS200(F α), TYR203(F α), ARG206(F α), ASP212(FG loop), ARG220(G β), LEU236(H β) (Figure 5.5 B(R1)). Residues HIS200, TYR203, and ARG206 of F α helix are found to have strong interactions with the heme group in both met-FixLH and met-FixLH-imd forms. The strong interaction between HIS200 and heme is due to the Fe-N coordination bonding. The proximal ARG206 was regarded as an important residue strongly influencing affinity and regulation through interaction with HIS214 of the FG loop in the *BjFixL* active state of the heme propionate 6 in its inactive state.^{156,158,160,168,177} Few studies investigated the role of TYR203 in the signaling process, but its potential role was mentioned in the structural analysis because it is as close to the heme as ARG206 is.¹⁷⁸ HIS214 in met-FixLH and ASP212 in met-FixLH-imd of FG loop are found to have strong contacts with heme. HIS214 forms a hydrogen bond with heme-propionate 7,¹⁵⁹ together with ARG206 to form a stable triangular interaction in met-FixLH. ASP212 forms a salt bridge with ARG206,¹⁵⁸ together with heme to form another stable triangular interaction in met-FixLH-imd. A distal ARG220 was found to have strong interaction with heme (and imidazole) both in met-FixLH and met-FixLH-imd. It has been reported that the guanidinium group of ARG220 of the ferrous *BjFixLH* changes its orientation upon oxygen binding due to its strong affinity to oxygen.^{46,155,159,168,177,179} As a result, its hydrogen bond with heme propionate 7 is broken and a new hydrogen bond is formed with the oxygen molecule stabilizing the bound state of the oxygen ligand.

Region (2) of residues having no direct contact with heme is also recognized. TYR207 and GLU240 form hydrogen bonding in the crystal structure and its strength varies between oxy and deoxy forms of *SmFixL*.¹⁸⁰ A stable triangle is formed among ARG-208(F α), HIS162(B β), and GLU-246(I β). It is noteworthy that a vibrational energy transfer pathway in met-FixLH (Figure 5.5 A) is spanned all the way from G β (VAL222) to the junction between the PAS core and the J α helix via H β (MET234, HIS235), I β (VAL253, ARG254), and the highly conserved DxT motif,¹⁸¹ (ASP-LEU-THR). It has been suggested that the hydrophobic residues, ILE215, VAL222, MET234, LEU236, ILE238, and VAL253 form a signal transduction pathway from heme to the J α helix. In fact, these hydrophobic residues have

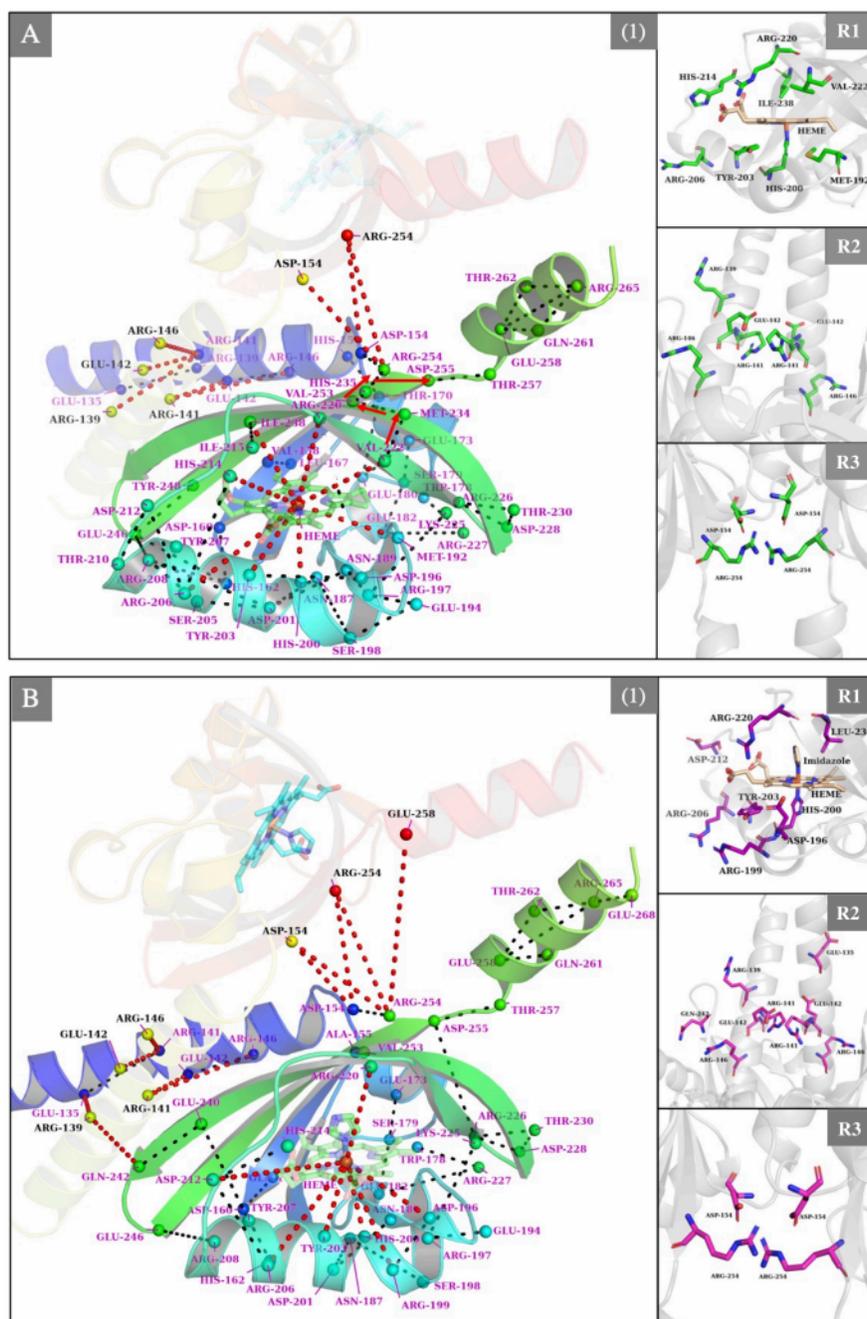


Figure 5.5: The strong interactions, represented by dash lines, are defined by the residue pairs with G values $> G_{\text{met}} + 2\sigma_{\text{met}} = 6.58 \times 10^{-3}$ [kcal²/(mol² fs)], where $G_{\text{met}} (= 2.42 \times 10^{-4})$ and $\sigma_{\text{met}} (= 3.17 \times 10^{-3})$ [kcal²/(mol² fs)] are the mean and the standard deviation of the G -values of all residue pairs, respectively. (A) met-FixL dimer, (B) met-FixLH-imd dimer. The interactions involved with the heme and the dimer interface are, shown in red and distinguished from the other interactions, shown in blue. Panels A/B, R1, R2, R3, show important amino acid residues in the regions around the heme ligand and on the dimer interface; close to the N-terminus; and close to the C-terminus, respectively. It should be noted that the connections between the imidazole ligand and residue HIS200, ARG206, ARG220 and LEU236 are not shown in Panel B1.

been found not only in *BjFixL* but also in the equivalent sites of the other PAS domains.^{178,182} In contrast, we cannot recognize such a pathway for met-FixLH-imd (Figure 5.5 B). Although the structural change between met-FixLH and met-FixLH-imd is not so large, the vibrational energy transfer pathways are markedly different between them. Only heme-ARG220 interaction is common between met-FixLH and met-FixLH-imd. The heme-protein interactions are almost completely localized in the F α helix and the FG loop.

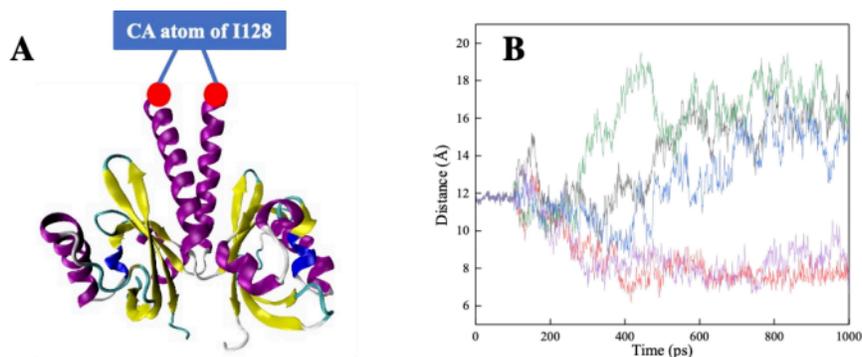


Figure 5.6: Molecular dynamic simulations of met-*BjFixLH* without force restraints on N-terminal of A' α helix. (A) The crystal structure model of MD simulations. (B) The time evolution of distance between carbon atom (CA) of residue ILE128s.

Strong interactions across the dimer interface are mainly found in two regions: (3) N terminus and (4) C terminus.

Region (3), includes charged residue pairs with ARG139, ARG141, and ARG146, for met-FixLH and met-FixLH-imd, respectively (Figure 5.5 R2 and R3). It should be noted that this region is close to neither PAS core nor heme. To save computational time, we imposed harmonic restraints on each C α atom of ILE128 of chains A and B, and excluded PAS-A domains in the calculations. Without such restraints, the pair of A' α helices undergo considerable fluctuations (Figure 5.6). Although the role of the *BjFixL* PAS-A domain remains unclear,¹⁶⁵ a possible role, speculated from the study of *SmFixL*,¹⁷² may be stabilization of the dimer interface of *BjFixL*.

In **Region (4)**, located in the coiled-coil linker, we recognize four residue-residue interactions among ($2 \times 2 =$) 4 charged residues, ASP154 and ARG254 of chain A and B (Figure 5.5 A(R3) and B(R3)). On the downstream of this region, we also recognized a densely connected intra-monomer network of vibrational energy transfer pathways with ASP255, THR257, GLU258, GLN261, THR262, ARG265, GLN267 and GLU268, in line with the site-directed mutagenesis on the coiled-coil linker that exhibited marked decrease of the kinase activity of *BjFixL*.¹⁶⁵

5.3 Reorganization of vibrational energy transfer pathways upon ligand binding

Superposition of the x-ray crystallographic structures of met-FixLH and met-FixLH-imd (Figure 5.7 B) shows a movement of the FG loop, with an average positional displacement of 1.4 Å,¹⁵⁶ together with a slight movement of the G β sheet and the reorientation of heme propionate 6 and 7 side chains. We speculate that this structural change induced

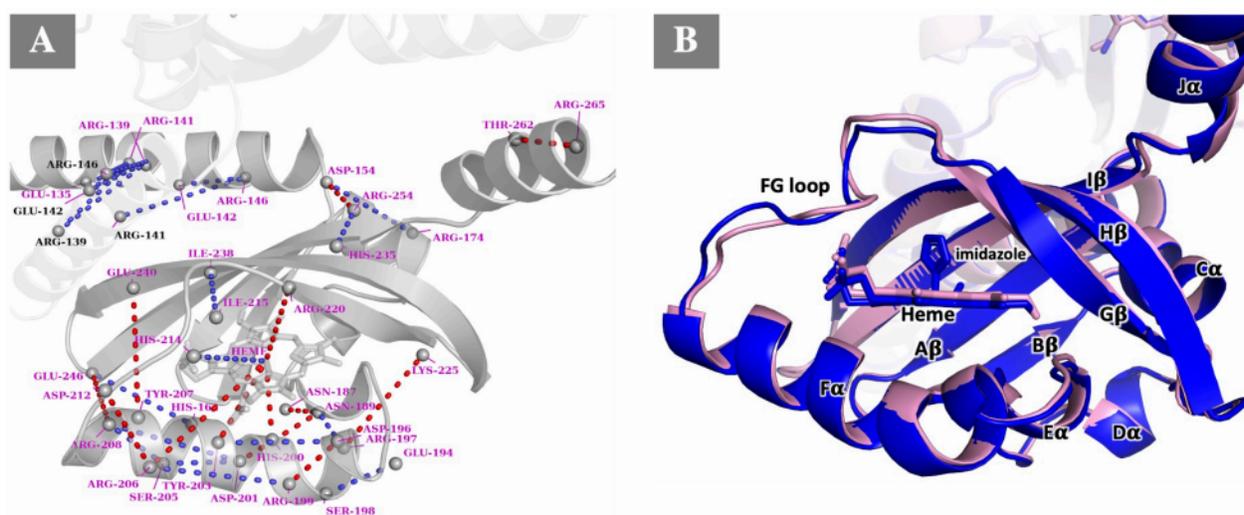


Figure 5.7: Reorganization of the network of vibrational energy transfer pathways. Strengthened (Weakened) paths upon ligand binding with $r\Delta G > (<) 0.5$ are indicated by dotted lines in red (blue). The residue name and residue number (colored in purple and black) represent the residues from chain A and chain B, respectively.

the reorganization of the vibrational energy transfer network between heme and the FG loop or $G\beta$ sheet, leading to the modulation of signal transmission to the downstream of PAS-B domain. To evaluate the effect of ligand binding quantitatively, we calculated the difference (ΔG) and the relative difference ($r\Delta G$) of the G -values for residue pairs in native contact between met-FixLH (G_{met}) and met-FixLH-imd (G_{imd}), where $\Delta G = G_{\text{imd}} - G_{\text{met}}$ and $r\Delta G = \Delta G / G_{\text{met}}$. In Figure 5.7 A, we illustrated the reorganization of the network of vibrational energy transfer pathways, where those with increased G are shown in red and distinguished from those with decreased G , which are shown in blue. Ligand binding to the heme weakened (enhanced) the interaction between heme and HIS214 (HIS200, TYR203, ARG206, and ARG220). The intra-helical interactions within $F\alpha$ were decreased, while those between the $F\alpha$ helix and the nearby secondary structural units, such as ARG208-GLU246, ARG206-ASP212, TYR207-GLU240, ASN189-HIS200, and ARG197-LYS225, were increased. Across the dimer interface, we recognized that the ligand binding induced considerable weakening of the monomer-monomer interactions involving charged residues ASP154, GLU135/142, ARG139/141/146/174 (ARG254) in the N(C)-terminus of the PAS-B domain. In addition, two residue pairs on the dimer interface, ASP154-ARG254 and THR262-ARG265, exhibit increased G -values by ligand binding, where the latter of which belongs to the coiled-coil linker. In summary, ligand binding enhanced the interactions between $F\alpha$ and heme or the terminal regions of $G/H/I \beta$ sheets, while it weakened those within the $F\alpha$ helix, β sheet, and those across the dimer interface. These observations suggest that the signal transmission from the heme core to the coiled-coil linker may be mainly blocked at the bridge of $G/H/I \beta$ -sheets, together with the synergetic reorganization of the compact networks constructed by multiple highly conserved amino acid residues.

Chapter 6

Structural and dynamical characterization of CP12 protein

Unlike folded domains or regions with well-defined three dimensional structures, intrinsically disordered regions (IDRs) in proteins, exhibit high flexibility under physiological conditions. Thereby, these regions lack stable secondary structures such as alpha helices or beta sheets. Structural analysis of IDRs within proteins is a long-standing challenge in structural biology due to their considerable flexibility and spatiotemporal heterogeneity.

In this chapter, we present a novel method for characterizing the conformational ensemble of IDRs and illustrate how it was applied to the homodimer of the chloroplast protein (CP12) derived from *Thalassosira pseudonana*. The computational process is divided into three steps: (1) AlphaFold-Multimer is used to obtain an initial structure guess based on the amino acid sequence; (2) we use harmonic-restrained MD simulations to obtain model that fit to the experimental SAXS and EPR/DEER data; and (3) refine the conformational ensemble using restrained-ensemble MD simulations based on the DEER and SAXS data.

6.1 AlphaFold2 model

The figures in Figure 6.1 display the top five models of CP12 homodimer for the WT, S46C, and S56C mutants. We recognize elongated forms (WT-1, 3-5) and an over-folded triangular form (WT-2). According to Shao et al.,¹⁰⁵ the CP12 dimer assumes an elongated cylindrical shape with kinks, as evidenced by the small angle X-ray scattering analysis. A representative WT model was chosen from the five candidates, as shown in Figure 6.2: one monomer is shown in cartoon representation, with each residue colored according to its per-residue confidence score (pLDDT), while the other is in surface representation. The helical regions were modeled with a high level of confidence, with the exception of the C-terminal helices and some segments of the second helices. In contrast, the random coils, which are likely to be disordered regions, were modeled with low level of confidence scores.

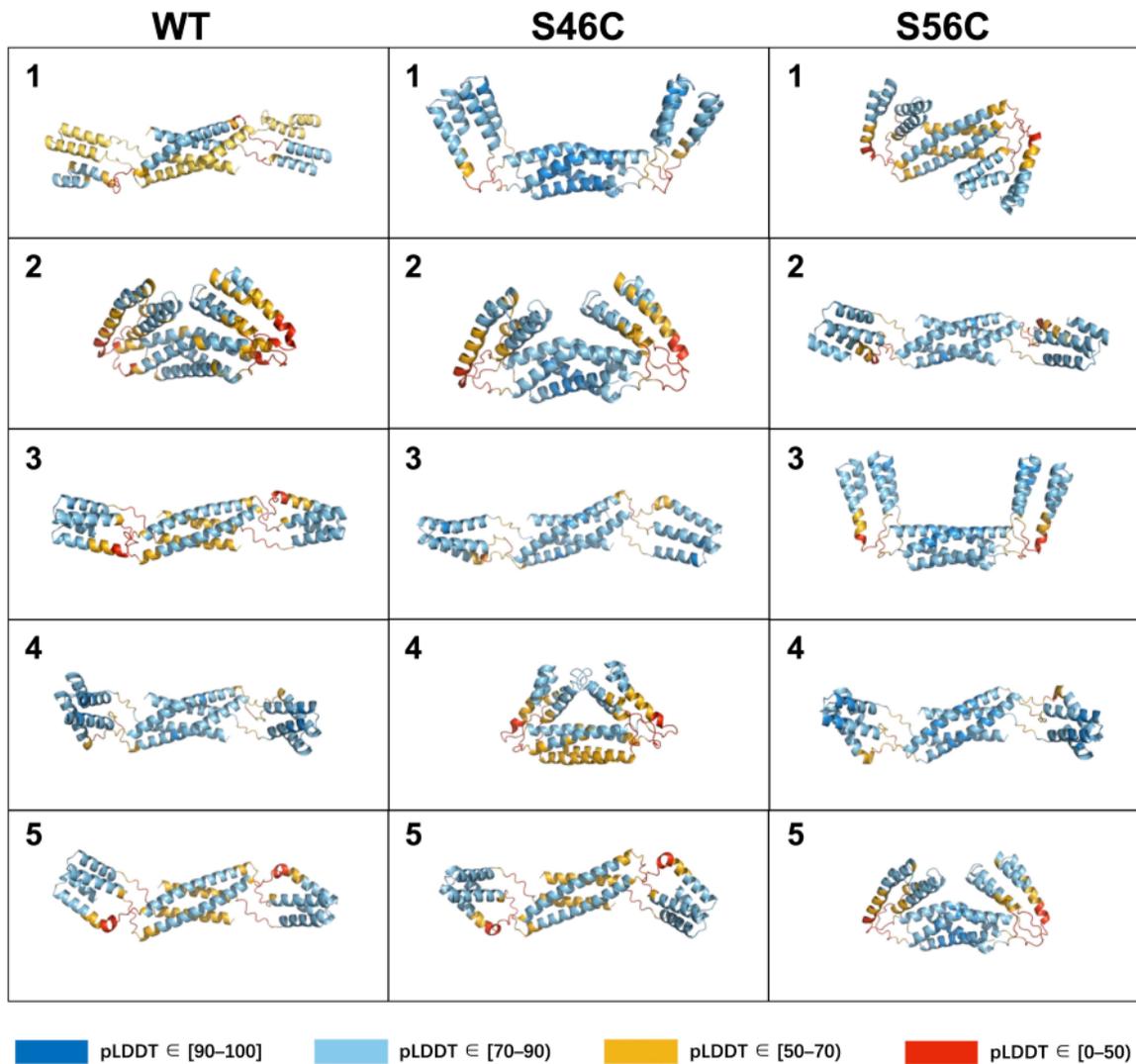


Figure 6.1: Top 5 homodimer models for WT, S46C and S56C mutants are shown with each residue being colored according to its per-residue confidence score (pLDDT): Blue (high), cyan (high medium), yellow (low medium) and red (low).

The presence of both globular folded and potentially disordered regions is consistent with the Kratky analysis conducted during the SAXS experiment.¹⁰⁵ The AF2 model predicted an antiparallel dimer with the characteristic coiled-coil region made up of residues 46-82, including 45-75 with particularly high pLDDT. It has been indicated that the area encompassing residues 46-82 has a strong tendency towards coiled-coil organization.¹⁰⁵ The secondary structure elements of the model were estimated using YASARA View.¹⁸³ The estimated percentages were 79.1% for helix, 2.5% for turn, 18.4% for random coil, and 0% for β sheet. It should be noted, however, that the measurement of circular dichroism spectra reported a 10% occurrence of β -sheet.¹⁰⁵ The two AWD_VEEL motifs and two cysteine residues are considered to be highly conserved, and they are marked with sticks and spheres, respectively. The AF2 model does not contain any inter- or intra-molecular disulfide bridges between two cysteine residues, in line with the experimental results.¹⁰⁵

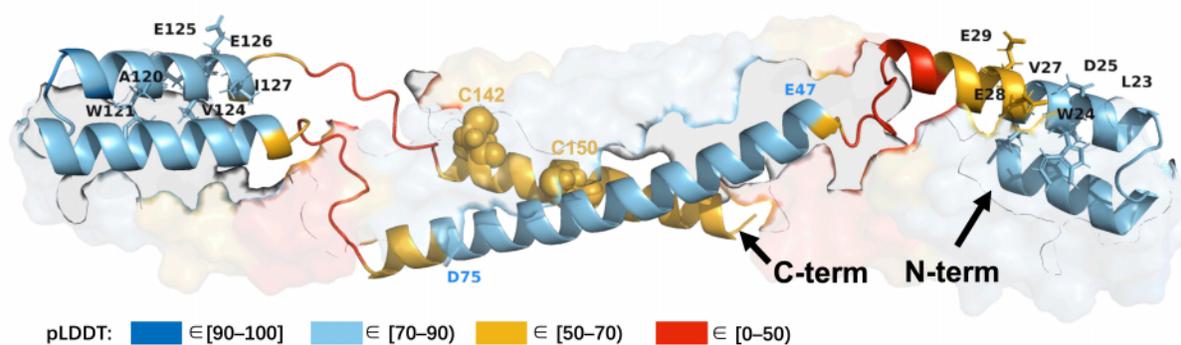


Figure 6.2: AlphaFold2 model of wild-type *Thalassosira pseudonana* CP12 dimer. One monomer is shown in cartoon representation using a color scheme based on confidence measure (blue: high, cyan: high medium, yellow: low medium, red: low), and the other monomer presents a translucent surface format. Two AWD_VEEL motifs and two cysteine residues (C142 and C150) are shown with sticks and spheres, respectively.

The experimental data (Figure 6.3 A) was compared with the estimated SAXS curve based on the AF2 model. The agreement of the intensity in the low- q region, ($q < 0.15 \text{ \AA}^{-1}$), was satisfactory, however that in the high- q region, ($q > 0.15 \text{ \AA}^{-1}$) was unsatisfactory. According to the SAXS measurement¹⁰⁵, the AF2 model's radius of gyration (R_g) was found to be 35.7 \AA , which is smaller than the experimental value of $R_g (38.2 \pm 0.4 \text{ \AA})$. The MTSL-MTSL distance distribution was analyzed using the AF2 model for various CP12 variants, as shown in Figure 2.8 C of the Amber MD analysis. The prediction model and the DEER experiment did not show any overlap for the distance distribution of S39R1 and C150R1 spin pairs, indicating a limited applicability of AF2 to such proteins with IDRs.

6.2 Molecular dynamic simulations

The AF2 model underwent further refinement through restrained molecular dynamics (MD) simulations. Before we start the harmonic restrained molecular dynamics (MD) simulations, we modified the AF2 model: The two C-termini of the dimer were pulled apart from its coiled-coil region using PyMOL software (Figure 2.8 B). After the harmonic restrained molecular dynamics (MD) simulations, we conducted the restrained-ensemble MD (reMD) simulations

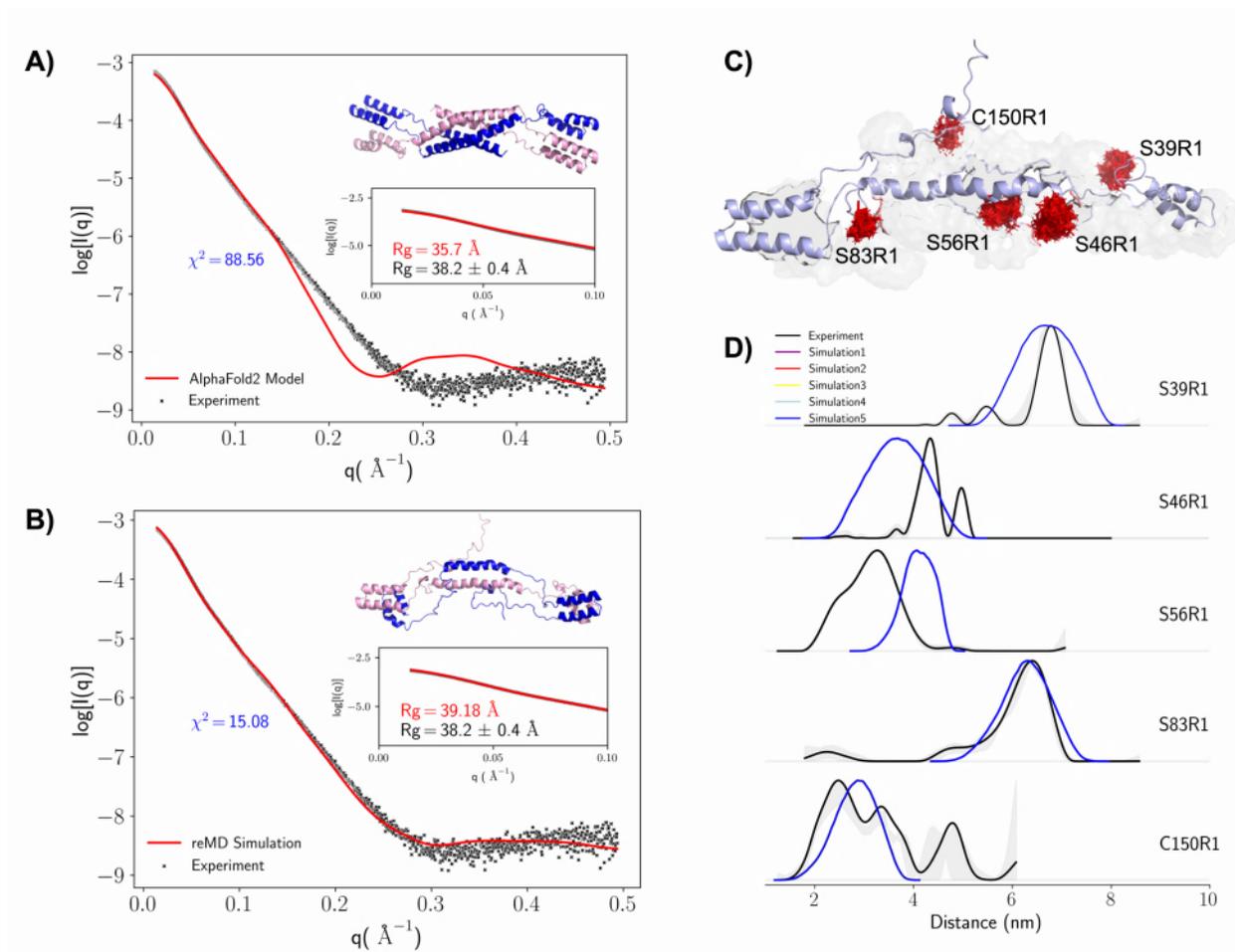


Figure 6.3: (A) Comparison of SAXS curves between AlphaFold2 model and experiment; (B) Comparison of SAXS curves between initial structure of restrained-ensemble molecular dynamics (reMD) simulation and experimental data; In both (A) and (B), an insert figure shows the SAXS curve at low q region ($0 - 0.1 \text{\AA}^{-1}$) and gives the R_g for both experiment and predicted model. (C) Initial structure of CP12 with 25 copies of R1 spin labels for the production run of reMD simulations; (D) Comparison of experimental and calculated distance distribution between the spin labels on 5 the residue pairs. The SAXS discrepancy between simulated and experimental data, χ^2 , for both (A) and (B), are marked in blue text.

(Figure 2.8 A). The estimated SAXS curve based on the initial model for the reMD simulations exhibits better alignment with the experimental curve than the AF2 model (Figure 2.8 B). The model will be hereafter referred to as the reMDini model. The YASARA View¹⁸³ was used to determine the estimated percentages of the secondary structure elements in the reMDini model: The model consisted of 50.6% α -helix, 2.1% β -sheet, 12.0% turn, and 35.3% random coil. The circular dichroism spectrum measurements indicated that the contents of α -helix and random coil were in good agreement, with values of 32-50% and 27-48%, respectively. The C-termini of the AF2 model exhibited significant disorder, while the well-folded long α -helices in the AF2 model underwent partial unfolding to become short α -helices in the reMDini model. Twenty-five copies of the all-atom R1 spin labels were affixed to residues S39, S46, S56, S83, and C150 in both monomers of the reMDini model (Figure 6.3 C). Subsequently, five independent reMD simulations were performed with different random number seeds. The distance distributions of 5 spin label pairs were analyzed using these reMD trajectories and compared with experimental ESR/DEER data (Figure 6.3 D). It should be noted that the distance distributions derived from the five distinct trajectories were remarkably similar to each other, indicating that the sampling of the distance distributions in the reMD simulations was well-converged. Each of the simulated distance distribution was, then, compared with the corresponding one obtained by the EPS/DEER experiment.

The simulated distributions for the spin pairs, S39R1, S83R1, and C150R1, exhibit significant overlap with the corresponding experimental distributions, whereas, only partial overlaps are observed for the spin pair, S46R1 and S56R1. The S56R1 spin pair exhibits a significantly broader distance distribution in the experiment than that obtained in the simulation. The experimental distance distribution of the S46R1 spin pair shows two peaks that are absent in the simulated distance distribution. Accordingly, the ESI-MS experiment¹⁰⁵, partial dissociation is likely to occur in the CP12 homodimer, indicating that the coiled-coil region may be much more flexible in an actual solution environment than in the simulations environment.

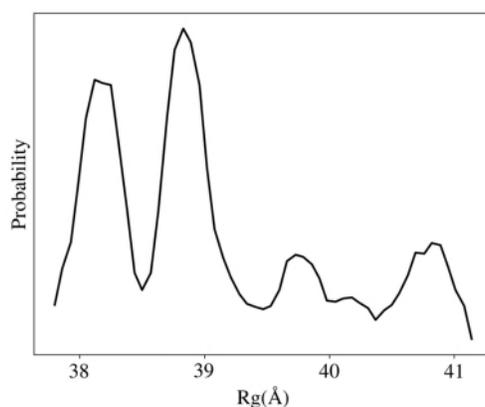


Figure 6.4: Probability distribution of radius of gyration (Rg). We used of 20, 000 sampled structures from reMD simulations.

To conduct a thorough assessment of the reMD sampling, a total of 20,000 structures were obtained as an ensemble structure pool with 4,000 structures extracted from each reMD trajectory. A GAJOE analysis was conducted to obtain the distribution of radius of gyration (Rg) for the 20,000 structures. The resulting distribution is shown in Figure 6.4.

GAJOE is a genetic algorithm-based program in ATSAS package used for selecting a set of models that accurately represent experimental small-angle X-ray scattering (SAXS) data.¹⁸⁴ The R_g distribution exhibits two major peaks at approximately 38.1 and 38.9 Å. These values are in closer agreement with the experimental data (38.2 ± 0.4 Å) than the initial model obtained by using AF2 (35.7 Å). We may conclude that MD simulations are effective in the modeling of proteins with IDRs.

In summary, the simulated SAXS curve and inter-label distance distributions of our CP12 model were in good agreement with the experimental data, indicating that a combination of deep learning based algorithms like AlphaFold-Multimer, MD simulations, SAXS, and EPR/DEER experiments would provide a new possibility for structural characterization of IDRs and their complexes.

Chapter 7

Conclusions and future Directions

7.1 Conclusions

This dissertation aims to develop a new formalism to study the energy transport, heat transport, and signaling in protein systems, based on the linear response theory using equilibrium MD simulations.

Proteins are crucial macromolecules in living organisms, performing various essential functions. In Chapter 1, the history of protein is reviewed, starting from the early discovery and initial understanding of proteins, to the development of modern scientific techniques that have enabled us to gain deeper insights into the structure and function of proteins. The structural composition rules of proteins and features were introduced. Then, the current methods, theories, and techniques used to investigate the energy and heat transfer in proteins was summarized and introduced, along with important conclusions regarding the energy transfer properties of short peptides and proteins.

In Chapter 2, starting from the Green-Kubo transport theory, the atomic expression of heat current and energy flow that is applicable to all density and situation materials under equilibrium state was derived. Then, to get a better understanding of the nature of transport properties in proteins, a linear-homopolymer-like model was introduced to divide the proteins into small pieces by residues. The local thermal transport properties at residue level was considered and calculated by the time-integrated ACFs of the partial heat currents. The interactions between two neighboring residues results in the independent movement and behavior between each other. Cross-correlation correction was introduced and employed to correct the overestimated contributions from partial heat currents to the total heat current for the entire protein. Two new concepts, *inter-residue thermal conductivity* and *inter-residue energy conductivity*, that can represent the amount of heat and energy transferred per unit time at steady-state.

In the Chapter 3 and Chapter 4, the thermal transport properties of α -helical proteins, using HP36 as an example, were investigated based on linear response theory using equilibrium molecular dynamics (MD) simulations. In the Chapter 3, the calculated thermal conductivity (λ) of the entire protein was found to be 0.26 ± 0.01 W/(m·K) using the

AMBER ff19SB protein force field and the rigid TIP3P water model. This value was in close agreement with the thermal conductivity obtained using the flexible TIP3P water model, suggesting that the choice of water model parameters does not significantly affect the inherent thermal transport properties of proteins. To explore the local heat transport properties within the protein's interior, the protein molecule was divided into 36 amino acid residues. A theoretical model known as the linear-homopolymer-like model was introduced for analysis. In this model, it was assumed that heat flow predominantly occurs along the polypeptide backbone and within each individual amino acid residue. Furthermore, it was assumed that the cross-correlation of partial heat currents between different regions is limited to short-range interactions. Remarkably, the model successfully reproduced the exact value of the protein's thermal conductivity, as derived from the total heat current, with an error of only 1%. Notably, the analysis revealed a distinct dependence of the residue-wise thermal conductivity on the type of amino acid residue. The thermal conductivity values decreased in the following order: charged residues, polar residues, and hydrophobic residues. This observation suggests that different residue types contribute differently to the overall heat transport within the protein.

In the Chapter 4, the thermal conductivity of peptide bonded residue pairs are found to be greater than those of nonbonded residue pairs. A strong correlation was observed between the thermal conductivity and energy conductivity of nonbonded residue pairs, despite they are representing different transport properties. Additionally, an empirical linear equation was derived to establish a relationship between them. In general, the $\lambda_{\alpha,\beta}$ values of nonbonded contacts in HP36 exhibited a decreasing trend in the following order: hydrogen bonds > π -stacking > electrostatic > hydrophobic. The hydrogen bond formation capacity during MD simulations demonstrated a significant role for $\lambda_{\alpha,\beta}$ values of hydrogen bonding contacts. In order to elucidate the factors influencing thermal transport ability in nonbonded contacts within the complex protein environment, a non-linear regression model was constructed for a wide range of contacts. This model employed random forest regression analysis to investigate the relationship between the thermal conductivity values ($\lambda_{\alpha,\beta}$) and various static and dynamic variables.

In the Chapter 5, the effects of imidazole binding and dimerization on an oxygen sensor protein, *BjFixLH*, were investigated by theoretical computations with the theory of energy flow and conductivity. Using the energy transport network model, the residue-residue and protein-protein interactions of dimeric FixL protein are described. A vibrational energy transfer pathway characterized in the active form of *BjFixLH* protein, could be a possible signaling pathway. Upon imidazole bound with heme, the allosteric effect induced from heme core is propagated starting from the FG loop, through G/H/I α strands bridge and coiled-coil linker region, to the kinase domain. The salt bridge, constructed by four charged residues (ASP154 and ARG254) and the DxT motif, might be as a gate controlling the signaling from PAS-B domain to a residue cluster of coiled-coil linker region with a compact local interaction network. A local residue cluster with compact interaction network is identified to be the signal transmitter connecting the PAS-B domain and HK domain to stimulus autophosphorylation. The interactions across the dimer interface plays a role in dimerization stabilization and the ligand binding effect will destabilize the dimer interface giving rise to a more independent behavior than before. Although the structural change between two states is very small, through the interaction changes in native contacts between the active form (met-*BjFixLH* dimer) and (met-FixLH-imd dimer), the significant changes in inter-residue interactions are characterized by the energy exchange network. The strengthened interactions between heme and F α

helix while the weakened interactions between heme-7-propionate-HIS214 and within G/H/I α antiparallel strands upon imidazole binding, indicates the signal regulation process is responsible by a synergetic effect of rearrangement of the compact interactions network constructed by multiple reserved residues, rather than by individual residue. The successful application of the energy exchange network model in the identification of functional residue network and protein-protein interactions shows the promising value in the area of solving the problem in biomolecules integrating with static structural information. This also highlighted the importance of dynamic behavior on biomolecular functions in uncovering the biomolecules functions.

In Chapter 6, we report a new strategy for characterizing the conformational ensemble of IDRs and its application to the *Thalassosira pseudonana* chloroplast protein (CP12) homodimer. The method consists of three stages: (1) Obtaining an initial guess structure based on the amino acid sequence using AlphaFold-Multimer; (2) Harmonic-restrained MD simulations to meet the experimental SAXS and EPR/DEER data; (3) Refinement of the conformational ensemble by using restrained-ensemble MD simulations based on the information from DEER and SAXS experiments. The simulated SAXS curve and inter-label distance distributions of our CP12 model were in good agreement with the experimental data, indicating that a combination of deep learning based algorithms like AlphaFold-Multimer, MD simulations, SAXS, and EPR/DEER experiments would provide a new possibility for structural characterization of IDRs and their complexes.

7.2 Future directions

Inspired by the results in this study, two possible directions are worthy to be investigated further in the future.

1. The secondary structure dependence of thermal transport in proteins.

In proteins, hydrogen bonds participate in stabilizing local structures to form different types of secondary structures. The hydrogen bonds between the main-chain NH and C'=O groups of residue α and $\alpha+4$ result in α -helix, while the hydrogen bonds between the main-chain NH and C'=O groups of different regions of the polypeptide chain result in β -sheet, either parallel or antiparallel types. It is unknown how the different types of secondary structures will affect the thermal transport in proteins.

2. The effect of solvent model on the thermal transport properties in protein. In this study, the thermal conductivity values of HP36 with a typical rigid TIP3P water model exhibited a smaller value (0.26 W/(m·K)) compared to the value using the same flexible TIP3P water model (0.3 W/(m·K)). It raises the interest of "How will the change of water solvent flexibility changes vibration frequency at the water-protein interface, accordingly to influence the thermal transport properties of protein itself"?

Contributions

Published articles

- Tingting Wang, Takahisa Yamato, Wataru Sugiura. Site-selective Heat Current Analysis of α -Helical Protein with Linear-homopolymer-like Model. (2023) *J. Chem. Phys.* 158, 214105.
- Takahisa Yamato, Tingting Wang, Wataru Sugiura, Olivier Lapr v te, and Takahiro Katagiri. Computational Study on the Thermal Conductivity of a Protein. (2022) *J. Phys. Chem. B.* 126, 3029–3036.

Conference talks

- Tingting WANG, Takahisa YAMATO, Wataru Sugiura. Local Thermal Transport in a α -helical Protein. The Physical Society of Japan 2023 Spring Meeting. March 22, 2023.
- Tingting WANG, Alessio BONUCCI, V ronique Receveur-Br chet, Val rie BELLE, Takahisa YAMATO. Structural and Dynamic Characterization of a Highly Flexible Protein by AlphaFold-Multimer and MD Simulations Integrating DEER and SAXS Data. The Physical Society of Japan 2023 Spring Meeting. March 22, 2023.
- Tingting WANG and Takahisa YAMATO. Computational Study on the Ligand Discrimination of Dimeric Sensory Domain of FixL Protein. The Physical Society of Japan 2022 Autumn Meeting. Sep. 14, 2022.
- Tingting WANG and Takahisa YAMATO. Computational Study on the Signal Transduction Mechanism of Dimeric Sensory Domain of FixL Protein. The Biophysical Society of Japan (Chubu). March 31, 2021.

Bibliography

- (1) Brändén, C.-I.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Pub: New York, 1999.
- (2) Toberer, E. S.; Baranowski, L. L.; Dames, C. Advances in Thermal Conductivity. *Annual Review of Materials Research* **2012**, *42* (1), 179–209. <https://doi.org/10.1146/annurev-matsci-070511-155040>.
- (3) Mizutani, Y.; Mizuno, M. Time-Resolved Spectroscopic Mapping of Vibrational Energy Flow in Proteins: Understanding Thermal Diffusion at the Nanoscale. *The Journal of Chemical Physics* **2022**, *157* (24), 240901. <https://doi.org/10.1063/5.0116734>.
- (4) Yamashita, S.; Mizuno, M.; Tran, D. P.; Dokainish, H.; Kitao, A.; Mizutani, Y. Vibrational Energy Transfer from Heme Through Atomic Contacts in Proteins. *Journal of Physical Chemistry B* **2018**, *122* (22), 5877–5884. <https://doi.org/10.1021/acs.jpccb.8b03518>.
- (5) Botan, V.; Backus, E. H. G.; Pfister, R.; Moretto, A.; Crisma, M.; Toniolo, C.; Nguyen, P. H.; Stock, G.; Hamm, P. Energy Transport in Peptide Helices. *Proceedings of the National Academy of Sciences* **2007**, *104* (31), 12749–12754. <https://doi.org/10.1073/pnas.0701762104>.
- (6) Deniz, E.; Valiño-Borau, L.; Löffler, J. G.; Eberl, K. B.; Gulzar, A.; Wolf, S.; Durkin, P. M.; Kaml, R.; Budisa, N.; Stock, G.; Bredenbeck, J. Through Bonds or Contacts? Mapping Protein Vibrational Energy Transfer Using Non-Canonical Amino Acids. *Nature Communications* **2021**, *12* (1), 1–8. <https://doi.org/10.1038/s41467-021-23591-1>.
- (7) Mizutani, Y.; Kitagawa, T. Direct Observation of Cooling of Heme Upon Photodissociation of Carbonmonoxy Myoglobin. *Science* **1997**, *278* (5337), 443–446. <https://doi.org/10.1126/science.278.5337.443>.
- (8) Fujii, N.; Mizuno, M.; Ishikawa, H.; Mizutani, Y. Observing Vibrational Energy Flow in a Protein with the Spatial Resolution of a Single Amino Acid Residue. *The Journal of Physical Chemistry Letters* **2014**, *5* (18), 3269–3273. <https://doi.org/10.1021/jz501882h>.
- (9) Kondoh, M.; Mizuno, M.; Mizutani, Y. Importance of Atomic Contacts in Vibrational Energy Flow in Proteins. *Journal of Physical Chemistry Letters* **2016**, *7* (11), 1950–1954. <https://doi.org/10.1021/acs.jpcclett.6b00785>.
- (10) Mizuno, M.; Mizutani, Y. Role of Atomic Contacts in Vibrational Energy Transfer in Myoglobin. *Biophysical Reviews* **2020**, *12* (2), 511–518. <https://doi.org/10.1007/s12551-020-00681-w>.

- (11) Leitner, D. M.; Buchenberg, S.; Brettel, P.; Stock, G. Vibrational Energy Flow in the Villin Headpiece Subdomain: Master Equation Simulations. *The Journal of Chemical Physics* **2015**, *142* (7), 075101. <https://doi.org/10.1063/1.4907881>.
- (12) Buchenberg, S.; Leitner, D. M.; Stock, G. Scaling Rules for Vibrational Energy Transport in Globular Proteins. *The Journal of Physical Chemistry Letters* **2016**, *7* (1), 25–30. <https://doi.org/10.1021/acs.jpcllett.5b02514>.
- (13) Buchenberg, S.; Sittel, F.; Stock, G. Time-Resolved Observation of Protein Allosteric Communication. *Proceedings of the National Academy of Sciences* **2017**, *114* (33), E6804–E6811. <https://doi.org/10.1073/pnas.1707694114>.
- (14) Helmer, N.; Wolf, S.; Stock, G. Energy Transport and Its Function in Heptahelical Transmembrane Proteins. *The Journal of Physical Chemistry B* **2022**, *126* (43), 8735–8746. <https://doi.org/10.1021/acs.jpccb.2c05892>.
- (15) Sagnella, D. E.; Straub, J. E. Directed Energy “Funneling” Mechanism for Heme Cooling Following Ligand Photolysis or Direct Excitation in Solvated Carbonmonoxy Myoglobin. *The Journal of Physical Chemistry B* **2001**, *105* (29), 7057–7063. <https://doi.org/10.1021/jp0107917>.
- (16) Bu, L.; Straub, J. E. Simulating Vibrational Energy Flow in Proteins: Relaxation Rate and Mechanism for Heme Cooling in Cytochrome c. *The Journal of Physical Chemistry B* **2003**, *107* (44), 12339–12345. <https://doi.org/10.1021/jp0351728>.
- (17) Bu, L.; Straub, J. E. Vibrational Energy Relaxation of “Tailored” Hemes in Myoglobin Following Ligand Photolysis Supports Energy Funneling Mechanism of Heme “Cooling.” *J. Phys. Chem. B* **2003**, *107* (38), 10634–10639. <https://doi.org/10.1021/jp034558f>.
- (18) Kobus, M.; Nguyen, P. H.; Stock, G. Coherent Vibrational Energy Transfer Along a Peptide Helix. *The Journal of Chemical Physics* **2011**, *134* (12), 124518. <https://doi.org/10.1063/1.3574395>.
- (19) Conti Nibali, V.; Morra, G.; Havenith, M.; Colombo, G. Role of Terahertz (THz) Fluctuations in the Allosteric Properties of the PDZ Domains. *The Journal of Physical Chemistry B* **2017**, *121* (44), 10200–10208. <https://doi.org/10.1021/acs.jpccb.7b06590>.
- (20) Kumawat, A.; Chakrabarty, S. Hidden Electrostatic Basis of Dynamic Allostery in a PDZ Domain. *Proceedings of the National Academy of Sciences* **2017**, *114* (29). <https://doi.org/10.1073/pnas.1705311114>.
- (21) Miño, G.; Barriga, R.; Gutierrez, G. Hydrogen Bonds and Heat Diffusion in α -Helices: A Computational Study. *The Journal of Physical Chemistry B* **2014**, *118* (34), 10025–10034. <https://doi.org/10.1021/jp503420e>.
- (22) Kurisaki, I.; Tanaka, S.; Mori, I.; Umegaki, T.; Mori, Y.; Tanaka, S. Thermal Conductivity and Conductance of Protein in Aqueous Solution: Effects of Geometrical Shape. *Journal of Computational Chemistry* **2023**, *44* (7), 857–868. <https://doi.org/10.1002/jcc.27048>.
- (23) *PROTEINS: Energy, Heat and Signal Flow*; CRC Press: Place of publication not identified, 2019.

- (24) Li, H.; Wu, S.; Ma, A. Origin of Protein Quake: Energy Waves Conducted by a Precise Mechanical Machine. *Journal of Chemical Theory and Computation* **2022**, *18* (9), 5692–5702. <https://doi.org/10.1021/acs.jctc.2c00514>.
- (25) Thompson, L. M.; Lasoroski, A.; Champion, P. M.; Sage, J. T.; Frisch, M. J.; Thor, J. J. van; Bearpark, M. J. Analytical Harmonic Vibrational Frequencies for the Green Fluorescent Protein Computed with ONIOM: Chromophore Mode Character and Its Response to Environment. *Journal of Chemical Theory and Computation* **2014**, *10* (2), 751–766. <https://doi.org/10.1021/ct400664p>.
- (26) Fujisaki, H.; Straub, J. E. Vibrational Energy Relaxation in Proteins. *Proceedings of the National Academy of Sciences* **2005**, *102* (19), 6726–6731. <https://doi.org/10.1073/pnas.0409083102>.
- (27) Pandey, H. D.; Leitner, D. M. Vibrational Energy Transport in Molecules and the Statistical Properties of Vibrational Modes. *Chemical Physics* **2017**, *482*, 81–85. <https://doi.org/10.1016/j.chemphys.2016.07.008>.
- (28) Ishikura, T.; Yamato, T. Energy Transfer Pathways Relevant for Long-Range Intramolecular Signaling of Photosensory Protein Revealed by Microscopic Energy Conductivity Analysis. *Chemical Physics Letters* **2006**, *432* (4–6), 533–537. <https://doi.org/10.1016/j.cplett.2006.10.092>.
- (29) Ota, K.; Yamato, T. Energy Exchange Network Model Demonstrates Protein Allosteric Transition: An Application to an Oxygen Sensor Protein. *Journal of Physical Chemistry B* **2019**, *123* (4), 768–775. <https://doi.org/10.1021/acs.jpcc.8b10489>.
- (30) Reid, K. M.; Yamato, T.; Leitner, D. M. Variation of Energy Transfer Rates Across Protein-Water Contacts with Equilibrium Structural Fluctuations of a Homodimeric Hemoglobin. *Journal of Physical Chemistry B* **2020**, *124* (7), 1148–1159. <https://doi.org/10.1021/acs.jpcc.9b11413>.
- (31) Poudel, H.; Leitner, D. M. Locating Dynamic Contributions to Allostery via Determining Rates of Vibrational Energy Transfer. *The Journal of Chemical Physics* **2023**, *158* (1), 015101. <https://doi.org/10.1063/5.0132089>.
- (32) Reid, K. M.; Leitner, D. M. Locating and Navigating Energy Transport Networks Energy Networks in Proteins. In *Allostery: Methods and Protocols*; Di Paola, L., Giuliani, A., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2021; pp 37–59. https://doi.org/10.1007/978-1-0716-1154-8_4.
- (33) Valiño Borau, L.; Gulzar, A.; Stock, G. Master Equation Model to Predict Energy Transport Pathways in Proteins. *The Journal of Chemical Physics* **2020**, *152* (4), 045103. <https://doi.org/10.1063/1.5140070>.
- (34) Reid, K. M.; Yamato, T.; Leitner, D. M. Scaling of Rates of Vibrational Energy Transfer in Proteins with Equilibrium Dynamics and Entropy. *Journal of Physical Chemistry B* **2018**, *122* (40), 9331–9339. <https://doi.org/10.1021/acs.jpcc.8b07552>.
- (35) Agbo, J. K.; Gnanasekaran, R.; Leitner, D. M. Communication Maps: Exploring Energy Transport Through Proteins and Water. *Israel Journal of Chemistry* **2014**, *54* (8–9), 1065–1073. <https://doi.org/10.1002/ijch.201300139>.
- (36) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, Calif, 2000.

- (37) Kubo, R. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems. *Journal of the Physical Society of Japan* **1957**, *12* (6), 570–586. <https://doi.org/10.1143/JPSJ.12.570>.
- (38) Kubo, R.; Yokota, M.; Nakajima, S. Statistical-Mechanical Theory of Irreversible Processes. II. Response to Thermal Disturbance. *Journal of the Physical Society of Japan* **1957**, *12* (11), 1203–1211. <https://doi.org/dj2m3c>.
- (39) Kubo, R. The Fluctuation-Dissipation Theorem. *Reports on Progress in Physics* **1966**, *29* (1), 255–284. <https://doi.org/fq47nr>.
- (40) Gottwald, G. A. Introduction to Focus Issue: Linear Response Theory: Potentials and Limits. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2020**, *30* (2), 020401. <https://doi.org/gsqctq>.
- (41) Zwanzig, R. Time-Correlation Functions and Transport Coefficients in Statistical Mechanics. *Annu. Rev. Phys. Chem.* **1965**, *16* (1), 67–102. <https://doi.org/10.1146/annurev.pc.16.100165.000435>.
- (42) Kaviany, M. *Heat Transfer Physics*; Cambridge University Press: Cambridge, 2008.
- (43) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Computational science series; Academic Press: San Diego, 2002.
- (44) Lipfert, J.; Doniach, S. Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36* (1), 307–327. <https://doi.org/10.1146/annurev.biophys.36.040306.132655>.
- (45) Boldon, L.; Laliberte, F.; Liu, L. Review of the Fundamental Theories Behind Small Angle X-Ray Scattering, Molecular Dynamics Simulations, and Relevant Integrated Application. *Nano Reviews* **2015**, *6* (1), 25661. <https://doi.org/gg7jhx>.
- (46) Buhrke, D.; Hildebrandt, P. Probing Structure and Reaction Dynamics of Proteins Using Time-Resolved Resonance Raman Spectroscopy. *Chemical Reviews* **2020**, *120* (7), 3577–3630. <https://doi.org/10.1021/acs.chemrev.9b00429>.
- (47) Torricella, F.; Pierro, A.; Mileo, E.; Belle, V.; Bonucci, A. Nitroxide Spin Labels and EPR Spectroscopy: A Powerful Association for Protein Dynamics Studies. *Biochimica et Biophysica Acta - Proteins and Proteomics* **2021**, *1869* (7). <https://doi.org/10.1016/j.bbapap.2021.140653>.
- (48) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **1998**, *102* (18), 3586–3616. <https://doi.org/10.1021/jp973084f>.

- (49) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995**, *117* (19), 5179–5197. <https://doi.org/10.1021/ja00124a002>.
- (50) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *Journal of Computational Chemistry* **2004**, *25* (13), 1656–1676. <https://doi.org/10.1002/jcc.20090>.
- (51) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996**, *118* (45), 11225–11236. <https://doi.org/fvftxj>.
- (52) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983**, *4* (2), 187–217. <https://doi.org/10.1002/jcc.540040211>.
- (53) Weiner, P. K.; Kollman, P. A. AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions. *Journal of Computational Chemistry* **1981**, *2* (3), 287–303. <https://doi.org/10.1002/jcc.540020311>.
- (54) Berendsen, H. J. C.; Van Der Spoel, D.; Van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1-3), 43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-e](https://doi.org/10.1016/0010-4655(95)00042-e).
- (55) Nelson, M. T.; Humphrey, W.; Gursoy, A.; Dalke, A.; Kalé, L. V.; Skeel, R. D.; Schulten, K. NAMD: A Parallel, Object-Oriented Molecular Dynamics Program. *The International Journal of Supercomputer Applications and High Performance Computing* **1996**, *10* (4), 251–268. <https://doi.org/10.1177/109434209601000401>.
- (56) Kobayashi, C.; Jung, J.; Matsunaga, Y.; Mori, T.; Ando, T.; Tamura, K.; Kamiya, M.; Sugita, Y. GENESIS 1.1: A Hybrid-parallel Molecular Dynamics Simulator with Enhanced Sampling Algorithms on Multiple Computational Platforms. *Journal of Computational Chemistry* **2017**, *38* (25), 2193–2206. <https://doi.org/10.1002/jcc.24874>.
- (57) Ishikura, T.; Hatano, T.; Yamato, T. Atomic Stress Tensor Analysis of Proteins. *Chemical Physics Letters* **2012**, *539-540*, 144–150. <https://doi.org/10.1016/j.cplett.2012.05.005>.
- (58) Ishikura, T.; Iwata, Y.; Hatano, T.; Yamato, T. Energy Exchange Network of Inter-Residue Interactions Within a Thermally Fluctuating Protein Molecule: A Computational Study. *Journal of Computational Chemistry* **2015**, *36* (22), 1709–1718. <https://doi.org/10.1002/jcc.23989>.
- (59) Yamato, T.; Wang, T.; Sugiura, W.; Lapr evote, O.; Katagiri, T. Computational Study on the Thermal Conductivity of a Protein. *The Journal of Physical Chemistry B* **2022**, *126* (16), 3029–3036. <https://doi.org/10.1021/acs.jpcc.2c00958>.

- (60) Suh, D.; Lee, J. W.; Choi, S.; Lee, Y. Recent Applications of Deep Learning Methods on Evolution- and Contact-Based Protein Structure Prediction. *International Journal of Molecular Sciences* **2021**, *22* (11), 6032. <https://doi.org/10.3390/ijms22116032>.
- (61) Pakhrin, S. C.; Shrestha, B.; Adhikari, B.; Kc, D. B. Deep Learning-Based Advances in Protein Structure Prediction. *International Journal of Molecular Sciences* **2021**, *22* (11), 5553. <https://doi.org/10.3390/ijms22115553>.
- (62) Morgunov, A. S.; Saar, K. L.; Vendruscolo, M.; Knowles, T. P. J. New Frontiers for Machine Learning in Protein Science. *J. Mol. Biol.* **2021**, *433* (20), 167232. <https://doi.org/10.1016/j.jmb.2021.167232>.
- (63) Greener, J. G.; Kandathil, S. M.; Moffat, L.; Jones, D. T. A Guide to Machine Learning for Biologists. *Nat. Rev. Mol. Cell Biol.* **2022**, *23* (1), 40–55. <https://doi.org/10.1038/s41580-021-00407-0>.
- (64) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; Velankar, S.; Kleywegt, G. J.; Bateman, A.; Evans, R.; Pritzel, A.; Figurnov, M.; Ronneberger, O.; Bates, R.; Kohl, S. A. A.; Potapenko, A.; Ballard, A. J.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Clancy, E.; Reiman, D.; Petersen, S.; Senior, A. W.; Kavukcuoglu, K.; Birney, E.; Kohli, P.; Jumper, J.; Hassabis, D. Highly Accurate Protein Structure Prediction for the Human Proteome. *Nature* **2021**, *596* (7873), 590–596. <https://doi.org/10.1038/s41586-021-03828-1>.
- (65) Bryant, P.; Pozzati, G.; Elofsson, A. Improved Prediction of Protein-Protein Interactions Using AlphaFold2. *Nature Communications* **2022**, *13* (1). <https://doi.org/10.1038/s41467-022-28865-w>.
- (66) Evans, R.; O’neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. **2021**. <https://doi.org/10.1101/2021.10.04.463034>.
- (67) Kashio, M.; Tominaga, M. TRP Channels in Thermosensation. *Current Opinion in Neurobiology* **2022**, *75*, 102591. <https://doi.org/10.1016/j.conb.2022.102591>.
- (68) Thompson, M. C.; Barad, B. A.; Wolff, A. M.; Sun Cho, H.; Schotte, F.; Schwarz, D. M. C.; Anfinrud, P.; Fraser, J. S. Temperature-Jump Solution X-Ray Scattering Reveals Distinct Motions in a Dynamic Enzyme. *Nature Chemistry* **2019**, *11* (11), 1058–1066. <https://doi.org/10.1038/s41557-019-0329-3>.
- (69) Riedel, C.; Gabizon, R.; Wilson, C. A. M.; Hamadani, K.; Tsekouras, K.; Marqusee, S.; Pressé, S.; Bustamante, C. The Heat Released During Catalytic Turnover Enhances the Diffusion of an Enzyme. *Nature* **2015**, *517* (7533), 227–230. <https://doi.org/10.1038/nature14043>.
- (70) Kiyonaka, S.; Kajimoto, T.; Sakaguchi, R.; Shinmi, D.; Omatsu-Kanbe, M.; Matsuura, H.; Imamura, H.; Yoshizaki, T.; Hamachi, I.; Morii, T.; Mori, Y. Genetically Encoded Fluorescent Thermosensors Visualize Subcellular Thermoregulation in Living Cells. *Nature Methods* **2013**, *10* (12), 1232–1238. <https://doi.org/10.1038/nmeth.2690>.

- (71) Weik, M.; Colletier, J. P. Temperature-Dependent Macromolecular X-Ray Crystallography. *Acta Crystallographica. Section D, Biological Crystallography* **2010**, *66* (Pt 4), 437–446. <https://doi.org/10.1107/S0907444910002702>.
- (72) Johnson, E.; Palmer III, A. G.; Rance, M. Temperature Dependence of the NMR Generalized Order Parameter. *Proteins: Structure, Function, and Bioinformatics* **2007**, *66* (4), 796–803. <https://doi.org/10.1002/prot.21274>.
- (73) Adams, E. M.; Pezzotti, S.; Ahlers, J.; Rüttermann, M.; Levin, M.; Goldenzweig, A.; Peleg, Y.; Fleishman, S. J.; Sagi, I.; Havenith, M. Local Mutations Can Serve as a Game Changer for Global Protein Solvent Interaction. *JACS Au* **2021**, *1* (7), 1076–1085. <https://doi.org/10.1021/jacsau.1c00155>.
- (74) Tanimoto, S.; Tamura, K.; Hayashi, S.; Yoshida, N.; Nakano, H. A Computational Method to Simulate Global Conformational Changes of Proteins Induced by Cosolvent. *Journal of Computational Chemistry* **2021**, *42* (8), 552–563. <https://doi.org/10.1002/jcc.26481>.
- (75) Devineau, S.; Inoue, K.; Kusaka, R.; Urashima, S.; Nihonyanagi, S.; Baigl, D.; Tsuneshige, A.; Tahara, T. Change of the Isoelectric Point of Hemoglobin at the Air/Water Interface Probed by the Orientational Flip-Flop of Water Molecules. *Physical Chemistry Chemical Physics* **2017**, *19* (16), 10292–10300. <https://doi.org/10.1039/C6CP08854F>.
- (76) Oroguchi, T.; Nakasako, M. Influences of Lone-Pair Electrons on Directionality of Hydrogen Bonds Formed by Hydrophilic Amino Acid Side Chains in Molecular Dynamics Simulation. *Scientific Reports* **2017**, *7* (1), 15859. <https://doi.org/10.1038/s41598-017-16203-w>.
- (77) Conti Nibali, V.; D'Angelo, G.; Paciaroni, A.; Tobias, D. J.; Tarek, M. On the Coupling Between the Collective Dynamics of Proteins and Their Hydration Water. *The Journal of Physical Chemistry Letters* **2014**, *5* (7), 1181–1186. <https://doi.org/10.1021/jz500023e>.
- (78) McMahon, B. H.; Frauenfelder, H.; Fenimore, P. W. The Role of Continuous and Discrete Water Structures in Protein Function. *The European Physical Journal Special Topics* **2014**, *223* (5), 915–926. <https://doi.org/10.1140/epjst/e2014-02125-y>.
- (79) Leitner, D. M.; Pandey, H. D.; Reid, K. M. Energy Transport Across Interfaces in Biomolecular Systems. *Journal of Physical Chemistry B* **2019**, *123* (45), 9507–9524. <https://doi.org/10.1021/acs.jpccb.9b07086>.
- (80) Poudel, H.; Reid, K. M.; Yamato, T.; Leitner, D. M. Energy Transfer Across Nonpolar and Polar Contacts in Proteins: Role of Contact Fluctuations. *Journal of Physical Chemistry B* **2020**, *124* (44), 9852–9861. <https://doi.org/10.1021/acs.jpccb.0c08091>.
- (81) Leitner, D. M.; Yamato, T. Recent Developments in the Computational Study of Protein Structural and Vibrational Energy Dynamics. *Biophysical Reviews* **2020**, *12* (2), 317–322. <https://doi.org/10.1007/s12551-020-00661-0>.
- (82) Poudel, H.; Leitner, D. M. Energy Transport in Class B GPCRs: Role of Protein–Water Dynamics and Activation. *The Journal of Physical Chemistry B* **2022**, *126* (42), 8362–8373. <https://doi.org/10.1021/acs.jpccb.2c03960>.

- (83) He, J.; Zhang, L.; Liu, L. The Hydrogen-Bond Configuration Modulates the Energy Transfer Efficiency in Helical Protein Nanotubes. *Nanoscale* **2021**, *13* (2), 991–999. <https://doi.org/10.1039/D0NR06031C>.
- (84) Zhang, L.; Chen, T.; Ban, H.; Liu, L. Hydrogen Bonding-Assisted Thermal Conduction in β -Sheet Crystals of Spider Silk Protein. *Nanoscale* **2014**, *6* (14), 7786. <https://doi.org/10.1039/c4nr01195c>.
- (85) Babaei, H.; Keblinski, P.; Khodadadi, J. M. Equilibrium Molecular Dynamics Determination of Thermal Conductivity for Multi-Component Systems. *Journal of Applied Physics* **2012**, *112* (5), 054310. <https://doi.org/10.1063/1.4749265>.
- (86) Li, P.; Merz, K. M. MCPB.py: A Python Based Metal Center Parameter Builder. *Journal of Chemical Information and Modeling* **2016**, *56* (4), 599–604. <https://doi.org/10.1021/acs.jcim.5b00674>.
- (87) Li, P.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chemical Reviews* **2017**, *117* (3), 1564–1686. <https://doi.org/10.1021/acs.chemrev.6b00440>.
- (88) Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Physical Chemistry Chemical Physics* **2011**, *13* (14), 6670. <https://doi.org/10.1039/c0cp02984j>.
- (89) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals. *The Journal of Physical Chemistry A* **2007**, *111* (42), 10439–10452. <https://doi.org/10.1021/jp0734474>.
- (90) Edler, E.; Stein, M. Spin-State-Dependent Properties of an Iron(III) Hydrogenase Mimic. *European Journal of Inorganic Chemistry* **2014**, *2014* (22), 3587–3599. <https://doi.org/10.1002/ejic.201402295>.
- (91) D.A. Case; I.Y. Ben-Shalom; S.R. Brozell; et al. AMBER 19, 2019.
- (92) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. NMR Structure of the 35-Residue Villin Headpiece Subdomain. *Nature Structural Biology* **1997**, *4* (3), 180–184. <https://doi.org/10.1038/nsb0397-180>.
- (93) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **2001**, *105* (43), 9954–9960. <https://doi.org/10.1021/jp003020w>.
- (94) Price, D. J.; Brooks, C. L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. *The Journal of Chemical Physics* **2004**, *121* (20), 10096–10103. <https://doi.org/10.1063/1.1808117>.
- (95) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Miguez, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained Against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16* (1), 528–552. <https://doi.org/10.1021/acs.jctc.9b00591>.
- (96) Shiro, Y.; Nakamura, H. Heme-Based Oxygen Sensor Protein FixL: Its Structure and Function. *International Congress Series* **2002**, *1233* (C), 251–257. [https://doi.org/10.1016/s0531-5131\(02\)00236-4](https://doi.org/10.1016/s0531-5131(02)00236-4).

- (97) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (98) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577–8593. <https://doi.org/10.1063/1.470117>.
- (99) Duan, Z.-H.; Krasny, R. An Adaptive Treecode for Computing Nonbonded Potential Energy in Classical Molecular Systems. *Journal of Computational Chemistry* **2001**, *22* (2), 184–195. [https://doi.org/10.1002/1096-987x\(20010130\)22:2<184::aid-jcc6>3.0.co;2-7](https://doi.org/10.1002/1096-987x(20010130)22:2<184::aid-jcc6>3.0.co;2-7).
- (100) Sagui, C.; Pedersen, L. G.; Darden, T. A. Towards an Accurate Representation of Electrostatics in Classical Force Fields: Efficient Implementation of Multipolar Interactions in Biomolecular Simulations. *The Journal of Chemical Physics* **2004**, *120* (1), 73–87. <https://doi.org/10.1063/1.1630791>.
- (101) Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chemical Reviews* **2013**, *113* (3), 1598–1613. <https://doi.org/10.1021/cr3002356>.
- (102) Di Paola, L.; Giuliani, A. Protein Contact Network Topology: A Natural Language for Allostery. *Current Opinion in Structural Biology* **2015**, *31*, 43–48. <https://doi.org/10.1016/j.sbi.2015.03.001>.
- (103) Di Paola, L.; Poudel, H.; Parise, M.; Giuliani, A.; Leitner, D. M. A Statistical Journey Through the Topological Determinants of the B2 Adrenergic Receptor Dynamics. *Entropy* **2022**, *24* (7), 998. <https://doi.org/10.3390/e24070998>.
- (104) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (105) Shao, H.; Huang, W.; Avilan, L.; Receveur-Bréchet, V.; Puppo, C.; Puppo, R.; Lebrun, R.; Gontero, B.; Launay, H. A New Type of Flexible CP12 Protein in the Marine Diatom *Thalassiosira Pseudonana*. *Cell Communication and Signaling* **2021**, *19* (1). <https://doi.org/10.1186/s12964-021-00718-x>.
- (106) Roux, B.; Islam, S. M. Restrained-Ensemble Molecular Dynamics Simulations Based on Distance Histograms from Double Electron–Electron Resonance Spectroscopy. *Journal of Physical Chemistry B* **2013**, *117* (17), 4733–4739. <https://doi.org/10.1021/jp3110369>.
- (107) Shen, R.; Han, W.; Fiorin, G.; Islam, S. M.; Schulten, K.; Roux, B. Structural Refinement of Proteins by Restrained Molecular Dynamics Simulations with Non-Interacting Molecular Fragments. *PLoS Computational Biology* **2015**, *11* (10). <https://doi.org/10.1371/journal.pcbi.1004368>.

- (108) Islam, S. M.; Stein, R. A.; McHaourab, H. S.; Roux, B. Structural Refinement from Restrained-Ensemble Simulations Based on EPR/DEER Data: Application to T4 Lysozyme. *Journal of Physical Chemistry B* **2013**, *117* (17), 4740–4754. <https://doi.org/10.1021/jp311723a>.
- (109) Qi, Y.; Lee, J.; Cheng, X.; Shen, R.; Islam, S. M.; Roux, B.; Im, W. CHARMM-GUI DEER Facilitator for Spin-Pair Distance Distribution Calculations and Preparation of Restrained-Ensemble Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2020**, *41* (5), 415–420. <https://doi.org/10.1002/jcc.26032>.
- (110) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D. AMBER 2020, 2020.
- (111) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **2014**, *5* (21), 3863–3871. <https://doi.org/10.1021/jz501780a>.
- (112) Duan, Z.-H.; Krasny, R. An Adaptive Treecode for Computing Nonbonded Potential Energy in Classical Molecular Systems. *Journal of Computational Chemistry* **2001**, *22* (2), 184–195. [https://doi.org/10.1002/1096-987x\(20010130\)22:2<184::aid-jcc6>3.0.co;2-7](https://doi.org/10.1002/1096-987x(20010130)22:2<184::aid-jcc6>3.0.co;2-7).
- (113) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Grand, S. L.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, *9* (9), 3878–3888. <https://doi.org/10.1021/ct400314y>.
- (114) Jo, S.; Cheng, X.; Islam, S. M.; Huang, L.; Rui, H.; Zhu, A.; Lee, H. S.; Qi, Y.; Han, W.; Vanommeslaeghe, K.; MacKerell, A. D.; Roux, B.; Im, W. CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Nonstandard Residues. In *Advances in Protein Chemistry and Structural Biology*; Academic Press Inc., 2014; Vol. 96, pp 235–265. <https://doi.org/10.1016/bs.apcsb.2014.06.002>.
- (115) Svergun, D.; Barberato, C.; Koch, M. H. J. CRY SOL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **1995**, *28* (6), 768–773. <https://doi.org/10.1107/s0021889895007047>.
- (116) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. *Journal of Applied Crystallography* **2017**, *50*, 1212–1225. <https://doi.org/10.1107/s1600576717007786>.
- (117) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; Groot, B. L. de; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods* **2017**, *14* (1), 71–73. <https://doi.org/10.1038/nmeth.4067>.
- (118) Islam, S. M.; Roux, B. Simulating the Distance Distribution Between Spin-Labels Attached to Proteins. *Journal of Physical Chemistry B* **2015**, *119* (10), 3901–3911. <https://doi.org/10.1021/jp510745d>.
- (119) Petersen, H. G. Accuracy and Efficiency of the Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, *103* (9), 3668–3679. <https://doi.org/10.1063/1.470043>.

- (120) LLC, S. The PyMOL Molecular Graphics System, Version 1.8, 2015.
- (121) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (122) Esque, J.; Léonard, S.; Brevern, A. G. de; Oguey, C. VLDP Web Server: A Powerful Geometric Tool for Analysing Protein Structures in Their Environment. *Nucleic Acids Research* **2013**, *41* (W1), W373–W378. <https://doi.org/10.1093/nar/gkt509>.
- (123) *Fundamentals of Heat and Mass Transfer*, 7th ed.; Bergman, T. L., Incropera, F. P., Eds.; Wiley: Hoboken, NJ, 2011.
- (124) Fischer, H.; Polikarpov, I.; Craievich, A. F. Average Protein Density Is a Molecular-Weight-Dependent Function. *Protein Science* **2009**, *13* (10), 2825–2828. <https://doi.org/10.1110/ps.04688204>.
- (125) Harpaz, Y.; Gerstein, M.; Chothia, C. Volume Changes on Protein Folding. *Structure (London, England : 1993)* **1994**, *2* (7), 641–649. [https://doi.org/10.1016/S0969-2126\(00\)00065-4](https://doi.org/10.1016/S0969-2126(00)00065-4).
- (126) Baud, F.; Karlin, S. Measures of Residue Density in Protein Structures. *Proceedings of the National Academy of Sciences* **1999**, *96* (22), 12494–12499. <https://doi.org/10.1073/pnas.96.22.12494>.
- (127) Lervik, A.; Bresme, F.; Kjelstrup, S.; Bedeaux, D.; Miguel Rubi, J. Heat Transfer in Protein–Water Interfaces. *Physical Chemistry Chemical Physics* **2010**, *12* (7), 1610. <https://doi.org/10.1039/b918607g>.
- (128) Henry, E. R.; Eaton, W. A.; Hochstrasser, R. M. Molecular Dynamics Simulations of Cooling in Laser-Excited Heme Proteins. *Proceedings of the National Academy of Sciences* **1986**, *83* (23), 8982–8986. <https://doi.org/10.1073/pnas.83.23.8982>.
- (129) Nguyen, P. H.; Park, S.-M.; Stock, G. Nonequilibrium Molecular Dynamics Simulation of the Energy Transport Through a Peptide Helix. *The Journal of Chemical Physics* **2010**, *132* (2), 025102. <https://doi.org/10.1063/1.3284742>.
- (130) Gekko, Kunihiko.; Noguchi, Hajime. Compressibility of Globular Proteins in Water at 25.Degree.C. *The Journal of Physical Chemistry* **1979**, *83* (21), 2706–2714. <https://doi.org/10.1021/j100484a006>.
- (131) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *The Journal of Chemical Physics* **2006**, *124* (2), 024503. <https://doi.org/10.1063/1.2136877>.
- (132) Takemura, K.; Kitao, A. Effects of Water Model and Simulation Box Size on Protein Diffusional Motions. *The Journal of Physical Chemistry B* **2007**, *111* (41), 11870–11872. <https://doi.org/10.1021/jp0756247>.
- (133) Xu, Y.; Leitner, D. M. Communication Maps of Vibrational Energy Transport Through Photoactive Yellow Protein. *The Journal of Physical Chemistry A* **2014**, *118* (35), 7280–7287. <https://doi.org/10.1021/jp411281y>.
- (134) Agbo, J. K.; Xu, Y.; Zhang, P.; Straub, J. E.; Leitner, D. M. Vibrational Energy Flow Across Heme–Cytochrome c and Cytochrome c–Water Interfaces. *Theoretical Chemistry Accounts* **2014**, *133* (7), 1504. <https://doi.org/10.1007/s00214-014-1504-7>.

- (135) Hamzi, H.; Rajabpour, A.; Roldán, É.; Hassanali, A. Learning the Hydrophobic, Hydrophilic, and Aromatic Character of Amino Acids from Thermal Relaxation and Interfacial Thermal Conductance. *The Journal of Physical Chemistry B* **2022**, *126* (3), 670–678. <https://doi.org/gr7k9p>.
- (136) Xu, Y.; Leitner, D. M. Vibrational Energy Flow Through the Green Fluorescent Protein–Water Interface: Communication Maps and Thermal Boundary Conductance. *The Journal of Physical Chemistry B* **2014**, *118* (28), 7818–7826. <https://doi.org/10.1021/jp412141z>.
- (137) Wang, T.; Yamato, T.; Sugiura, W. Site-Selective Heat Current Analysis of α -Helical Protein with Linear-Homopolymer-Like Model. *The Journal of Chemical Physics* **2023**, *158* (21). <https://doi.org/10.1063/5.0149362>.
- (138) Zschiedrich, C. P.; Keidel, V.; Szurmant, H. Molecular Mechanisms of Two-Component Signal Transduction. *Journal of Molecular Biology* **2016**, *428* (19), 3752–3775. <https://doi.org/10.1016/j.jmb.2016.08.003>.
- (139) Gilles-Gonzalez, M. A.; Gonzalez, G.; Perutz, M. F.; Kiger, L.; Marden, M. C.; Poyart, C. Heme-Based Sensors, Exemplified by the Kinase FixL, Are a New Class of Heme Protein with Distinctive Ligand Binding and Autoxidation. *Biochemistry* **1994**, *33* (26), 8067–8073. <https://doi.org/10.1021/bi00192a011>.
- (140) Gao, R.; Stock, A. M. Biological Insights from Structures of Two-Component Proteins. *Annual Review of Microbiology* **2009**, *63*, 133–154. <https://doi.org/10.1146/annurev.micro.091208.073214>.
- (141) Taylor, B. L.; Zhulin, I. B. PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. *Microbiology and Molecular Biology Reviews* **1999**, *63* (2), 479–506. <https://doi.org/10.1128/mnbr.63.2.479-506.1999>.
- (142) Semenza, L. Sukhamay. ; G. L.; Prabhakar, N. R. *Oxygen Sensing : Responses and Adaptation to Hypoxia*; New York: Marcel Dekker, 2003.
- (143) Chan, M. K. Recent Advances in Heme-Protein Sensors. *Current Opinion in Chemical Biology* **2001**, *5* (2), 216–222. [https://doi.org/10.1016/s1367-5931\(00\)00193-9](https://doi.org/10.1016/s1367-5931(00)00193-9).
- (144) Henry, J. T.; Crosson, S. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. *Annual Review of Microbiology* **2011**, *65* (1), 261–286. <https://doi.org/10.1146/annurev-micro-121809-151631>.
- (145) Rodgers, K. R.; Lukat-Rodgers, G. S. Insights into Heme-Based O₂ Sensing from Structure-Function Relationships in the FixL Proteins. *Journal of Inorganic Biochemistry* **2005**, *99* (4), 963–977. <https://doi.org/10.1016/j.jinorgbio.2005.02.016>.
- (146) Gong, W.; Hao, B.; Mansy, S. S.; Gonzalez, G.; Gilles-Gonzalez, M. A.; Chan, M. K. Structure of a Biological Oxygen Sensor: A New Mechanism for Heme-Driven Signal Transduction. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (26), 15177–15182. <https://doi.org/10.1073/pnas.95.26.15177>.
- (147) Gilles-González, M. A.; Gonzalez, G.; Perutz, M. F. Kinase Activity of Oxygen Sensor FixL Depends on the Spin State of Its Heme Iron. *Biochemistry* **1995**, *34* (1), 232–236. <https://doi.org/10.1021/bi00001a027>.

- (148) Bertolucci, C.; Ming, L.; Gonzalez, G.; Gilles-Gonzalez, M. A. Assignment of the Hyperfine-Shifted $^1\text{H-NMR}$ Signals of the Heme in the Oxygen Sensor FixL from *Rhizobium Meliloti*. *Chemistry & Biology* **1996**, *3* (7), 561–566. [https://doi.org/10.1016/s1074-5521\(96\)90147-7](https://doi.org/10.1016/s1074-5521(96)90147-7).
- (149) Honorio-Felício, N.; Carepo, M. S. P.; F. Paulo, T. de; França Lopes, L. G. de; Sousa, E. H. S.; Diógenes, I. C. N.; Bernhardt, P. V. The Heme-Based Oxygen Sensor *Rhizobium Etli* FixL: Influence of Auxiliary Ligands on Heme Redox Potential and Implications on the Enzyme Activity. *Journal of Inorganic Biochemistry* **2016**, *164*, 34–41. <https://doi.org/10.1016/j.jinorgbio.2016.08.009>.
- (150) Ayers, R. A.; Moffat, K. Changes in Quaternary Structure in the Signaling Mechanisms of PAS Domains. *Biochemistry* **2008**, *47* (46), 12078–12086. <https://doi.org/10.1021/bi801254c>.
- (151) Rodgers, K. R.; Lukat-Rodgers, G. S.; Barron, J. A. Structural Basis for Ligand Discrimination and Response Initiation in the Heme-Based Oxygen Sensor FixL. *Biochemistry* **1996**, *35* (29), 9539–9548. <https://doi.org/10.1021/bi9530853>.
- (152) Hiruma, Y.; Kikuchi, A.; Tanaka, A.; Shiro, Y.; Mizutani, Y. Resonance Raman Observation of the Structural Dynamics of FixL on Signal Transduction and Ligand Discrimination. *Biochemistry* **2007**, *46* (20), 6086–6096. <https://doi.org/10.1021/bi062083n>.
- (153) Winkler, W. C.; Gonzalez, G.; Wittenberg, J. B.; Hille, R.; Dakappagari, N.; Jacob, A.; Gonzalez, L. A.; Gilles-Gonzalez, M. A. Nonsteric Factors Dominate Binding of Nitric Oxide, Azide, Imidazole, Cyanide, and Fluoride to the Rhizobial Heme-Based Oxygen Sensor FixL. *Chemistry and Biology* **1996**, *3* (10), 841–850. [https://doi.org/10.1016/s1074-5521\(96\)90070-8](https://doi.org/10.1016/s1074-5521(96)90070-8).
- (154) Tuckerman, J. R.; Gonzalez, G.; Dioum, E. M.; Gilles-Gonzalez, M. A. Ligand and Oxidation-State Specific Regulation of the Heme-Based Oxygen Sensor FixL from *Sinorhizobium Meliloti*. *Biochemistry* **2002**, *41* (19), 6170–6177. <https://doi.org/10.1021/bi025628w>.
- (155) Shimizu, T.; Huang, D.; Yan, F.; Stranova, M.; Bartosova, M.; Fojtíková, V.; Martínková, M. Gaseous O_2 , NO , and CO in Signal Transduction: Structure and Function Relationships of Heme-Based Gas Sensors and Heme-Redox Sensors. *Chemical Reviews* **2015**, *115* (13), 6491–6533. <https://doi.org/10.1021/acs.chemrev.5b00018>.
- (156) Gong, W.; Hao, B.; Chan, M. K. New Mechanistic Insights from Structural Studies of the Oxygen-Sensing Domain of *Bradyrhizobium Japonicum* FixL. *Biochemistry* **2000**, *39* (14), 3955–3962. <https://doi.org/10.1021/bi992346w>.
- (157) Jasaitis, A.; Hola, K.; Bouzahir-Sima, L.; Lambry, J. C.; Balland, V.; Vos, M. H.; Liebl, U. Role of Distal Arginine in Early Sensing Intermediates in the Heme Domain of the Oxygen Sensor FixL. *Biochemistry* **2006**, *45* (19), 6018–6026. <https://doi.org/10.1021/bi060012i>.
- (158) Dunham, C. M.; Dioum, E. M.; Tuckerman, J. R.; Gonzalez, G.; Scott, W. G.; Gilles-Gonzalez, M. A. A Distal Arginine in Oxygen-Sensing Heme-PAS Domains Is Essential to Ligand Binding, Signal Transduction, and Structure. *Biochemistry* **2003**, *42* (25), 7701–7708. <https://doi.org/10.1021/bi0343370>.

- (159) Balland, V.; Bouzahir-Sima, L.; Anxolabéhère-Mallart, E.; Boussac, A.; Vos, M. H.; Liebl, U.; Mattioli, T. A. Functional Implications of the Propionate 7-Arginine 220 Interaction in the FixLH Oxygen Sensor from *Bradyrhizobium Japonicum*. *Biochemistry* **2006**, *45* (7), 2072–2084. <https://doi.org/10.1021/bi051696h>.
- (160) Gilles-Gonzalez, M. A.; Caceres, A. I.; Sousa, E. H. S.; Tomchick, D. R.; Brautigam, C.; Gonzalez, C.; Machius, M. A Proximal Arginine R206 Participates in Switching of the Bradyrhizobium Japonicum FixL Oxygen Sensor. *Journal of Molecular Biology* **2006**, *360* (1), 80–89. <https://doi.org/10.1016/j.jmb.2006.04.054>.
- (161) Miyatake, H.; Mukai, M.; Park, S. Y.; Adachi, S. I.; Tamura, K.; Nakamura, H.; Nakamura, K.; Tsuchiya, T.; Iizuka, T.; Shiro, Y. Sensory Mechanism of Oxygen Sensor FixL from *Rhizobium Meliloti*: Crystallographic, Mutagenesis and Resonance Raman Spectroscopic Studies. *Journal of Molecular Biology* **2000**, *301* (2), 415–431. <https://doi.org/10.1006/jmbi.2000.3954>.
- (162) Mukai, M.; Nakamura, K.; Nakamura, H.; Iizuka, T.; Shiro, Y. Roles of Ile209 and Ile210 on the Heme Pocket Structure and Regulation of Histidine Kinase Activity of Oxygen Sensor FixL from *Rhizobium Meliloti*. *Biochemistry* **2000**, *39* (45), 13810–13816. <https://doi.org/10.1021/bi001184x>.
- (163) Salvi, M.; Schomburg, B.; Giller, K.; Graf, S.; Unden, G.; Becker, S.; Lange, A.; Griesinger, C. Sensory Domain Contraction in Histidine Kinase CitA Triggers Transmembrane Signaling in the Membrane-Bound Sensor. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (12), 3115–3120. <https://doi.org/10.1073/pnas.1620286114>.
- (164) Möglich, A.; Ayers, R. A.; Moffat, K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure (London, England : 1993)* **2009**, *17* (10), 1282–1294. <https://doi.org/10.1016/j.str.2009.08.011>.
- (165) Wright, G. S. A.; Saeki, A.; Hikima, T.; Nishizono, Y.; Hisano, T.; Kamaya, M.; Nukina, K.; Nishitani, H.; Nakamura, H.; Yamamoto, M.; Antonyuk, S. V.; Hasnain, S. S.; Shiro, Y.; Sawai, H. Architecture of the Complete Oxygen-Sensing FixL-FixJ Two-Component Signal Transduction System. *Science Signaling* **2018**, *11* (525), 1–13. <https://doi.org/10.1126/scisignal.aag0825>.
- (166) Bhate, M. A. P.; Molnar, K. A. S.; Goulian, M.; Degrado, W. F. Signal Transduction in Histidine Kinases: Insights from New Structures. *Structure (London, England : 1993)* **2015**, *23* (6), 981–994. <https://doi.org/10.1016/j.str.2015.04.002>.
- (167) Yamawaki, T.; Ishikawa, H.; Mizuno, M.; Nakamura, H.; Shiro, Y.; Mizutani, Y. Regulatory Implications of Structural Changes in Tyr201 of the Oxygen Sensor Protein FixL. *Biochemistry* **2016**, *55* (29), 4027–4035. <https://doi.org/10.1021/acs.biochem.6b00405>.
- (168) Tanaka, A.; Nakamura, H.; Shiro, Y.; Fujii, H. Roles of the Heme Distal Residues of FixL in O₂ Sensing: A Single Convergent Structure of the Heme Moiety Is Relevant to the Downregulation of Kinase Activity. *Biochemistry* **2006**, *45* (8), 2515–2523. <https://doi.org/10.1021/bi051989a>.
- (169) Diensthuber, R. P.; Bommer, M.; Gleichmann, T.; Möglich, A. Full-Length Structure of a Sensor Histidine Kinase Pinpoints Coaxial Coiled Coils as Signal Transducers and Modulators. *Structure (London, England : 1993)* **2013**, *21* (7), 1127–1136. <https://doi.org/10.1016/j.str.2013.04.024>.

- (170) Leitner, D. M.; Hyeon, C.; Reid, K. M. Water-Mediated Biomolecular Dynamics and Allostery. *Journal of Chemical Physics* **2020**, *152* (24). <https://doi.org/10.1063/5.0011392>.
- (171) Skalova, T.; Lengalova, A.; Dohnalek, J.; Harlos, K.; Mihalcin, P.; Kolenko, P.; Stranova, M.; Blaha, J.; Shimizu, T.; Martínková, M. Disruption of the Dimerization Interface of the Sensing Domain in the Dimeric Heme-Based Oxygen Sensor AfGcHK Abolishes Bacterial Signal Transduction. *Journal of Biological Chemistry* **2020**, *295* (6), 1587–1597. <https://doi.org/10.1074/jbc.ra119.011574>.
- (172) Yamawaki, T.; Mizuno, M.; Ishikawa, H.; Takemura, K.; Kitao, A.; Shiro, Y.; Mizutani, Y. Regulatory Switching by Concerted Motions on the Microsecond Time Scale of the Oxygen Sensor Protein FixL. *The Journal of Physical Chemistry B* **2021**, *125* (25), 6847–6856. <https://doi.org/10.1021/acs.jpcc.1c01885>.
- (173) Leitner, D. M.; Straub, J. E.; Group, F. *Proteins*; Leitner, D. M., Straub, J. E., Eds.; CRC Press, 2009. <https://doi.org/10.1201/9781420087048>.
- (174) Leitner, D. M. Energy Flow in Proteins. *Annual Review of Physical Chemistry* **2008**, *59*, 233–259. <https://doi.org/10.1146/annurev.physchem.59.032607.093606>.
- (175) Poudel, H.; Leitner, D. M. Activation-Induced Reorganization of Energy Transport Networks in the β 2Adrenergic Receptor. *Journal of Physical Chemistry B* **2021**, *125* (24), 6522–6531. <https://doi.org/10.1021/acs.jpcc.1c03412>.
- (176) Neiditch, M. B.; Federle, M. J.; Pompeani, A. J.; Kelly, R. C.; Swem, D. L.; Jeffrey, P. D.; Bassler, B. L.; Hughson, F. M. Ligand-Induced Asymmetry in Histidine Sensor Kinase Complex Regulates Quorum Sensing. *Cell* **2006**, *126* (6), 1095–1108. <https://doi.org/10.1016/j.cell.2006.07.032>.
- (177) *The Smallest Biomolecules: Diatomics and Their Interactions with Heme Proteins*; Elsevier, 2008. <https://doi.org/10.1016/B978-0-444-52839-1.X5001-3>.
- (178) Perutz, M. F.; Paoli, M.; Lesk, A. M. Fix L, a Haemoglobin That Acts as an Oxygen Sensor: Signalling Mechanism and Structural Basis of Its Homology with PAS Domains. *Chemistry and Biology* **1999**, *6* (11). [https://doi.org/10.1016/s1074-5521\(99\)80121-5](https://doi.org/10.1016/s1074-5521(99)80121-5).
- (179) Key, J.; Moffat, K. Crystal Structures of Deoxy and CO-Bound bjFixLH Reveal Details of Ligand Recognition and Signaling. *Biochemistry* **2005**, *44* (12), 4627–4635. <https://doi.org/10.1021/bi047942r>.
- (180) Yano, S.; Ishikawa, H.; Mizuno, M.; Nakamura, H.; Shiro, Y.; Mizutani, Y. Ultraviolet Resonance Raman Observations of the Structural Dynamics of Rhizobial Oxygen Sensor FixL on Ligand Recognition. *Journal of Physical Chemistry B* **2013**, *117* (49), 15786–15791. <https://doi.org/10.1021/jp406709e>.
- (181) Möglich, A.; Ayers, R. A.; Moffat, K. Design and Signaling Mechanism of Light-Regulated Histidine Kinases. *Journal of Molecular Biology* **2009**, *385* (5), 1433–1444. <https://doi.org/10.1016/j.jmb.2008.12.017>.
- (182) Bashford, D.; Chothia, C.; Lesk, A. M. Determinants of a Protein Fold. *Journal of Molecular Biology* **1987**, *196* (1), 199–216. [https://doi.org/10.1016/0022-2836\(87\)90521-3](https://doi.org/10.1016/0022-2836(87)90521-3).

BIBLIOGRAPHY

- (183) Krieger, E.; Vriend, G. YASARA View—Molecular Graphics for All Devices—from Smartphones to Workstations. *Bioinformatics (Oxford, England)* **2014**, *30* (20), 2981–2982. <https://doi.org/10.1093/bioinformatics/btu426>.
- (184) Tria, G.; Mertens, H. D. T.; Kachala, M.; Svergun, D. I. Advanced Ensemble Modelling of Flexible Macromolecules Using X-Ray Solution Scattering. *IUCrJ* **2015**, *2*, 207–217. <https://doi.org/10.1107/s205225251500202x>.