2024

MOLECULAR DYNAMICS STUDY OF ALPHA-SYNUCLEIN REPEAT MUTATIONS AND SARS-COV-2 NUCLEOCAPSID LINKER PHOSPHORYLATION

Coby Yeung
Saint John's University, Jamaica New York

Recommended Citation

Yeung, Coby, "MOLECULAR DYNAMICS STUDY OF ALPHA-SYNUCLEIN REPEAT MUTATIONS AND SARS-COV-2 NUCLEOCAPSID LINKER PHOSPHORYLATION" (2024). *Theses and Dissertations*. 755.
https://scholar.stjohns.edu/theses_dissertations/755

MOLECULAR DYNAMICS STUDY OF ALPHA-SYNUCLEIN REPEAT MUTATIONS AND SARS-COV-2 NUCLEOCAPSID LINKER PHOSPHORYLATION

A thesis submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

to the faculty of the

DEPARTMENT OF CHEMISTRY

of

ST. JOHN’S COLLEGE OF LIBERAL ARTS AND SCIENCES

at

ST. JOHN'S UNIVERSITY
New York
by
Coby Yeung

Date Submitted 4/18/2024

Date Approved 4/18/2024

Coby Yeung

Dr. Francisco Vazquez

ABSTRACT

MOLECULAR DYNAMICS STUDY OF ALPHA-SYNUCLEIN REPEAT MUTATIONS AND SARS-COV-2 NUCLEOCAPSID LINKER PHOSPHORYLATION

Coby Yeung

The study aimed to utilize molecular dynamics (MD) simulation to investigate intrinsically disordered protein (IDP) molecules and regions, focusing on IDPs such as alpha-synuclein and the SAR-CoV-2 nucleocapsid linker segment and their variants. Methodologically, the research employed MD simulation with implicit solvent methods, GaMD enhanced sampling, Python-based analysis, and VMD visualization. The study's significance lies in revealing structural dynamics insights of IDP regions. Key findings include the average structures of target sequences from k-means clustering analysis. The impact of mutations in alpha-synuclein KTKEGV repeat mutations is also revealed. Mutations introducing hydrophobic residues, like KGV and KLK, increased surface hydrophobicity, with KGV over-stabilized due to added positive charges. The research also explored the rigidity fluctuations of SAR-CoV-2 linker segments, comparing those from the Wuhan strain to those with r203k/g204r mutation and various phosphorylation levels.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Francisco Vazquez, whose unwavering guidance, mentorship, and encouragement have been a huge support to my academic journey at St. John's University. This thesis study opportunity would not have been possible without him. I am also very thankful to Dr. Erica Jacobs and Dr. Joseph M. Serafin for their invaluable insights and contributions as committee members, which have greatly enriched my work. Additionally, I would like to extend my appreciation to the entire chemistry department, including Dr. Richard Rosso, Dr. Alison Hyslop, and Ms. Joanne Belcon, for their collective dedication to fostering excellence in educational experience.

Lastly, I must express a special thanks to the Vazquez research group for their role in maturing the simulation method. Their expertise and collaborative efforts have been important in refining the methodology employed in this thesis study, enhancing its accuracy and reliability.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Proteins, the major components of the cell, play critical roles in every biological process. Proteins serve not only as structural components but also as enzymes, carriers, and hormones.[1] The aggregation of proteins often leads to the formation of insoluble fibrils or plaques that are toxic to cells. Such aggregates can interfere with cellular function in multiple ways.[2] Among various types of proteins, intrinsically disordered proteins (IDPs) stand out due to their lack of a fixed or rigid three-dimensional structure under physiological conditions. The flexible nature of IDPs enables them to bind to multiple partners and play crucial roles in complex signaling networks. However, this flexibility also renders IDPs prone to misfolding and aggregation, contributing to their association with various diseases.[3]

IDPs have emerged as a significant field of study due to their unique structural features. Investigating IDPs poses significant challenges due to their dynamic nature. Traditional structural biology techniques often fall short to investigate the transient conformations and complex dynamics of these proteins, demanding novel approaches and methodologies for a comprehensive understanding.[4]

1.1 Objectives

This thesis aims to delve into the molecular dynamics of two significant biological systems: alpha-synuclein and the SR-rich linker segments of SARS-CoV-2 nucleocapsid protein. It is important to understand the behavior of alpha-synuclein due to its role in neurodegenerative processes, where aggregation mechanisms contribute to cellular dysfunction. Similarly, investigating specific regions of SARS-CoV-2 is critical in the aftermath of the global pandemic, as it is essential to develop targeted therapeutic interventions via comprehensive analysis of the structural and functional dynamics of viral components.

Through MD simulations, the intermolecular interactions driving aggregation will be unraveled, providing computational insights that complement experimental data.

1.2 Literature Review

1.2.1 Alpha-synuclein

Alpha-synuclein (RCSB PDB: 1XQ8)[5] is composed of 140 amino acids residues, and its structure is shown in **Figure 1**. It is characterized by its intrinsically disordered nature in its monomeric state, which contributes to its solubility as a protein and its prevalent presence in the cytoplasm. It is distinguished by three primary domains: the N-terminal region, the non-amyloid component (NAC) region, and the C-terminal region.[6] Alpha-synuclein adopts a helical confirmation when it binds to lipid membranes, particularly within micelles. This Helical arrangement is important for its physiological functions. Due to its amphipathic nature, alpha-synuclein can interact with lipid bilayers, with its hydrophobic region embedded into the lipid core, while leaving its hydrophilic part exposed to the solvent environment. This interaction stabilizes the helical structure of alpha-synuclein, facilitating its bind-ability to membranes.[7]

The aggregation of alpha-synuclein into oligomers and fibrils is a critical pathological feature of several neurodegenerative conditions. These protein aggregates emerge through mechanisms such as cross-seeding with other protein types, leading to what are known as abnormal intracellular inclusions.[8–10]

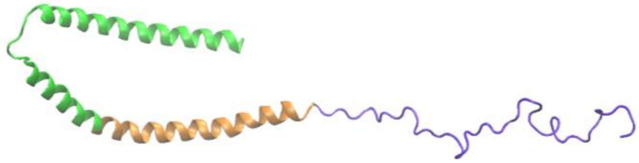

Figure 1. Structure of alpha-synuclein. Visualized in New Cartoon representation with the N-terminal domain (residue number 1 to 60) colored in green, NAC domain (residue number 61 to 95) colored in orange, and C-terminal domain (residue number 96 to 140) colored in purple.[5]

1.2.2 KTKEGV Repeat Mutants Related

In this thesis, alpha-synuclein variants with mutated KTKEGV repeat regions will be studied.

The KTKEGV repeats within the N-terminal domain adopt helical configurations when interacting with lipid membranes. The NAC domain tends to aggregate due to its substantial hydrophobic areas, thereby increasing protein toxicity. Conversely, The C-terminal exhibits a disordered structure, marked by regions with low hydrophobicity and high net charge. This disordered characteristic of the C-terminal domain plays an important role in preventing the NAC domain from assembling into fibrils and aggregates, tightly linked to the development of Lewy bodies.[11,12]

Numerous studies have explored the native states of alpha-synuclein in human brains. The critical balance maintained by alpha-synuclein in its monomeric and multimeric forms has been underscored.[13] Identifying alpha-synuclein conformers enriches the understanding of alpha-synuclein's physiological roles and aggregation propensity, aligning with the notion that disruption in alpha-synuclein tetramerization contributes to its pathological aggregation.[14]

Recent investigations into alpha-synuclein have unveiled the nature of KTKEGV repeat motifs in its tetramerization process. Disruption of these motifs through targeted mutations, influencing its solubility, assembly, and interaction with cellular membranes.[15] These mutations lead to a dramatic shift towards neurotoxic alpha-synuclein monomers, highlighting a potential molecular pathway to neurodegeneration observed in Parkinson's disease and similar disorders. This advancement in understanding the structural basis of alpha-synuclein functionality showcases the significance of its tetrameric form in maintaining neuronal integrity.[8]

The KTKKGV repeat mutant was specifically noted, in one literature, to abolish tetramers and induce toxicity. This mutation caused alpha-synuclein to become insoluble, disrupting normal homeostasis and function. Signs of neurotoxicity were exhibited in cells, evidenced by increased cell death, detection of adenylate kinase release, and cleaved PARP. In cellular models, the KGV mutations aggregated into a comparatively round and dense form of alpha-synuclein compared to the wild type. All of these findings strongly suggest that maintaining a balance between alpha-synuclein's multimeric and monomeric forms is crucial for neuronal survival and health.[8]

1.2.3 SARS-CoV-2 Nucleocapsid Protein

This study also investigated SARS-CoV-2 N-protein (RCSB PDB: 8FD5)[16], as shown in **Figure 2**, with a focus on the original Wuhan strain. SARS-CoV-2, shown in **Figure 3**, is a virus with single-stranded RNA and large genome, within a membrane. This membrane consists of the spike and envelop proteins, surrounds the nucleocapsid that encapsulates the genome. The encapsulation protects the viral genome from the host's cellular environment. The role of N-protein is critical for the life cycle of viruses. It facilitates the assembly of virion, boost the efficiency of RNA transcription, and being essential for viral replication.[17] Both Wuhan WT and its r203k/g204r variant[18] have been studied for their potential impact on transmission and vaccine effectiveness during and after the COVID-19 pandemic.[19]

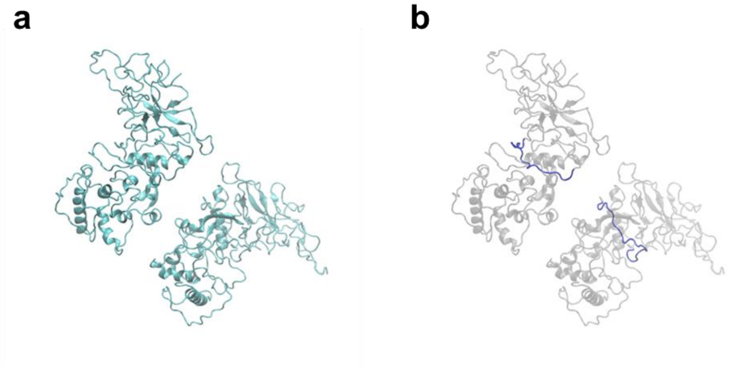

Figure 2. Structures of SARS-CoV-2 nucleocapsid protein. Wuhan strain in pairs shown as (a) Full proteins in New Cartoon representation.[16] (b) Full proteins in New Cartoon with the SR-rich IDP linker region highlighted in deep blue.

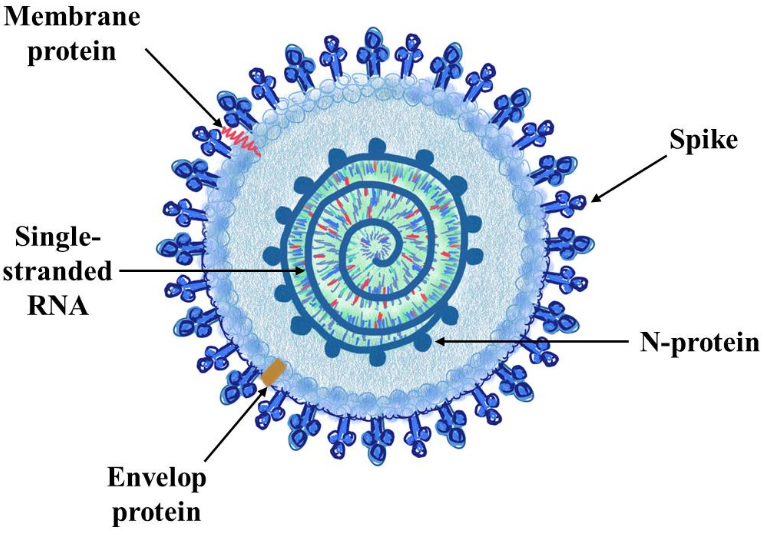

Figure 3. Schematic diagram illustrating SARS-CoV-2 virus.

Rather than analyzing the N-protein in its entirety, the research focuses on the linker segment in the Wuhan strain, referred to as the intrinsically disordered region, which includes amino acids 175 to 212. The specific segment undergoes several simulations under different conditions: a control scenario (presumably without alterations), a scenario where it is partially phosphorylated (phosphorylation on S188 and S206).

1.2.4 SR-rich Linker Segments

The SR-rich region of the N-protein is filled with charged residues and phosphorylation sites, affecting its interactions within the virus and human cells. This linker segment displays significant redundancy and flexibility in its phosphorylation sites while preserving local density, showcasing a high degree of variability. The r203 mutation in the SR-rich region was identified and linked to other variants early in the pandemic. Specifically, the r203k/g204r mutation increases the N-protein's capacity to form condensates, which are important for viral assembly and replication. This highlights the major position of the SR-rich region in the virus's pathogenicity and, of course, its potential as a target for therapeutic intervention.[20]

1.2.5 Amino Acids Charge Dynamics and Hydrophobicity

Each protein residue possesses specific characteristics including its formula, weight, charge status, and hydropathy index. Alanine, with a formula of $C3H5NO$ and a weight of 71.08 amu, is neutral with a hydropathy index of 1.8, indicating a relatively hydrophobic nature. In contrast, Glutamic acid is negatively charged, with a more complex formula ($C5H7NO3$), a higher molecular weight of 129.12 amu, and a hydropathy index of -3.5, making it more hydrophilic. Glycine, the simplest amino acid with a formula of $C2H3NO$

and the lowest weight of 57.05 amu, is also neutral but slightly hydrophilic. Lysine and Arginine are both positively charged, with Lysine having a slightly lower weight than Arginine, but both exhibit strong hydrophilic tendencies, as reflected by their negative hydropathy indices. Leucine stands out with a high hydropathy index of 3.8, indicating strong hydrophobic properties, whereas Glutamine, similar to Glutamic acid in charge and molecular weight, shares its hydrophilic nature. Serine, Threonine, and Valine, though neutral, vary in their hydrophilicity and hydrophobicity, with Valine being the most hydrophobic among them.[21] **Table 1** effectively summarizes the physical and chemical information reported above, focusing on the components within the scope of this thesis.

Table 1. Table of interested residues in this study. Includes residue formula, weight (in atomic mass units, amu) charge status, and hydropathy information.[21]

NAME	RESIDUE FORMULA	RESIDUE WEIGHT (AMU)	CHARGE STATUS	HYDROPATHY INDEX
ALANINE (ALA/A)	C_3H_5NO	71.08	Neutral	1.8
GLUTAMIC ACID (GLU/E)	$C_5H_7NO_3$	129.12	Negatively charged	-3.5
GLYCINE (GLY/G)	C_2H_3NO	57.05	Neutral	-0.4
LYSINE (LYS/K)	$C_6H_{12}N_2O$	128.18	Positively charged	-3.9
LEUCINE (LEU/L)	$C_6H_{11}NO$	113.16	Neutral	3.8
GLUTAMINE (GLN/Q)	$C_5H_8N_2O_2$	128.13	Neutral	-3.5
ARGININE (ARG/R)	$C_6H_{12}N_4O$	156.19	Positively charged	-4.5
SERINE (SER/S)	$C_3H_5NO_2$	87.08	Neutral	-0.8
THREONINE (THR/T)	$C_4H_7NO_2$	101.11	Neutral	-0.7
VALINE (VAL/V)	C_5H_9NO	99.13	Neutral	4.2

Charge distribution and hydrophobicity in IDPs can significantly alter their interactions with other biomolecules, for example, other proteins and membranes. Electrostatic interaction with the oppositely charged molecules or binding sites relies on the presence of charged residues.[22] Additionally, the presence of hydrophobic regions within IDPs promote contacts with lipid bilayers, facilitating membrane localization.

Hydrophobic interaction within cells also contribute to providing structural stability and functional diversity.[23]

1.2.6 Protein Phosphorylation Related

Phosphorylation impacts both the charge dynamics and hydrophilicity of proteins. Adding a phosphate group to the amino acid residues serine, threonine, or tyrosine significantly alters their structure, function, and interaction within the cell due to the negatively charged moiety. This change in charge can promote or inhibit interactions with other proteins or DNA.[24] Phosphorylation also increases the hydrophilicity of a protein, enhancing the solubility of proteins and facilitating their localization to specific cellular compartments. This further affects their ability to translocate to the nucleus and bind DNA.[25]

1.3 Methodology

1.3.1 Molecular Dynamic Overview

MD simulation is a computational strategy to explore and study the dynamic behavior of atoms and molecules. The basics of MD simulation involve the application of Newton's second law of motion,

$$F = ma \tag{1}$$

which predicts how particles interact over time. Other fundamental principles and components of MD simulations are the concepts of force fields, integration algorithms, periodic boundary conditions, ensemble and temperature control, and analysis of simulation outcomes.[26,27]

To initiate MD simulation, one must verify the starting positions of atoms and their initial velocities, considering a specified direction.[28] The positions and velocities of the atoms at a future time point can be determined using the following equations:

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \mathbf{v}(t_0)\Delta t + \frac{1}{2}\mathbf{a}(t_0)\Delta t^2 \tag{2}$$

$$\mathbf{v}(t) = \mathbf{v}(t_0) + \mathbf{a}(t_0)\Delta t \tag{3}$$

where $\mathbf{r}(t)$ is the time dependent position of the atom, $\mathbf{r}(t_0)$ is the initial time dependent position, $\mathbf{v}(t_0)$ is the initial time dependent velocity, Δt is the timestep, $\mathbf{a}(t_0)$ is the initial time dependent acceleration, and $\mathbf{v}(t)$ refers to the time dependent velocity.

To minimizes noise and errors in the computation in MD simulations, the Verlet algorithm[29] below can be utilized:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 \tag{4}$$

where Δt is the time step, $\mathbf{r}(t + \Delta t)$ is the position of a particle at time $t + \Delta t$, $\mathbf{r}(t)$ and $(t - \Delta t)$ are the position at times t and $t - \Delta t$ respectively, $\mathbf{a}(t)$ is the acceleration of the particle at time t, calculated from the forces acting on it.[29] Newer Verlet algorithms developed maintain the simplistic and time-reversible nature of the original methods but achieve higher precision at lower computational costs.[30]

These formulas highlight the necessity of accounting for acceleration in the model set up. The simulation is an iterative process, where calculations are executed repeatedly, often exceeding millions of iterations, to adapt to the alterations in both the atomic positions and their corresponding potential energy.

1.3.2 Force Field

The force field is a mathematical construct for potential energy. Essentially, it is a set of equations and parameters used to calculate the forces acting on each particle in the simulation. Its computation relies on atomic positions as expressed in equation below[28,29,31]:

$$U = U_{bonded} + U_{non-bonded} \tag{5}$$

$$= \sum_{bonds} k_l \left(l_{ij} - l_{eq}\right)^2 + \sum_{angles} k_\theta \left(\theta_{ijk} - \theta_{eq}\right)^2$$

$$+ \sum_{dihedrals} V_\phi \left(1 + \cos\left(n\phi_{ijkl} - \gamma_n\right)\right)$$

$$+ \sum_{nonbonded} \left[\frac{k_e q_i q_j}{r_{ij}} + 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)\right]$$

where k_l is the bond length (l) dependent spring constant; k_θ is the angle (θ) dependent spring constant; V_ϕ is a dihedral angle (ϕ) dependent constant; n refers to an integer and γ_n is the n dependent constant; k_e is the coulomb constant; q_i and q_j are referring to the charges, while r_{ij} is the distance between the charges; ε_{ij} is the potential well depth; σ_{ij} is the distance between the charges at which the potential energy is zero, and

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right) \tag{6}$$

where σ_i and σ_j represent potential interaction parameters.

Equation (5) encapsulates both bonded and nonbonded interactions among atoms. The initial component addresses bond elongation or compression, employing a harmonic potential model. The angle between atoms is accounted for and denoted by the second term.

The third term considered the dihedral angles employing a periodic potential approach. Finally, the fourth term accounts for nonbonded interactions, including long-range attraction and short-range repulsion, encompassed through the Lennard-Jones potential. Attraction scales with $\left(\frac{1}{r}\right)^6$ and repulsion with $\left(\frac{1}{r}\right)^{12}$. Notably, repulsive interactions dominate at proximity to the environment, leading to exponential energy escalation.[28,29,31]

Various modern force field equations are suitable for accommodating globular proteins; however, they lack the accuracy to represent the diverse ensembles of IDPs due to their tendency to overestimate structural stability. Recent developments in specially refined force fields designed for IDPs prioritize the adjustment of backbone dihedrals to improve the generation of reliable IDP ensembles.[23]

In this thesis, the ff14SBonlysc force field was selected for its compatibility with IDPs. This modeling tool is used in simulations to explore proteins behavior under different conditions. The name is originated from its foundation, the "ff99SB" model, with updated parameter for side chains, represented by "sc". The ff14SBonlysc model is designed based on quantum mechanics, providing precise methods to understand molecular behavior. One significant improvement of this force field compares to its foundation is the ability to troubleshoot problems encountered in its previous versions, such as "stickiness" issues with polar group in simulations, by replacing improvable terms with new parameters and rules to bring balance of interactions between proteins and water.[32]

1.3.3 Gaussian Accelerated Molecular Dynamic

To address the issue of sampling in simulations, Gaussian Accelerated Molecular Dynamic (GaMD) was employed. GaMD effectively addresses the challenge of preserving energy wells by applying a Gaussian-shaped boost potential to lower the energy barriers across the entire potential energy landscape, as shown in **Figure 4**. The boost potential is calculated based on the statistical distribution of the potential energies of the system. This allows for smoother and consistent acceleration of dynamics across different energy states, thereby avoiding the typical hindrance in conventional MD simulations.[33,34]

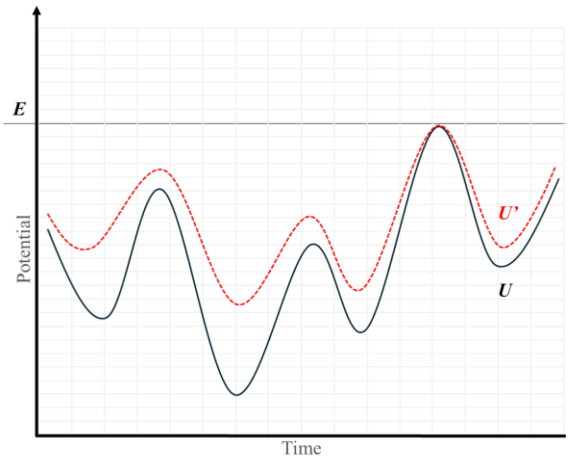

Figure 4. GaMD schematic diagram. Representing gaussian-shaped boost potential across potential energy landscape.[33] *U* is the original potential energy curve, where *U'* represents the new curve after lowering the energy barrier based on calculation.

Using a system with N atoms at position $\vec{r} = \{\vec{r}_1, \dots \vec{r}_N\}$ as an example, when the system potential $V(\vec{r})$ is lower than the threshold energy E, as shown in **Figure 5**, a boost potential can be added as:

$$\Delta V(\vec{r}) = \frac{1}{2}k\big(E - V(\vec{r})\big)^2 \tag{7}$$

where k is the harmonic force constant. One of the major sampling principles applied to the boost potential in GaMD is that, for any arbitrary $V_1(\vec{r})$ and $V_2(\vec{r})$ found on the original energy landscape, if one is greater than the other, the boost potential should not be changing the relative order of the original system potential data point.[32,35]

The capability brought by GaMD is extremely beneficial as it proves particularly valuable for comprehensively sampling the entire free energy landscape and accurately simulating complex biological systems.[35] Given that IDPs undergo conformational changes on a nanosecond scale, the application of GaMD becomes imperative for capturing diverse structures without the need to define collective variables, such as reaction coordinates. GaMD does not work by directly providing exact well depth information, but it offers necessary correct insights into the location and ranking of minima and maxima.[36,37]

1.3.4 Implicit Solvent Methods

To accurately replicate the behavior of the targeted IDP and regions, within its native neural-solution environment, MD simulations often incorporate a solvation box step. This step, while important for realism, significantly increases computational costs due to the increased number of solvent molecules involved.[38] To estimate the total solvation free energy, ΔG_{solv}, of a molecule, one can assume the expression can be decomposed into two parts:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel} \tag{8}$$

where ΔG_{nonel} is the non-electrostatic part, the free energy to solvate a molecule has all charges removed; ΔG_{el} is the electrostatic part, the free energy of removing all charges in the first vacuum step, followed by the re-addition in the continuum solvent environment.[32]

The analytical generalized Born (GB) method is an implicit solvent model (ISM) approach developed as an alternative to effectively simulate the influence of the solvent on the solute — in this case, protein — without the need to simulate each solvent molecule, which significantly speed up the folding landscape sampling in MD simulation.[39] The GB method approximates ΔG_{el} by below equation:

$$\Delta G_{el} \approx -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{GB}\left(r_{ij}, R_i, R_j\right)}\left(1-\frac{\exp[-\kappa f_{GB}]}{\varepsilon}\right) \tag{9}$$

where q_i and q_j are the charges at the center of the corresponding atoms i and j, r_{ij} is the distance between atom i and j, R_i and R_j are the Born radii, $f_{GB}(\,)$ represents the smooth function of its argument, and κ is the Debye-Huckel screening parameter, assumed zero for water.[40] A common selection of $f_{GB}(\,)$ is:

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2/4R_i R_j\right)\right]^{1/2} \tag{10}$$

which help to obtain a famous yet simple form of Born's single ion solvation energy equation:

$$\Delta G_{el} = -\frac{q_i^2}{2\rho_i}\left(1-\frac{1}{\varepsilon}\right) \tag{11}$$

where ρ_i refers to the large radii of deeply buried atoms.[41] This method substantially reduces computational costs compared to explicit solvent methods, offering a more efficient way to study the dynamics and interactions of the protein.[32,42]

CHAPTER 2: EXPERIMENTAL METHOD

All alpha-synuclein variants, including the WT and the SARS-CoV-2 N-protein IDP linker segments simulations in this study were performed using the Amber 21 software package[32]. The alpha-synuclein variants and first attempt of linker segment simulation utilized the ff14SBonlysc force field. Hydrogen mass repartitioning was employed for all simulations at a time step of 4 fs.

Fully extended structures were first constructed from the FASTA files containing the protein sequences. Specific regions of the SARS-CoV-2 protein underwent a phosphorylation step. The proteins and regions were then energy minimized and collapsed *in vacuo* using MD at 500K. The resulting atoms counts were as reported in **Table 2**.

Table 2: Summary of atoms counts for each MD simulation.

Alpha-synuclein and its Variants	Number of Atoms	SARS-CoV-2 Nucleocapsid Protein's Linker Segments	Number of Atoms
WT	2016	from the Wuhan strain	508
GTK	1881	from the phosphorylated variant	514
KLK	2052		
KGV	2056		
EGR	2070		

Once the systems were relaxed, a 300ns equilibration at the elevated temperature was performed for each molecule individually. The completely unfolded structures were explored in GaMD improved sampling simulations. Trajectory analyses and graphical representations utilized Python-based analytical tools. K-means algorithm-based clustering analysis was employed to predict the most possible folding structures, followed by molecules visualization via VMD.[43] Assessments of secondary structure were performed using the Dictionary of Secondary Structure of Proteins (DSSP).

The second attempt of linker segment simulation utilized the ff14SB force field and expanded the scope to include more phosphorylated variants. Hydrogen mass repartitioning was not employed for this part of the simulations. Fully extended structures were also constructed from the FASTA files containing the protein sequences and underwent the phosphorylation steps. The linker segments were then energy minimized but not collapsed *in vacuo*. Once the system was relaxed, the same equilibration and sampling simulations were conducted, followed by similar analysis.

CHAPTER 3: RESULTS AND DISCUSSION

3.1 Alpha-synuclein and its Variants

3.1.1 Overview and Charge Status Theory

Figure 5 displays the sequences for both the wild-type (WT) and its KTKEGV mutants, showcasing the repeated mutations characterized by reoccurring residues swaps in four distinct pattern arrangements. Residues 1 to 60 correspond to the N-terminal domains, exhibiting a range of 3 to 5 mutations; residues 61 to 95 correspond to the NAC domains, exhibiting a range of 2 to 4 mutations; residues 96 to 140 comprise the C-terminal domains, mostly remaining unaltered in the mutants. By characterizing these mutations and their associated structural changes, insights into the molecular mechanisms of IDPs can be gained, laying the foundation for understanding related disease pathology and informing the development of targeted therapeutic interventions.

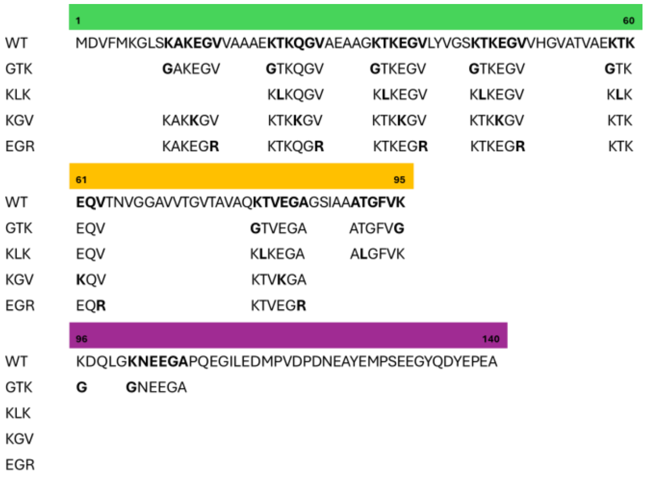

Figure 5. Sequence of alpha-synuclein and its KTKEGV repeat mutants. The N-terminal region (residue number 1 to 60) is highlighted in green, the NAC region (residue number 61 to 95) is highlighted in orange, and the C-terminal region (residue number 96 to 140) is highlighted in purple. Repeat units in the WT sequence are identified and bolded.

The individual analysis of alpha-synuclein is showed in **Figure 6**, including its interdomain hydrophobic contacts, hydrogen bonds, and salt bridge contents. This profile serves as a reference for later analysis by building a thorough understanding of the WT as a control to draw comparisons with the repeat mutants.

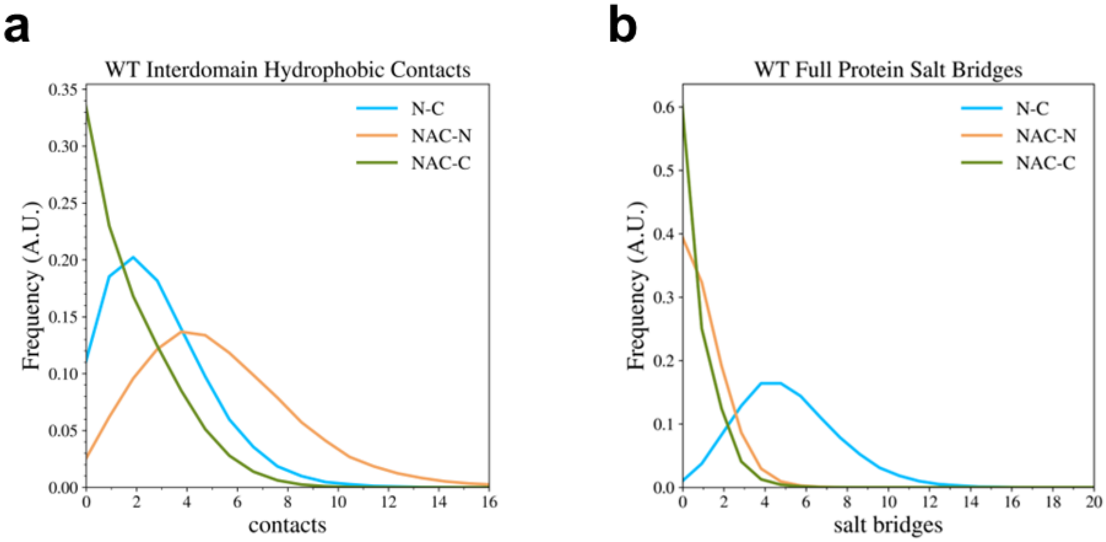

Figure 6. Alpha-synuclein's profile. Individual analysis of (a) interdomain hydrophobic contacts and (b) salt bridges. Color coding indicates interactions between N-terminal and C-terminal domains in cyan, between NAC and N-terminal domains in orange, and between NAC and C-terminal domains in green.

The theoretical impacts of four specific KTKEGV repeat mutations on the sequences' physical and chemistry properties are summarized in **Table 3** using **Table 1** as reference.

Table 3. Summary of the theoretical change in KTKEGV repeat mutation sites. Uses **Table 1** as reference.

		Lysine (Lys/K)	**to**	**Glycine (Gly/G)**
GTK	Number of Mutation Site		9	
	Size (amu)	128.18	→	57.05
	Charge Status	Positively charged	→	Neutral
	Hydropathy Index	-3.9	→	-0.4
KLK		**Threonine (Thr/T)**	**to**	**Leucine (Leu/L)**
	Number of Mutation Site		7	
	Size (amu)	101.11	→	113.16
	Charge Status	Neutral	→	Neutral
	Hydropathy Index	-0.7	→	3.8
KGV		**Glutamic acid (Glu/E) or Glutamine (Gln/Q)**	**to**	**Lysine (Lys/K)**
	Number of Mutation Site		6	
	Size (amu)	129.12 or 128.13	→	128.18
	Charge Status	Negatively charged or Neutral	→	Positively charged
	Hydropathy Index	-3.5 or -3.5	→	-3.9
EGR		**Valine (Val/V) or Alanine (Ala/A)**	**to**	**Lysine (Lys/K)**
	Number of Mutation Site		6	
	Size (amu)	99.13 or 71.08	→	128.18
	Charge Status	Neutral	→	Positively charged
	Hydropathy Index	4.2 or 1.8	→	-3.9

This study characterized four specific amino acid mutations by examining changes in size, charge status, and hydropathy index across various mutation sites. The first mutation, GTK, involved substituting Lysine with Glycine, impacting 9 sites, and resulting in a significant decrease in molecular size from 128.18 amu to 57.05 amu. A shift from a positively charged to a neutral charge status occurred, and there was a change in the hydropathy index from -3.9 to -0.4. The second mutation, KLK, involved substituting Threonine with Leucine across 7 sites, resulting in an increase in size from 101.11 amu to 113.16 amu. Despite maintaining a neutral charge, this mutation significantly altered the hydropathy index from -0.7 to 3.8, indicating increased hydrophobicity. The third mutation, KGV, involved converting Glutamic acid or Glutamine to Lysine at 6 sites. This mutation shifted from a negatively charged or neutral state to a positively charged state while maintaining a hydropathy index of close to -3.9. The final mutation, EGR, involved converting Valine or Alanine with Lysine at 6 sites, resulting in a significant increase in size. This mutation shifted the charge from neutral to positively charged and altered the hydropathy index from positive values to -3.9, indicating a transition to less hydrophobic or more hydrophilic conditions.

Anticipated alterations in molecular compactness and interdomain contacts would reflect changes in size, charge status, and hydrophobicity. The GTK mutation was projected to induce a reduction in molecular size, potentially leading to a smaller structure. Interdomain contacts, particularly between the N-terminal and C-terminal regions, might decrease as a result. This reduction in contacts could be attributed to the transition of positively charged basic regions to a more neutral state, altering their interaction with the acidic regions. Additionally, the increased hydrophobicity associated with this mutation might promote the formation of fibrils and aggregates. The KLK mutation was projected to either maintain or slightly increase the molecular size, with minimal changes expected in the charge status of basic regions. However, a significant increase in hydrophobicity was anticipated. As a result, formation of fibrils and aggregates is likely to happen, promoting alterations in interdomain contacts. The KGV mutation was anticipated to lead to a reduction in molecular size, primarily due to significant changes in charge status, resulting in a transition to a positively charged state. This shift was expected to promote attraction to the acidic region rather than alterations in size or hydrophobicity, which remained similar. Changes in interdomain contacts were anticipated to primarily occur through electrostatic changes. The EGR mutation was projected to result in an increase in molecular size. The transition in charge status from neutral to positively charged was expected to potentially enhance interdomain contacts through the attraction of basic and acidic regions. Additionally, the significant decrease in hydrophobicity might reduce the likelihood of fibril and aggregate formation.

The most probable structures, or "top frame", of the alpha-synuclein proteins, WT and its four KTKEGV repeat mutants, were determined via k-means clustering analysis of MD simulation data. This analysis enables the isolation and visualization of the key conformations that dominated the structural landscape. To enhance the clarity and understanding of these complex structures, a color-coding scheme was employed for the distinct domains of the protein, as shown in **Figure 7**. All structures were captured using the same scaling, enabling direct comparison of the compactness between the wild-type and repeat mutants.

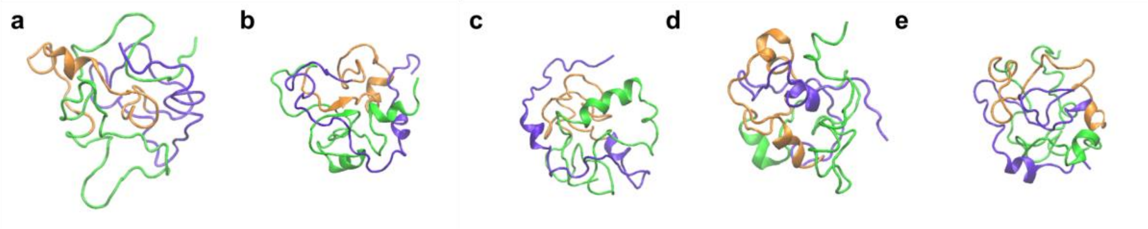

Figure 7. The top frame of alpha-synuclein and its KTKEGV repeat mutants. Rendering were captured via k-means clustering analysis. Panels show (a) the WT, (b) GTK, (c) KLK, (d) KGV, and (e) EGR in New Cartoon representation with the N-terminal domain (residue number 1 to 60) colored in green, NAC domain (residue number 61 to 95) colored in orange, and C-terminal domain (residue number 96 to 140) colored in purple.

The Dictionary of Secondary Structure of Proteins (DSSP) algorithm was utilized to outline the secondary structure contents within the full protein, as well as specifically within its N-terminal, NAC, and C-terminal domains as shown in **Figure 8**.

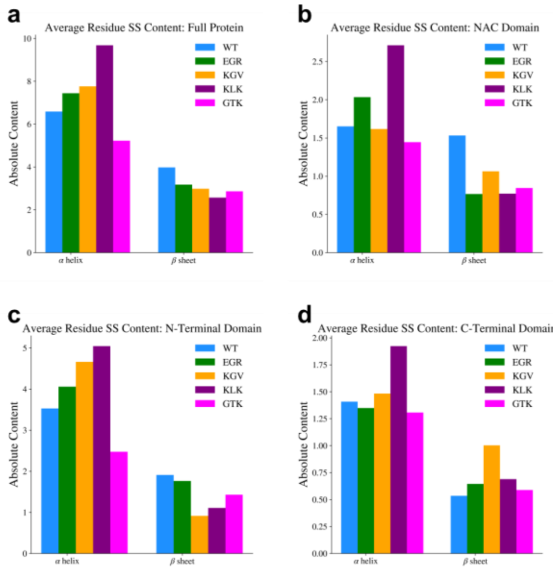

Figure 8. DSSP profile of alpha-synuclein and its KTKEGV repeat mutants. Analysis in (a) full protein, (b) NAC domains, (c) N-terminal domains, and (d) C-terminal domains. The WT is shown in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

In the analysis of the full protein (**Figure 8a**), the KLK variant exhibits the highest α-helix content. This is very likely due to the huge α-helix observed in N-terminal (**Figure 7c**), while the GTK variant shows the least α-helix content. Conversely, the β-sheet content remains similar across all variants. Within the N-terminal domain (**Figure 8b**), KLK again displays the highest α-helix content, with GTK continuing to exhibit the least. Additionally, the β-sheet content in the N-terminal domain is the highest for the WT, while the KGV variant shows the lowest content. Regarding the NAC domain (**Figure 8c**), the KLK variant still displays the highest α-helix content. Lastly, in the C-terminal domain (**Figure 8d**), KLK maintains the highest α-helix content, while KGV exhibits the highest β-sheet content.

These DSSP results serve as a complementary insight to further accurately analyze the structures shown in **Figure 7**.

Polymeric properties were examined across the WT and its KTKEGV mutants. The end-to-end distance (R_{ee}) distribution profile is shown in **Figure 9**; the radius of gyration (R_g) distribution profile is shown in **Figure 10**. These analyses covered both the full protein and its individual domains.

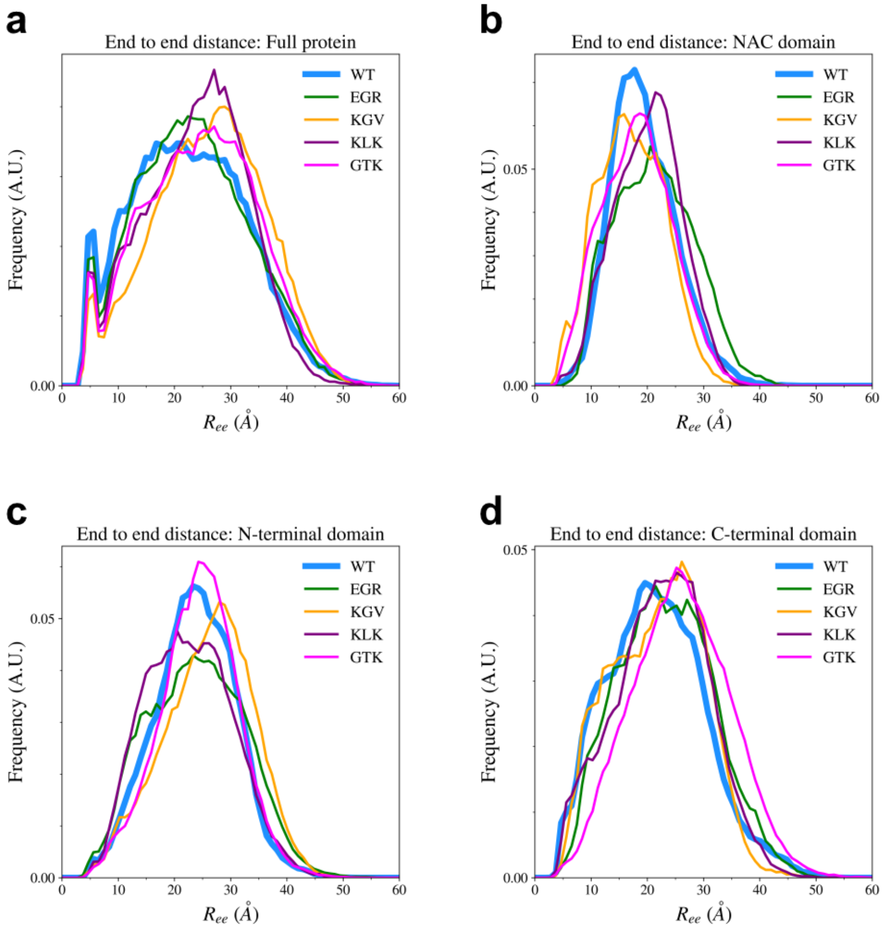

Figure 9. End-to-end distance of alpha-synuclein and its KTKEGV repeat mutants. Analysis in (a) full protein, (b) NAC domains, (c) N-terminal domains, and (d) C-terminal domains. The WT is shown in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

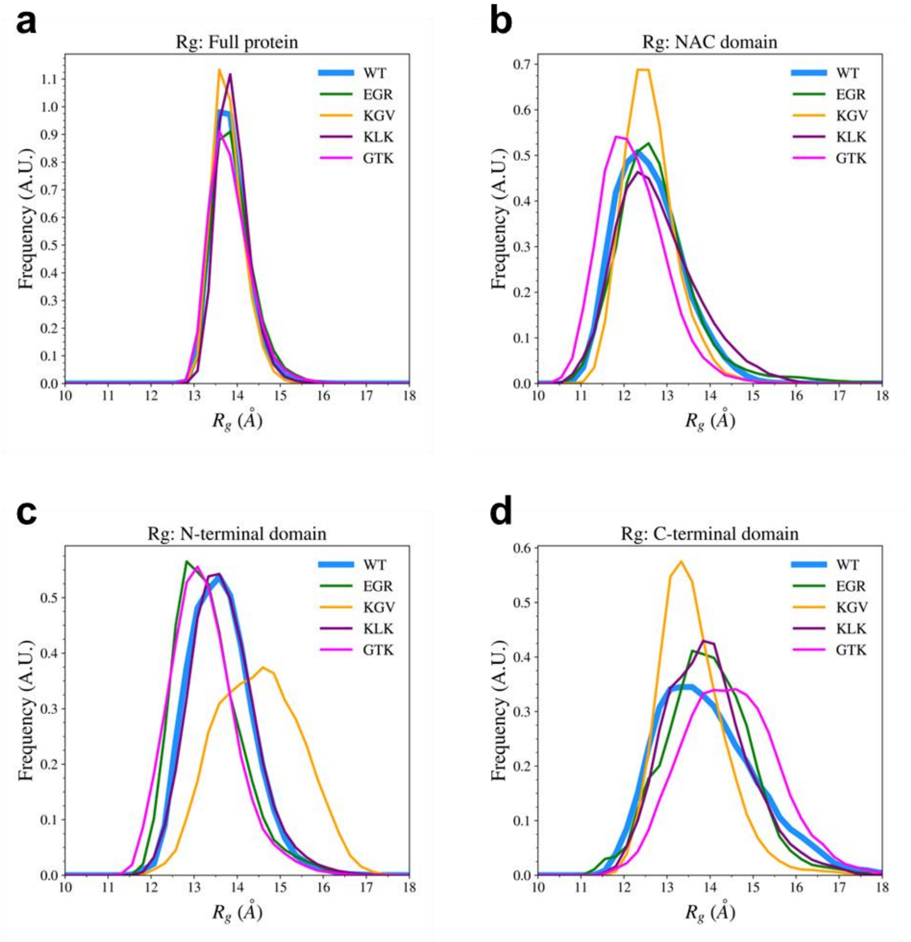

Figure 10. Radius of gyration of alpha-synuclein and its KTKEGV repeat mutants. Analysis in (a) full protein, (b) NAC domains, (c) N-terminal domains, and (d) C-terminal domains. The WT is show in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

Figure 9 shows the R_{ee} of all variants remained similar. This similarity is suggested by the overlapping peaks across full protein, NAC, N-terminal and C-terminal domains. (**Figure 9a** to **d**). On the other hand, the R_g, offers insight into the overall compactness of the protein structure post-mutation, while the R_{ee} provides information on spatial configuration. Together, these two measurements delivered the big picture of protein toxicity association, a key component to evaluate and observe aggregation. They revealed how each variant and domain responded to structural residue modifications. This analysis provides critical understanding of the dimensional flexibility and intrinsic stability of the protein.

Figure 10c, the R_g profile of the N-terminal domains of alpha-synuclein WT and mutants, shows a significant rightward shift for KGV mutant. **Figure 10a, 10b and 10d**, on the other hand, shows the peaks overlapping with each other. This finding suggests that the overall compactness of the variants are similar across all perspectives: full protein, just NAC domains, or just C-terminal domains.

While maintaining the compactness in the full protein, KGV mutant elongates in its N-terminal domain. **Figure 10d** also shows a comparatively more "relaxed" N-terminal region (green) versus a more "packed" NAC (orange) and C-terminal (purple) regions cluster. Linking this back to **Table 3**, KGV's repeat units remain similar size and hydrophobicity. Its charge status underwent a significant shift from negatively charged to positively charged, which was the most pronounced change among all mutants. This is evidence that the compactness change occurred due to the charge status change. Previously, it was assumed that if the mutation shifted towards a more positive charge in the N-terminal, the interaction with C-terminal domains (acidic region) would increase due to charge attraction, resulting in increased compactness. However, experimentally, the KGV mutant experienced the opposite effect, likely due to repulsion between the positive charges within its N-terminal.

Continuing with the association between charge status change and experimental data analysis, salt bridge contents are the primary components to evaluate. In the context of biochemical studies, salt bridges represent an important point for analyzing interactions within protein structures. This electrostatic phenomenon involves the attraction of oppositely charged residues, which are related to protein conformations and folding processes. **Figure 11** displays a detailed profile of salt bridges between domains. In the NAC-N salt bridges distribution graph (**Figure 11a**), KGV and GTK exhibits a decrease in salt bridge counts, while KLK shows similar counts to the WT. Additionally, EGR has slightly more salt bridges, as evidenced by its rightward shift comparing to the WT. In the NAC-C salt bridges distribution graph (**Figure 11b**), a similar decreasing trend is observed for GTK. However, KGV, KLK, and EGR variants exhibit an increasing trend compared to the WT.

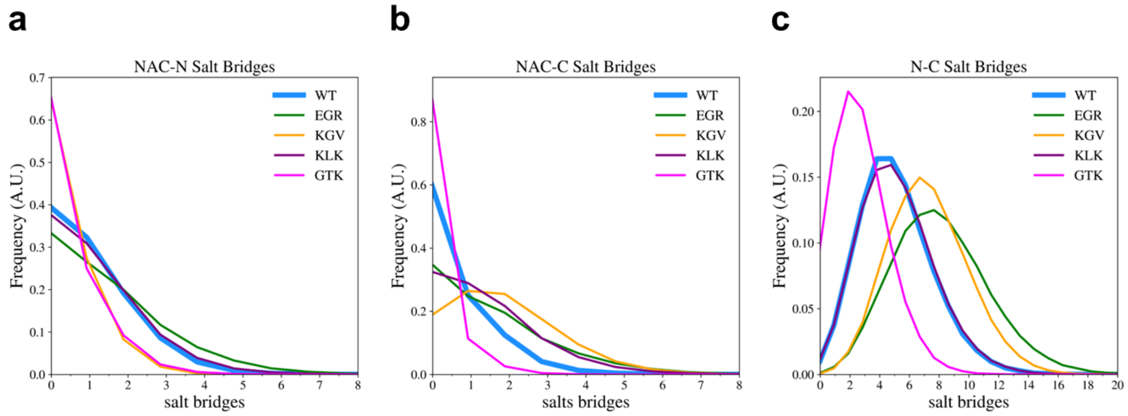

Figure 11. Salt bridges of alpha-synuclein and its KTKEGV repeat mutants. Analysis of salt bridge contents (a) between the NAC and N-terminal domains, (b) between NAC and C-terminal domains and (c) between N-terminal and C-terminal domains. The WT is shown in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

The profile of N-C salt bridges (**Figure 11c**) reveals a markedly different pattern, with a pronounced peak at an intermediate number of salt bridges. Initially, in the WT, N-C has the most salt bridge contents compared to the other two distribution graphs mentioned. Interestingly, the distribution correlates with the change in charge status caused by mutations, as listed in **Table 3**. The GTK mutation, indicated by the pink line, exhibits a left-skewed distribution, suggesting reduced salt bridge formation between the N-terminal and C-terminal domains. This mutation was the only one that decreased among all mutations. The charge status in GTK mutations was changed from positively charged to neutral across 9 mutation sites. Applying a similar interpretation for the other four mutations reveals that the KLK peak overlapped with the WT, indicating no change in salt bridge formation between the domains. Additionally, the mutation itself involved a neutral-to-neutral residue swap at 7 sites. On the contrary, KGV transitioned from negatively charged or neutral to positively charged. This change in charge status affected 6 sites, with the data showing a significant increase in salt bridge formation due to the mutation. EGR underwent a significant change in charge status across 6 mutation sites, shifting from neutral to positively charged. This change, coupled with a notable increase in residue size, resulted in a slightly more significant increase in salt bridge formation.

The observed trends in salt bridge formation from N to C terminal aligned with the proposed theory in related to charge status (**Table 3)**. At the molecular level, a salt bridge represents an attraction between an anion and a cation. **Figure 11** strongly suggests that the theoretical changes resulting from residue swapping are carried over to the molecules via mutation, and the effects are particularly evident in the interaction between the N and C terminals.

3.1.2 More on Charge Status Theory: Interdomain Contacts

In this section, interdomain contacts, as well as hydrophobic interactions, are analyzed and reported in this section. The general analysis, shown in **Figure 12**, covers the distribution of interdomain contacts between the domains across the WT protein and its mutants: EGR, KGV, KLK, and GTK. These contacts include those between the NAC and

N-terminal (NAC-N) domains, NAC to C-terminal (NAC-C) domains, and N-terminal to C-terminal (N-C) domains. By integrating the mentioned polymeric properties and salt bridge analysis, which were proposed to be highly related to the theoretical charge status change (**Table 3**) resulting from KTKEGV repeat mutations, an approach to understanding structural behavior was constructed.

In **Figure 12a**, the WT displays a prominent peak at the center of the distribution, while all mutants show peaks shift toward the left side of the graph, suggesting a general decrease in NAC-N interdomain contacts due to the mutations. In the NAC-C interdomain contact graph (**Figure 12b**), the KGV mutant shows a peak more centrally located, while the other variants, including the WT, are positioned toward the left end of the graph. This suggests that the KGV mutation increases the interdomain contact of NAC-C specifically., aligning with the previously mentioned elongation of this domain in the KGV mutant. Lastly, in the N-C interdomain contact graph (**Figure 12c**), the peaks of all proteins overlap with each other, with the mutation showing a slight shift to the right side of the graph compared to everything else. This suggests that the mutation subtly increases the N-C interdomain contacts.

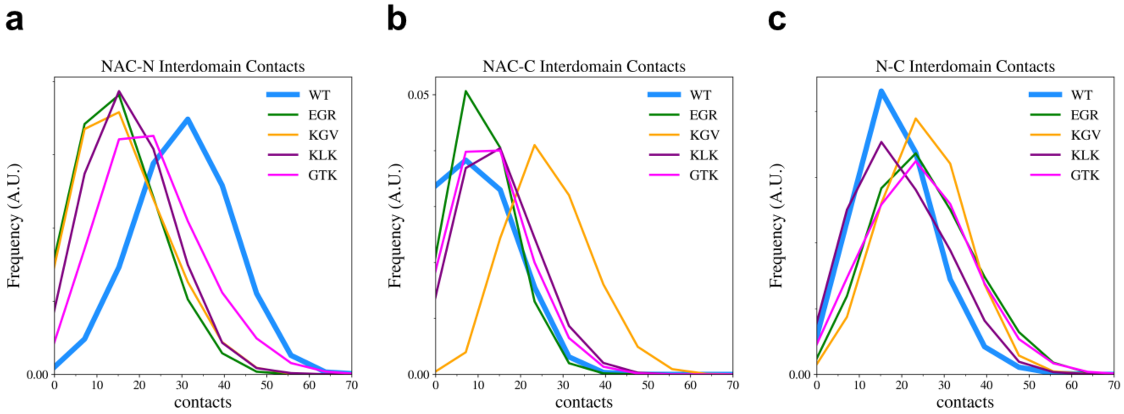

Figure 12. Interdomain contacts of alpha-synuclein and its KTKEGV repeat mutants. Analysis of general interdomain contacts (a) between the NAC and N-terminal domains, (b) between NAC and C-terminal domains and (c) between N-terminal and C-terminal domains. The WT is shown in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

A contact map is a powerful tool for visualizing structural and contact changes. **Figure 13** shows contact maps for the alpha-synuclein WT and all mutants (GTK, KLK, KGV, and EGR). These maps represent the pairwise residue interactions within the protein molecules. The axes of the map correspond to the residue numbers along the protein's sequence. Clusters of data points off the main diagonal revealed interaction patterns that denote secondary structural features, such as the formation of α-helices or β-sheets. The contact maps in **Figure 13** illustrate the interdomain contacts of all variants, providing visual support for the previously reported data.

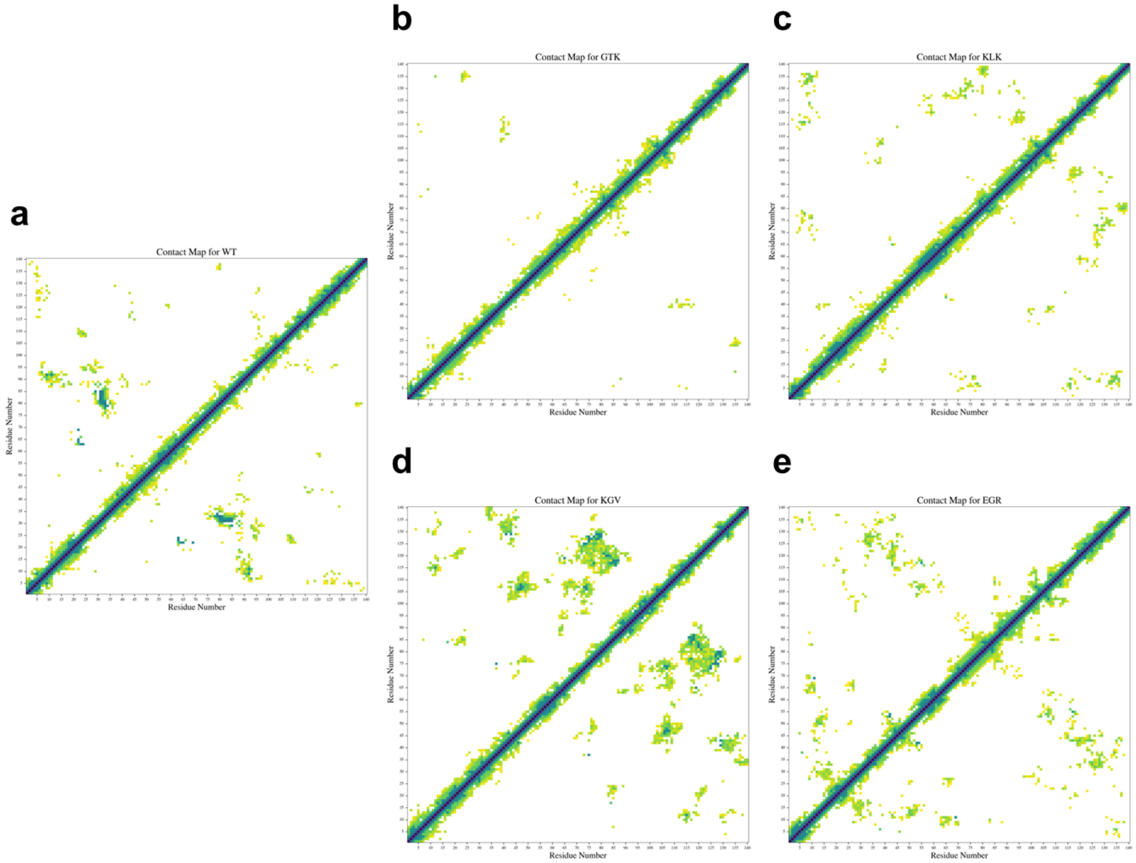

Figure 13. Contact map of alpha-synuclein and its KTKEGV repeat mutants. The maps are illustrating protein structure interactions within (a) the WT, (b) GTK, (c) KLK, (d) KGV, and (e) EGR. Interdomain contact details are interpreted by aligning residue numbers along the x and y-axes. (The N-terminal region comprises residue numbers 1 to 60; the NAC region comprises residue numbers 61 to 95; the C-terminal region comprises residue numbers 96 to 140.)

Figure 13a serves as the contact map for alpha-synuclein, providing a control for deriving interpretations on mutant comparisons. The map displays a prominent cluster in the NAC to N-terminal regions ({x = 70-100, y = 0-40}, and its reflection along the diagonal line). This aligns with the interdomain contact profiles (**Figure 12**), indicating that for WT, the majority of contacts were in the NAC-N regions. Across all mutants (**Figure 13b**, **13c**, and **13d**), there are no clusters as significant as the one highlighted in this particular region for the WT, supporting the overall left-skewed trends observed in

Figure 12a. The KGV mutant exhibits elongation in the NAC-C and N-C interdomain contacts profiles (**Figure. 12b** and **12c**), while the other mutants are overlapped with the WT. This phenomenon is visualized in the KGV contact map (**Figure 13d**), where multiple dense clusters appeared in a large area encompassing all three regions ({x = 90-140, y = 30-90}, and its reflection along the diagonal line).

Figure 14 shows the distribution of interdomain hydrophobic contacts between the domains across the WT protein and its KTKEGV mutants. Interestingly, these hydrophobic contacts do not correlate closely with the theoretical hydrophobicity changes listed in **Table 3**. Instead, the trends observed in **Figure 14a**, **14b** and **14c** aligned closely with those in **Figure 12**. This suggests that the hydrophobic contacts are highly related to the general interdomain contacts for alpha-synuclein WT and its KTKEGV repeat mutants. Secondly, charge status appears to be the major factor affecting the compactness-related polymeric properties, followed by altered salt bridge contents triggering shifts in interdomain contacts. **Figure 15** provides another complementary analysis, presenting hydrogen bonds profiles, to support the proposed theory. The distribution of hydrogen bonds forms the foundation for determining higher structures of proteins, significantly influencing interdomain contacts and playing a crucial role in stabilizing IDP conformations.

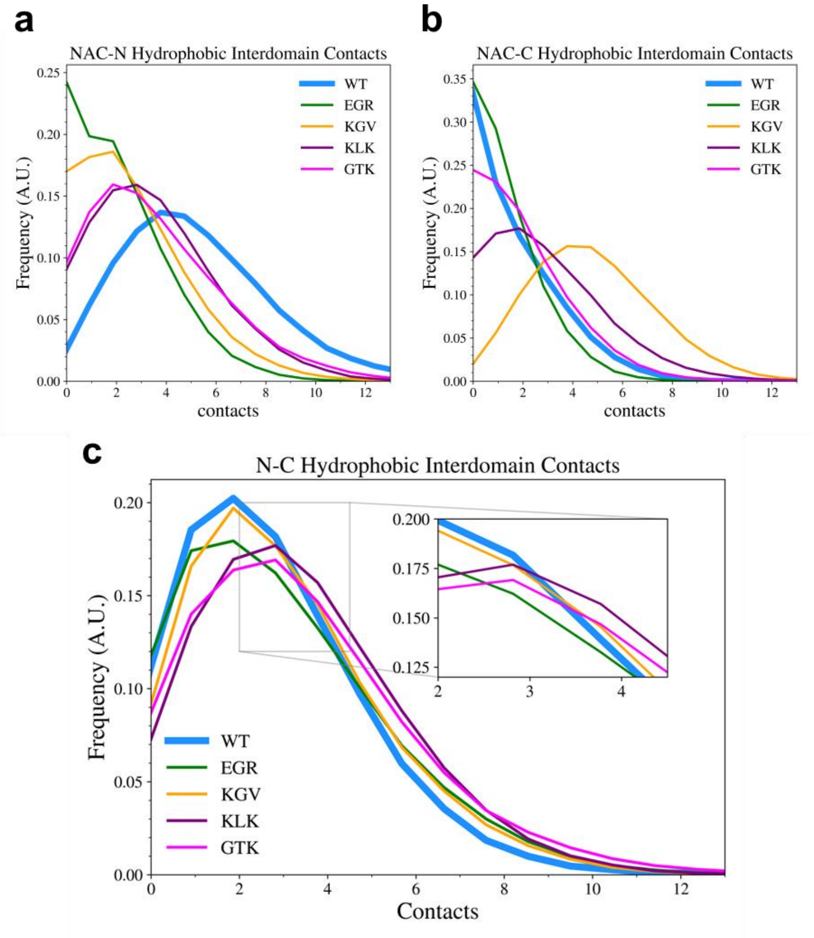

Figure 14. Hydrophobic interdomain contacts of alpha-synuclein and its KTKEGV repeat mutants. Analysis of hydrophobic interdomain contacts (a) between the NAC and N-terminal domains, (b) between NAC and C-terminal domains and (c) between N-terminal and C-terminal domains. The WT is shown in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

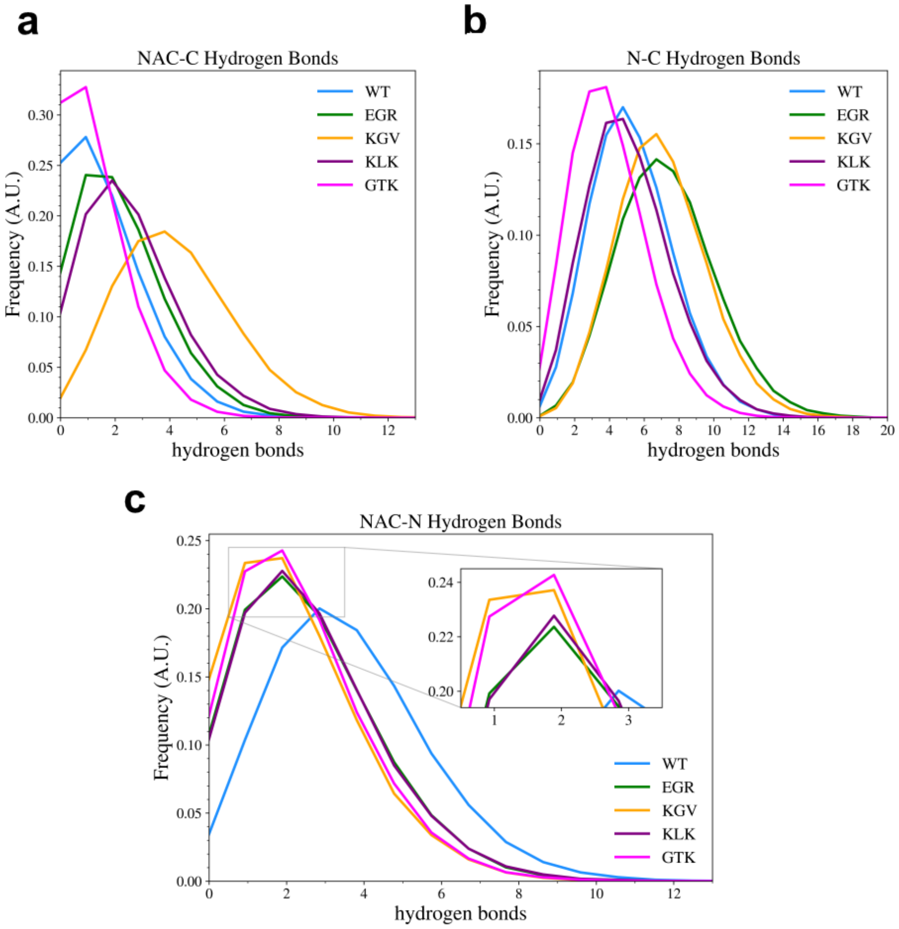

Figure 15. Hydrogen bonds of alpha-synuclein and its KTKEGV repeat mutants. Analysis of hydrogen bonds (a) between the NAC and C-terminal domains, (b) between N-terminal and C-terminal domains and (c) between NAC and N-terminal domains. The WT is shown in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

Figure 15 compares the amounts of hydrogen bonds formed within the NAC-N and NAC-C regions among the WT and its variants: EGR, KGV, KLK, and GTK. The trends shown in **Figure 15a** and **15c** aligned with the previously discussed interdomain contacts profiles (**Figure 12a** and **12b**): KGV mutant elongates in the NAC to C-terminal regions, while there is a significant decrease in the NAC to N-terminal regions for all mutants compared to the WT. This analysis of hydrogen bonds further supports the proposed theory related to residues' charge status change. **Figure 15b**, on the other hand, matches with the trends observed in salt bridge analysis (**Figure 11c**).

3.1.3 Solvent-Accessible Surface Area

SASA offers insights into the dynamic nature of protein folding, stability, and interactions. **Figure 16** shows a comparative analysis of the general SASA across all protein variants, including the WT and its KLKEGV repeat mutants (GTK, KLK, KGV, and EGR). Each graph represents different aspects of the protein structure: the full protein, N-terminal, NAC regions, and C-terminal.

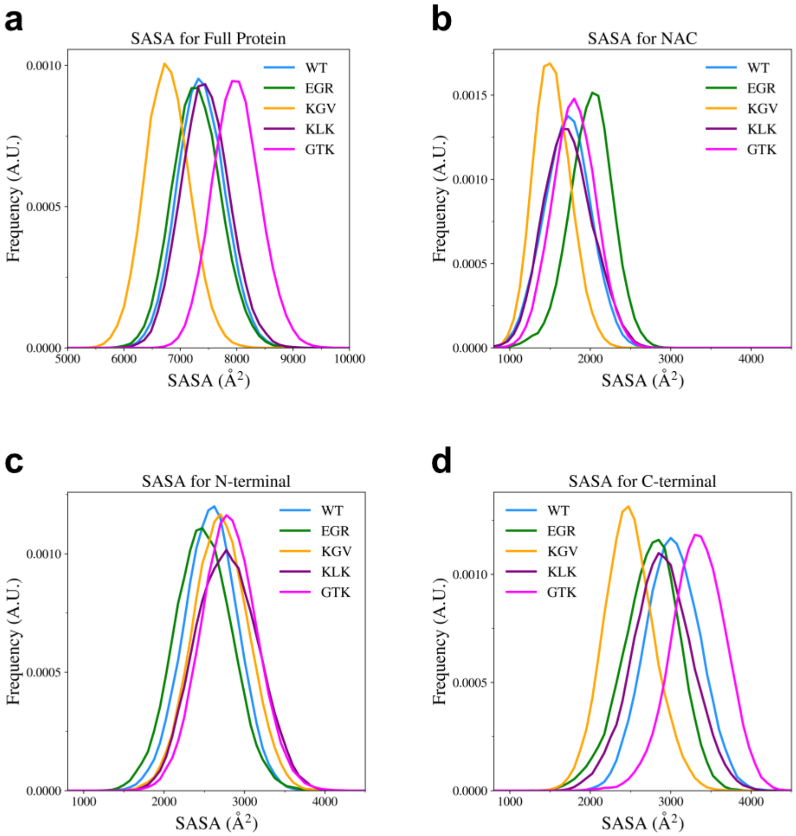

Figure 16. General SASA of alpha-synuclein and its KTKEGV repeat mutants. Analysis in (a) full protein, (b) NAC domains, (c) N-terminal domains, and (d) C-terminal domains with the WT in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

Figures 16a and **16d** show a leftward shift for the KGV variant and a rightward shift for the GTK variant, indicating a change in accessible surface area compared to the WT. These shifts are significant, while **Figure 15b** and **15c**, on the other hand, show overlapping of all mutants and WT with minor differences.

According to the previously interpreted charge status theory, KGV exhibits a significantly elongated R_g in N-terminal domains (**Figure 10c**) while maintaining the overall R_{ee} (**Figure 9**). This suggests a decrease in compactness in the N-terminal region and an increase in compactness in others. **Figure 17** shows the individual interdomain hydrophobic contacts (**Figure 17a**) and salt bridge contents (**Figure 17b**) to comparison with the corresponding data from alpha-synuclein WT (**Figure 6**). The difference observed aligns with the analyses indicating a shift in major contacts from NAC-N to NAC-C (**Figure 12a**, **12b**, **13a**, and **13d**), as well as an increase in the salt bridge contents in N-C (**Figure 11c**). This suggests that, instead of strengthening the attraction between mutation-introduced charges in N-terminal domain with the acidic C-terminal, the charges repulsed among themselves, elongating N-terminal. This forced the contact between the NAC and C-terminal domain. **Figure 17** also shows the KGV repeat mutant with the mutated residues colored in black, aligning with the theory. Their positive charge status is also confirmed and highlighted in blue (**Figure 17c**).

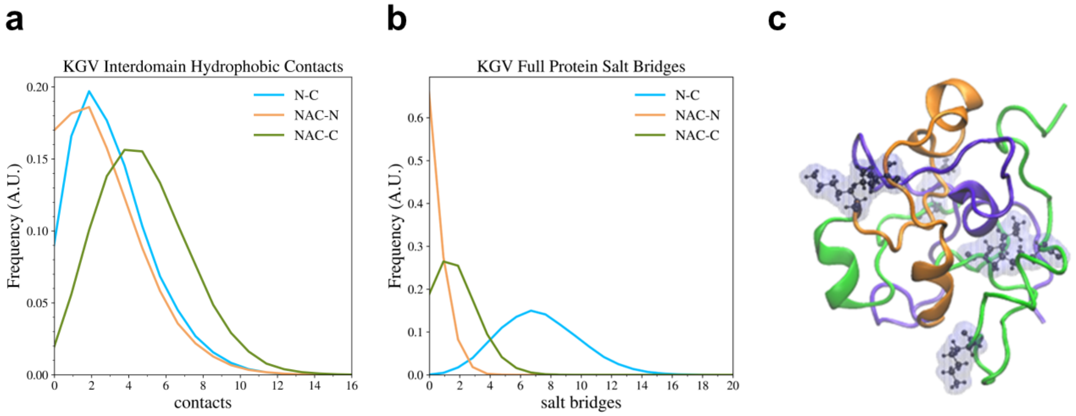

Figure 17. KGV repeat mutant's profile. Individual analysis of (a) interdomain hydrophobic contacts and (b) salt bridges, with the interactions between N-terminal and C-terminal domains shown in cyan, between NAC and N-terminal domains in orange, and between NAC and C-terminal domains in green. Additionally, (c) displays the top frame with mutated residues colored in black in CPK representation. Types of mutated residues (basic/positive) are identified and highlighted in Surf representation with Glass filter.

Combining the SASA analysis with other evidence, there is support for the idea that the charge status change via KGV repeat mutation alters the protein structure and salt bridge contents. This further affects the interdomain contact activities, whether general or hydrophobic interaction, ultimately resulting in a decrease in the solvent-accessible surface area. An extreme case from the other end of the spectrum, as shown in **Figure 16a** and **16d**, is the GTK repeat mutant, which interestingly underwent an "opposite" mutation from the charge status perspective. It transitioned from positively charged to neutral, triggering a slight elongation in C-terminal domain's R_g profile (**Figure 10d**) while maintaining the overall R_{ee} (**Figure 9**). This indicates an increase of compactness in the C-terminal, and a decrease in compactness in others, which, again, is opposite to KGV repeat mutant.

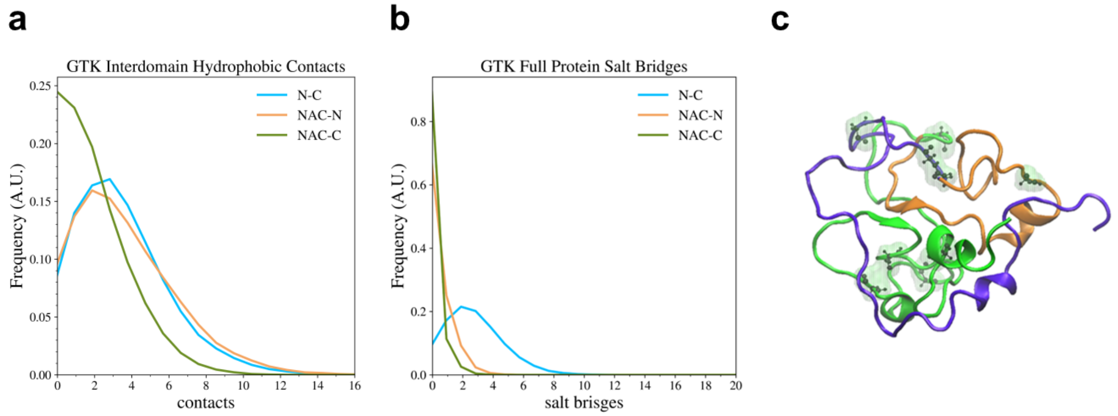

Figure 18. GTK repeat mutant's profile. Individual analysis of (a) interdomain hydrophobic contacts and (b) salt bridges, with the interactions between N-terminal and C-terminal domains shown in cyan, between NAC and N-terminal domains in orange, and between NAC and C-terminal domains in green. Additionally, (c) displays the top frame with mutated residues colored in black in CPK representation. Types of mutated residues (polar/uncharged) are identified and highlighted in Surf representation with Glass filter.

Figure 18 shows the individual interdomain hydrophobic contacts (**Figure 18a**) and salt bridge contents (**Figure 18b**) to compare with the corresponding data from the alpha-synuclein WT (**Figure 6**). The difference observed aligns with the analysis that a shift in major contacts from NAC-N, distributes almost evenly to N-C and NAC-C (**Figure 12a** and **12b**). Additionally, the salt bridge contents are significantly decreased in N-C (**Figure 11c**). This provides sufficient evidence to support the proposed charge status theory, despite the contact map (**Figure 13b**) not showing significant dense clusters. This could be attributed by the overall lower secondary structure contents mentioned in **Figure 8. Figure 18** also shows the GTK repeat mutant, with the mutated residues colored in black, aligning with the theory. Their neutral charge status is also confirmed and highlighted in green (**Figure 18c**).

Hydrophobic SASA (hSASA) quantifies the hydrophobic surface area of a biomolecule accessible to solvents, helping to understand the contribution of hydrophobic effects to protein aggregation. **Figure 19** shows a comparative analysis of the hSASA across all protein variants, once again including the WT and its mutated forms (GTK, KLK, KGV, and EGR). Each graph also represents a different aspect of the protein structure: the full protein, N-terminal, NAC regions, and C-terminal.

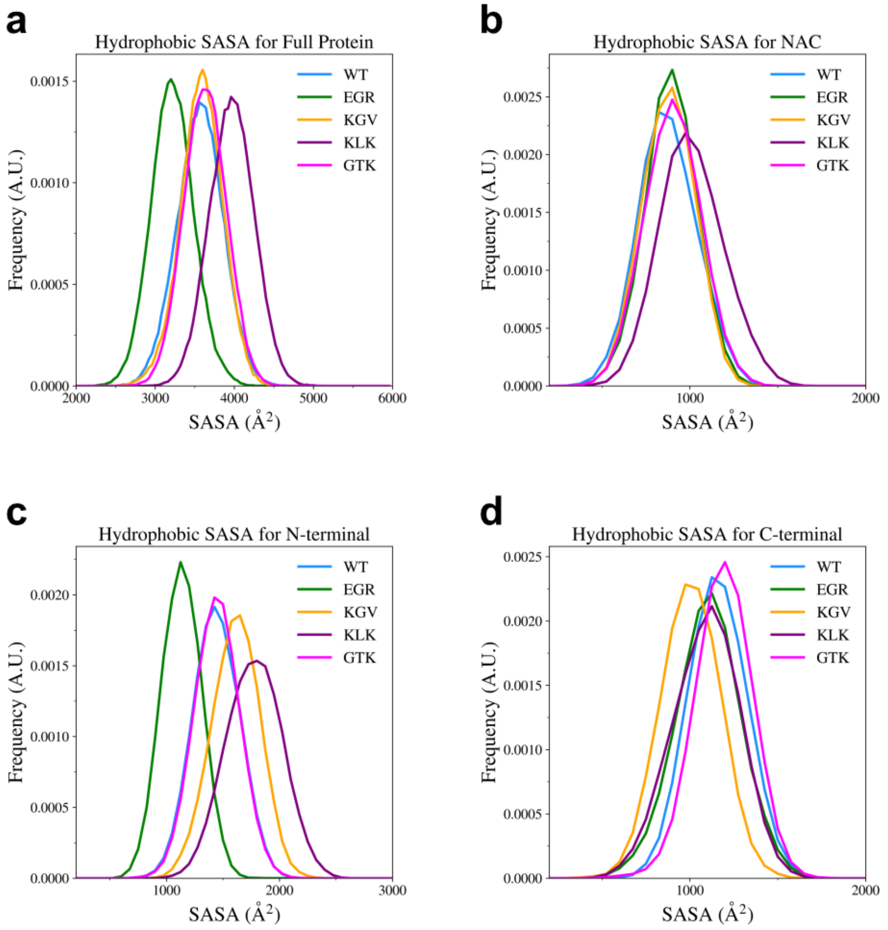

Figure 19. Hydrophobic SASA of alpha-synuclein and its KTKEGV repeat mutants. Analysis in (a) full protein, (b) NAC domains, (c) N-terminal domains, and (d) C-terminal domains with the WT in blue, EGR mutant in green, KGV mutant in orange, KLK mutant in purple, and GTK mutant in pink.

Figures 19b and **19d** display the overlapping of the alpha-synuclein WT and its KTKEGV repeat mutants, suggesting minimal changes in hSASA among the NAC and C-terminal regions. **Figure 19a**, on the other hand, shows a decreased hSASA value for EGR, and an increased value for KLK, as compared to the WT. This suggests that, when evaluating the variants as a full protein, the hSASA did not change much for GTK and KGV repeat mutants, but EGR and KLK repeat mutation produced an effect. Looking at the N-terminal hSASA profile, it reveals similar trends for EGR and KLK, with slightly more information on KGV being very slightly elongated, and GTK being overlapping with the WT.

No correlation has been found between these hSASA trends and the proposed charge status theory. However, the changes in hSASA in N-terminal (**Figure 19c**) are closely related to the hydrophobicity changes listed in **Table 3**. **Figure 20** shows the changes in hydrophobicity of each repeat mutation, as well as the WT, plotted against their post-MD simulation hSASA values in N-terminal. It is critical to note that the degree of hydrophobicity change was represented by delta in residues' hydropathy index. Very importantly, if the mutated residues did not undergo hydrophobicity switching (from hydrophobic to hydrophilic or vice versa), the change in hydrophobicity is 0.

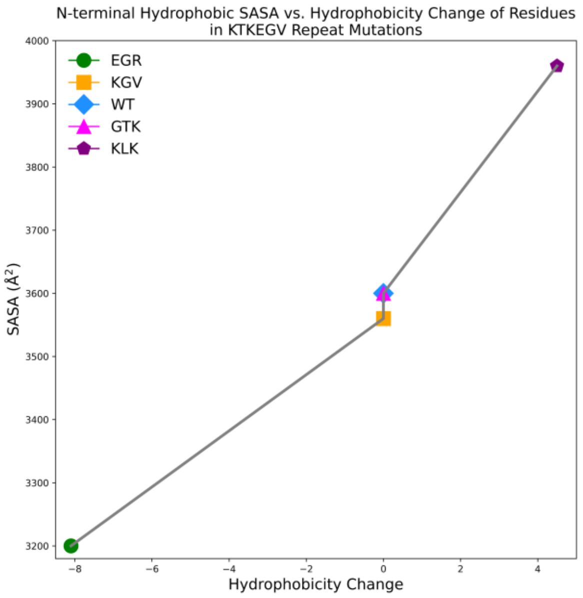

Figure 20. Plot of change in hydrophobic SASA value vs. change in hydrophobicity across alpha-synuclein and its KTKEGV repeat mutants. If the repeat mutation did not involve switching in hydrophobicity of residues (hydrophobic-to-hydrophilic or hydrophilic-to-hydrophobic), the change in hydrophobicity is 0. Otherwise, the hydropathy index was used to calculate the degree of change.

A nearly linear correlation is observed between the hSASA values and the hydrophocity characteristics, which aligns with the hSASA profile (**Figure19c**). Overall, this is an important finding in this study, suggesting that KTKEGV mutations can affect alpha-synuclein's hSASA based on the hydrophobicity switching of the involved residues. Another major finding, reiterated throughout this section, is the charge status theory: alterations in the charge status of residues via KTKEGV repeat mutations have significant effects on the structure and compactness of the mutants. These alterations impact all types of interdomain contacts, ultimately contribute to changes in SASA. SASA and hSASA are keys to determine IDP's aggregation propensity given the solvent-rich physiological environment (within neurons in the brain). The structural and interdomain contacts analyses also lay the foundation for investigating the intradomain contacts, which are closely tied to protein-protein interaction, in the future.

3.2 SARS-CoV-2 Nucleocapsid Protein: Wuhan and its Variant

SR-rich linker segment has important role in regulating pre-mRNA splicing, which is a fundamental process in gene expression. Insights into splicing mechanisms and disease pathways can be gained by understanding the structure-function correlations. Phosphorylation, specifically, can affect the splicing mechanisms by altering the structures by introducing bulky and negatively charged phosphate groups.

3.2.1 The First Attempt

The comprehensive overview of phosphorylation of interest is captured in **Table 4**. This table revealed significant theoretical molecular changes involving the size of residues, their charge status, and hydrophobicity. Phosphorylation on S188 and S206, would result in an increase in the size of the targeted residues, a shift in charge status from neutral to negatively charged, and increased hydrophilicity.

Table 4. Summary of the theoretical change in phosphorylation sites. Uses **Table 1** as reference.

		Phosphorylation on S188 and S206		
SARS-CoV-2 Wuhan Strain	Number of Phosphorylation Site		2	
	Size (amu)	87.08	→	181.08
	Charge Status	Neutral	→	Negatively charged
	Hydrophobicity	Slightly Hydrophilic	→	Highly hydrophilic

The top frame of the SR-rich IDP linker segments were determined via k-means clustering analysis of MD simulation data. To enhance the clarity and understanding of these complex structures, a color-coding scheme was employed for the mole, as shown in **Figure 21**. All structures were captured using the same scaling, enabling direct comparison of the compactness between the Wuhan WT and its phosphorylated variant.

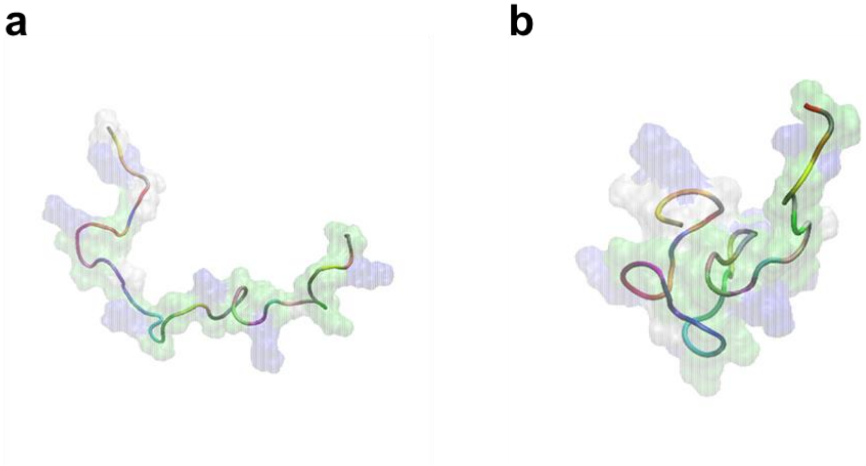

Figure 21. First attempt: top frame of the SR-rich IDP linker segments. Segments are belonging to the SARS-CoV-2 (a) Wuhan strain and (b) its variant, captured post-molecular dynamics (MD) simulation via k-means clustering analysis. Rendered in New Cartoon representation, with residue types highlighted in Surf representation and Glass filter (hydrophobic residues are white, polar/uncharged residues are green, acidic/negative residues are red, and basic/positive residues are blue).

In this study, MD simulations were conducted employing the exact experimental procedure used for alpha-synuclein proteins. Similar structural DSSP, polymeric properties, and SASA analysis were performed and shown in **Figure 22.**

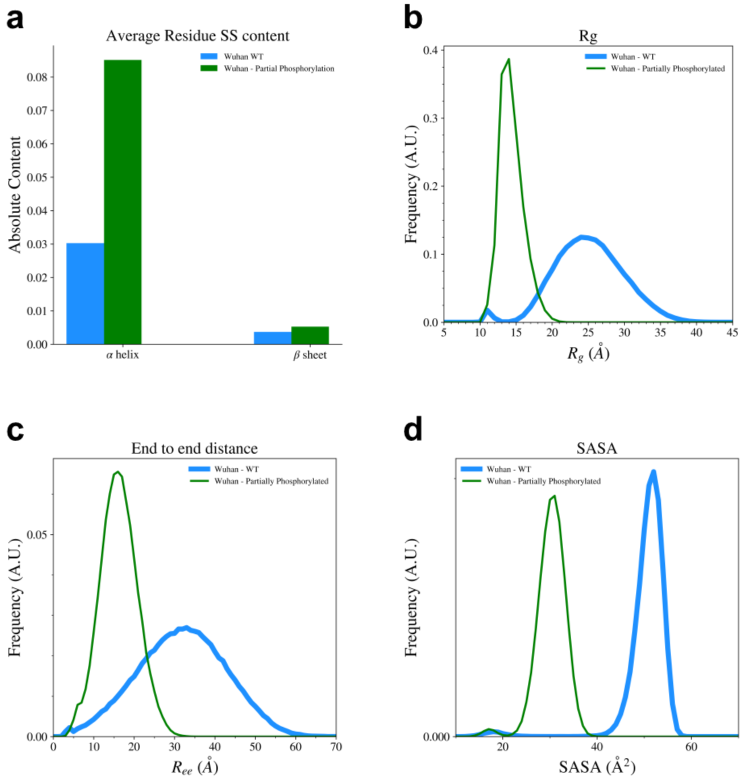

Figure 22. First attempt: DSSP, radius of gyration, end-to-end distance, and general SASA of the SR-rich IDP linker segment from the Wuhan strain and its phosphorylated (S188 and S206) variant. The WT is shown in blue, and the variant in green.

The DSSP analysis (**Figure 22a**) shows that there was an increase in alpha helix contents in the phosphorylated variant. Both the (R_g) and (R_{ee}) show significant decrease in the variant as compared to the WT, suggesting significantly increase in compactness, which aligned with **Figure 21a**. The SASA profile (**Figure 22d**) shows the decrease in solvent-accessible surface area, which is reasonable due to the very compact structure.

The phosphorylation details are captured and shown in **Figure 23**, with the phosphorylation site highlighted (**Figure 23a**) and zoomed in (**Figure 23b**). The two phosphate groups are in close contact with each other, which might be the reason why the structure has become so compact. However, this is not a phenomenon that obeys general chemistry, because phosphate groups are negatively charged so they should be repelling each other. The MD stimulation method obviously needs improvement to take care of these abnormal behaviors between the phosphate groups before assessing results. The first steps would be to evaluate the force field selection and simulation parameters.

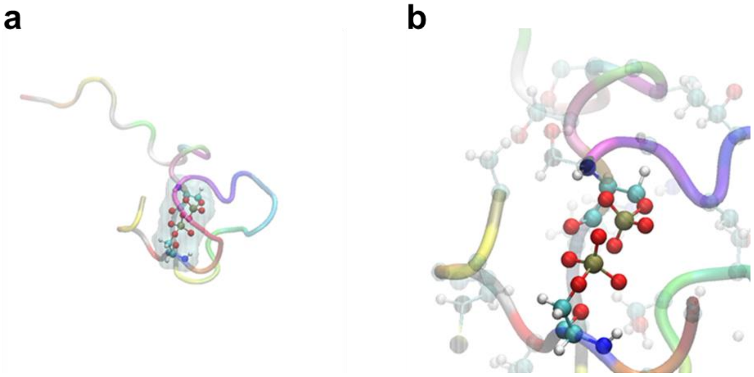

Figure 23. First attempt: phosphorylation details on residues S188 and S206. The rendering of (a) the full segment and (b) a zoomed-in view in Surf representation, with phosphate groups displayed in Element representation. Residue types are colored according to standard conventions.

3.2.2 The Second Attempt

The second attempt involves fine-tuning the simulation parameters, considering the significant differences in the number of residues and searching for a more suitable force field for phosphorylation. This adjustment is necessary due to the transition from full alpha-synuclein proteins to smaller SR-rich IDP linker segments from the SARS-CoV-2 nucleocapsid protein. Additionally, the r203k/g204r variant from the Wuhan strain, developed later in the pandemic, is included in this part.

Figure 24 shows the top frames of all six molecules: the Wuhan and r203k/g204r WT, along with their S188+S206 phosphorylated variants, and all SER+THR phosphorylated vatiants. **Figure 25** presents a DSSP analysis, showing the secondary structure contents in the variants. **Figuer 26** provides the contact maps showcasing the interaction within the variants.

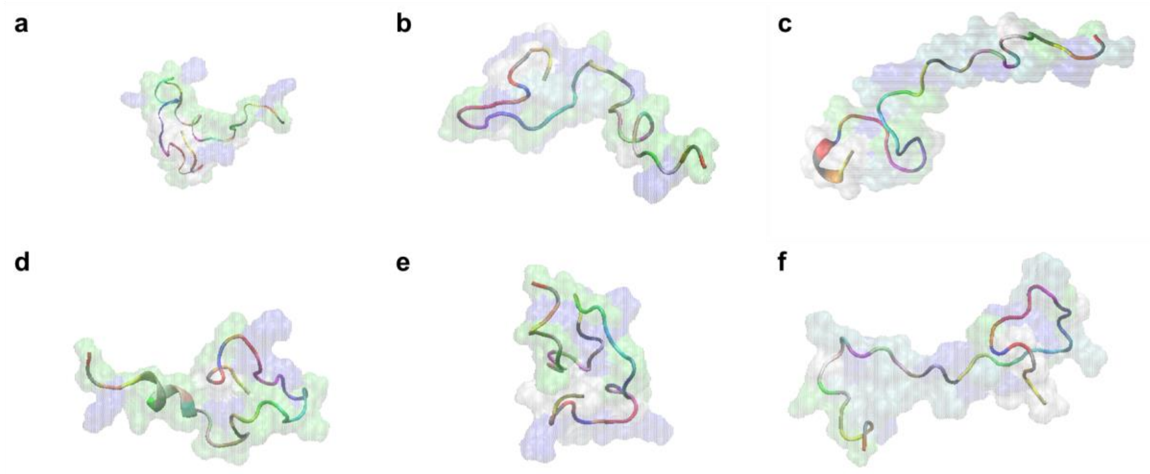

Figure 24. Second attempt: top frame of the SR-rich IDP linker segments. Segments are belonging to the SARS-CoV-2 (a) Wuhan strain WT, (b) Wuhan S188+S206 phosphorylated variant, (c) Wuhan all SER+THR phosphorylated variant, (d) r203k/g204r WT, (e) r203k/g204r S188+S206 phosphorylated variant, and (f) r203k/g204r all SER+THR phosphorylated variant, captured post-molecular dynamics (MD) simulation via k-means clustering analysis. Rendered in New Cartoon representation, with residue types highlighted in Surf representation and Glass filter (hydrophobic residues are white, polar/uncharged residues are green, acidic/negative residues are red, and basic/positive residues are blue).

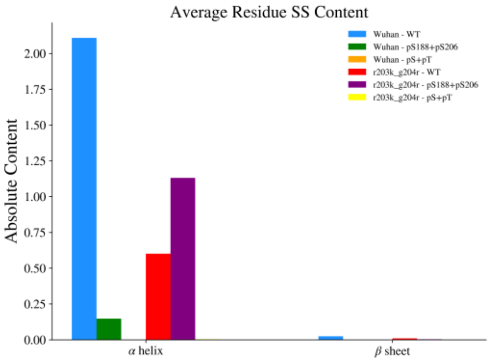

Figure 25. Second attempt: DSSP profile of the SR-rich IDP linker segments. Analysis in segments of SARS-CoV-2 (a) Wuhan strain WT in *blue*, (b) Wuhan S188+S206 phosphorylated variant in *green*, (c) Wuhan all

SER+THR phosphorylated variant in *orange*, (d) r203k/g204r WT in *red*, (e) r203k/g204r S188+S206 phosphorylated variant in *purple*, and (f) r203k/g204r all SER+THR phosphorylated variant in *yellow*.

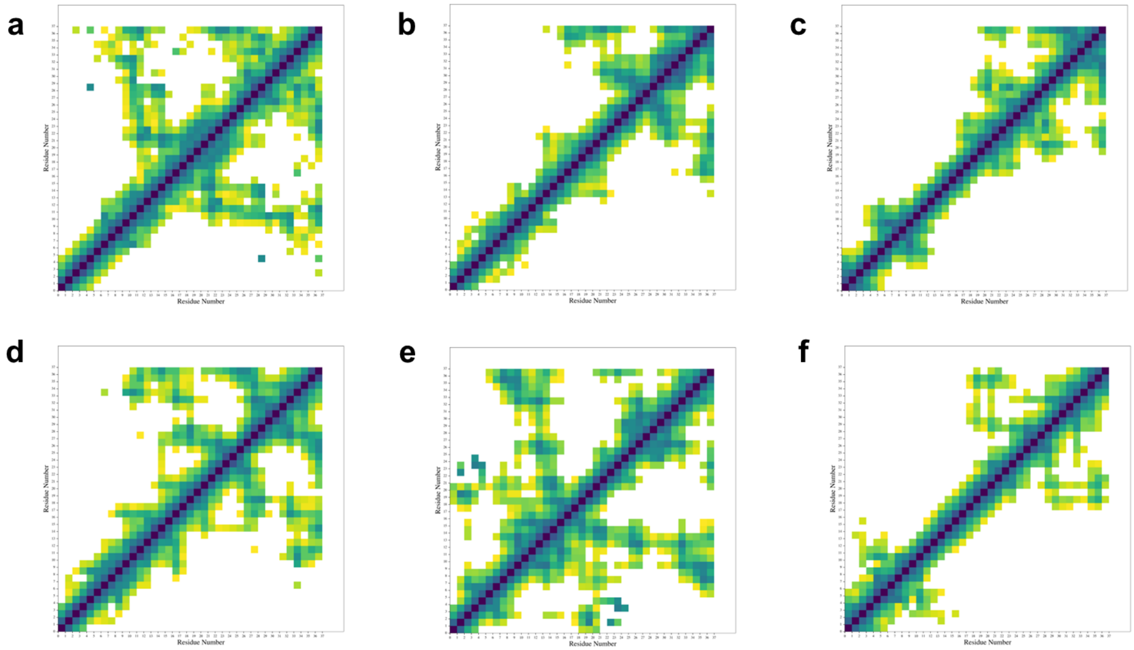

Figure 26. Second attempt: contact maps of the SR-rich IDP linker segments. The maps are illustrating protein structure interactions within segments of SARS-CoV-2 (a) Wuhan strain WT, (b) Wuhan S188+S206 phosphorylated variant, (c) Wuhan all SER+THR phosphorylated variant, (d) r203k/g204r WT, (e) r203k/g204r S188+S206 phosphorylated variant, and (f) r203k/g204r all SER+THR phosphorylated variant. Interaction details are interpreted by aligning residue numbers along the x and y-axes.

In **Figure 24**, significant differences are observed among all variants. The Wuhan and r203k/g204r all SER+THR phosphorylated variants (**Figure 24c** and **24f**) appear to be the most streched, which alignsm with the DSSP analysis in **Figure 25**. These two variants show no detected secondary structure contents. **Figure 26** also confirms similar trends with the most open contact maps (**Figure 26c** and **26f**). The remaining variants are ranked as follow: r203k/g204r S188+S206 phosphorylated variant (**Figure 26e**), Wuhan WT (**Figure 26a**), r203k/g204r WT (**Figure 26d**), and Wuhan S188+S206 phosphorylated variant

(**Figure 26b**), from most contacts to least contacts among the four. Their corresponding top frame structures are shown in **Figure 24a**, **24b**, **24d**, and **24e**.

Figure 27 shows the R_{ee}, R_g, general SASA and hSASA of the SR-rich IDP linker segment from the Wuhan strain and r203k/g204r WT, as well as their variants. These analyses contribute to a quantitative understanding of the changes in compactness and solvent-accessible surface area resulting from phosphorylation. Both R_{ee} (**Figure 27a**) and R_g(**Figure 27b**) exhibit changes, making the assessment of compactness somewhat tricky. This observation is consistent with the nature of these IDP linker segments, which lack the bulkiness of alpha-synuclein structures. The SASA (**Figure 27c**) and hSASA (**Figure 27d**) findings, on the other hand, align with the previous contact map analysis (**Figure 26**), suggesting that the all SER+THR phosphorylated variants have the highest general or hydrophobic solvent-accesible surface area.

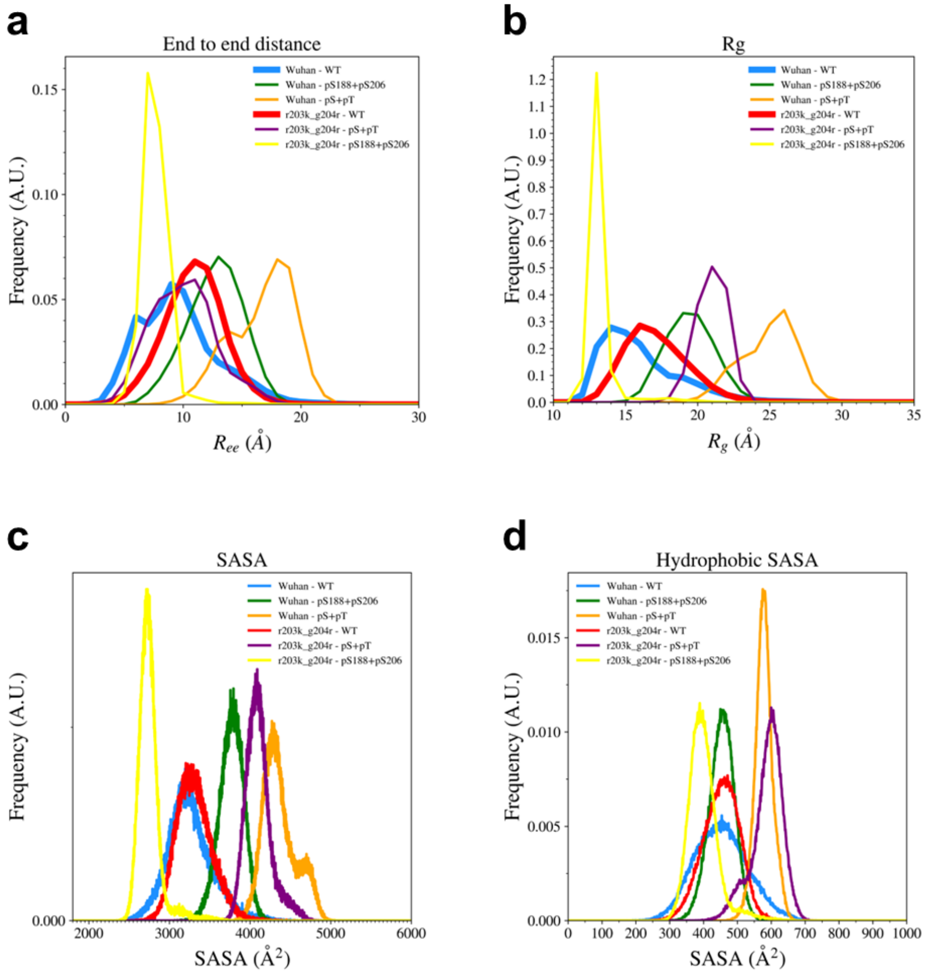

Figure 27. Second attempt: end-to-end distance, radius of gyration, general SASA and hydrophobic SASA of the SR-rich IDP linker segment from the Wuhan strain and r203k/g204r WT, as well as their variants. Wuhan strain WT is in *blue*, Wuhan S188+S206 phosphorylated variant is in *green*, Wuhan all SER+THR phosphorylated variant is in *orange*, r203k/g204r WT is in *red*, r203k/g204r S188+S206 phosphorylated variant is in *purple*, and r203k/g204r all SER+THR phosphorylated variant is in *yellow*.

Figure 28 and **29** show the phosphorylation details of the linker segments. This second attempt of MD simulation has successfully generated a comprehensive analysis of the SR-rich linker segments of SAR-CoV-2 Wuhan strain and r203k/g204r, including both WT and phosphorylated variants.

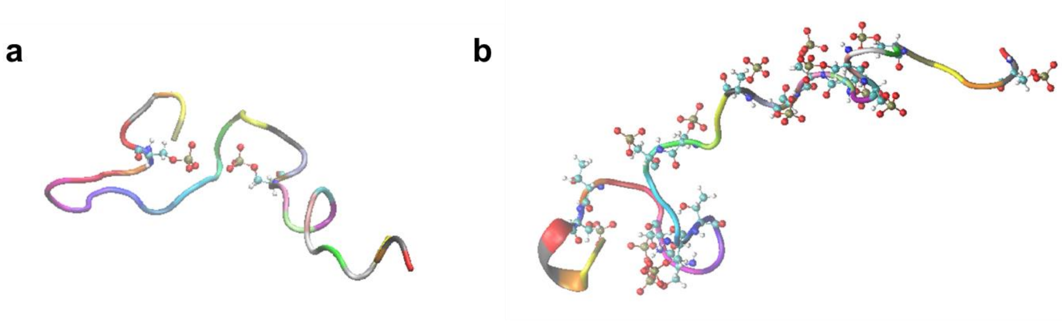

Figure 28. Second attempt: phosphorylation details on the SR-rich IDP linker segments of the Wuhan strain. The rendering of (a) S108+S206 phosphorylated variant and (b) all SER+THR phosphorylated variant in New Cartoon representation, with phosphate groups displayed in Element representation. Residue types are colored according to standard conventions.

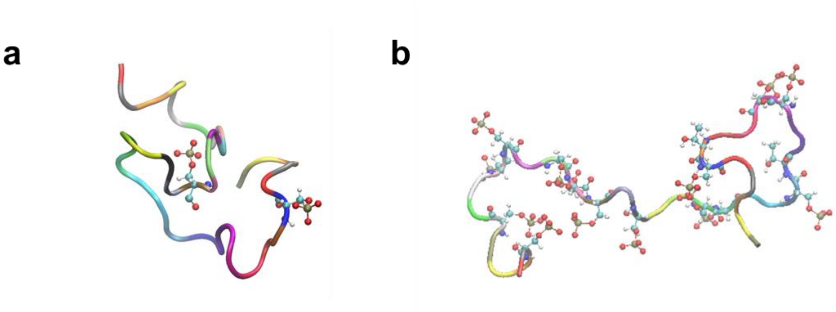

Figure 29. Second attempt: phosphorylation details on the SR-rich IDP linker segments of the r203k/g204r variant. The rendering of (a) S108+S206 phosphorylated variant and (b) all SER+THR phosphorylated variant in New Cartoon representation, with phosphate groups displayed in Element representation. Residue types are colored according to standard conventions.

CHAPTER 4: CONCLUSION

In summary, this comprehensive investigation has explained the structural, interdomain contacts, and solvent-accessible area transformations induced by repeat KTKEGV mutations within alpha synuclein, an Intrinsically Disordered Proteins (IDP). Through analyses, the study highlights the significance of changes in charge status and hydrophobicity switch associated with specific mutations. These modifications profoundly alter the physical and chemical properties of proteins, offering insights into protein folding, stability, and interactions.

KTKEGV repeat mutations consistently reflect the expected theoretical changes in residue swapping concerning charge status and the hydropathy index, thereby serving as indicators for hydrophobicity level. Employing k-means clustering analysis and visualizing top frame structures using VMD visualization software, with DSSP aiding in understanding secondary structure contents, pairing with end-to-end distance (R_{ee}) and radius of gyration (R_g) analyses, revealed that while the overall protein sizes and compactness remain relatively stable, individual domains exhibit diverse alterations depending on the mutations. Additionally, the distribution of salt bridges in alpha-synuclein and its variants exhibits intriguing trends: the observed patterns in salt bridge formation from N to C-terminal closely reflect the theoretical changes in charge status outlined, suggesting a molecular reflection of theoretical alterations due to residue swapping, particularly evident in the interaction between N and C-terminals.

Furthermore, interdomain hydrophobic and general contacts are also detailed, supported by corresponding contact maps and followed by hydrogen bond analysis. Hydrophobic and general Solvent Accessible Surface Area (hSASA and SASA) analyses provide insights into the surface area exposed to the solvent, with hSASA results aligning with the theoretical change in hydrophobicity switch expected from residue swapping. Additionally, the general SASA analysis indicates that the accessible surface area in N-terminal domains follows the proposed charge status theory via KTKEGV mutations.

Two representative repeat mutants in the study, KGV and GTK, which represent extreme cases in terms of SASA (focusing on the N-terminal domains), were individually analyzed, highlighting the contrasting effects of the charge status theory. The KGV repeat mutation induces repulsion among introduced charges in the N-terminal domain, elongating the domain and shifting major contacts towards NAC-C, while the GTK repeat mutation neutralizes charges in the N-terminal domain, compacting the C-terminal domain, and redistributing interdomain contacts towards N-C and NAC-C. These findings highlight the critical role of charge status in altering protein structure and interdomain interactions in alpha-synuclein mutants.

On the other hand, hydrophobicity switching contributes significantly to the influence of KTKEGV mutations on alpha-synuclein's hSASA exhibiting a nearly linear relationship. These findings deepen the understanding of the structural consequences of KTKEGV mutations in alpha-synuclein, providing important building blocks to predict IDPs' behavior from the physical and chemical perspectives, with potential implications for protein function and dysfunction.

Shifting focus to the SARS-CoV-2 nucleocapsid, this chapter highlights the phosphorylation of the SR-rich linker segment as a key modifier of structural and functional properties. Understanding the structure-function correlations provides insights into splicing mechanisms and disease pathways, particularly regarding the impact of phosphorylation. Phosphorylation, by introducing bulky and negatively charged phosphate groups, can significantly alter splicing mechanisms by affecting the structures of proteins involved.

In the initial attempt of MD simulations on the SR-rich IDP linker segments of the Wuhan strain and its S188+S206 phosphorylated variant, their top frames and structural alterations were identified. This includes changes in secondary structure contents, polymeric properties, and SASA are analyzed. The phosphorylated variant exhibited increased alpha helix contents and compactness, as evidenced by decreased R_{ee} and R_g, and SASA. However, closer examination of phosphorylation details reveals abnormal behaviors between the phosphate groups, evidencing the necessity for refinement in simulation parameters.

Motivated by these findings, a second attempt of MD simulations was done to fine-tune simulation parameters, considering variations in residue numbers, and exploring more suitable force fields for phosphorylation, as compared to alpha-synuclein. In this attempt, an additional variant, r203k/g204r, was also included for a broader exploration of phosphorylation.

Significant differences were observed among the variants. The Wuhan and r203k/g204r all SER+THR phosphorylated variants appeared to be the most stretched, lacking detected secondary structures. Contact maps further confirm these trends, showing the most open interactions in these variants. General SASA and hSASA analyses also aligned with contact map findings, indicating that all SER+THR phosphorylated variants have the highest solvent-accessible surface area.

Phosphorylation details of linker segments, with representations of element interactions, concluded the investigation. This provided a comprehensive understanding of the SR-rich linker segments of the SAR-CoV-2 nucleocapsid protein, including WT and phosphorylated variant. These efforts deepen the understanding of the structural consequences of phosphorylation in SR-rich IDP linker, offering valuable insights into disease mechanisms and potential therapeutic targets.

REFERENCES

(1) Wen, J.-H.; He, X.-H.; Feng, Z.-S.; Li, D.-Y.; Tang, J.; Liu, H.-F. Cellular Protein Aggregates: Formation, Biological Effects, and Ways of Elimination. *Int. J. Mol. Sci.* **2023**, *24*, 8593. https://doi.org/10.3390/ijms24108593.

(2) Sheng, J.; Olrichs, N.; Gadella, B.; Kaloyanova, D.; Helms, J. Regulation of Functional Protein Aggregation by Multiple Factors: Implications for the Amyloidogenic Behavior of the CAP Superfamily Proteins. *Int. J. Mol. Sci.* **2020**, *21*. https://doi.org/10.3390/ijms21186530.

(3) Yu, J.-F.; Wu, E.-S.; Wang, C.-L.; Wang, H.-M.; Wang, J.-H. Classification of Ordered/Disordered Regions of Intrinsically Disordered Proteins Based on Comprehensive Sequence Analysis and Chou's Pseudo Amino Acid Composition Method. *Match* **2016**, *75*, 417–430.

(4) Zhang, Z.; Zhao, Q.; Gong, Z.; Du, R.; Liu, M.; Zhang, Y.; Zhang, L.; Li, C. Progress, Challenges and Opportunities of NMR and XL-MS for Cellular Structural Biology. *JACS Au* **2024**, *4* (2), 369–383. https://doi.org/10.1021/jacsau.3c00712.

(5) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10* (12), 980–980. https://doi.org/10.1038/nsb1203-980.

(6) Bernal-Conde, L. D.; Ramos-Acevedo, R.; Reyes-Hernández, M. A.; Balbuena-Olvera, A. J.; Morales-Moreno, I. D.; Argüero-Sánchez, R.; Schüle, B.; Guerra-Crespo, M. Alpha-Synuclein Physiology and Pathology: A Perspective on Cellular Structures and Organelles. *Front. Neurosci.* **2020**, *13*.

(7) Jo, E.; Mclaurin, J.; Yip, C.; George-Hyslop, P.; Fraser, P. α-Synuclein Membrane Interactions and Lipid Specificity. *J. Biol. Chem.* **2000**, *275*, 34328–34334. https://doi.org/10.1074/jbc.M004345200.

(8) Dettmer, U.; Newman, A. J.; von Saucken, V. E.; Bartels, T.; Selkoe, D. KTKEGV Repeat Motifs Are Key Mediators of Normal α-Synuclein Tetramerization: Their Mutation Causes Excess Monomers and Neurotoxicity. *Proc. Natl. Acad. Sci.* **2015**, *112* (31), 9596–9601. https://doi.org/10.1073/pnas.1505953112.

(9) Oliveira, L. M. A.; Gasser, T.; Edwards, R.; Zweckstetter, M.; Melki, R.; Stefanis, L.; Lashuel, H. A.; Sulzer, D.; Vekrellis, K.; Halliday, G. M.; Tomlinson, J. J.; Schlossmacher, M.; Jensen, P. H.; Schulze-Hentrich, J.; Riess, O.; Hirst, W. D.; El-Agnaf, O.; Mollenhauer, B.; Lansbury, P.; Outeiro, T. F. Alpha-Synuclein Research: Defining Strategic Moves in the Battle against Parkinson's Disease. *Npj Park. Dis.* **2021**, *7* (1), 65. https://doi.org/10.1038/s41531-021-00203-9.

(10) Xu, L.; Bhattacharya, S.; Thompson, D. Re-Designing the α-Synuclein Tetramer. *Chem. Commun.* **2018**, *54*. https://doi.org/10.1039/C8CC04054K.

(11) Deleersnijder, A.; Gerard, M.; Debyser, Z.; Baekelandt, V. The Remarkable Conformational Plasticity of Alpha-Synuclein: Blessing or Curse? *Trends Mol. Med.* **2013**, *19*. https://doi.org/10.1016/j.molmed.2013.04.002.

(12) Iyer, A.; Claessens, M. *Disruptive Membrane Interactions of Alpha-Synuclein Aggregates*; 2018; Vol. 1867. https://doi.org/10.1016/j.bbapap.2018.10.006.

(13) Zarbiv, Y.; Haham, D.; Israeli, E.; Abd Elhadi, S.; Grigoletto, J.; Sharon, R. Lysine Residues at the First and Second KTKEGV Repeats Mediate α-Synuclein Binding

to Membrane Phospholipids. *Neurobiol. Dis.* **2014**, *70*. https://doi.org/10.1016/j.nbd.2014.05.031.

(14) Gould, N.; Mor, D.; Lightfoot, R.; Malkus, K.; Giasson, B.; Ischiropoulos, H. Evidence of Native α-Synuclein Conformers in the Human Brain. *J. Biol. Chem.* **2014**, *289*. https://doi.org/10.1074/jbc.C113.538249.

(15) Brontesi, L.; Imberdis, T.; Ramalingam, N.; Dettmer, U. The Effect of KTKEGV Repeat Motif and Intervening ATVA Sequence on α-Synuclein Solubility and Assembly. *J. Neurochem.* **2023**, *165*. https://doi.org/10.1111/jnc.15763.

(16) Casasanta, M. A.; Jonaid, G. M.; Kaylor, L.; Luqiu, W. Y.; DiCecco, L.-A.; Solares, M. J.; Berry, S.; Dearnaley, W. J.; Kelly, D. F. Structural Insights of the SARS-CoV-2 Nucleocapsid Protein: Implications for the Inner-Workings of Rapid Antigen Tests. *Microsc. Microanal.* **2023**, *29* (2), 649–657. https://doi.org/10.1093/micmic/ozac036.

(17) Savastano, A.; Ibáñez de Opakua, A.; Rankovic, M.; Zweckstetter, M. Nucleocapsid Protein of SARS-CoV-2 Phase Separates into RNA-Rich Polymerase-Containing Condensates. *Nat. Commun.* **2020**, *11*. https://doi.org/10.1038/s41467-020-19843-1.

(18) Yang, H.; Rao, Z. Structural Biology of SARS-CoV-2 and Implications for Therapeutic Development. *Nat. Rev. Microbiol.* **2021**, *19* (11), 685–700. https://doi.org/10.1038/s41579-021-00630-8.

(19) Cubuk, J.; Alston, J. J.; Incicco, J. J.; Singh, S.; Stuchell-Brereton, M. D.; Ward, M. D.; Zimmerman, M. I.; Vithani, N.; Griffith, D.; Wagoner, J. A.; Bowman, G. R.; Hall, K. B.; Soranno, A.; Holehouse, A. S. The SARS-CoV-2 Nucleocapsid Protein

Is Dynamic, Disordered, and Phase Separates with RNA. *Nat. Commun.* **2021**, *12* (1), 1936. https://doi.org/10.1038/s41467-021-21953-3.

(20) Zhao, H.; Nguyen, A.; Wu, D.; Li, Y.; Hassan, S.; Chen, J.; Shroff, H.; Piszczek, G.; Schuck, P. Plasticity in Structure and Assembly of SARS-CoV-2 Nucleocapsid Protein. *PNAS Nexus* **2022**, *1*. https://doi.org/10.1093/pnasnexus/pgac049.

(21) Kyte, J.; Doolittle, R. A Simple Method for Displaying the Hydropathic Character of a Protein. *J Mol Biol* **1982**, *157*, 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.

(22) Uversky, V. N.; Dunker, A. K. Understanding Protein Non-Folding. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2010**, *1804* (6), 1231–1264. https://doi.org/10.1016/j.bbapap.2010.01.017.

(23) Liu, H.; Song, D.; Lu, H.; Luo, R.; Chen, H.-F. Intrinsically Disordered Protein-Specific Force Field CHARMM36IDPSFF. *Chem. Biol. Drug Des.* **2018**, *92*. https://doi.org/10.1111/cbdd.13342.

(24) Nolen, B.; Taylor, S.; Ghosh, G. Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation. *Mol. Cell* **2004**, *15*, 661–675. https://doi.org/10.1016/j.molcel.2004.08.024.

(25) Diella, F.; Haslam, N.; Chica, C.; Budd, A.; Michael, S.; Brown, N.; Trave, G.; Gibson, T. Understanding Eukaryotic Linear Motifs and Their Role in Cell Signaling and Regulation. *Front. Biosci. J. Virtual Libr.* **2008**, *13*, 6580–6603. https://doi.org/10.2741/3175.

(26) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929. https://doi.org/10.1063/1.1755656.

(27) Karplus, M.; McCammon, J. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652. https://doi.org/10.1038/nsb0902-646.

(28) Allen, M. P. Introduction to Molecular Dynamics Simulation. In *NIC series*; John von Neumann Institute for Computing: Jülich, 2004; Vol. 23, pp 1–28.

(29) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2004.

(30) Omelyan, I.; Mryglod, I.; Reinhard, F. Optimized Verlet-like Algorithms for Molecular Dynamics Simulations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2002**, *65*, 056706. https://doi.org/10.1103/PhysRevE.65.056706.

(31) Allen, M. P.; Tildesley, D. J.; Tildesley, D. J. *Computer Simulation of Liquids*, Reprinted.; Oxford science publications; Clarendon Pr: Oxford, 2009.

(32) Case, D.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Giambasu, G.; Gilson, M.; Gohlke, H.; Götz, A.; Harris, R.; Izadi, S.; Izmailov, S.; Jin, C.; Kasavajhala, K.; Kollman, P. *Amber 2021*; 2021. https://doi.org/10.13140/RG.2.2.15902.66881.

(33) Oshima, H. 12.4 Gaussian Accelerated Molecular Dynamics (GaMD), 2022. https://www.r-ccs.riken.jp/labs/cbrt/tutorials2022/tutorial-12-4/.

(34) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J.*

Chem. Theory Comput. **2015**, *11* (8), 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.

(35) Wang, J.; Arantes, P.; Bhattarai, A.; Hsu, R.; Pawnikar, S.; Huang, Y.; Palermo, G.; Miao, Y. Gaussian Accelerated Molecular Dynamics: Principles and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**. https://doi.org/10.1002/wcms.1521.

(36) Liao, Q. Enhanced Sampling and Free Energy Calculations for Protein Simulations. In *Progress in molecular biology and translational science*; 2020; Vol. 170. https://doi.org/10.1016/bs.pmbts.2020.01.006.

(37) Pang, Y.; Miao, Y.; Wang, Y.; McCammon, J. Gaussian Accelerated Molecular Dynamics in NAMD. *J. Chem. Theory Comput.* **2016**, *13*. https://doi.org/10.1021/acs.jctc.6b00931.

(38) Onufriev, A.; Case, D. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 1–22. https://doi.org/10.1146/annurev-biophys-052118-115325.

(39) Zagrovic, B.; Pande, V. Solvent Viscosity Dependence of the Folding Rate of a Small Protein: Distributed Computing Study. *J. Comput. Chem.* **2003**, *24*, 1432–1436. https://doi.org/10.1002/jcc.10297.

(40) Srinivasan, J.; Trevathan, M.; Beroza, P.; Case, D. Application of a Pairwise Generalized Born Model to Proteins and Nucleic Acids: Inclusion of Salt Effects. *Theor. Chem. Acc. Theory Comput. Model. Theor. Chim. Acta* **1999**, *101*, 426–434. https://doi.org/10.1007/s002140050460.

(41) Zou, X.; Kuntz, I. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. *J. Am. Chem. Soc. - J AM CHEM SOC* **1999**, *121*. https://doi.org/10.1021/ja984102p.

(42) Norjmaa, G.; Ujaque, G.; Lledós, A. Beyond Continuum Solvent Models in Computational Homogeneous Catalysis. *Top. Catal.* **2022**, *65* (1), 118–140. https://doi.org/10.1007/s11244-021-01520-2.

(43) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

Vita

Name	*Coby Yeung*
Baccalaureate Degree	*Bachelor of Science*, *Stony Brook University,* *Stony Brook* *Major: Chemistry*
Date Graduated	*May 2018*
Other Degrees and Certificates	*Master of Business Administration*, *Stony Brook University,* *Stony Brook* *Major: Marketing*
Date Graduated	*December 2023*