A Molecular Dynamics Simulation Study of Peptides Confined in Aqueous Nanodroplets

by

© **Yiming Huang**

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science.

Department of Physics & Physical Oceanography

Memorial University

May 2023

St. John's, Newfoundland and Labrador, Canada

Abstract

We present a molecular simulation study of five peptides, CBS-5, CBS-9, KKKDDD, DKDKDK and GAD-1, in aqueous nanodroplets. The simulation is performed through the GROMACS/2021.4 package, employing the OPLS-AA force field. Hydrophobic and amphipathic peptides prefer to reside near the surface while the extremely hydrophilic DK peptides are kept deep inside the droplet at higher temperatures. At low temperature, the nanodroplet preserves a shell-like structure with higher density in the subsurface region as previously observed for pure water nanodroplets. Peptides confined in such heterogeneous nanodroplets show a temperature-dependent transition in conformation and spatial distribution. Together with the nanodroplet environment, a few counter ions (Na^+ or Cl^-) added to electrically neutralize the peptide significantly affect the unfolding process of GAD-1 at 300 K.

Acknowledgements

I would express gratitude to my supervisors Dr. Ivan Saika-Voivod and Dr. Shahrazad Malek who generously spent time to guide me through every complicated aspect in this research. I enjoy every discussion on Slack and in the honours room, where I don't have to explain too much for understanding. I would acknowledge that Dr. Shahrazad Malek wrote the centering code and the density code that facilitates subsequent analysis, including the various distribution plots presented in the thesis. I would thank my family who support me financially and academically.

I also thank anyone else who had ever helped me.

Table of contents

List of figures

Chapter 1

Introduction

Water is an indispensable element in the origin of life. Every characteristic of water allows for the existence of all diverse, vigorous, and fragile living creatures on Earth. The immense heat capacity of water regulates a stable temperature for our earth; the density anomaly prevents rivers and oceans from freezing in cold winter; and the polarity propels organic compounds to array against the entropy. The cell is considered the basic structural and functional unit of life. With crowded cell organelles inside, 70% mass of a cell is constituted by water. Koga et al. [1] redefined the assembly of a peptide and nucleotide-containing microdroplet as a membrane-free protocell model in 2011. They discovered through experiments that the assembly is stable under changes in temperature and promotes α-helical structure. Furthermore, recent work has revealed that a heterogeneous structure in water nanodroplets emerges at low temperatures with a relatively tetrahedrally well-ordered core and a more disordered subsurface [2]. Subsequent work showed that specific ions are expelled from the core at low temperature [3]. All the previous research shows a potential correlation between water confinement and biomolecular activities. To explore the interactions between

water and proteins, we performed molecular dynamics simulations for five peptide-nanodroplet systems and the corresponding bulk peptide-water systems under a wide range of temperatures. We anticipate that the temperature-dependent structure of water nanodroplets may significantly affect the structure and location of peptides within nanodroplets. Our work may serve in the rational design of drugs if new mechanisms to control peptide folding are confirmed, which include changing the temperature or the size of water droplets.

1.1 Protein Secondary Structure

Peptides are chains of amino acids connected by *trans* peptide bonds -CO-NH-. The resonance between C=O and C-N, which brings rigidity between two amino acids, is the reason why proteins have relatively stable structures. In nature, most organic compounds coexist with their corresponding stereoisomers, but this is not true with amino acids. Most native proteins are composed of L-amino acids. Without the complexity of chirality and random rotation around peptide bonds, the structure of a protein is chiefly determined by its 1-D sequence of amino acids. The secondary structures of the protein are promoted by hydrogen bonds between C=O and H-N, while the right-hand α-helix is the most prevalent form of the secondary structure. In α-helices, the i-th H-N group forms a hydrogen bond with the O=C group 4 residues earlier. This repeated $i + 4 \rightarrow i$ hydrogen bonding is the most crucial characteristic shared by proteins. β-sheets are multiple strands of peptides that are connected laterally by the hydrogen bonds between backbone C=O and H-N. The connection can be antiparallel (the N-terminus of one strand is adjacent to the C-terminus of the next) or parallel (all the N-terminus of successive strands are oriented in the same

direction). In antiparallel structures, the hydrogen bonds form between i-th and j-th amino acids in adjacent strands. In parallel structures, the hydrogen bonds form between i-th and (j+1)-th or i-th and (j-1)-th amino acids in adjacent strands.

1.2 Recent Relevant Research

Nanoscale water droplets in the real world are applied in electrospray ionization (ESI) and membrane vesicle trafficking. Electrospray ionization is a mass spectrometry technique for macromolecules (e.g. proteins). Vesicles are containers of enzymes or organelle secretions, which help giant molecules pass through the membrane by merging with the lipid membrane. The water droplet size in ESI ranges from 100 nm to 10 µm, and vesicles are cytoplasm enclosed by a lipid bilayer with a diameter ranging from 30 nm to 1 µm.

El-Baba et al. [4] stated that ubiquitin within electrosprayed nanodroplets displays a temperature or laser-induced unfolding transition similar to bulk water. Their experiments discovered a general dependence on temperature and droplet sizes. The ratio of compact conformers decreases upon increasing the temperature of the solution or the laser power. The *trans*-conformers are greatly diminished in 0.05 µm droplets compared to 1 µm at high temperature (96 °C) and high laser power (13 W).

Kim et al. [5] simulated the dehydration process of charged nanodroplets with a peptide inside. The hydronium and Cl^- ions present in the nanodroplets were ejected when the solvent evaporated. As droplet evaporated and decreased in size, the peptide adapted it through conformational changes before being kinetically trapped in a compact conformation. The authors suggested that intramolecular interactions involving the glutamines and charged sites, as well as the droplet volume, were responsible for

the formation of the compact conformer.

Zhong et al. [6] simulated the ·NH2 (amino radical) in a nanodroplet composed of 191 water molecules. The ·NH2 was always driven to the surface after 200 ps regardless of its initial location. The authors studied the orientation and probability of typical hydrogen-bonding complexes between $\cdot NH_2$ and H_2O, and concluded that the $HOH\cdots NH_2$ hydrogen bond dominates over $HNH\cdots OH_2$ and the N atom forms a loose hydrogen-bonding network with the surrounding water. Due to the weak hydrogen bond, the ·NH2 in the outer-layer of the droplet has a very small radial diffusion coefficient compared to the angular diffusion coefficient.

1.3 Peptides Studied

The peptides in our research are CBS-5 and CBS-9 [7], GAD-1 [8], KKKDDD and DKDKDK; all are small enough to be loaded into a water droplet of 776 or 2880 molecules.

The CBS peptides disrupt calmodulin from binding to Cyclin E during phase G1 of the cell cycle, thereby blocking cancer cells or smooth muscle cells from proliferation. Their high selectivity to Cyclin E and low toxicity to other cells make CBS peptides potential therapeutic agents for cancers [9].

Peptide	1	4	7	10	13	16	19
CBS-5	VTV	FL					
CBS-9	ANV	TVF	LQD				
KKKDDD	KKK	DDD					
DKDKDK	DKD	KDK					
GAD-1	FIH	HII	GWI	SHG	VRA	IHR	AIH-NH2

Table 1.1: Amino acids sequence of our peptides

The two KD peptides are constructed for this thesis to study the interaction between the water nanodroplet and charged sites on the peptides. KKKDDD is the most polar sequence of peptides consisting of three positively charged lysines (K) and three negatively charged aspartates (D). DKDKDK has the same composition, but the charge distribution is less blocky. They are both charge-rich and perfectly hydrophilic peptides (which never crossed the subsurface region).

Gaduscidin-1, or GAD-1, is an antimicrobial peptide (AMP) from the cod's immune system, containing multiple histidine residues that trigger sensitive responses towards electrical perturbation and pH variation. It is also considered a therapeutic candidate for tumours in the skin-like acidic pH environment. A previous simulation study illustrated that its regions, including sequential histidine residues, prefer to interact closely with bilayer pores. It shows structural variability when placed within a lipid bilayer membrane, and is the most helical when in close proximity to water.

Chapter 2

Methods

2.1 Molecular Dylamics Simulation

MD simulations were performed through the GROMACS/2021.4 package [10] [11] [12] [13] [14] [15] [16] [17], within the OPLS-AA all-atom force field. Initial structures of peptides were introduced from PDB files generated through SWISS-MODEL. SWISS-MODEL is a modelling program from Swiss Pdb-Viewer that automatically generates a reliable stationary model based on the sequence information of peptides [18]. Next, we created basic geometric configurations of peptides through a GROMACS built-in function $pdb2gmx$ and placed water molecules (2880 for GAD-1 and 776 for all other peptides) with *solvate* using the TIP4P/2005 water model [19]. Peptides with net charges were neutralized by replacing Na^+ or Cl^- ions at random positions. The simulation protocol started with an energy minimization using the steepest descent algorithm. Following the first stage of NPT equilibration (2 ns) in a small cubic box (5nm for GAD-1 and 4nm otherwise) at 300K, we relocated the peptide to the box center using GROMACS $trjconv$. Then the box containing water-peptide bulk was

expanded (20 nm for GAD-1 and 15 nm otherwise) in order to allow the system to form a droplet during a short NVT equilibration (2 ns). A random cluster of water molecules forms a sphere naturally in the canonical ensemble. This is because the box is big enough so that periodic boundary condition cannot connect clusters in adjacent units to form bulk water. Lastly, we copied the droplet system to seven temperatures ranging from 180 K to 300 K and performed a long NVT production run ($\sim$ 330 ns), during which peptides can make significant conformational changes.

For a comparison between the nanodroplets and the bulk water, we carried out a long NPT equilibration (2000 ns) corresponding to each of the NVT runs. The time step was set to 0.002 ps for all NVT and NPT simulations. The pressure in NPT simulations was controlled with the Parrinello-Rahman method, and the temperature in both NVT and NPT simulations was constrained with the Nosé–Hoover thermostat. Covalent bonds connected with hydrogen atoms were treated as constraints using the LINCS algorithm [20].

We visualized and monitored the simulation process using the visualization program VMD [21]. In what follows, we describe the water model used and the various quantities we use to characterize the system.

2.2 The TIP4P and TIP4P/2005 Models of Water

Molecular dynamics simulations use classical forces and Newton's Laws to propagate atomic positions and velocities in time. An atom is represented by a single point mass interacting through the Lennard-Jones potential:

$$V_{LJ} = 4\sum_{i=1}^{n-1} \epsilon_i \left[\left(\frac{\sigma}{r_i} \right)^{12} - \left(\frac{\sigma}{r_i} \right)^{6} \right] \tag{2.1}$$

where n is the number of interacting particles, r_i is the distance between the assigned atom and the i^{th} interacting particle, ϵ_i is the well depth and a measure of how strongly the assigned two particles attract each other, and σ is the distance at which the particle-particle potential energy V_{LJ} is zero. Energy minimization is a necessary process responsible for excluding the singularity at $r = 0$. Unlike what we learned from quantum molecular orbitals, the electrons are not modelled as a cloud around the nucleus. Instead, the partial charges with a rigid geometry are calculated for atoms to exhibit an average electrical polarization. Coulomb's law calculates the electrostatic potential of a point charge:

$$U_E = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{n-1} q \frac{Q_i}{r_i} \tag{2.2}$$

where n is the number of interacting particles, ϵ_0 is the permittivity of free space, r_i is the distance between the point charges q and Q_i, q is the assigned point charge, and Q_i is the i^{th} interacting charge. To generate a fixed temperature at the start of the simulation, the kinetic energy is randomly allocated to satisfy the Maxwell-Boltzmann distribution.

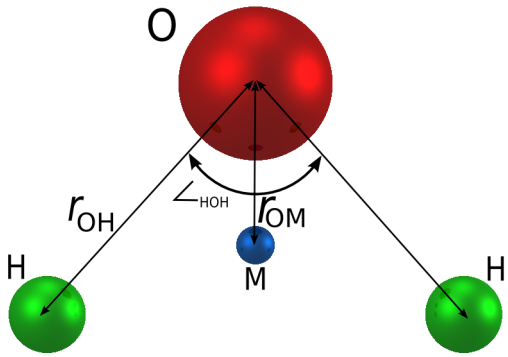


Figure 2.1: TIP4P/2005 Model. Carl McBride, via SklogWiki, 2009

The TIP4P/2005 model of water is an updated version of the TIP4P model. The density anomaly of the TIP4P model under a pressure of 1 bar occurred at 253 K, while TIP4P/2005 gives a much more accurate value of ∼270 K under 1 bar [19]. A

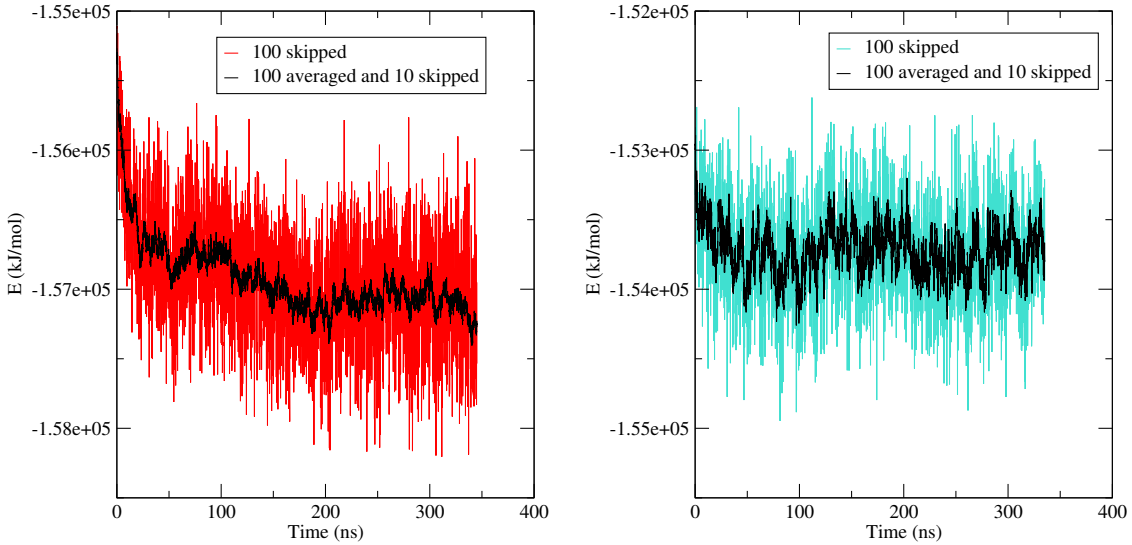
schematic of TIP4P/2005 is shown in figure 2.1. Both models employ a fourth massless site M for better tetrahedral liquid water hydrogen bonding network structures. Site M is located at the bisector of the two H atoms, coplanar with the oxygen and hydrogen atoms. The oxygen is chargeless and interacts via the Lennard-Jones force, while positive hydrogens and the negative M interact via electrostatic forces.

2.3 Energy

The energy time series illustrates the result of calculating the sum of kinetic and potential energy at every output step in the simulation,

$$E = K + PE = K + V_{LJ} + U_E \tag{2.3}$$

where K is the kinetic energy and PE is the potential energy. For clarity of figures,



(a) Unequilibrated (b) Equilibrated

Figure 2.2: Examples of potential energy time series. Plotted are the instantaneous energy every 100 ps and the running average of the energy using 100 points that we sampled every 100 ps.

instantaneous values of the energy are reported every 100 ps for the droplets and 500 ps for bulk. To determine whether the temperature is stabilized, we plotted every energy curve separately as shown in figure 2.2. Panel (a) shows an example of an unequilibrated system with a significant portion of the time series clearly trending down. Panel (b) shows an equilibrated system with the energy fluctuating about the mean. The energy curves not showing repeated fluctuations about the mean were determined as unequilibrated.

2.4 Radius of Gyration

When all particles in a system are equal in mass, the radius of gyration is the calculated root mean square distance from the center of mass. For systems with different particles, R_g is calculated from:

$$R_g = \left(\frac{\sum_{i=1}^{n} m_i {r_i}^2}{\sum_{i=1}^{n} m_i} \right)^{\frac{1}{2}} \tag{2.4}$$

where r_i is the i^{th} particle's distance from the center of mass. It is often a measurement of the compactness of mass distribution. A higher radius of gyration suggests an extended conformer of a protein.

2.5 Alpha Carbon

Amino acids share a standard structure where the so called alpha carbon, C_α, is connected to amino and carboxylic acid functional groups and the side chain residue, see figure 2.3. To facilitate the calculation of distances from the coordinates in GROMACS output files, we wrote a centering code to place the droplet's center of mass

at the origin. This removes technical difficulties associated with periodic boundary conditions. Frequencies of C_α radial distance from the center of mass in the last 2/3 frames were plotted in a histogram and then normalized.

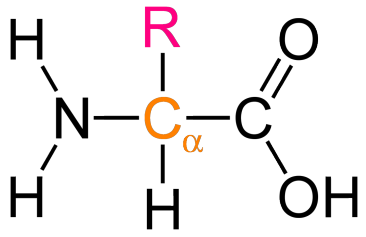


Figure 2.3: General structure of an alpha amino acid. Benjah-bmm27, Public domain, via Wikimedia Commons

2.6 Radial Distribution of Charged Sites

We are interested in the location of charges within droplets. For peptides in this research, positive charges are carried by -NH_3^+, and negative charges are brought by -COO^- either from the first/last amino acid or the R groups. The R group is a side chain connected with C_α. We selected the N atoms from -NH_3^+ for locating positive charges because of the tetrahedral structure of protonated amino groups and the relatively small radius of H atoms. The two oxygen atoms in carboxylate share the same electronegativity, so we averaged their coordinates for assigning the location of a negative charge.

2.7 Secondary Structure

We used VMD for secondary structure calculations. VMD takes a simulation trajectory to calculate distances between possible hydrogen bonding groups and determines

the type of secondary structure. The last 2/3 of output data were averaged to show the probability of helical structures (α-helix, π-helix, and 3_{10} helix) for all amino acids.

2.8 Density Profile

Our code to calculate the density profile divides the sphere into 0.05 nm thick shells and counts the number of water molecules in the shell. The density is then calculated by dividing m_{water} by V_{shell}.

Chapter 3

Results

We first simulate the CBS peptides from the droplet center at $T = 300$ K and $P = 1$ bar for a trajectory of 2 ns. CBS-5's (VTVFL) both termini and the N-terminus of CBS-9 (ANVTVFLQD) were almost immediately pushed to the surface of the droplet, corresponding to the fact that valine (V), leucine (L) and alanine (A) are hydrophobic amino acids. In order to study the phenomenon that CBS peptides are absent from the droplet center, we constructed two charge-rich peptides, KKKDDD and DKDKDK, with positive lysine (K) and negative aspartate (D) under a neutral pH. Consequently, they remained deep inside the droplet during this short 2 ns simulation. The comparison between CBS peptides and DK peptides indicates the hydrophobicity of amino acids may lead a peptide to the droplet surface. A natural peptide, GAD-1 (FIHHIIGWISHGVRAIHRAIH-NH_2), was introduced for further comparison. During the 2-ns trajectory, the hydrophobic N-terminus (F) along with other hydrophobic amino acids were pushed to the vapour phase, and the polar terminus (H) remained inside the droplet. This is consistent with our judgement on

the hydrophobicity despite GAD-1's relatively complex structure. Therefore, we conclude that the hydrophobicity of amino acids plays a leading role in a peptide's initial absence from the droplet center.

After this short 2-ns simulation, we attempted to equilibrate all systems in bulk and nanodroplets and we succeeded at most temperatures during the NPT or NVT production runs. The following results are analyzed from production runs.

3.1 Energy

Our nanodroplet systems are simulated for a shorter trajectory than the bulk system due to the faster NPT simulations compared to NVT. The water molecules in smaller (776 water molecules) droplet-peptide systems are equilibrated within ∼600 ns except at $T = 180$ K for KKKDDD system. The equilibration process of the peptide itself is relatively slow because the positive lysine residues are observed to approach the arginine. GAD-1, the longest peptide in our research, is held in a bigger nanodroplet with 2880 water molecules. The GAD-1 system is equilibrated at most temperatures despite having more molecules. However, the droplet system is not equilibrated at 180 K and the bulk is not equilibrated at 180 K and 200 K correspondingly. Detailed lengths of simulation for nanodroplet and bulk systems are listed separately in table 3.1 and 3.2.

Peptide	180 K	200 K	220 K	240 K	260 K	280 K	300 K
CBS-5	374.4	377.7	380.6	372.5	374.9	331.1	378.7
CBS-9	575.5	559.2	540.3	561.3	531.4	535.6	538.7
KKKDDD	603.0*	603.9	580.9	582.6	607.2	603.0	572.3
DKDKDK	487.8	466.3	468.3	459.4	485.7	476.9	505.7
GAD-1	345.2*	335.2	346.0	343.9	333.7	352.2	357.2

Table 3.1: Length of simulation for nanodroplet-peptide systems (ns). The asterisked state points are not equilibrated.

Peptide	180 K	200 K	220 K	240 K	260 K	280 K	300 K
CBS-5	2000	2000	2000	2000	2000	2000	2000
CBS-9	2000	2000	2000	2000	2000	2000	2000
KKKDDD	2000	2000	2000	2000	2000	2000	2000
DKDKDK	2000	2000	2000	2000	2000	2000	2000
GAD-1	2000*	2000*	2000	2000	2000	2000	2000

Table 3.2: Length of simulation for bulk-peptide systems (ns). The asterisked state points are not equilibrated.

Figure 3.1 and 3.2 illustrate that the average energy per water molecule decreases nearly linearly when cooling nanodroplet systems from 300 K to 200 K or from 300 K to 260 K for bulk systems, before the rate of descending reduces. The heat capacity can be estimated from the slope of energy curves by:

$$C_v = \frac{\delta E}{\delta T} = slope \tag{3.1}$$

where δE and δT are the difference in energy and in temperature. As shown in the figures, δE barely depend on the peptide, such that the heat capacity is able to characterize water molecules. Molecules being in a liquid state can explore anharmonicity in their vibrations and many different bonding configurations, leading to a larger heat capacity compared to a solid. The inflection in $E(T)$ near 240 K for the bulk systems indicates the presence of a maximum heat capacity, one of the anomalies of water. However, the inflection temperature of nanodroplet systems is much lower than in bulk systems, since the nanodroplets interior is at a higher pressure (due to the Laplace pressure). Furthermore, the energy of nanodroplet-peptide assemblies are higher than for bulk. This is because water molecules at the surface cannot form open, low energy structures as they seek to maximize hydrogen bonding.

Energy ranking for systems with different peptides is as follows: CBS-5 > GAD-1 > CBS-9 > KKKDDD = DKDKDK. In the 776-molecule nanodroplet, CBS-5 only has

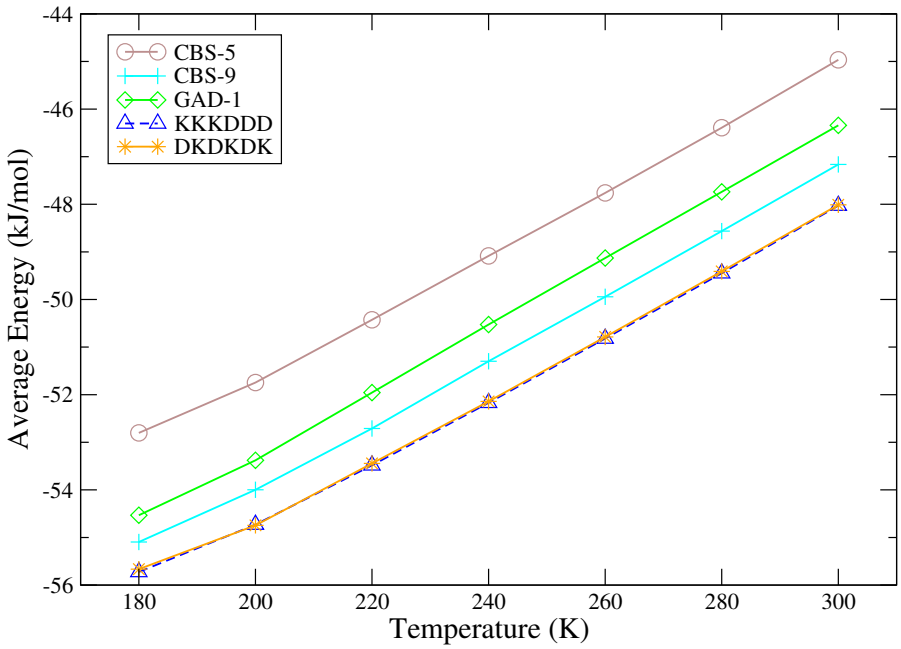


Figure 3.1: Average potential energy per water molecule for nanodroplet-peptide systems.

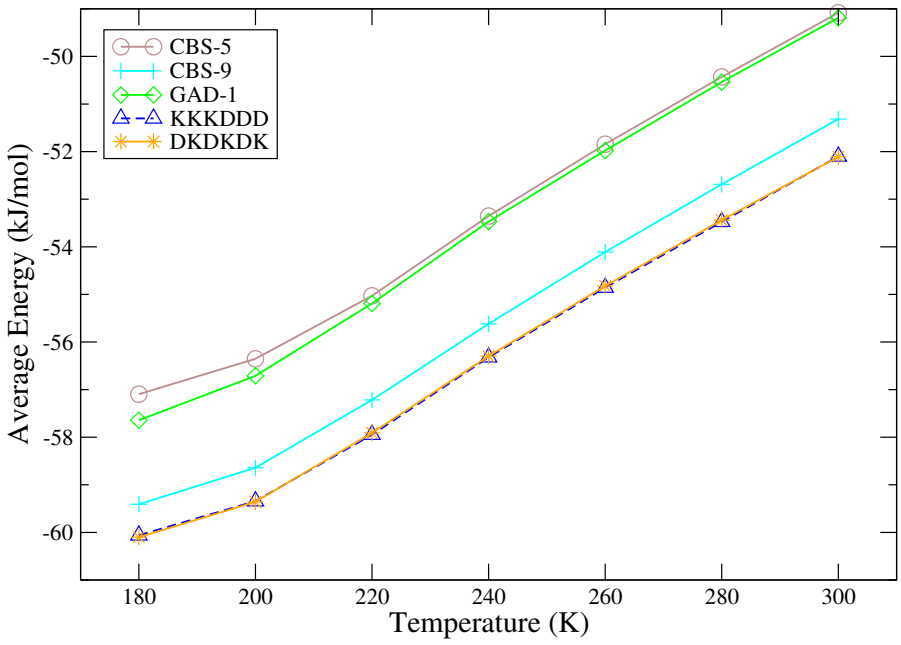


Figure 3.2: Average potential energy per water molecule for bulk-peptide systems.

3 hydrophilic residues; CBS-9 has 6 hydrophilic residues along with the chloride ion; and both DK peptides have 8 hydrophilic residues. In the 2880-molecule nanodroplet, GAD-1 carries 10 hydrophilic residues along with 3 chloride ions. The number of hydrophilic sites from peptides and ions divided by the number of water molecules determines their energy rankings to a great extent. This ratio for CBS-5 is 0.0039; for CBS-9 is 0.0090; and is 0.0045 for GAD-1. However, the bulk-peptide assembly of GAD-1 displays relatively higher energy than the corresponding nanodroplet-peptide assembly compared to other peptides. For example, the energy of CBS-5, the extremely short and neutral peptide that has the minimal secondary structure, is about the same in the bulk system as GAD-1 and clearly lower in the nanodroplet. Therefore, GAD-1 may be more stabilized in nanodroplets after building secondary structures through intrapeptide bonding.

3.2 Mean Square Displacement

Mean square displacement (MSD) is a measure of diffusive random motion. The linearity between MSD and time t of equation 3.2 allows for graphical methods to determine the diffusion coefficient D. Equilibrated water molecules in bulk have a temperature-dependent diffusion coefficient, obtained from:

$$\left\langle (x(t) - x_0)^2 \right\rangle = 2Dt, \tag{3.2}$$

where x_0 is the reference position, $x(t)$ is the position at time t, and the average is taken over all water molecules. Previously, we showed the energy time series and determined that water molecules in the CBS-5 bulk system at 180 K are equilibrated while the GAD-1 droplet system at 180 K and 200 K is not, within the limited

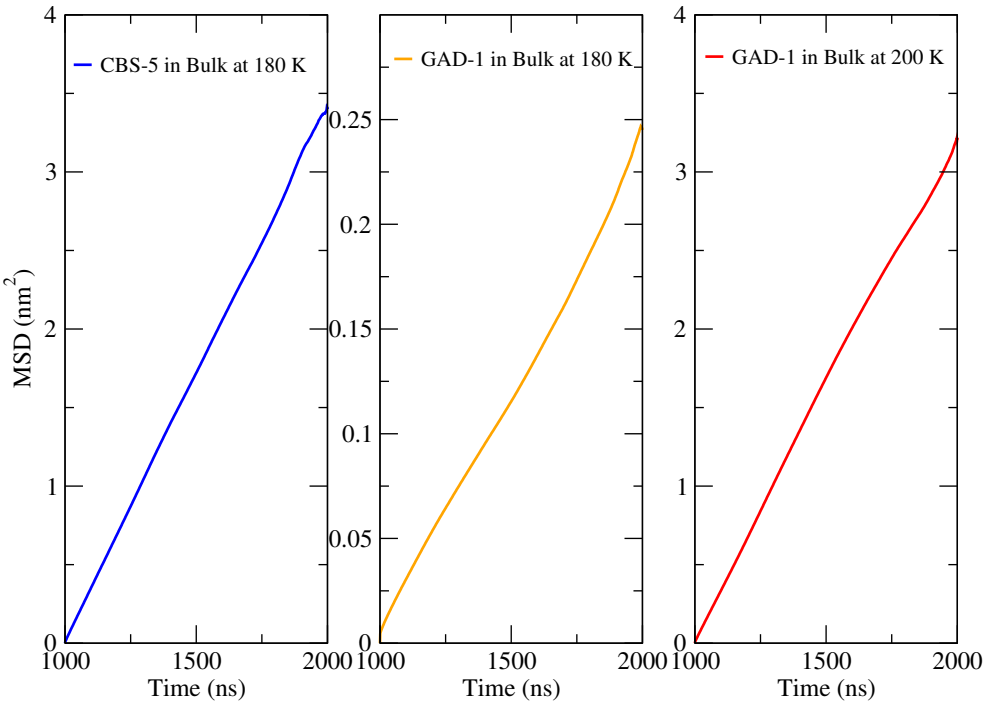


Figure 3.3: Mean square displacement with a time origin at 1000 ns as a function of time for CBS-5 in bulk at 180 K (left), GAD-1 in bulk at 180 K (middle), and GAD-1 in bulk at 200 K (right).

length of simulation (see table 3.2). Therefore, we inspect how far they are from the equilibrium by comparing the MSD of water in the other two systems with CBS-5. Figure 3.3 shows MSD plots of water molecules in the CBS-5 bulk system at 180 K and GAD-1 bulk systems at 180 K and 200 K. The plot of CBS-5 illustrates a linear relation between MSD and time, indicating a clearly equilibrated system. Both GAD-1 systems are fairly linear, but the 180 K displays very slow diffusion. The 200 K shows a slope ten times greater in the last 250 ns, but the simulation is not long enough to complete a significant number of fluctuations.

3.3 Density Profile

Here we report on the density profiles, i.e. the local density as a function of radial distance from the center of mass of the water molecules in the nanodroplet. From the density profiles in figure 3.4, we see a spherical nanodroplet composed of 776 (or 2880) water molecules has a radius of ∼2 (or 3) nm before the density decreases to zero. Moreover, a shell-like structure of nanodroplets in both sizes is revealed by the density profile, in which the density is higher than 1 g/cm^3 in the subsurface region from 180 K to 240 K. In contrast, the interior region shares the typical density of bulk water. Compared with pure water density plots in figures 3.5 and 3.6, we see that all peptides hardly affect the occurrence of density anomaly in the subsurface region below 260 K. However, the heterogeneous environment of nanodroplets may influence a peptide's equilibrated conformation and location.

An extremely low water density in the nanodroplet core shown for both DK-peptide systems results from both peptides residing near the droplet core. Technically, peptides are not counted as water molecules, and the core region has a relatively small shell volume, $dV = 4\pi r^2 dr$, when r is small. The difference in density of DK systems compared to pure water nanodroplets, $\Delta\rho = \frac{\Delta m}{4\pi r^2 dr}$, is greater than for systems for which the peptide resides on the surface, where Δm is the mass of water that is displaced by the peptide in the shell volume. The greater slope of curves at small radii at higher temperatures indicates that the DK peptides are deep inside the droplet but then move toward the surface on cooling. The lack of smoothness in curves of low temperatures (180 K and 200 K) is also a result of poor sampling due to slow water diffusion.

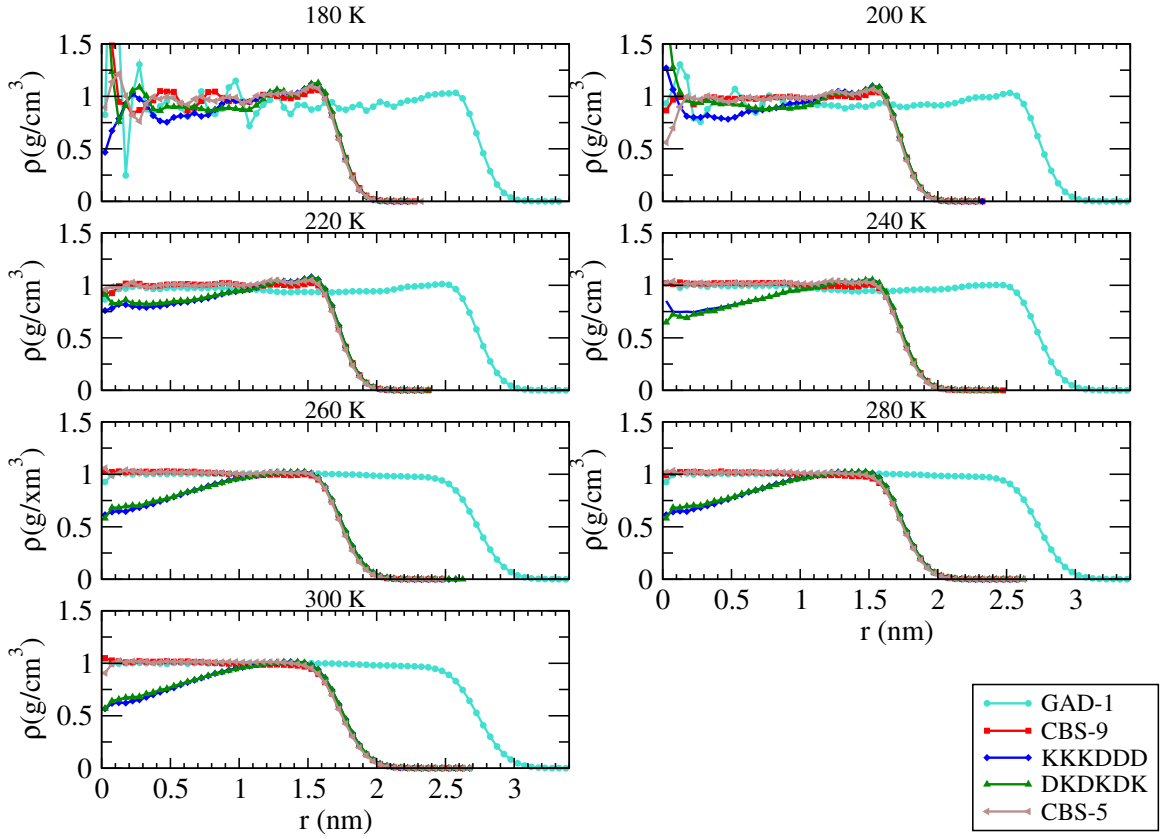


Figure 3.4: Water density profiles for GAD-1, CBS-9, KKKDDD, DKDKDK and CBS-5 nanodroplet systems. Each panel illustrates density profile for all systems at the same temperature.

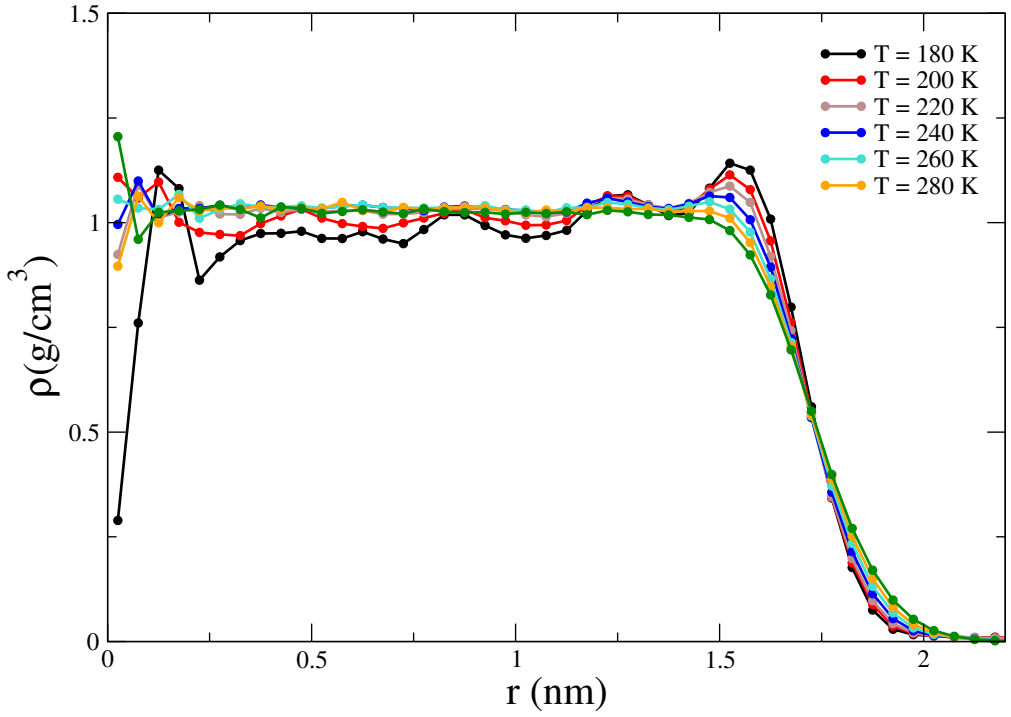


Figure 3.5: Density profile of a 776-molecule nanodroplet.

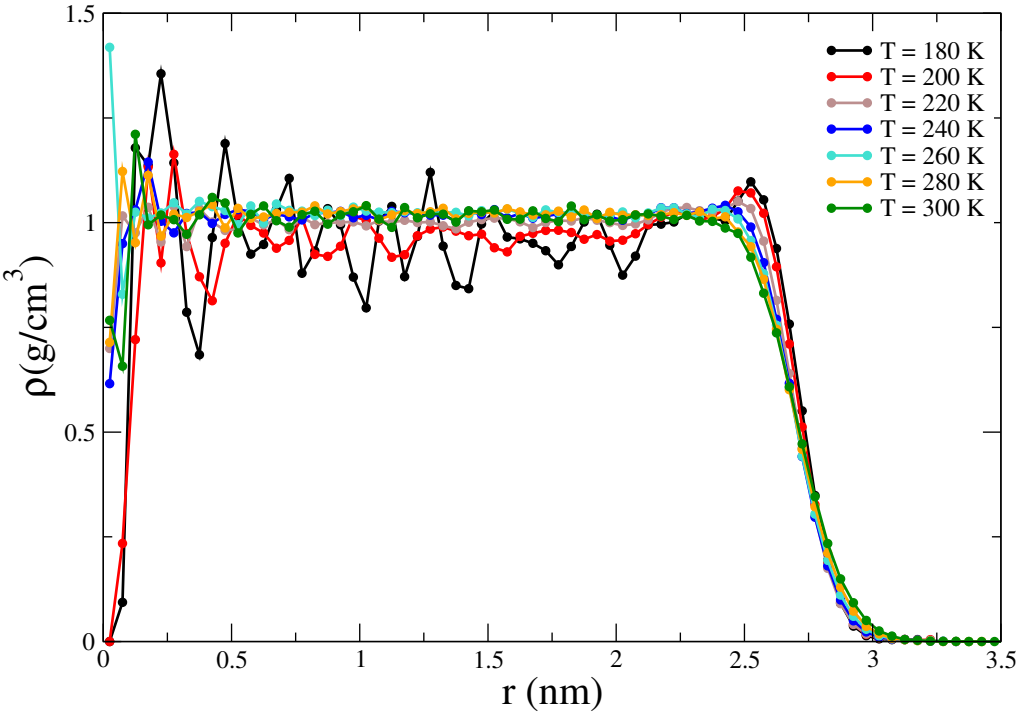


Figure 3.6: Density profile of a 2880-molecule nanodroplet.

3.4 Alpha Carbon

The alpha carbon (C_α) probability density is a measure of the radial distribution of each amino acid. For CBS-5, CBS-9 and GAD-1, a series of similar distributions observed at temperatures ranging from 180 to 240 K is followed by an abrupt transition at 260 K. Due to this transition, the widest probability distribution appears in peptide systems at 260 K. However, the width of thermodynamic distributions should typically grow since molecules gain mobility as the temperature increases.

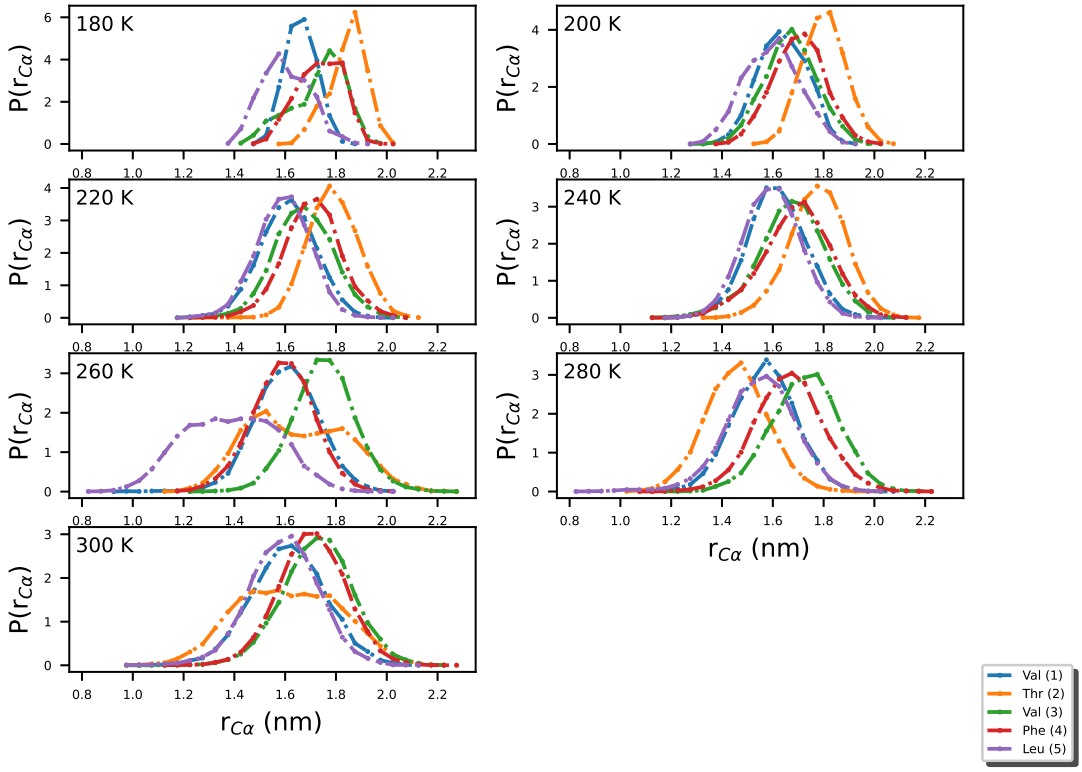


Figure 3.7: Radial distribution function of C_α for CBS5 in droplets.

CBS-5 is a highly hydrophobic peptide residing on the droplet surface. The only hydrophilic residue, Thr 2, is at an outermost radius (see figure 3.7) below 260 K. While for 260 K, it attempts to transit toward the core. Threonine finally leaves the

surface ($r = 1.8$ nm) and gets merged into the subsurface ($r = 1.5$ nm) at 280 K, where the subsurface region spans $r = 1.4 - 1.6$ nm.

CBS-9 is an amphipathic peptide with a negatively charged aspartate residue at the C-terminus. The sodium ion sometimes stays closely to the C-terminus on account of the double charges. C_αs are tightly grouped across the subsurface region at low temperatures, see figure 3.8. On heating to 260 K, Asp 9 and Ala 1 attempt to move toward the surface and Thr 4 goes to the opposite direction, which leads CBS-9 to alternate between a compact and an extended conformer. The overlapped distribution of the two conformers is responsible for the abnormally large width at 260 K.

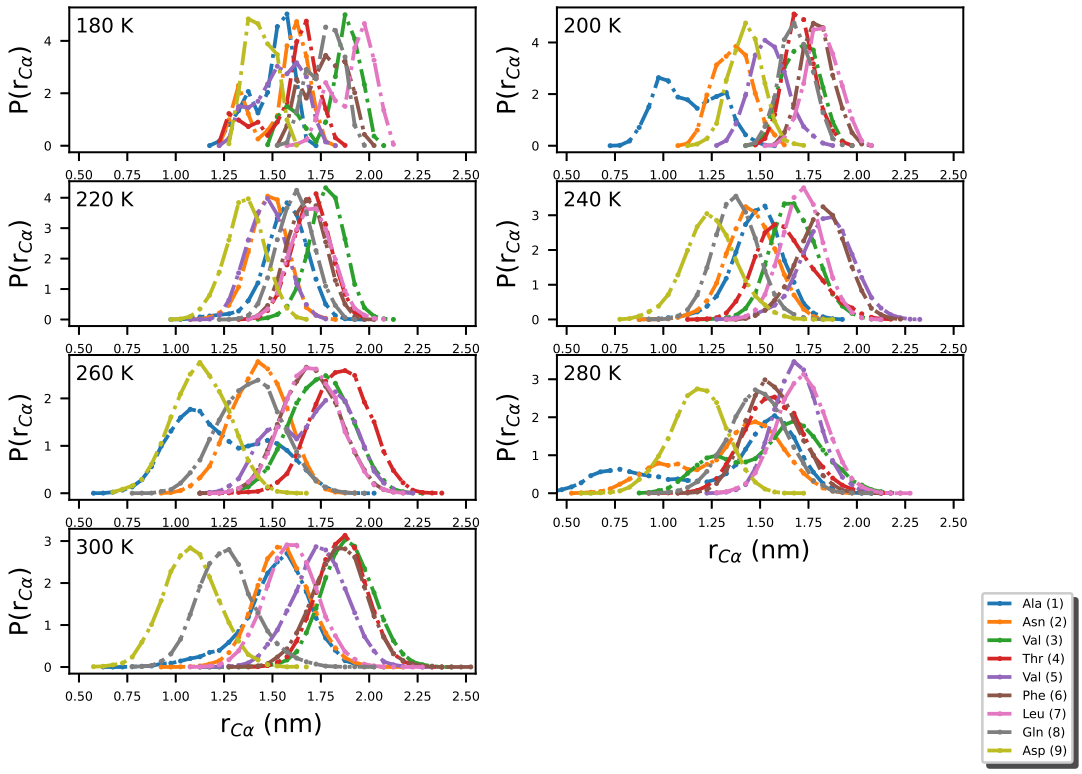


Figure 3.8: Radial distribution function of C_α for CBS9 in droplets.

KKKDDD and DKDKDK (see figure 3.9 and figure 3.10) are composed of charge-rich and hydrophilic lysine and aspartate, which would tend to keep them deep inside the nanodroplet. On cooling from 300 to 240 K, they moved outward slightly without reaching the subsurface region. We notice that the doubly charged N-terminal lysine from KKKDDD (at high temperature) and the neutral N-terminal aspartate from DKDKDK are located at a greater radius than any other residues. The N-terminus shows a strong tendency to leave the droplet center.

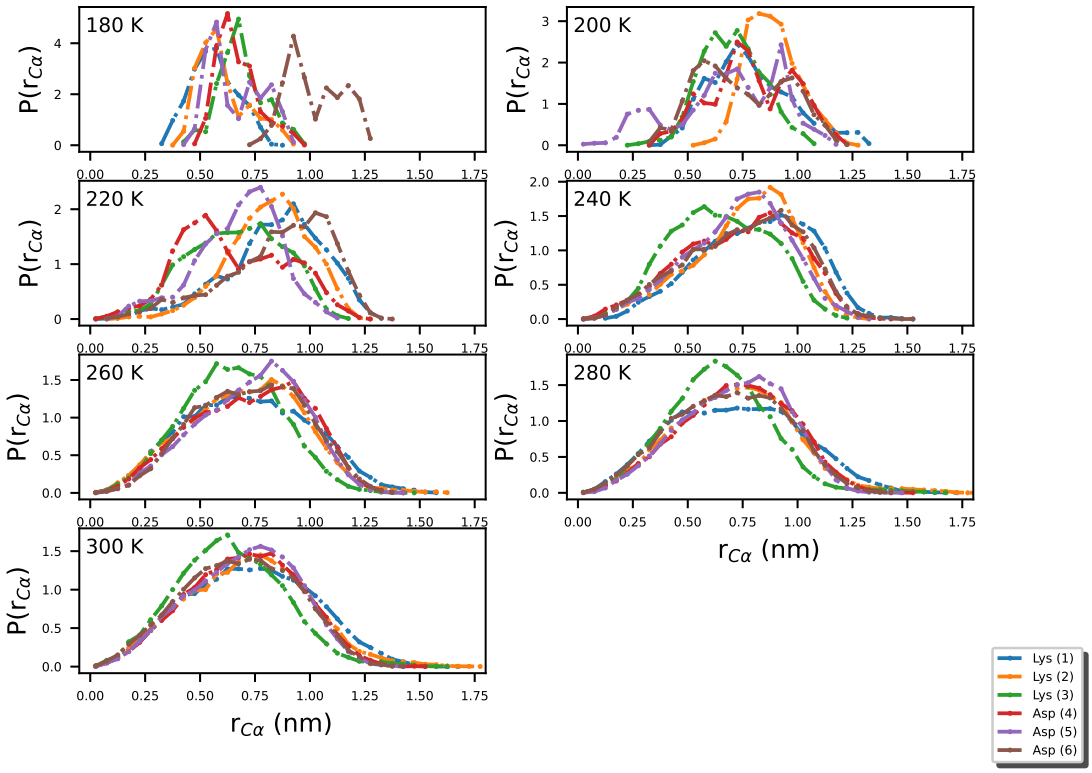


Figure 3.9: Radial distribution function of C_α for KKKDDD in droplets.

GAD-1 is an amphipathic peptide consisting of eight hydrophilic and thirteen hydrophobic amino acids. For reference, GAD-1's 2880 nanodroplet subsurface region is around $r = 2.2 - 2.7$ nm. At high temperatures, the peptide (see figure 3.11)

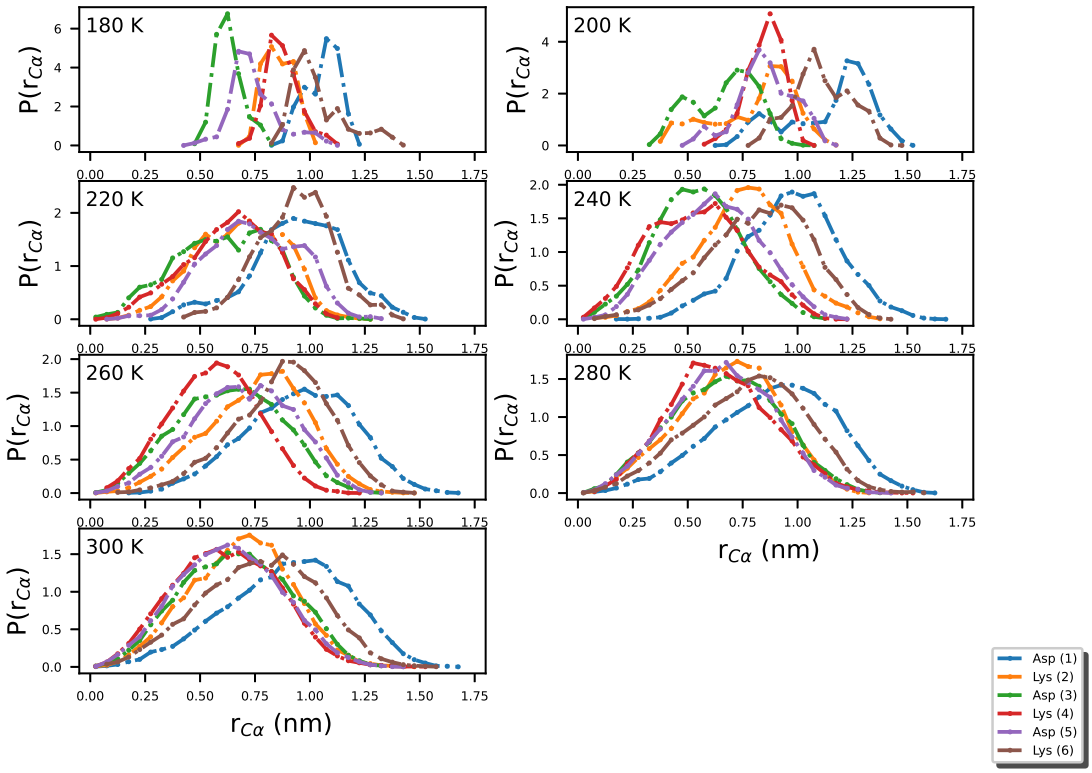


Figure 3.10: Radial distribution function of C_α for DKDKDK in droplets.

resides across the subsurface region around 2.0-3.0 nm from the droplet center before the distributions dramatically shift to the center around $r = 1.0 - 3.0$ nm at 240 K, with some hydrophobic amino acids moving to the innermost radius, around 1.0 nm from the center. Specifically, the Ala15, Ala19 and Ile16 near the C-terminus move into the interior region, while the Ile2, Ile5 and Ile6 near the N-terminus moved toward the surface.

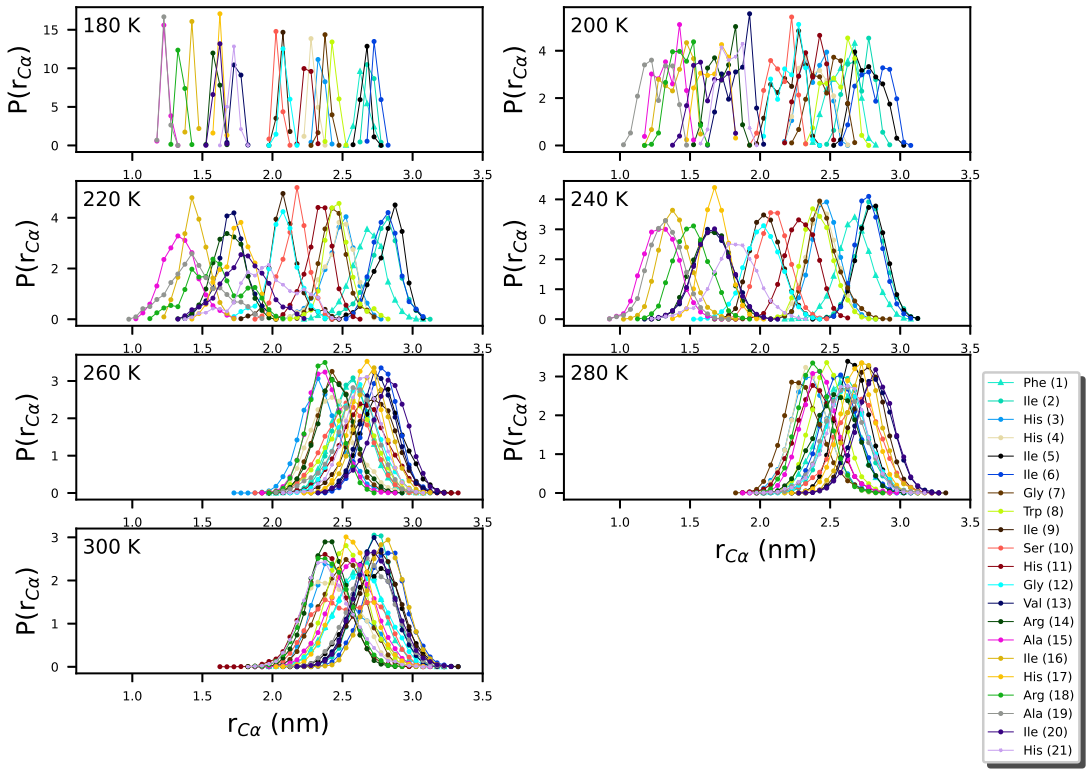


Figure 3.11: Radial distribution function of C_α for GAD-1 in droplets.

3.5 Charged Sites

The charged sites we study include the N-terminus, the C-terminus, residues of charged amino acids and the sodium or chloride ions, all of which carry one unit of positive or negative charge (-e or +e).

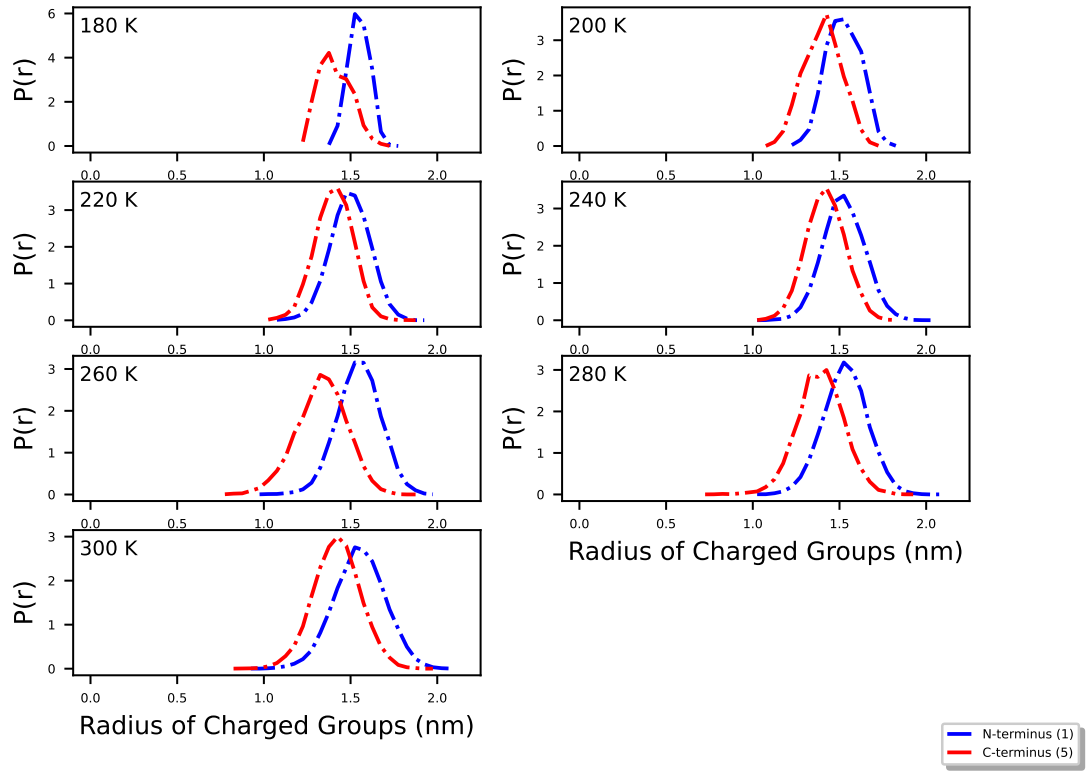


Figure 3.12: Charged groups distribution function of CBS-5 simulated in droplets.

CBS-5 (figure 3.12) displays an orientation in which the N-terminus points outward and the C-terminus points inward. CBS-9 (figure 3.13) and GAD-1 (figure 3.14) share the same orientation. KKKDDD and DKDKDK in figure 3.15 and figure 3.16 did not display such an explicit orientation, but all charged sites share about the same distribution. The probability of charged lysine and aspartate residues occurring at a small radius increases at higher temperatures, corresponding to the previously

studied sodium and fluoride ions [3]. For example, the probability density for residues to appear at 0.5 nm grows with temperature. Approximately, $P(r = 0.5)$ is 0.5 at 220 K, 0.6 at 260 K and 0.75 at 300 K. This is consistent with the peptide being more centrally located at higher temperatures. A deviation between the positive (K) and negative (D) sites of DKDKDK is observed at low temperatures, revealing a negative outer shell and a positive interior region of the nanodroplet itself which arises from a preferred orientation of water molecules.

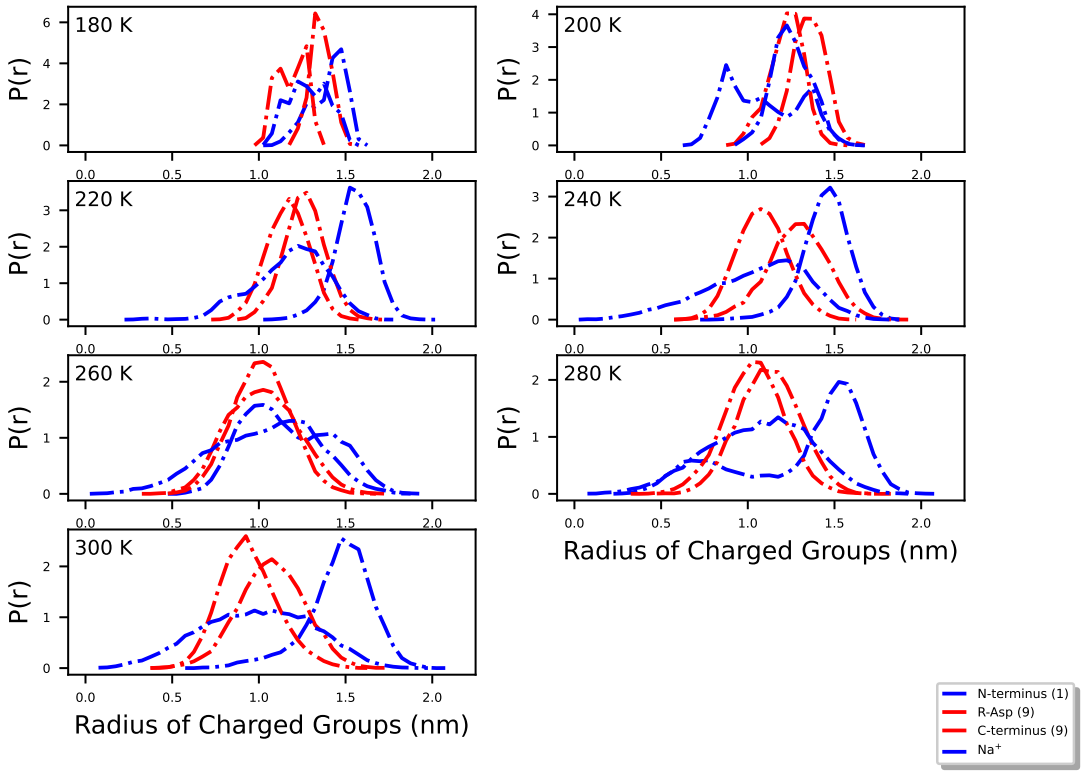


Figure 3.13: Charged groups distribution function of CBS-9 simulated in droplets.

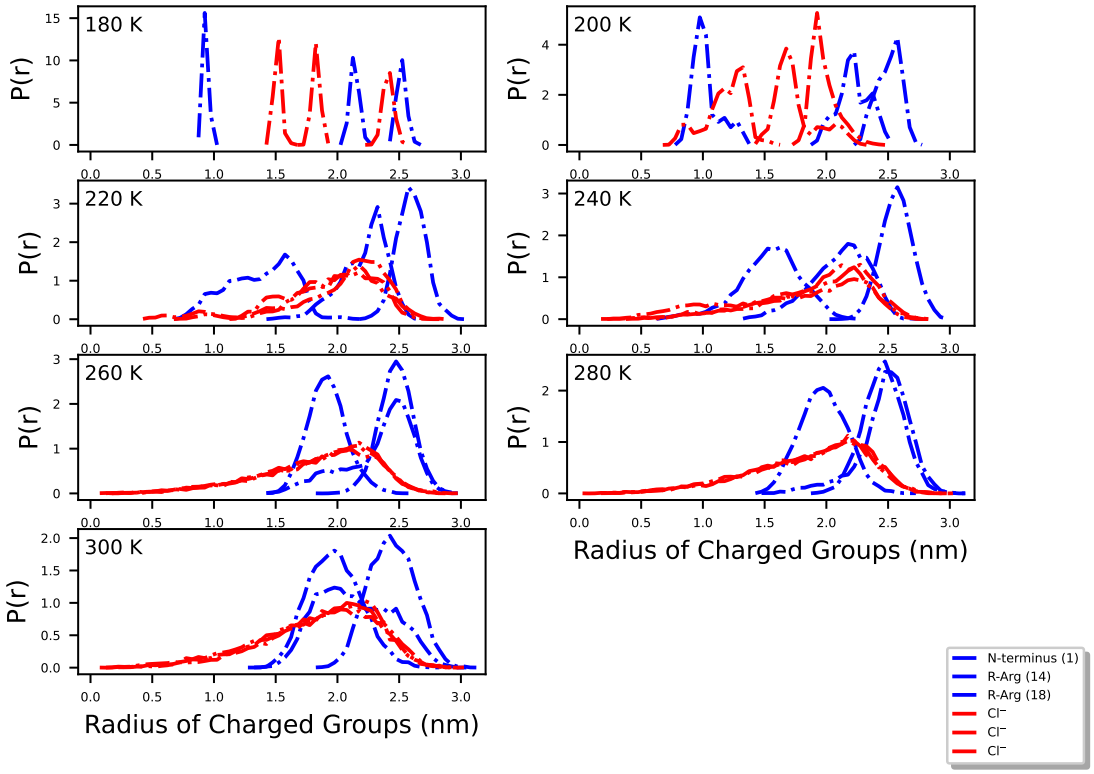


Figure 3.14: Charged groups distribution function of GAD-1 simulated in droplets.

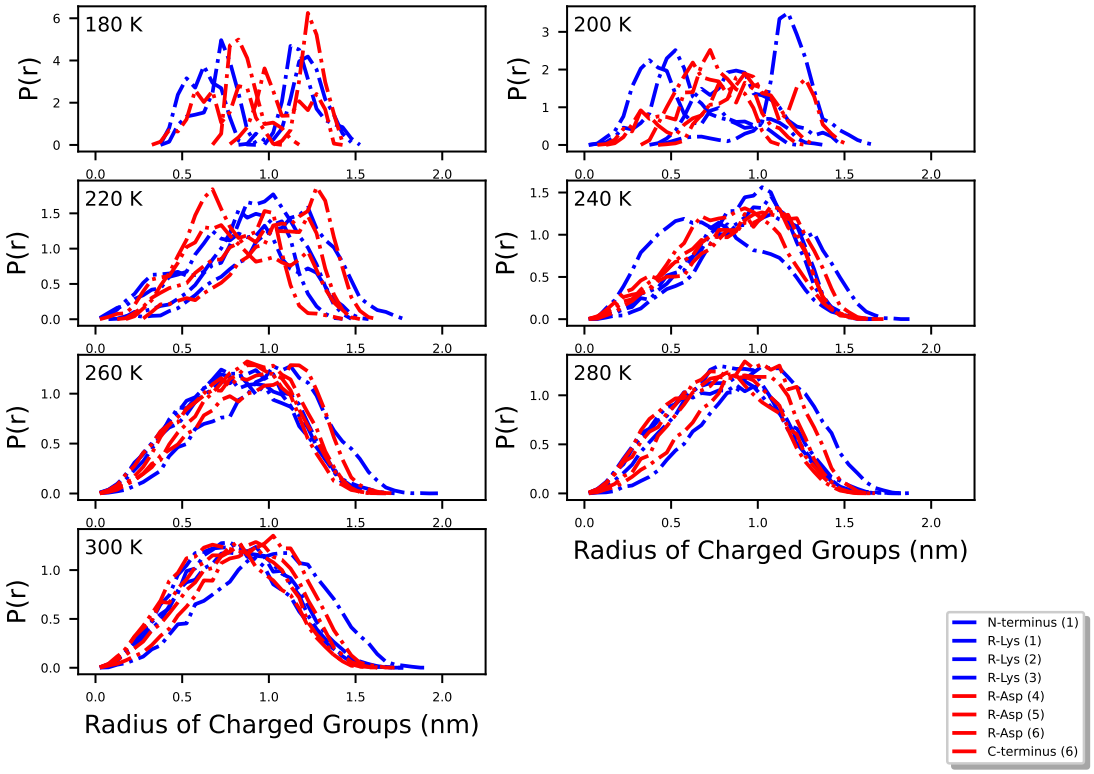


Figure 3.15: Charged groups distribution function of KKKDDD simulated in droplets.

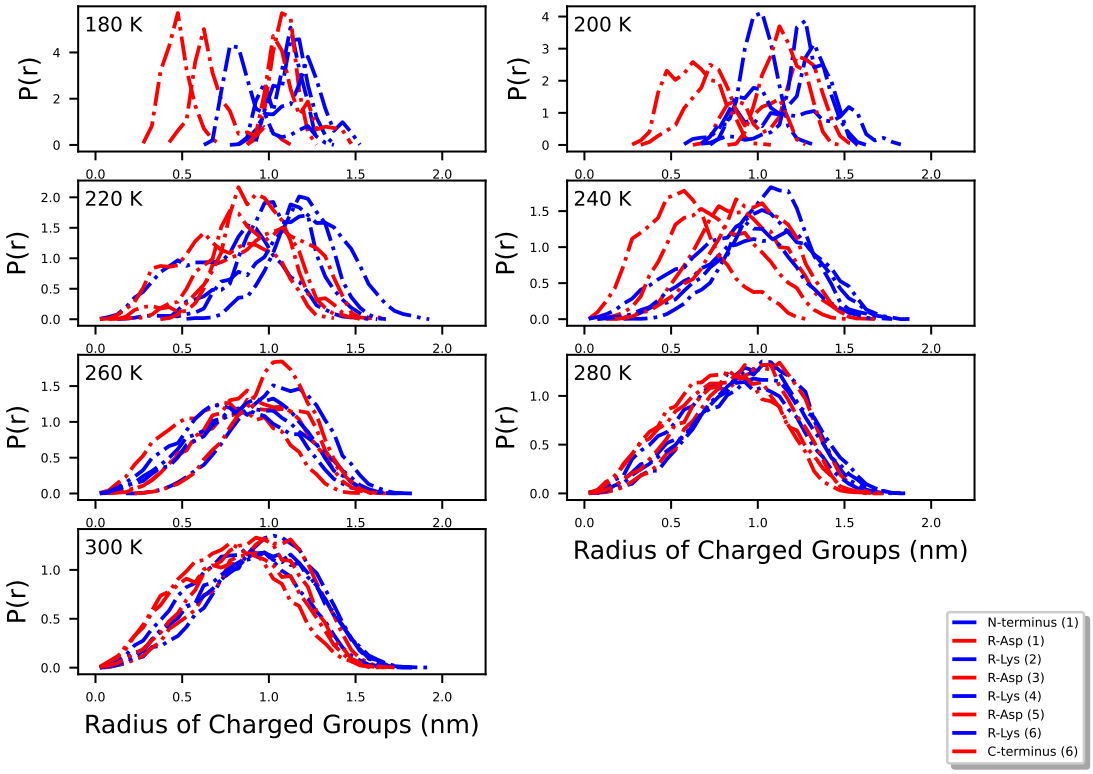


Figure 3.16: Charged groups distribution function of DKDKDK simulated in droplets.

3.6 Radius of Gyration

The radius of gyration (R_g) characterizes the compactness of peptides, which should have an increasing trend on heating because the thermodynamic force may dominate electrical forces which maintain the folded state. For CBS-5, DKDKDK and GAD-1, R_g in bulk differs from the nanodroplet counterpart at lower temperatures. At T higher than the transition temperature of 240 K, the R_g of peptides in nanodroplets tends towards the bulk value.

The R_g of CBS peptides in bulk water shares common features upon increasing the temperature. Both CBS-5 and CBS-9 remain in a compact shape for $T \leq 200$ K, while for $T > 200$ K, they start to unfold, see figures A.1, A.5 in the appendix for R_g time series and 3.17 for the average R_g. CBS-5 reaches and stays at the maximal radius of gyration after 220 K, and CBS-9 continues to increase until reaching its maximum at 240 K, after which a slight fluctuation was observed. They would both transit from an unfolded state (high R_g) to a folded state (low R_g) on cooling from above 240 K, with CBS-5 folding at a lower temperature (200 K) than CBS-9 (220 K). In the nanodroplet, the folding transition is higher (280 K) than in bulk. This is an indication that the two ionized termini come together more efficiently at the surface.

In bulk water, the R_g of KKKDDD is approximately constant at all temperatures. For KKKDDD in the nanodroplet, the high-temperature ($T \geq 240$ K) state has a lower R_g when the peptide is in the centre of the nanodroplet. R_g increases as temperature decreases, and the peptide moves toward the surface. It is worth exploring more why R_g for bulk and droplet are different even when the peptide is inside the droplet where it might be bulk-like. For DKDKDK at low temperatures ranging from 180 K to 260 K, when the peptide is more at the droplet's surface, R_g is systematically higher in droplet than bulk. In contrast to the hydrophobic CBS peptides, DKDKDK has a

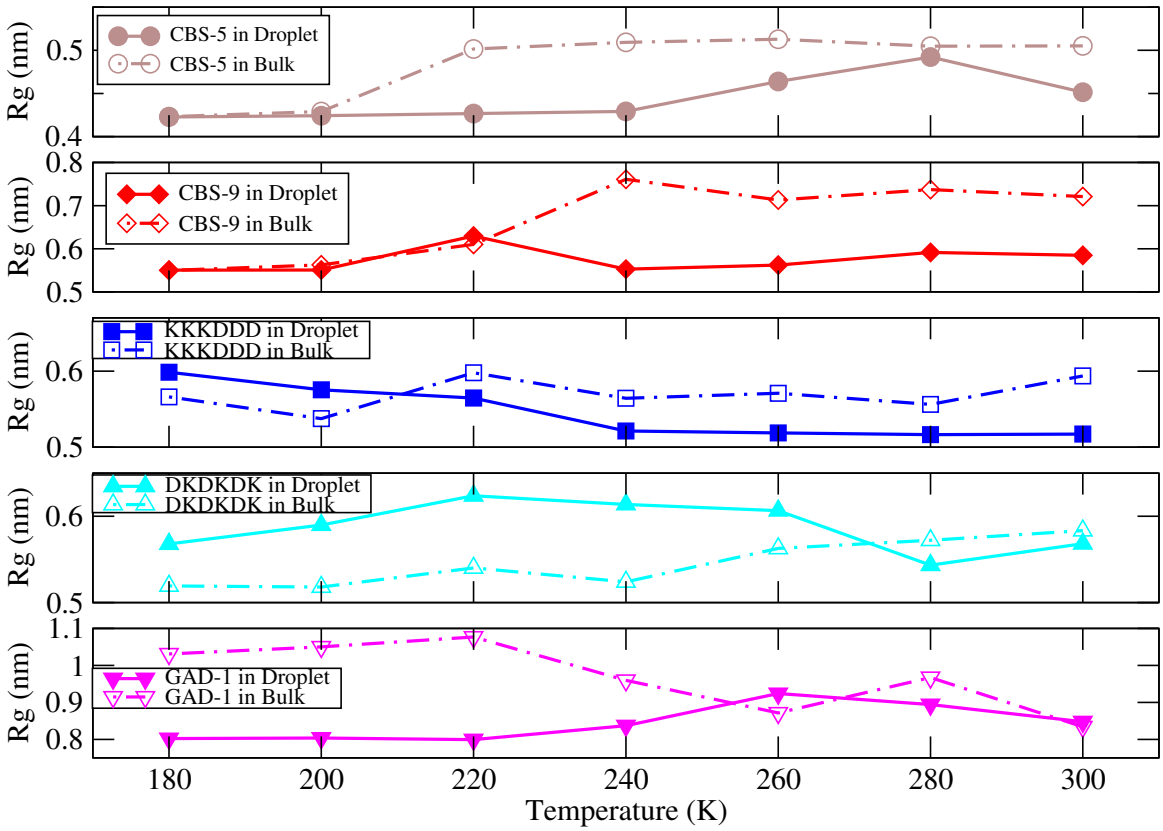
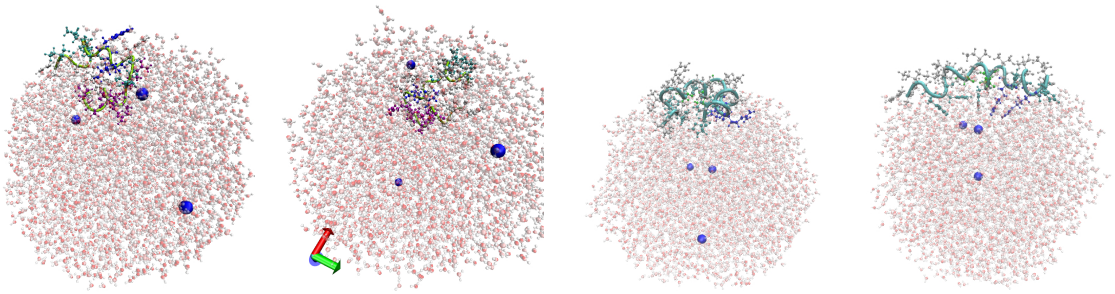


Figure 3.17: Average radius of gyration of peptides simulated in nanodroplets as a function of temperature.



(a) Low T1=180K (b) Low T2=240K (c) High T1=300K (d) High T1=300K

Figure 3.18: Snapshots of GAD-1 in nanodroplet

larger R_g when closer to the surface. The two DK peptides need further study.

At higher temperatures, the R_g of GAD-1 is roughly the same for bulk and droplet. When GAD-1 starts probing the droplet interior below 260K (see figure 3.18(b)), R_g decreases in the droplet and increases in the bulk. We observed a significant phenomenon when watching the visualized trajectory of the 300 K GAD-1 system. As a snapshot from the simulation, Figure 3.18(c) shows a curly GAD-1 residing on the droplet surface in the first half trajectory. This compact conformation unexpectedly unfolds into a linearly helical conformer with hydrophilic histidine and arginine residues stretching straight toward the approaching ions, see figure 3.18(d). The R_g time series in figure A.19 confirms the story by showing a stably low R_g before an instantaneous opening-up appears at ∼190 ns, which strongly suggests that ions play a vital role in the unfolding process of GAD-1. The R_g time series of GAD-1 sinks to the bottom when ions move away from the peptide. We think that ions again regulate the folding process.

3.7 Secondary Structure

The nanodroplet environment promotes α-helical structure of the antimicrobial peptide GAD-1. GAD-1 is quite helical in the nanodroplet, see figure 3.19, except for 260 K when a chloride ion is close to the peptide. In bulk, it is not helical; when $T \geq 260$ K, there is roughly zero helicity in every residue in figure 3.20. Besides, the 180 and 200 K systems are not counted into the comparison because they are not equilibrated, and likely stuck at the initial configuration. The helicity (G, H and I, see table 3.3 for reference), in the droplet decreases on the N-terminal side on cooling but partially reemerges (see figure 3.20) as ions get away from the surface and parts of GAD-1 probe the interior of the droplet, shielding the N-terminus from chloride ions, see figure 3.18(a).

Histidine in both bulk and droplet systems breaks helicity. The amino acids which lead the peptide to move outside the droplet at an increasing temperature, Ala15, Ala19 and Ile16, maintain their helicity in the droplet but lose it in the bulk water. E and B, the only two secondary structures concerning the β-sheet, never emerge in the droplet. In bulk water, the isoleucine near both N and C terminus and the middle residues (tryptophan, isoleucine, serine, glycine, and valine) show a tendency for the development of β-sheet structures at high temperature, see figure 3.20.

Structures	Description
T	Hydrogen bonded turn (3, 4 or 5 turn).
E	Extended strand in parallel and/or anti-parallel β-sheet conformation.
B	Residue in isolated β-bridge.
H	4-turn helix (α helix). Minimum length 4 residues
G	3-turn helix (3_{10} helix). Minimum length 3 residues.
I	5-turn helix (π helix). Minimum length 5 residues.
C	Coil (residues which are not in any of the above conformations).

Table 3.3: Description of protein secondary structure

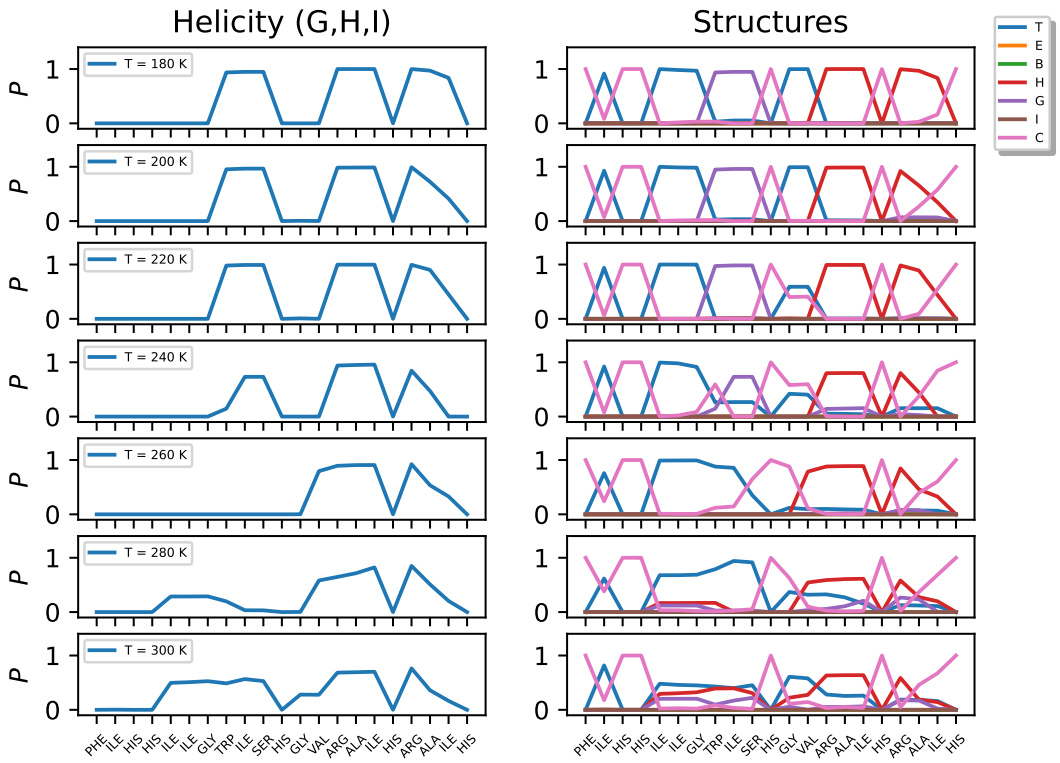


Figure 3.19: Per residue secondary structure of GAD-1 in the droplet

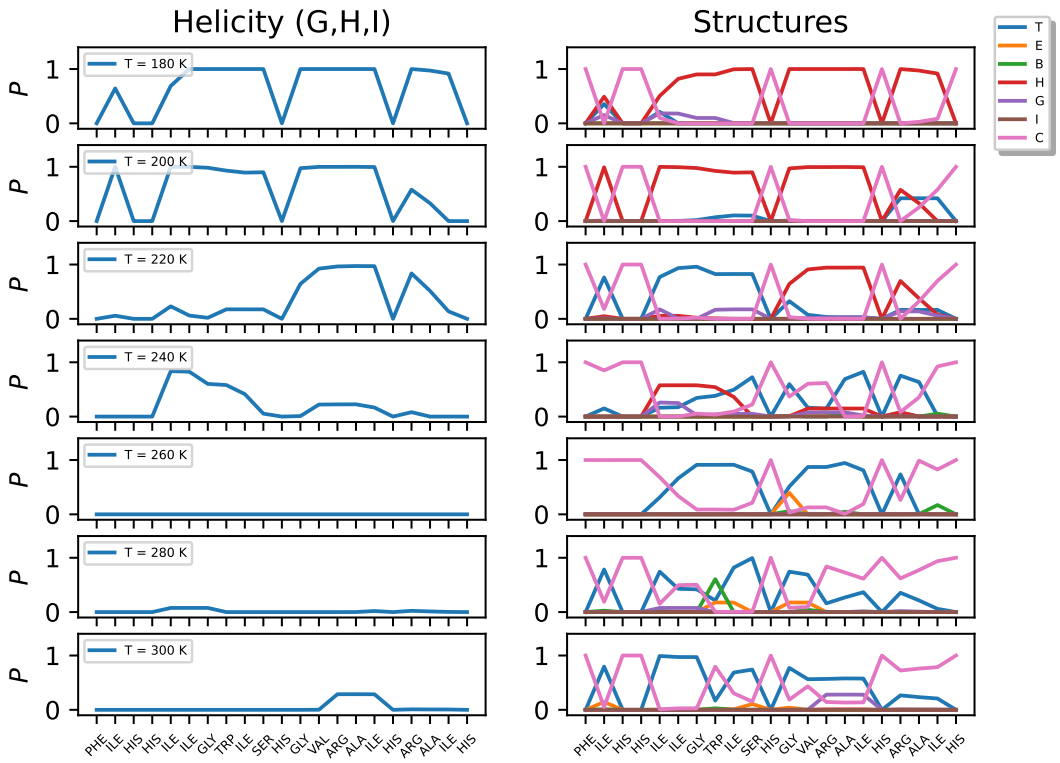


Figure 3.20: Per residue secondary structure of GAD-1 in the bulk

Chapter 4

Discussion, Conclusions and Future Work

In contrast to bulk water, the simulated peptides displayed different conformations and behaviours when interacting with the nanodroplet. Previous research suggested that the nanodroplet is under immense surface tension and forms a loose hydrogen bonding network. In this research, we discovered hydrophobic and amphipathic peptides would be expelled to the surface during the early stage of the droplet-forming process in a canonical ensemble. During a following lengthy equilibration, peptides adjust conformations and locations with the assistance of a unique nanodroplet environment and counter ions, surprisingly. CBS-5 and CBS-9 more easily adopted a folded state on the droplet surface; KKKDDD and DKDKDK were tucked inside the droplet at high temperatures; GAD-1 was folded near the surface at low temperatures and adopted a more helical structure at high temperatures. In general, peptides are more compact under the confinement of a nanodroplet. The counter ions we initially added to neutralize the system may actually boost an extended conformation. The

role of ions in this regard warrants further study. Another surprising discovery in our research is a strange transition of peptide structures and locations between 240 K and 260 K observed from C_α distributions. Some hydrophilic residues (Thr from CBS-5) could migrate towards the vapour phase, and some hydrophobic residues (Ala and Ile from GAD-1) might dwell in the droplet interior on cooling from 260 K, below which the density anomaly intrinsic to the water nanodroplet emerges. The orientation of peptides was highly uniform, with the N-terminus residing at a larger radius than the C-terminus. This is probably the consequence of an electrostatic force from the negative outer shell (and a positive interior). Analyzing the secondary structure of GAD-1, we see that the nanodroplet environment promotes helicity. The reasons behind the outcomes need further characterization.

The ideally amphipathic peptide LK_{22} (Dns-KLLKLLLKLLKLLLKLLLKLLK) would be a great comparison to GAD-1 because of their similar length and function. GAD-1 is an antimicrobial peptide composed of 21 amino acids, while LK_{22} is an artificial cytotoxic peptide 5-10 times more hemolytic than melittin [22]. It consists of repetitive three or four lysine and leucine residues and exhibits a perfect α-helical structure at the air/water interfaces [23]. If the environment promotes helical structures for a less helical peptide GAD-1, what will it do to LK_{22} with a perfect α-helical structure? We will know the answer after simulating LK_{22} in nanodroplet.

In the nanodroplets, the N-terminus was the preferred side going out in CBS-9 and GAD-1 systems and the outermost charged residue in CBS-5, CBS-9 and GAD-1. Meanwhile, the C-terminus might be the particular residue attracted by the center. However, the mechanics of dragging the N-terminus toward the surface is left unexplored. To verify whether the N-terminus indeed leads peptides outward, we must find direct evidence to exclude the influence from side chains and electrostatic forces.

Bibliography

[1] S. Koga, D.S. Williams, A.W. Perriman, and S. Mann. Peptide–nucleotide microdroplets as a step towards a membrane-free protocell model. *Nat. Chem.*, 3(9):720–724, 2011.

[2] S. M. A. Malek, P. H. Poole, and I. Saika-Voivod. Thermodynamic and structural anomalies of water nanodroplets. *Nat. Commun.*, 9(1):2041–1723, 2019.

[3] S. M. A. Malek, V. Kwan, I. Saika-Voivod, and S. Consta. Low density interior in supercooled aqueous nanodroplets expels ions to the subsurface. *J. Am. Chem. Soc*, 143(33):13113–13123, 2021.

[4] Tarick J. El-Baba, Daniel R. Fuller, Daniel W. Woodall, Shannon A. Raab, Christopher R. Conant, Jonathan M. Dilger, Yoni Toker, Evan R. Williams, David H. Russell, and David E. Clemmer. Melting proteins confined in nanodroplets with 10.6 μm light provides clues about early steps of denaturation. *Chemical Communications*, 54(26):3270 – 3273, 2018.

[5] Doyong Kim, Nicole Wagner, Kerry Wooding, David E. Clemmer, and David H. Russell. Ions from solution to the gas phase: A molecular dynamics simulation of the structural evolution of substance p during desolvation of charged nanodroplets generated by electrospray ionization. *Journal of the American Chemical Society*, 139(8):2981 – 2988, 2017.

[6] J. Zhong, Y. Zhao, L. Li, H. Li, J.S. Francisco, and X.C. Zeng. Interaction of the nh2 radical with the surface of a water droplet. *Journal of the American Chemical Society*, 137(37):12070–12078, 2015.

[7] A.M. Almudallal, I. Saika-Voivod, and J.M. Stewart. Folding and binding energy of a calmodulin-binding cell antiproliferative peptide. *Journal of Molecular Graphics and Modelling*, 61:281–289, 2015.

[8] M.H. Khatami, M. Bromberek, I. Saika-Voivod, and V. Booth. Molecular dynamics simulations of histidine-containing cod antimicrobial peptide paralogs in self-assembled bilayers. *Biochimica et Biophysica Acta - Biomembranes*, 1838(11):2778–2787, 2014.

[9] M. Husain and J. Choi. Calmodulin-binding peptides that reduce cell proliferation in cancer and smooth muscle proliferation diseases. International Application Number PCT/CA2008/001474, International Publication Number WO 2009/023959 A1, 2009.

[10] P. Bauer, B. Hess, and E. Lindahl. Gromacs 2022.1 source code (2022.1). Zenodo., 2022.

[11] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindah. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.

[12] S. Páll, M.J. Abraham, C. Kutzner, B. Hess, and E. Lindahl. Tackling exascale software challenges in molecular dynamics simulations with gromacs. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8759:3–27, 2015. cited By 575.

[13] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. Van Der Spoel, B. Hess, and E. Lindahl. Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.

[14] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl. Grgmacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[15] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen. Gromacs: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005.

[16] E. Lindahl, B. Hess, and D. van der Spoel. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7(8):306–317, 2001.

[17] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, 1995.

[18] Nicolas Guex and Manuel C. Peitsch. Swiss-model and the swiss-pdbviewer: An environment for comparative protein modeling. *Electrophoresis*, 18(15):2714 – 2723, 1997.

[19] J.L.F. Abascal and C. Vega. A general purpose model for the condensed phases of water: Tip4p/2005. *Journal of Chemical Physics*, 123(23), 2005.

[20] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[21] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[22] I. Cornut, K. Büttner, J.-L. Dasseux, and J. Dufourcq. The amphipathic α-helix concept. application to the de novo design of ideally amphipathic leu, lys peptides with hemolytic activity higher than that of melittin. *FEBS Letters*, 349(1):29–33, 1994.

[23] Sabine Castano, Bernard Desbat, Michel Laguerre, and Jean Dufourcq. Structure, orientation and affinity for interfaces and lipids of ideally amphipathic lytic l(i)k(j)(i=2j) peptides. *Biochimica et Biophysica Acta - Biomembranes*, 1416(1-2):176 – 194, 1999.

Appendix A

R_g Time Series and Snapshots

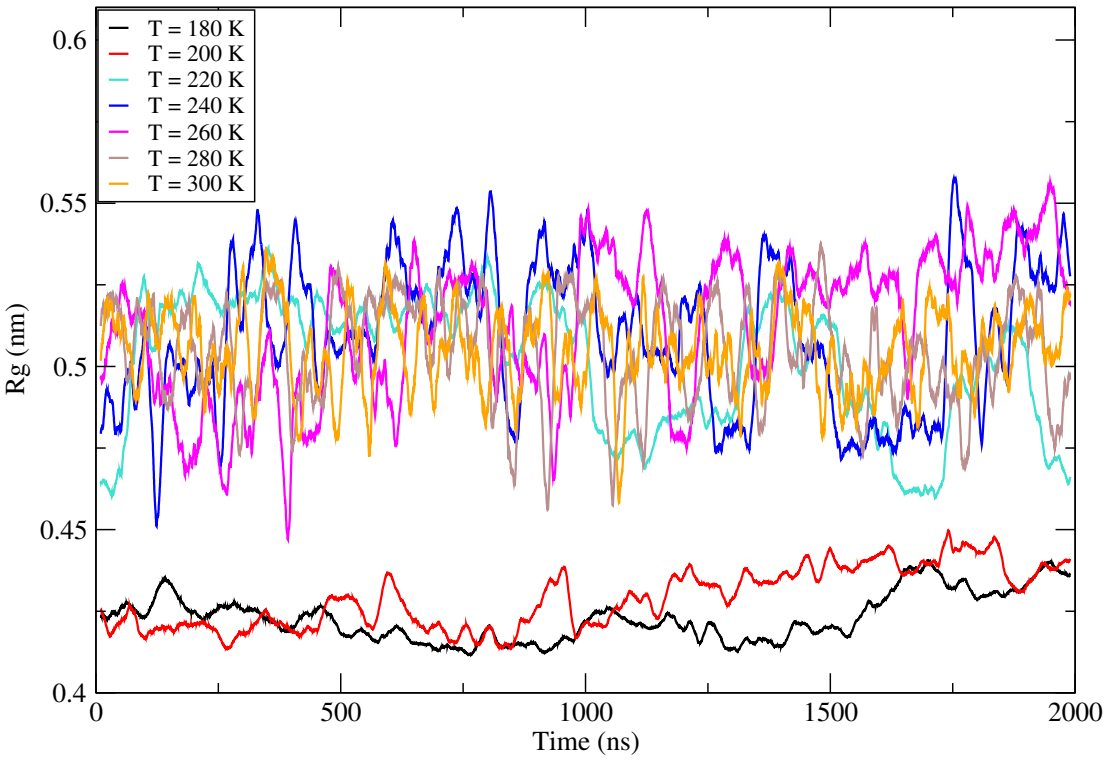
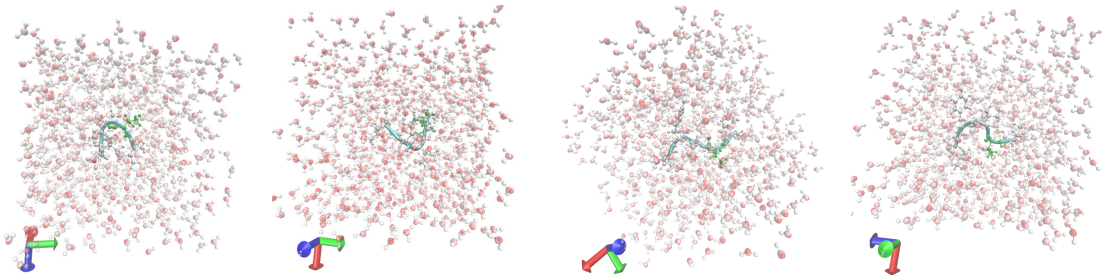


Figure A.1: 100 steps averaged Rg time series of CBS-5 in bulk water



(a) $T = 200$ K (b) $T = 220$ K (c) $T = 280$ K (d) $T = 300$ K

Figure A.2: Snapshots of CBS-5 in bulk

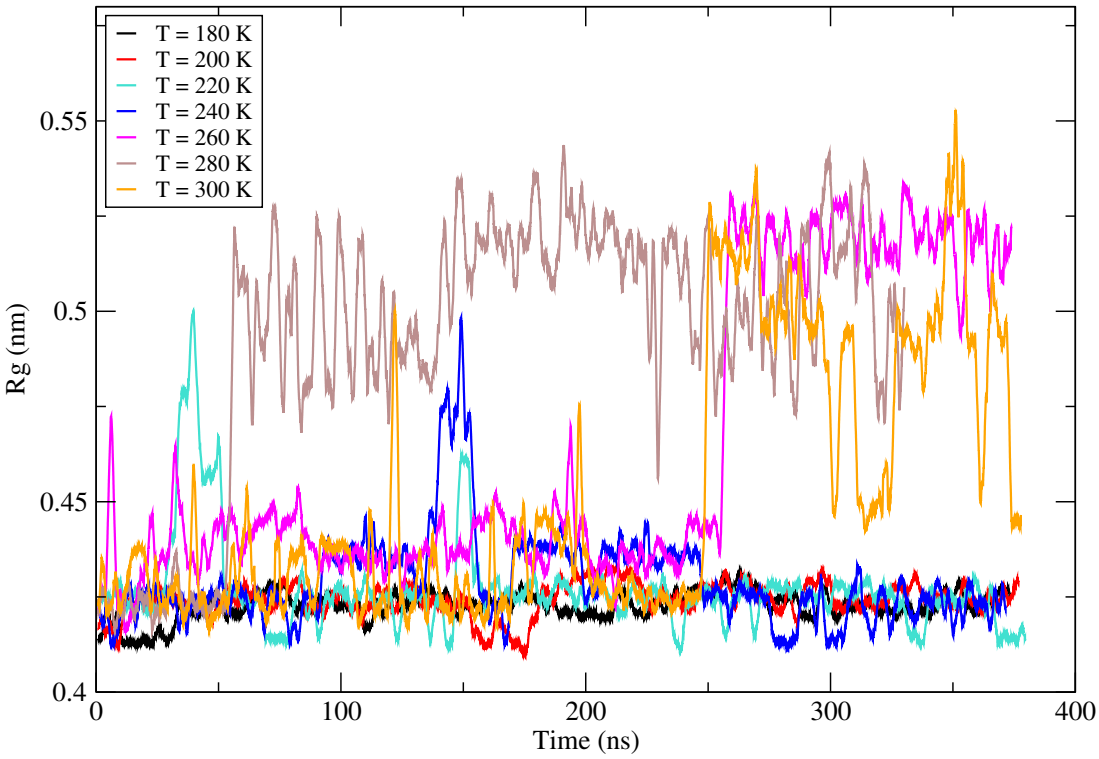
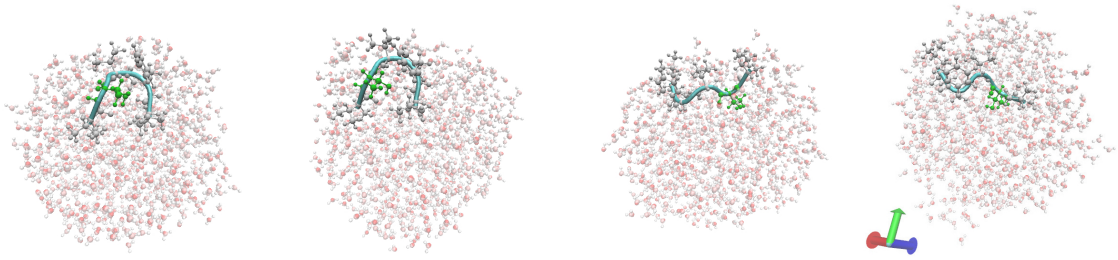


Figure A.3: 100 steps averaged Rg time series of CBS-5 in nanodroplets



(a) $T = 180$ K (b) $T = 220$ K (c) $T = 280$ K (d) $T = 280$ K

Figure A.4: Snapshots of CBS-5 in nanodroplet

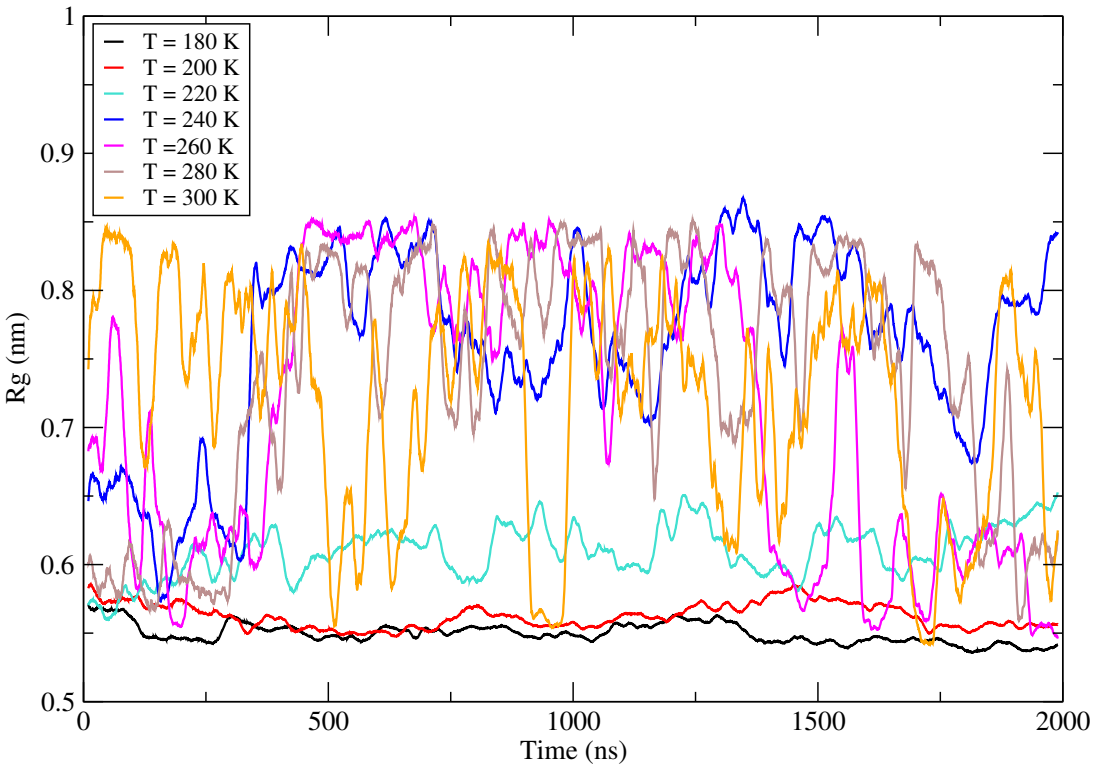
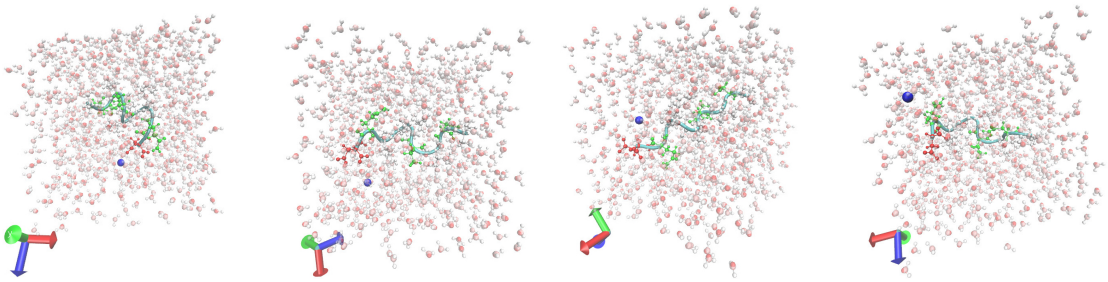


Figure A.5: 100 steps averaged Rg time series of CBS-9 in bulk water



(a) $T = 200$ K (b) $T = 220$ K (c) $T = 280$ K (d) $T = 300$ K

Figure A.6: Snapshots of CBS-9 in bulk

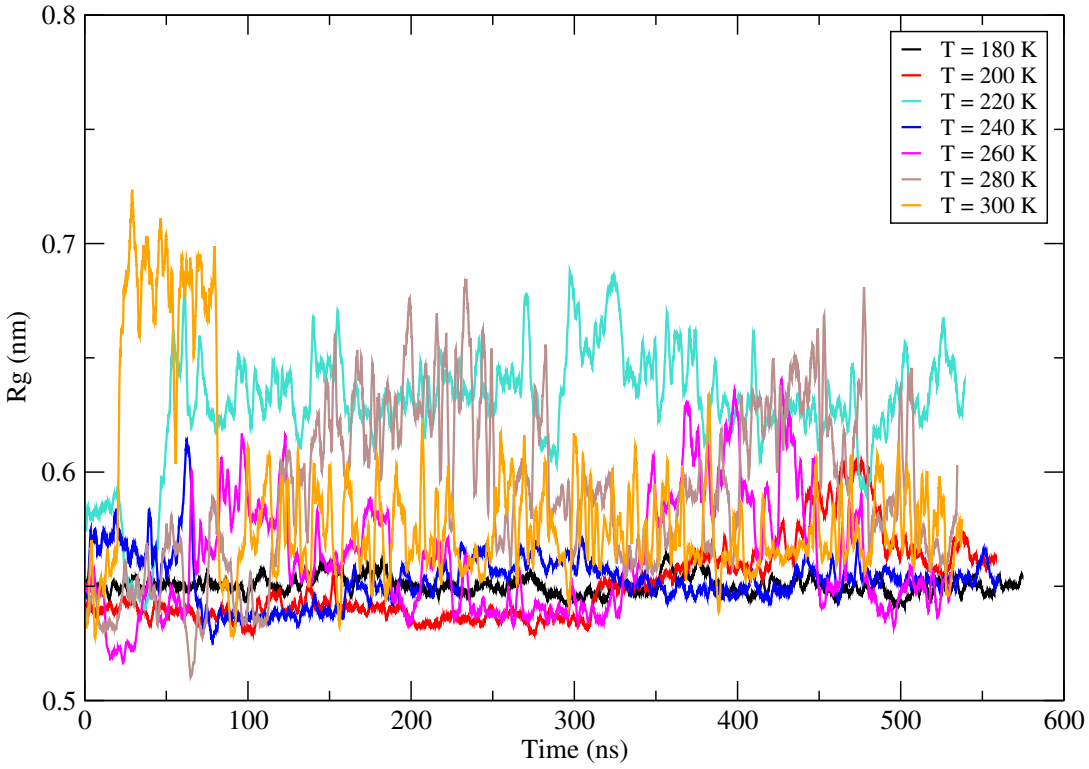
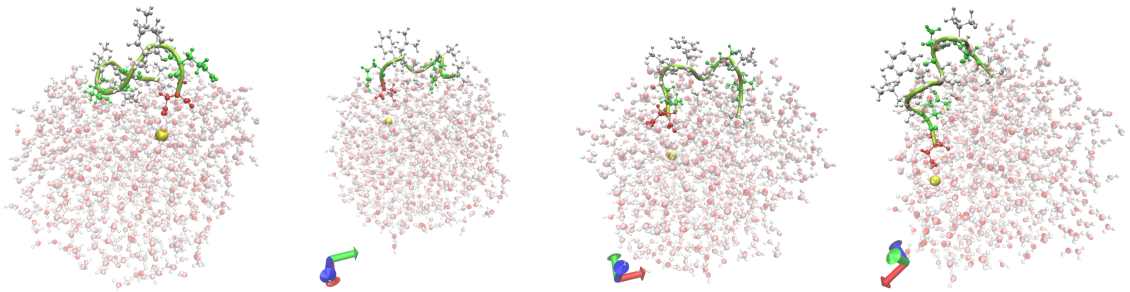


Figure A.7: 100 steps averaged Rg time series of CBS-9 in nanodroplets



(a) $T = 180$ K (b) $T = 220$ K (c) $T = 260$ K (d) $T = 360$ K

Figure A.8: Snapshots of CBS-9 in nanodroplet

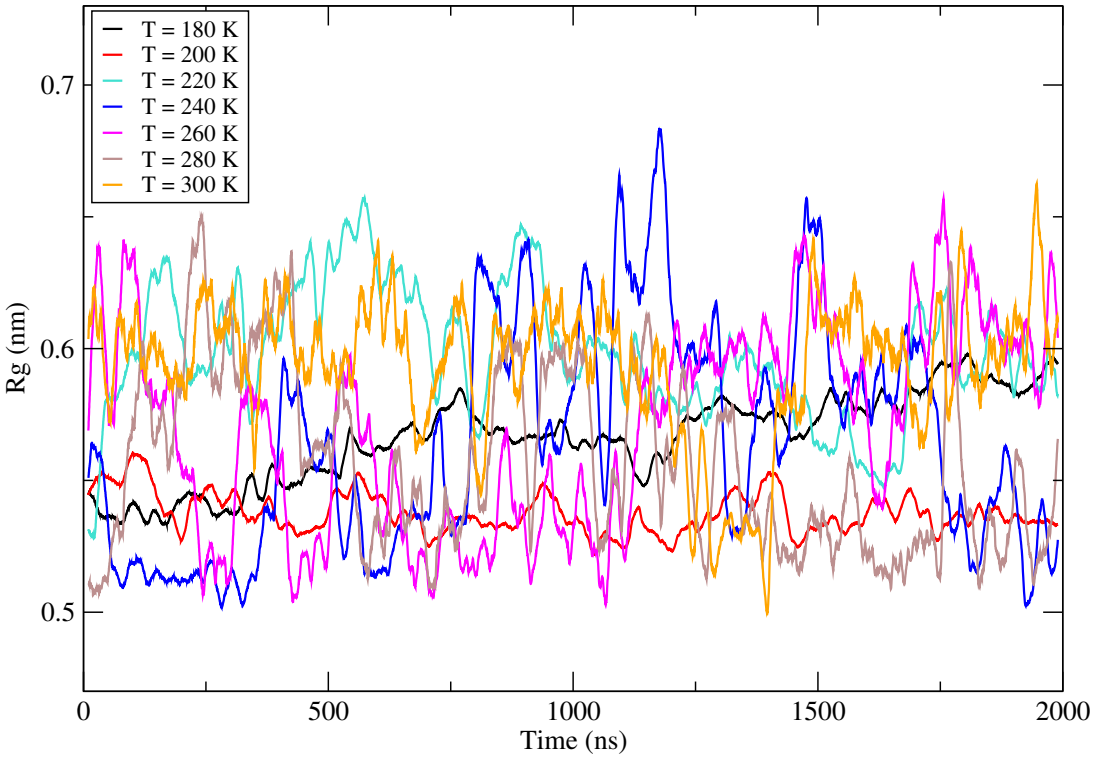
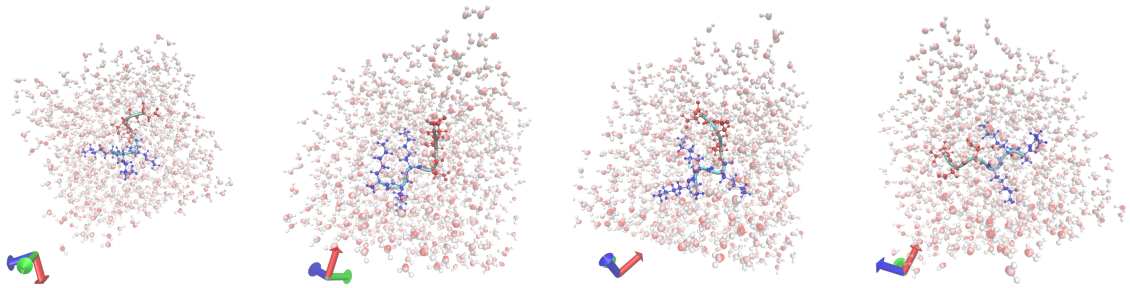


Figure A.9: 100 steps averaged Rg time series of KKKDDD in bulk water



(a) $T = 180$ K (b) $T = 220$ K (c) $T = 300$ K (d) $T = 300$ K

Figure A.10: Snapshots of KKKDDD in bulk

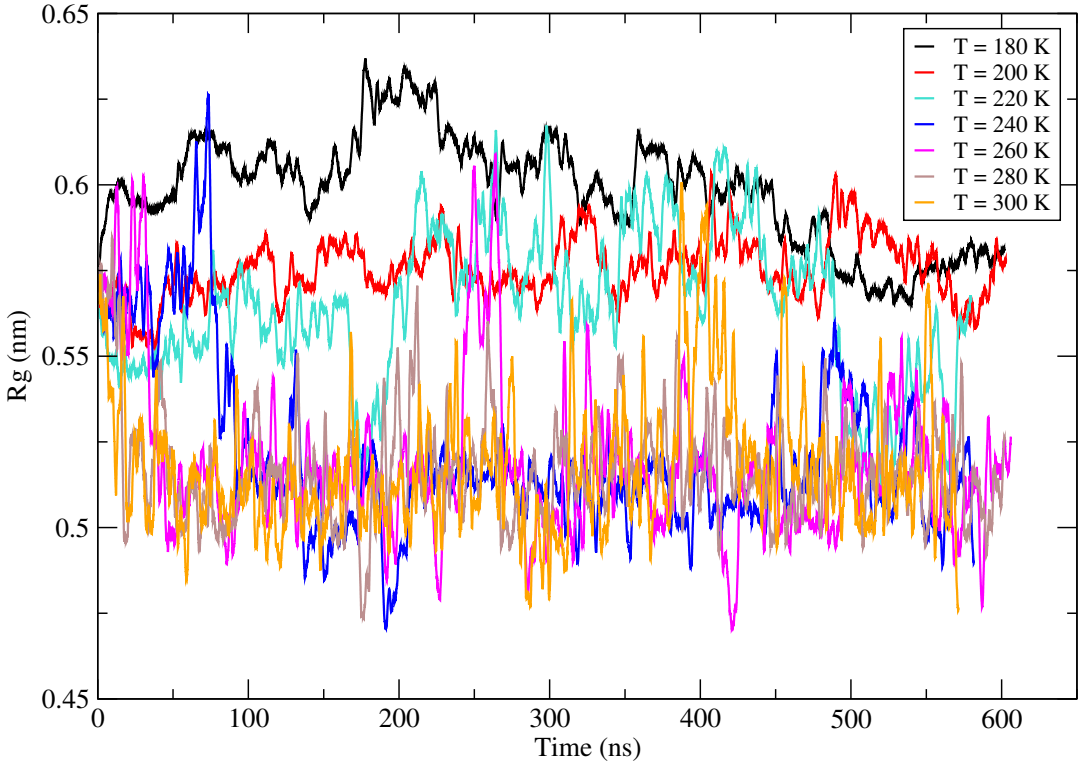
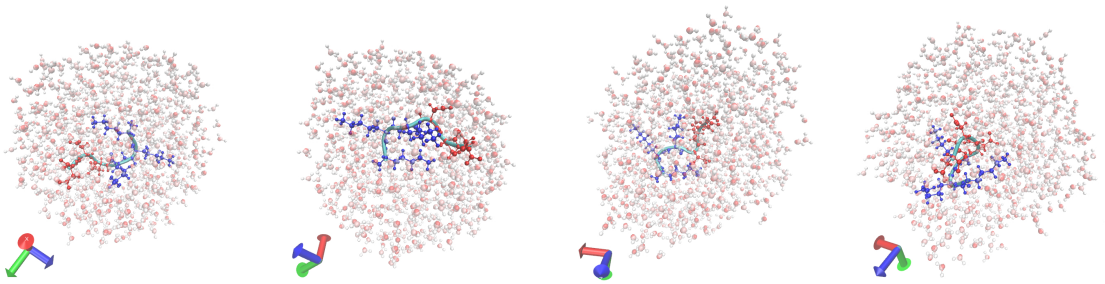


Figure A.11: 100 steps averaged Rg time series of KKKDDD in nanodroplets



(a) $T = 180$ K (b) $T = 180$ K (c) $T = 300$ K (d) $T = 300$ K

Figure A.12: Snapshots of KKKDDD in nanodroplet

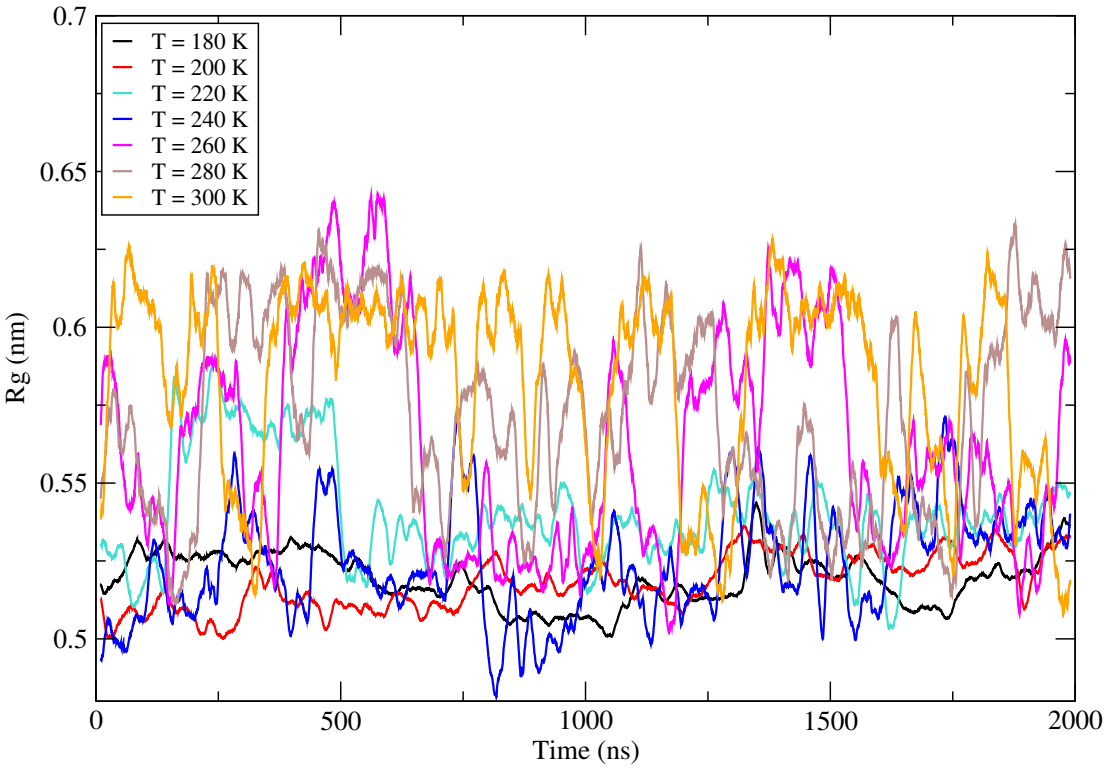
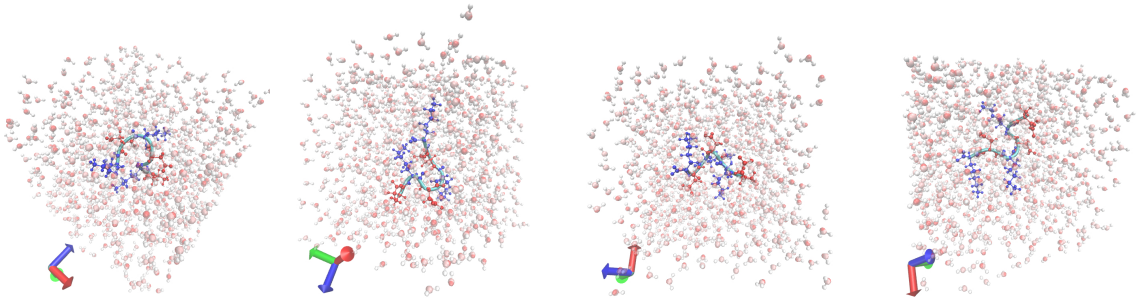


Figure A.13: 100 steps averaged Rg time series of DKDKDK in bulk water



(a) $T = 180$ K (b) $T = 240$ K (c) $T = 280$ K (d) $T = 300$ K

Figure A.14: Snapshots of DKDKDK in bulk

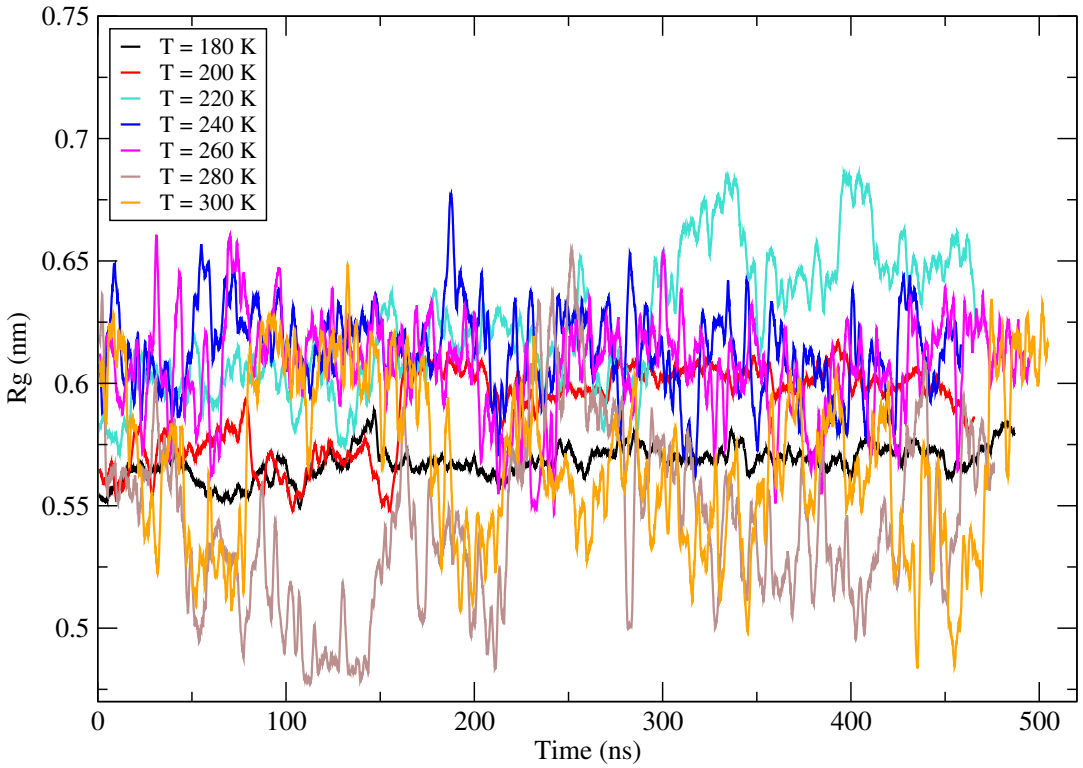
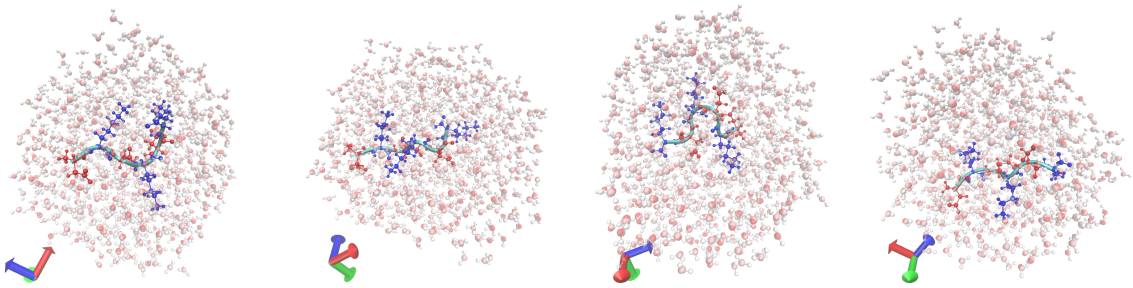


Figure A.15: 100 steps averaged Rg time series of DKDKDK in nanodroplets



(a) $T = 180$ K (b) $T = 240$ K (c) $T = 280$ K (d) $T = 300$ K

Figure A.16: Snapshots of DKDKDK in nanodroplet

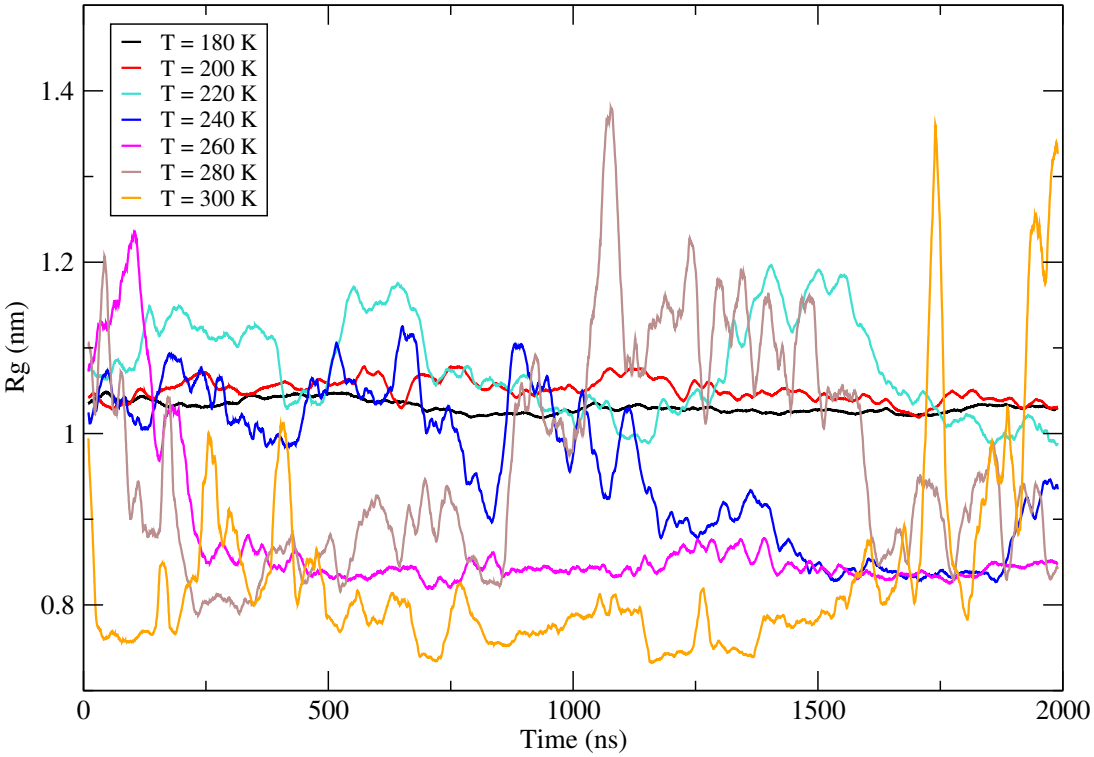
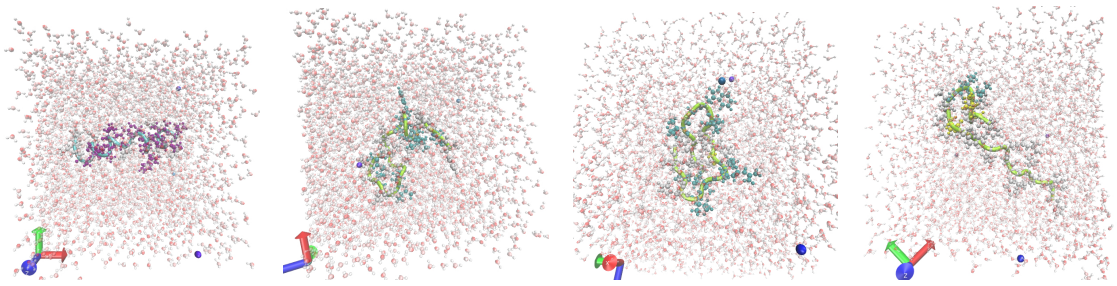


Figure A.17: 100 steps averaged Rg time series of GAD-1 in bulk water



(a) $T = 180$ K (b) $T = 240$ K (c) $T = 260$ K (d) $T = 300$ K

Figure A.18: Snapshots of GAD-1 in bulk

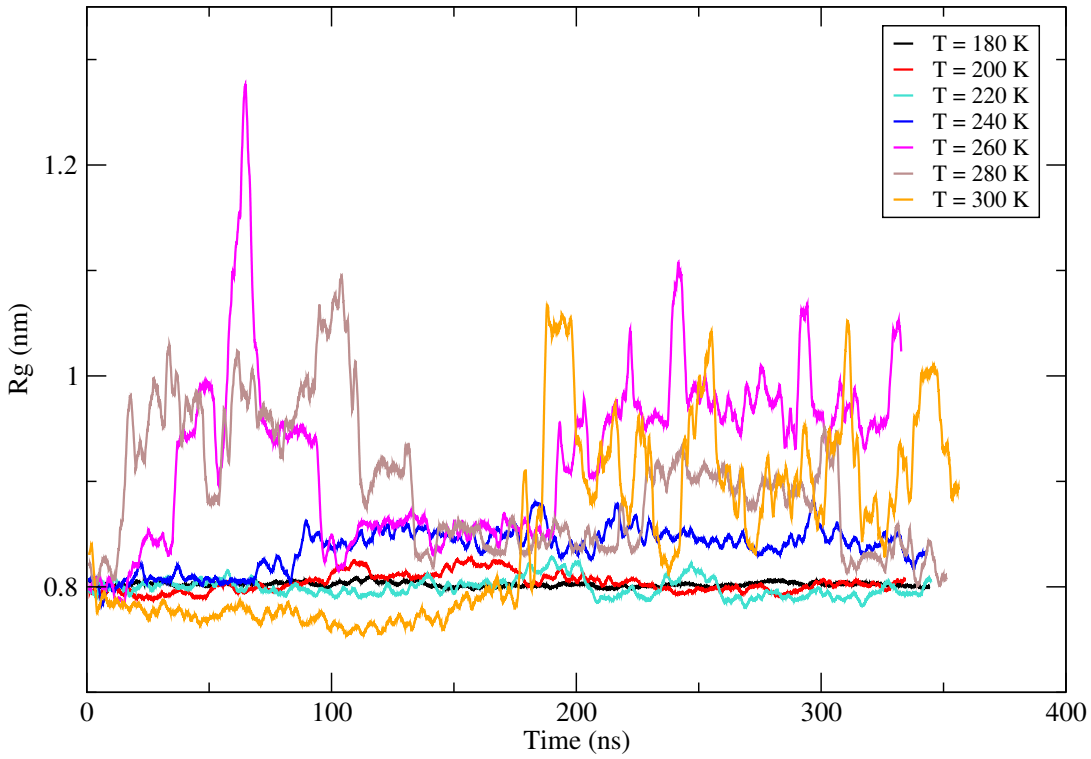


Figure A.19: 100 steps averaged Rg time series of GAD-1 in nanodroplets

Appendix B

The N-terminus

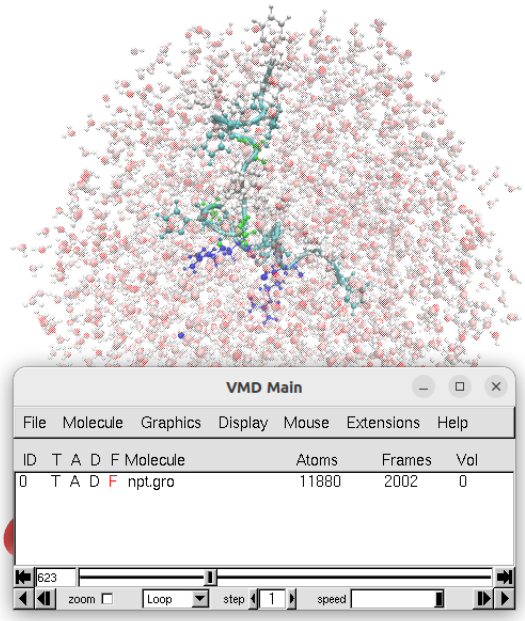


Figure B.1: N-terminus of GAD-1 emerging out of the droplet

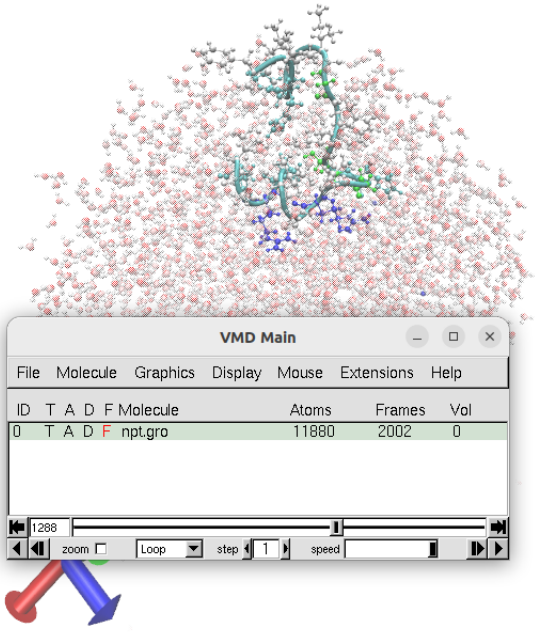


Figure B.2: Other hydrophobic residues followed the N-terminus