From folding to misfolding: Conformational properties of biomolecules studied by computational simulations

Marijana Ugrina Barun

Angaben zur Veröffentlichung / Publication details:

Ugrina Barun, Marijana. 2026. “From folding to misfolding: Conformational properties of biomolecules studied by computational simulations.” Augsburg: Universität Augsburg.

From Folding to Misfolding: Conformational Properties of Biomolecules Studied by Computational Simulations

Dissertation

zur Erlangung des akademischen Grades
Dr. rer. nat.

eingereicht an der
Mathematisch-Naturwissenschaftlich-Technischen Fakultät
der Universität Augsburg

von
Marijana Ugrina Barun

Augsburg, Oktober 2025

Universität
Augsburg
University

1. Erstgutachter: Prof. Dr. Nadine Schwierz

2. Zweitgutachter: Prof. Dr. Cristoph Weber

Tag der mündlichen Prüfung: 17. 04. 2026

Contents

List of Abbreviations **ix**

List of Figures **xii**

List of Tables **xiii**

Zusammenfassung **xx**

Summary **xxv**

1 Introduction **1**

1.1 Nucleic acids structure and function 1

1.1.1 RNA G-quadruplexes 2

1.2 Protein structure and misfolding 5

1.2.1 Misfolding of proteins: amyloid fibrils 7

1.3 Outline of the thesis 8

2 Computational Methods **11**

2.1 Molecular dynamics simulations 11

2.1.1 Newton's equations of motion 11

2.1.2 Initiating an MD simulation 13

2.1.3 Force fields for biomolecules 14

2.1.4 Enhanced sampling MD simulations 18

2.2 Langevin dynamics simulations 19

2.2.1 Langevin equations of motion 19

2.2.2 Coarse-grained models for RNA simulations 20

2.3 Root mean square deviation 23

2.4 Root mean square fluctuation 24

3 Coarse-grained and MD simulation studies of folding pathways of RNA G-Quadruplex **25**

3.1 Introduction 25

3.2 Methods 27

3.2.1 Simulation Models 27

3.2.2 All-Atom MD Simulations 29

3.2.3 CG Simulations Using the TIS Model 30

3.2.4 Time Scaling and Kinetic Analysis in Coarse-Grained Simulations . 31

3.2.5 Gibbs Free Energy Calculation 32

3.2.6 Radius of Gyration and Native Contact Analysis 32

3.2.7 Order Parameters 33

3.2.8 Concentration Matching Between Atomistic and Coarse-Grained Simulations . . . 34
3.2.9 Estimating the Fraction of Folded Structures from CD Melting Curves 35
3.3 Results and discussion . . . 36
3.3.1 Quantitative comparison between simulated and experimental fractions of folded structures . . . 36
3.3.2 The folding and unfolding of TERRA25 proceed sequentially through multiple pathways . . . 37
3.3.3 Conformational entropy is a hallmark of rG4 folding . . . 43
3.4 Conclusion . . . 46
3.5 Appendix . . . 47
3.5.1 Data archive . . . 47

4 Towards artificial ion channels: probing the ion specificity of TERRA G-quadruplex by enhanced sampling simulations 51
4.1 Introduction . . . 51
4.2 Methods to analyze ion binding patterns and ion transport . . . 53
4.2.1 Straightforward MD simulations . . . 53
4.2.2 TREMD simulations . . . 53
4.2.3 Pulling simulations . . . 54
4.3 Results and discussion . . . 55
4.3.1 Straightforward simulations provide an initial insight into the ionic transport . . . 55
4.3.2 TREMD simulations reveal energetics of cation binding to TERRA G4 . . . 58
4.3.3 Force-dependent ionic conductivity and selectivity in TERRA G4 from pulling simulations . . . 61
4.4 Conclusion . . . 63
4.5 Appendix . . . 65
4.5.1 Data archive . . . 65

5 Misfolding of proteins into amyloid fibrils 67
5.1 Introduction: Misfolding of proteins into fibrils . . . 67
5.2 Methods: MD simulations of fibrils . . . 68
5.3 Results 1: Comparison of ex-vivo and in-vitro Aß amyloid fibril structures using cryo-EM and MD simulations . . . 68
5.3.1 Methods for simulating Aß amyloid fibrils . . . 69
5.3.2 Results and discussion: molecular determinants of stability in seeded and unseeded Aß fibrils . . . 69
5.3.3 Summary and conclusion . . . 72
5.4 Results 2: MD simulation study into structural properties of amyloid resistance 72
5.4.1 Methods for simulating ApoA-II amlyoid fibrils . . . 73

5.4.2 Results and discussion: Impact of Point Mutations on the Structural and Energetic Stability of ApoA-II Amyloid Fibrils 73
5.4.3 Summary and conclusion 77
5.5 Results 3: Impact of point mutations on stability of lysosyme-derived amyloid fibril 78
5.5.1 Methods for simulating lysozyme fibrils 78
5.5.2 Results and discussion: Impact of the D87G mutation on the structural stability of lysozyme amyloid fibrils 79
5.5.3 Summary and conclusion 83
5.6 Appendix 84
5.6.1 Data archive 84

6 Conclusions and Outlook **89**

List of Publications **93**

Bibliography **93**

Acknowledgments **111**

List of Abbreviations

MD	Molecular dynamics
CG	Coarse-grained
LD	Langevin dynamics
TIS	Three Interaction Site (CG model)
TREMD	Temperature replica exchange molecular dynamics
ML	Machine Learning
AI	Artificial Intelligence
AMBER	Assisted model building with energy refinement
CHARMM	Chemistry at Harvard macromolecular mechanics
GAFF	General amber force field
TIP3P	Transferable intermolecular potential 3-point
TIP4P	Transferable intermolecular potential 4-point
LINCS	Linear constraint solver
vdW	Van der Waals
QM	Quantum mechanical
COM	Center of mass
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
G4	G-quadruplex
rG4	RNA G-quadruplex
dG4	DNA G-quadruplex
mRNA	Messenger RNA
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
tRNA	Transfer RNA
rRNA	Ribosomal RNA
siRNA	Small interfering RNA
miRNA	MicroRNA
ncRNA	Non-coding RNA
lncRNA	Long non-coding RNA
TERRA	Human telomeric RNA
cryoEM	Cryogenic Electron Microscopy
smFRET	Single-molecule Fröster Resonance Energy Transfer
NMR	Nuclear magnetic resonance
ssNMR	Solid state NMR

TEM	Transmission electron microscopy
AFM	Atomic force microscopy
CD	Circular dichroism
MFPT	Mean first passage time
CV	Collective variable
PDF	Probability density function
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
e.g.	Latin *exempli gratia*; for example

List of Figures

1.1 RNA and DNA structure representation 1
1.2 RNA G-quadruplex structure representation and guanine syn- and anti-angle definition. 4
1.3 Structure of the alanine (ALA) amino acid and ubiquitin 6
1.4 Schematic representation of an amyloid fibril structure. 8

2.1 Schematic representation of the interaction potential within a molecule. . . 15
2.2 A schematic representation of periodic boundary conditions. 15
2.3 Schematic representation of parameters of the TIS force-field 22
2.4 Schematic representation of the highly charged rod-like polymer. 23

3.1 Monomeric 25-mer TERRA G4 used in simulations. 27
3.2 rG4 hairpin models simulated in all-atom simulations. 28
3.3 Concentration matching procedure. 35
3.4 Comparison of the fraction of folded states n_f from simulations and experiments. 36
3.5 Representative rG4 simulation trajectory showing multiple folding and unfolding events together with snapshots and free-energy landscape of folding and unfolding transitions. 38
3.6 Representative folding pathway of TERRA G4. 38
3.7 Folding and unfolding pathways of TERRA25 from the TIS simulations. . . 40
3.8 Folding and unfolding pathways of TERRA25 from the TIS simulations, with double-hairpin intermediates. 41
3.9 Transient non-native states observed on the TERRA25 folding pathway. . . 42
3.10 Order parameters describing the number of hydrogen bonds within hairpins simulated in all-atom simulations. 43
3.11 Free energy profiles and thermodynamic stabilities of different rG4 systems. 44
3.12 Free energy landscapes of TERRA25 folding projected onto the composite order parameters s_1 and s_2 45
3.13 Free energy landscapes of rG4 folding projected onto the composite order parameters s_1 and s_2 46

4.1 Initial set-up for constant-force pulling simulations. 55
4.2 Order parameters $\eta^{i}_{q_n}$ describing the ionic binding to rG4 quartets. 57
4.3 RMSD of the TERRA G4 central pore calculated from the simulations in NaCl, KCl, CsCl, LiCl and NH_4Cl solutions. 58
4.4 Free energy profiles of cations along the z-axis of the TERRA G4 central pore. 58
4.5 2D histograms of cationic locations in the TERRA G4 central pore according to the xy-plane perpendicular to the direction of the cationic flow. 59

4.6 Histograms representing number of water molecules in cationic hydration shells along the z-axis of the TERRA G4 channel pore. 60
4.7 Mean first passage times (τ) of cations along the TERRA G4 central pore under the influence of the pulling force 61
4.8 MFPT of K^+ and Na^+ ions along the TERRA G4 central pore as a function of the force constants. 62
4.9 I-V plot for K^+ and Na^+ conductance through the TERRA G4 ion channel. 63
4.10 Average time that K^+ and Na^+ ions spend above quartets q_1, q_2 and q_3. . . 63

5.1 Cartoon representation of the Cryo-EM-resolved structures of seeded and unseeded morphologies with sequences. 70
5.2 Cross sections representing two interacting chains from two protofibrils of simulated β-fibril morphologies. 70
5.3 Structures of simulated β-fibril morphologies and their RMSDs from simulations. 71
5.4 Cryo-EM structure of ApoA-IIC fibril Morphology I. 74
5.5 Cross-section of ApoA-IIC and ApoA-IIF amyloid structures 75
5.6 RMSDs calculated from 200 ns simulations of ApoA-IIC structure and homology models. 76
5.7 Chains from even tip, center and odd tip of the ApoA-IIF model fibril structure colored according to the RMSFs in simulations. 76
5.8 Interaction energies and hydrogen bond analysis in the protein and between protein and solvent, calculated from ApoA-IIC and ApoA-IIF simulations. . 77
5.9 Three-dimensional structure of the lysozyme fibril structure, resolved by cryo-EM spectroscopy. 79
5.10 Cross-section of a single chain of the patient-derived and wild-type lysozyme fibrils. 80
5.11 RMSD analysis of the patient-derived and wild-type structure during simulations with snapshots. 81
5.12 RMSF per residue represented as the color gradient on the average coordinates of two outer chains, and a middle chain for the WT fibril structure. . 82
5.13 Free energy calculations and hydrogen-bond analysis of the patient and the WT fibril structure 82
5.14 Maximal distances per reside in the patient and wild type sequence between the fibril core and outer chains. 83

List of Tables

3.1 Sequences of simulated rG4 structures. 29
3.2 Melting temperatures (T_m) of TERRA25 at various bulk KCl concentrations obtained from CD experiments and coarse-grained (CG) simulations. 37
3.3 Populations and lifetimes of the folded, unfolded, and intermediate states from CG simulations. 39
3.4 Rates of transitions between intermediates on the TERRA25 folding pathway. 42

Zusammenfassung

Nukleinsäuren dienen in allen Organismen und Viren als Träger der genetischen Information. Desoxyribonukleinsäure (DNA) enthält die genetischen Anweisungen, die die Entwicklung, Funktion, Reproduktion und Aufrechterhaltung von Lebewesen steuern. Sie wird in Ribonukleinsäuren (RNAs) transkribiert, welche wiederum Proteine kodieren (kodierende RNA), die Genexpression regulieren oder strukturelle und katalytische Funktionen in der Zelle ausüben (nicht-kodierende RNA) [1]. Ein wichtiges Beispiel für eine nicht-kodierende RNA ist die lange nicht-kodierende RNA (lncRNA). lncRNAs spielen eine bedeutende Rolle bei der Genregulation und sind an Krebsmechanismen beteiligt [2]. Sie sind lange RNA Strukturen mit repetitiven Elementen, die für ihre Funktion entscheidend sind. Eine bekannte lncRNA Struktur, die menschliche telomerische RNA (TERRA), enthält eine charakteristische guaninreiche Sequenz r(UAGGG), die RNA G-Quadruplexe (rG4) bildet und in die Telomerregulation involviert ist [3, 4]. G-Quadruplexe sind nicht-kanonische Strukturen, die durch nicht-kanonische Bindung von Guaninen in planaren Quartetten von DNA und RNA Sequenzen gebildet werden. DNA G-Quadruplexe (dG4s) wurden aufgrund experimenteller Nachweise für ihre Existenz in vivo und langlebiger Zwischenzustände, die für experimentelle Techniken zugänglich sind, intensiver untersucht [5, 6]. Im Gegensatz dazu sind rG4s in Zellen hochdynamisch, und ihre Existenz in vivo wurde erst vor kurzem durch experimentelle Belege gestützt [7]. Das Forschungsinteresse an rG4s stieg jedoch stark an, als ihre potenzielle Rolle in wichtigen zellulären Prozessen wie Transkription, Translation, prä-mRNA-Prozessierung und Telomererhaltung entdeckt wurde [3, 4, 8, 9].

G4s wurden mit einer Vielzahl experimenteller Methoden untersucht. Die Kernspinresonanzspektroskopie (NMR) und die Röntgenkristallographie gehören zu den am häufigsten verwendeten Methoden, um die Strukturen von rG4s und dG4s auf atomarer Ebene aufzulösen. Die Röntgenkristallographie ermöglicht zudem die Visualisierung von Kationen im G4-Kern, die für deren Stabilität essentiell sind [10]. Zeitaufgelöste NMR-Messungen haben langlebige Zwischenzustände entlang des dG4-Faltungsweges aufgezeigt [11–13] sowie Unterschiede in den Faltungswegen zwischen dem TERRA rG4 und seinem DNA Gegenstück [14].

UV-Schmelzexperimente und die Zirkulardichroismus (CD) Spektroskopie werden verwendet, um G4-Topologien anhand ihrer charakteristischen Spektralsignaturen und thermischen Stabilitäten zu unterscheiden. Diese Methoden werden oft mit Techniken kombiniert, die G4s in vivo untersuchen, wie etwa die Live-Cell-Bildgebung. Kürzlich wurde eine Kombination aus Live-Cell-Bildgebung und CD-Spektroskopie eingesetzt, um die Faltung und Entfaltung von rG4s in menschlichen Zervixkarzinomzellen zu beobachten [7]. Dennoch bleiben kurzlebige Zwischenzustände im Faltungsprozess von rG4s aufgrund ihrer dynamischen und kurzlebigen Natur schwer zu charakterisieren.

Kleinere Zeitskalen sind mit computergestützten Methoden wie Molekulardynamik (MD) und Langevin- Dynamik (LD) Simulationen zugänglich. In all-atomaren MD-Simulationen wird jedes Atom explizit dargestellt, und ihre Dynamik wird durch Lösung

der Newtonschen Bewegungsgleichungen simuliert. Die Wechselwirkungen zwischen den Atomen werden durch Kraftfelder beschrieben, die in Abhängigkeit von Atom- und Molekültyp parametrisiert sind. Auf diese Weise können MD-Simulationen Dynamiken im Mikrosekunden- bis Millisekundenbereich auflösen. Darüber hinaus können in gröberen Simulationen mehrere Atome zu einer Wechselwirkungsstelle zusammengefasst werden, um Prozesse auf größeren Zeitskalen zu beobachten, wobei dennoch ein ausreichendes Maß an Detail erhalten bleibt. Die Langevin Dynamik, ein verwandter Ansatz, integriert stochastische und Reibungskräfte, um Lösungsmittelwirkungen effizienter nachzubilden, was oft längere Simulationen spezifischer Prozesse ermöglicht.

Im ersten Teil dieser Arbeit verwendete ich ein Drei-Wechselwirkungsstellen (TIS) LD-Modell zusammen mit all-atomaren MD-Simulationen und CD Experimenten, um den Faltungsweg des TERRA-G4 zu untersuchen. Die Forschung wurde in Zusammenarbeit mit den Wissenschaftlerinnen und Wissenschaftlern aus der Gruppe von Prof. Harald Schwalbe an der Goethe-Universität Frankfurt am Main durchgeführt, die die CD-Messungen bereitstellten. Ich simulierte TERRA-G4 und G4s, die von Pflanzen-RNA gebildet wurden, sowie einige synthetische RNA-Sequenzen mit unterschiedlichen Schleifenlängen.

Zunächst wurde, um die im TIS-Modell angenommene uniforme Ionendistribution zu korrigieren und die für RNAs charakteristische Bildung einer Ionen-Doppelschicht zu berücksichtigen, ein Konzentration Anpassungsverfahren entwickelt, das auf all-atomaren MD-Simulationen basiert. Damit konnten wir den experimentell gemessenen Anteil gefalteter Zustände bei variierenden Salzkonzentrationen und Temperaturen reproduzieren. Die Ergebnisse von Simulationen bei 25°C zeigten, dass TERRA-G4 einen verzweigten Faltungsweg mit mehreren kurzlebigen Zwischenzuständen aufweist. Die Faltung ist entropiegetrieben, weshalb die Zwischenzustände durch Guanin-Stapelung und sequentielle Bindung von Guanin Wiederholungen gebildet werden. Die in den Simulationen beobachteten Haarnadel, Triplex und Doppelhaarnadel Zwischenzustände bestätigen Vorhersagen mehrerer NMR-Experimente an dG4s [11–13] sowie experimentelle Arbeiten an rG4s [15, 16]. Zusätzlich zeigten Simulationen anderer rG4-Systeme dieselben Zwischenzustände, wobei sich die Populationsverteilungen je nach Sequenz unterschieden. Die Berechnungen der Gibbs'schen freien Energie ergaben außerdem eine interessante Eigenschaft: die Stabilität von rG4s nimmt mit zunehmender Schleifenlänge ab. Dies unterstreicht die Bedeutung der Entropie bei der rG4-Faltung.

Zusammenfassend konnten wir in dieser Arbeit Zwischenzustände im Faltungsweg von rG4s auflösen, die experimentellen Methoden aufgrund ihrer transienten Natur meist nicht zugänglich sind. Die Ergebnisse machten zudem Schwächen in der Definition von Observablen bei der Untersuchung der rG4-Faltung deutlich und zeigten, dass einfache Metriken wie die End-zu-End-Distanz nicht geeignet sind, die strukturell komplexen Zwischenzustände aufzulösen.

G4s binden Kationen an ihren Guanosin-Kern, die für ihre Stabilität entscheidend sind. Aufgrund von Eigenschaften wie Hydratationsfreier Energie, dem van-der-Waals-Radius der Kationen und der Bindungsfreien Energie an die Guanosin O6-Atome haben nicht alle Ionen denselben stabilisierenden Effekt auf G4s [17]. Daher zeigen G4s Ionspezifität

und bilden sich bevorzugt in Gegenwart bestimmter einwertiger Kationen wie K^+ oder Na^+. Dies macht G4s zu attraktiven Zielen in der Forschung zu künstlichen Ionenkanälen. Insbesondere rG4s stechen aufgrund ihrer höheren Stabilität im Vergleich zu dG4s hervor [18].

Künstliche Ionenkanäle finden Anwendung bei der Behandlung von Ionenkanaldysfunktionen, in der Krebstherapie, in antibakteriellen Behandlungen, beim Biosensing und in der Nanotechnologie [19, 20]. Das Hauptziel der Forschung ist es, einen Ionenkanal zu entwerfen, der in die hydrophobe Membran integriert werden kann und Ionen-Erkennungsstellen besitzt. Frühere NMR- und Röntgenexperimente haben gezeigt, dass G4s Kationen in ihren Kernbindungsstellen koordinieren und diese gerichtet transportieren können [21–24]. Darüber hinaus wurde eine Studie mit Voltage-Clamp-Messungen und MD-Simulationen durchgeführt, in der ein dG4 erfolgreich mithilfe eines gestapelten lipophilen Guanosins in eine Membran eingefügt wurde, wodurch ein künstlicher Ionophor selektiv für K^+-Ionen entstand [25].

Im zweiten Teil meiner Arbeit verwendete ich MD-Simulationen, um die Ionentransporteigenschaften eines durch das TERRA-Molekül gebildeten G4 zu untersuchen. Dasselbe Molekül wurde in dieser Forschung wie in den RNA-Faltungsstudien aus dem ersten Teil der Arbeit verwendet. Es wurde in fünf verschiedenen Systemen simuliert, die sich durch die Salzlösungen KCl, NaCl, LiCl, CsCl und NH_4Cl unterschieden.

Die Ergebnisse von direkten und TREMD-Simulationen bestätigten die Spezifität des rG4 für K^+ und Na^+ Ionen. Diese hatten in den direkten Simulationen die höchste durchschnittliche Anzahl an Ionen im Kanal und in den TREMD-Simulationen die niedrigsten freien Energiebarrieren in den freien Energieprofilberechnungen. Außerdem gaben nur Na^+ und K^+ Ionen alle Wassermoleküle aus ihrer Hydrathülle frei, während sie sich in der Kanalpore befanden. Die Berechnungen zeigten jedoch auch Unterschiede in der Bindung dieser beiden Ionen an das rG4. Während das K^+ Ion die höchste Bindungsaffinität zu den zentralen Quartetten aufwies, verbrachte das Na^+ Ion die meiste Zeit am Kanalausgang und wanderte kumulativ durch den Kanal.

Die Zugkraft-Simulationen mit konstanter Kraft, die zur Messung der mittleren ersten Durchtrittszeit (MFPT) für den Ionentransport durch die G4-Pore eingesetzt wurden, bestätigten die Ergebnisse der MD- und TREMD-Simulationen, wobei Na^+ und K^+ Ionen die niedrigsten MFPTs zeigten. Sie offenbarten jedoch ein interessantes Verhalten der MFPTs von Na^+ und K^+ Ionen. Bei höheren Kraftkonstanten war die MFPT des Na^+-Ions geringer als die des K^+-Ions. Dieses Verhältnis kehrte sich bei niedrigeren Kraftkonstanten um, sodass die K^+ Ionen eine niedrigere MFPT aufwiesen. Eine detailliertere Analyse der MFPTs entlang der Bindungsstellen ergab, dass Na^+ Ionen die meiste Zeit am Kanalausgang verweilten, was mit den freien Energieprofilen aus den TREMD-Simulationen übereinstimmt, während K^+ Ionen die meiste Zeit im Kanalzentrum gebunden blieben. Die Ergebnisse für niedrigere Kraftkonstanten stimmen mit früheren Untersuchungen zur Ionenleitfähigkeit von dG4s überein, bei denen das dG4 die höchste Leitfähigkeit für K^+ Ionen zeigte [25]. Daher ist es möglich, dass die überhängenden Uridine den Transport von Na^+ Ionen am Kanalausgang behindern, dieser Effekt jedoch bei höheren Kraftkonstanten

weniger stark ausgeprägt ist.

Insgesamt lieferten diese Simulationen wertvolle Einblicke in die Unterschiede der Ionenbindung und Transporteigenschaften von rG4s. Sie bestätigten frühere Ergebnisse, die die Spezifität von rG4s für K^+ Ionen zeigten. Dennoch ist weitere Forschung notwendig, um die freien Energiebarrieren beim Ionentransport durch den rG4 Kanal vollständig aufzuklären und die wichtige Rolle der überhängenden Nukleoside in diesem Prozess besser zu verstehen.

Ein dritter Teil meiner Arbeit konzentriert sich auf den Einsatz von MD-Simulationen zur Untersuchung der Bedingungen, die die Fehlfaltung von Proteinen und die Bildung von Amyloidfibrillen beeinflussen. Die Forschung wurde in Zusammenarbeit mit der Gruppe von Prof. Marcus Fändrich vom Institut für Proteinbiochemie an der Universität Ulm durchgeführt.

Proteine sind essentiell für die Mechanismen des Lebens und übernehmen eine Vielzahl von Funktionen, die für alle zellulären Prozesse entscheidend sind. Ihre Funktion wird durch die Anordnung der Aminosäuren im gefalteten Zustand bestimmt. Wechselwirkungen innerhalb eines Proteins und zwischen Protein und Umgebung werden durch elektrostatische Wechselwirkungen, Wasserstoffbrückenbindungen, van-der-Waals (vdW) und hydrophobe Wechselwirkungen bestimmt. Mutationen und Umweltbedingungen können dazu führen, dass Proteine sich entfalten und ihre hydrophoben Bereiche dem Lösungsmittel exponieren. Dies führt zur Aggregation von Proteinen, die in der Bildung von Amyloidfibrillen resultieren kann.

Amyloidfibrillen sind geordnete, unlösliche Strukturen, die von Protofilamenten gebildet werden, welche sich umeinander winden und sogenannte sterische Reißverschlussmotive bilden. Die Hauptbausteine der Protofilamente sind β-Faltblätter, die durch Wasserstoffbrückenbindungen zwischen Rückgrat Amidgruppen aufeinander gestapelt sind [26]. Die Kryo-Elektronenmikroskopie (cryo-EM) ist eine leistungsstarke Methode zur Untersuchung von Amyloidfibrillen, da sie hochauflösende Einblicke in deren Struktur bietet [27]. Diese strukturellen Informationen können in MD-Simulationen verwendet werden, um Amyloidfibrill-Interaktionen und Stabilität auf atomarer Ebene weiter zu untersuchen.

Die Bildung von Amyloidfibrillen beginnt mit der Nukleation, bei der geordnete β-Stränge gestapelt werden, um einen Fibrillkern zu bilden. Dieser Prozess ist thermodynamisch ungünstig, doch ein gebildeter Kern ermöglicht anschließend einen schnelleren Elongationsprozess [28]. Vorgebildete Fibrillfragmente, die in eine Lösung monomerer Vorläuferproteine eingebracht werden, können das Fibrillwachstum beschleunigen und die Eigenschaften des Samens übertragen [28]. Im ersten Projekt untersuchten wir die Unterschiede in der strukturellen Stabilität zwischen ex-vivo- und in-vitro Aβ-Amyloidfibrillen mithilfe von cryo-EM und MD-Simulationen [29].

Die ungeimpften Fibrillen zeigten strukturellen Polymorphismus mit einer linkshändigen Konformation. Eine Morphologie der geimpften Fibrillen wurde mit einer rechtshändigen Konformation aufgelöst, und die strukturellen Eigenschaften des Samens wurden weitergegeben. Darüber hinaus zeigten geimpfte Fibrillen einen erhöhten β-Faltblattgehalt, eine höhere Auflösung in den cryo-EM-Rekonstruktionen und eine

verbesserte Resistenz gegen Proteolyse, was auf eine höhere Stabilität der geimpften Fibrillen hinweist.

MD-Simulationen der durch cryo-EM aufgelösten geimpften und ungeimpften Fibrillenstrukturen bei mehreren Temperaturen bestätigten, dass die geimpfte Morphologie konformationell günstiger ist. Root-Mean-Square-Deviation (RMSD) Berechnungen zeigten, dass ungeimpfte Fibrillen stärkere konformationelle Umordnungen durchliefen, insbesondere in den terminalen Regionen. Darüber hinaus beeinflusste der Protonierungszustand von Histidinresten die Stabilität der geimpften Morphologie erheblich, was die Rolle von Umweltfaktoren wie pH-Wert oder Ligandenbindung für die Fibrillenstruktur und stabilität verdeutlicht.

Diese Ergebnisse bestätigen die Ergebnisse der cryo-EM-Messungen, die eine höhere konformationelle Stabilität der geimpften Fibrillen zeigten, und klären die molekularen Mechanismen, die die Fibrillenstabilität beeinflussen.

Das zweite Projekt bezieht sich auf die strukturellen Eigenschaften amyloidresistenter Fibrillen [30]. Amyloidresistenz bezieht sich auf spezifische Mutationen, die Proteine vor der Fibrillenbildung schützen [30]. Diese natürlich vorkommenden Mutationen bieten Schutz vor Krankheiten wie Alzheimer, Kuru und systemischer Amyloidose, indem sie Schlüsselstrukturen stören, die für die Fibrillenausbreitung notwendig sind. In dieser Studie wurde die molekulare Grundlage der Amyloidresistenz in Apolipoprotein A-II (ApoA-II), einem Plasmaprotein, das an der systemischen AApoAII-Amyloidose beteiligt ist, untersucht. Zwei natürliche Varianten, ApoA-IIC (amyloidogen) und ApoA-IIF (resistent), unterscheiden sich durch vier Mutationen. Um die strukturellen Determinanten der Resistenz zu identifizieren, wurden MD-Simulationen an durch cryo-EM aufgelösten Fibrillen von ApoA-IIC, einem vollständigen ApoA-IIF-Homologiemodell und vier Einzelpunktmutanten durchgeführt.

Die Ergebnisse der RMSD-Berechnungen und der Wechselwirkungsenergien zeigten, dass die ApoA-IIF-Variante weniger stabil ist als ApoA-IIC. Unter den Einzelpunktmutationsmodellen wies das Modell mit der Asn62Lys-Mutation die geringste konformationelle Stabilität im Vergleich zum ApoA-IIC-Modell auf, was den Einfluss dieser Mutation auf die Stabilität der Fibrillenstruktur verdeutlicht. Berechnungen der Wechselwirkungsenergien zeigten, dass die geringere Stabilität der ApoA-IIF-Variante aus einer Abnahme der Wechselwirkungsenergien innerhalb des Proteins resultiert.

Diese Ergebnisse verdeutlichen, wie kernnahe Mutationen, die die lokale Ladung verändern, Amyloidfibrillen destabilisieren und liefern mechanistische Einblicke in die Amyloidresistenz. Die Ergebnisse unterstreichen das Potenzial gezielter Mutationen oder der Unterdrückung aggregationsanfälliger Varianten als therapeutische Strategien bei Amyloiderkrankungen.

Humanes Lysozym ist ein antimikrobielles Protein, das in verschiedenen Körperflüssigkeiten und Geweben vorkommt [31]. Mutationen im Lysozym, insbesondere Punktmutationen, können seine native Faltung destabilisieren und die Fehlfaltung in Amyloidfibrillen fördern, die in Geweben abgelagert werden und eine hereditäre systemische Amyloidose verursachen [32, 33]. Eine solche Mutation, D87G, wurde bei Patienten

identifiziert und mit Amyloidbildung in Verbindung gebracht [33]. In dieser Studie wurde die cryo-EM-Struktur einer patientenabgeleiteten Lysozymfibrille mit der D87G-Mutation analysiert, um zu untersuchen, wie diese Punktmutation die Fibrillenstabilität beeinflusst [34].

Mit Hilfe von MD-Simulationen wurde die Stabilität der patientenabgeleiteten Fibrille mit einem Wildtyp (WT) Fibrillenmodell verglichen, das durch Rückmutation des Glycins an Position 87 in die Wildtyp-Asparaginsäure erstellt wurde. Die patientenabgeleitete Fibrille blieb während der Simulationen strukturell stabil, wie durch niedrigere RMSD Werte und höheren β-Faltblattgehalt angezeigt wurde. Im Gegensatz dazu zeigte die WT-Fibrille stärkere strukturelle Abweichungen, insbesondere in den lösungsexponierten äußeren Ketten und in der Region um Rest 87. RMSF-Analysen zeigten höhere Fluktuationen in diesen Bereichen der WT-Struktur, was darauf hinweist, dass das geladene Asp87-Rest die hydrophobe Fibrillenkernstruktur destabilisiert.

Weitere Analysen der Wechselwirkungsenergien und Wasserstoffbrückenbindungen zeigten, dass die WT-Fibrille stärker mit dem Lösungsmittel wechselwirkt, was die internen Wasserstoffbrückenbindungen verringert. Diese erhöhten Protein-Lösungsmittel-Wechselwirkungen sind mit einer Destabilisierung und teilweisen Entfaltung der Fibrille verbunden, insbesondere an den Rändern. Im Gegensatz dazu behielt die patientenabgeleitete Fibrille mit Gly87 eine engere Packung und geringere Lösungsmittelexposition bei. Dies deutet darauf hin, dass die D87G-Mutation sterisch besser in den Fibrillenkern passt und zudem eine thermodynamisch günstigere Konfiguration begünstigt, indem sie ungünstige elektrostatische Wechselwirkungen reduziert.

Insgesamt zeigt diese Studie, dass Punktmutationen die Protein-Protein- und Protein-Lösungsmittel-Wechselwirkungen innerhalb einer Fibrille tiefgreifend verändern können. Die Ergebnisse unterstreichen die entscheidende Rolle der Sequenzzusammensetzung im Fibrillenkern für die Aufrechterhaltung der strukturellen Stabilität. Darüber hinaus erweitern diese Befunde unser Verständnis der molekularen Mechanismen amyloider Erkrankungen und heben das Potenzial hervor, native Proteinstrukturen durch kleine Moleküle oder Liganden zu stabilisieren, was eine vielversprechende therapeutische Strategie darstellt.

Summary

Nucleic acids serve as carriers of genetic information in all organisms and viruses. Deoxyribonucleic acid (DNA) contains the genetic instructions that govern the development, functioning, reproduction, and maintenance of living organisms. It is transcribed into rybonucleic acids (RNAs), which than encode proteins (coding RNA), regulate gene expression or perform structural and catalytic functions in the cell (non-coding RNA) [1]. An important example of a non-coding RNA is long non-coding RNA (lncRNA). lncRNAs have important roles in gene regulation and are implicated in cancer mechanisms [2]. They are long RNA structures with repetitive elements that are crucial for their function. An lncRNA structure known as human telomeric RNA (TERRA) contains a characteristic guanine-rich sequence r(UAGGG) that forms RNA G-quadruplex (rG4) and is implicated in telomere regulation [3, 4]. G-quadruplexes are non-canonical structures formed by DNA and RNA sequences by non-canonical binding of guanines into planar quartets. DNA G-quadruplexes (dG4s) have been studied more extensively due to the available evidence of their existence *in vivo* and long-lived intermediates that are accessible to the experimental techniques [5, 6]. In contrast, rG4s are highly dynamic in cells and their existence *in vivo* has only recently been supported by experimental evidence [7]. However, interest revolving around research of rG4s has increased once their potential role in important cellular processes like transcription, translation, pre-mRNA processing and telomere maintenance was uncovered [3, 4, 8, 9].

G4s have been investigated using a wide range of experimental techniques. Nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography are among the most commonly used methods for resolving the structures of rG4s and dG4s at atomic resolution. X-ray crystallography also enables visualization of cations within the G4 core, which are essential for their stability [10]. Time-resolved NMR measurements have revealed long-lived intermediates along the dG4 folding pathway [11–13], as well as contrast in folding pathways between the TERRA rG4 and its DNA counterpart [14].

UV-melting and circular dichroism (CD) spectroscopy are used to distinguish G4 topologies based on their characteristic spectral signatures and thermal stabilities. These methods are often combined with techniques aimed at studying G4s *in vivo*, such as live-cell imaging. Recently, a combination of live-cell imaging and CD spectroscopy was employed to observe rG4 folding and unfolding in human cervical cancer cells [7]. However, transient intermediates in the rG4 folding landscape remain difficult to characterize due to their dynamic and short-lived nature.

Smaller timescales are accessible to computational techniques such as molecular dynamics (MD) and Langevin dynamics (LD) simulations. In all-atom MD simulations, each atom is explicitly represented and their dynamics is simulated by solving Newton's equations of motion. The interatomic interactions are represented by force fields, which are parametrized with respect to the atom and molecule type. In this way, MD simulations can resolve dynamics on the microsecond to millisecond timescale. Additionally, more atoms

can be combined in one interaction site, like in coarse-grained simulations, to observe interactions taking place at larger timescales, while still maintaining sufficient amount of detail. Langevin dynamics, a related approach, incorporates stochastic and frictional forces to mimic solvent effects more efficiently, often enabling longer simulations of specific processes.

In the first part of the thesis, I used a three-interaction site (TIS) LD model alongside all-atom MD simulations and CD experiments to study the folding pathway of the TERRA G4. The research was conducted in collaboration with scientists from the group of prof. Harald Schwalbe at the Goethe University in Frankfurt am Main, who provided the CD measurements. I simulated TERRA G4 and G4 formed by plant-derived RNA, as well as few synthetic RNA sequences with varying loop lengths.

Firstly, in order to correct the uniform ionic distribution modeled in the TIS model and account for the ionic double layer formation characteristic of RNAs, the concentration-matching procedure was developed based on all-atom MD simulations. It allowed us to reproduce the experimentally measured fraction of folded states across varying salt concentrations and temperatures. The results from simulations at 25°C revealed that TERRA G4 exhibits a branched folding pathway, with multiple transient intermediates. The folding is entropy-driven and therefore the intermediates are formed by guanine stacking and sequential binding of pairs of guanine repeats. The hairpin, triplex and double-hairpin intermediates that were resolved in the simulations confirm predictions from multiple NMR experiments on dG4s [11–13] as well as experimental work on rG4s [15, 16]. Additionally, simulations of other rG4 systems revealed the same on-pathway intermediates, with populations differing across different sequences. The Gibbs free energy calculations also revealed an interesting property: stability of rG4s decreases with increasing loop length. This finding underscores the importance of entropy in rG4 folding.

Taken together, in this work we were able to resolve the intermediates along the rG4 folding pathway, which are usually inaccessible to experimental methods due to their transient nature. The results also highlighted weaknesses in observable definition while studying the rG4 folding, indicating that simple metrics such as end-to-end distance aren't appropriate to resolve the structurally complex intermediates.

G4s bind cations in their guanosine core which are crucial for their stability. Due to properties like hydration free energy, cationic van der Waals radius and binding free energy to the guanosine O6 atoms, not all ions have the same stabilizing effect on G4s [17]. Therefore, G4s display ion specificity, forming preferentially in the presence of certain monovalent cations such as K^+ or Na^+. This makes G4s attractive targets in the research of artificial ion channels. Specifically, rG4s stand out due to their higher stability in comparison with dG4s [18].

Artificial ion channels are finding application in treating ion channel dysfunction, cancer and anti-bacterial treatment, bio-sensing and nanotechnologies [19, 20]. The main goal of research is to design an ion channel that can be incorporated into the hydrophobic membrane and possess ion recognition sites. Previous NMR and X-ray experiments have shown that G4s coordinate cations in their core binding sites and are able to transport

them directionally [21–24]. Moreover, a study using voltage-clamp measurements and MD simulations was performed, where a dG4 was successfully inserted into a membrane using a stacked lipophylic guanosine, therefore creating an artificial ionophore selective to K^+ ions [25].

In the second part of my thesis, I used MD simulations to study ion conductance properties of a G4 formed by the TERRA molecule. Same molecule was used in this research as in the RNA folding studies from the first part of the thesis. It was simulated in five different systems differing in salt solutions: KCl, NaCl, LiCl, CsCl and NH_4Cl.

Results from straightforward and temperature replica exchange MD (TREMD) simulations confirmed rG4 specificity for K^+ and Na^+ ions. They had the largest average number of ions inside the channel in the straightforward simulations and lowest free energy barriers in free energy profile calculations from TREMD simulations. Additionally, only Na^+ and K^+ ions released all waters from their hydration shells while inside the channel pore. However, the calculations also highlighted differences in binding of these two ions to the rG4. Namely, while the K^+ ion exhibited the highest binding affinity for the central quartets, Na^+ ion spent most of its time at the channel exit and traveled cumulatively through the channel.

The constant-force pulling simulations used to measure the mean first passage time (MFPT) for ionic transport through the G4 pore confirmed the results from MD and TREMD simulations, with Na^+ and K^+ ions exhibiting the lowest MFPTs. However, they revealed an interesting behavior of the MFPTs of Na^+ and K^+ ions. For higher force constants the MFPT of Na^+ ion was lower than that of K^+ ion. This relationship inverted by lowering the force constant, resulting in the lower MFPT for the K^+ ions. More detailed exploration of the MFPTs along the binding sites revealed that Na^+ ions spend most of the time at the channel exit, consistent with the free energy profiles from TREMD simulations, while K^+ ions stay bound at the channel center most of the time. The results for lower force constants are in line with previous research on dG4 ionic conductance, where dG4 demonstrated the highest conductance for K^+ ions [25]. Therefore, it is possible that the overhanging uridines hinder the transport of Na^+ ions at the channel exit, but at higher force constants this effect is not as prominent.

Overall, these simulations provided valuable insights into the differences in ion binding and transport properties of rG4s. They confirmed previous findings demonstrating the specificity of rG4s toward K^+ ions. However, further research is needed to fully resolve the free energy barriers associated with ion transport through the rG4 channel and to better understand the important role played by the overhanging nucleosides in this process.

A third part of my thesis focuses on using MD simulations to study conditions that impact the misfolding of proteins and formation of amyloid fibrils. The research was performed in collaboration with the group of Prof. Marcus Fändrich from the Institute of Protein Biochemistry at the Ulm University.

Proteins are essential for the mechanism of life and they perform a vast array of functions crucial for all cellular processes. Their function is determined by the arrangement of the amino acids in the folded state. Interactions within a protein and between protein

and its surrounding are governed by electrostatic interactions, hydrogen bonds, van der Waals (vdW) and hydrophobic interactions. Mutations and environmental conditions may cause proteins to unfold and expose the hydrophobic part to the solvent. This results in aggregation of proteins that may result in a formation of amyloid fibrils.

Amyloid fibrils are ordered insoluble structures formed by protofilaments that twist around each-other, forming steric zipper motifs. The main building block of protofilaments are β-sheets that stack on each-other by forming hydrogen bonds over backbone amide groups [26]. Cryogenic electron microscopy (cryo-EM) is a powerful technique for studying the amyloid fibrils as it provides high resolution insight into their structure [27]. This structural information can be used in MD simulations to further study the amyloid fibril interactions and stability at the atomistic level.

Amyloid fibril formation starts with nucleation, where ordered β-strands stack to form a fibril nucleus. This process is thermodynamically unfavorable but a formed nucleus than enables faster elongation process [28]. Pre-formed fibril fragments inserted in a solution of monomeric precursor proteins can accelerate the fibril growth process, and propagate the properties of the seed [28]. In the first project, we investigated the structural stability differences between ex-vivo and in-vitro Aβ amyloid fibrils using cryo-EM and MD simulations [29].

The unseeded fibrils demonstrated structural polymorphism with a left-handed conformation. One morphology of the seeded fibril was resolved with a right-handed conformation, and the structural properties of the seed were propagated. Moreover, seeded fibrils demonstrated increased β-sheet content, greater resolution in cryo-EM reconstructions and enhanced resistance to proteolysis, suggesting higher stability of the seeded fibrils.

MD simulations of cryo-EM resolved seeded and unseeded fibril structures at multiple temperatures confirmed that the seeded morphology is conformationally more favorable. Root-mean-square deviation (RMSD) calculations showed that unseeded fibrils underwent greater conformational rearrangements, specifically at the terminal regions. Furthermore, the protonation state of histidine residues significantly affected the structural integrity of the seeded morphology, implicating the role of environmental factors such as pH or ligand binding in fibril structure and stability.

These findings confirm the results from cryo-EM measurements demonstrating the higher conformational stability of seeded fibrils and resolve the molecular-level mechanisms impacting fibril stability.

Second project pertains to structural properties of amyloid resistant fibrils [30]. Amyloid resistance refers to specific mutations that protect the proteins from fibril formation [30]. These naturally occurring mutations allow protection against diseases such as Alzheimer's, kuru, and systemic amyloidosis by disrupting key structural features necessary for fibril propagation. In this study, the molecular basis of amyloid resistance was investigated in apolipoprotein A-II (ApoA-II), a plasma protein implicated in systemic AApoAII amyloidosis. Two natural variants, ApoA-IIC (amyloidogenic) and ApoA-IIF (resistant), differ by four mutations. To identify the structural determinants of resistance, MD simulations were performed on cryo-EM resolved fibrils of ApoA-IIC, a full ApoA-IIF homology model,

and four single-point mutants.

The results from RMSD calculations and interaction energies revealed that the ApoA-IIF variant is less stable than the ApoA-IIC. Of the single-point-mutation models, the one containing Asn62Lys mutation had the lowest conformational stability in comparison to the ApoA-IIC model, which indicates the impact of this mutation on the fibril fold stability. Calculations of interaction energies revealed that the lower stability of ApoA-IIF variant stems from decrease in the interaction energies within the protein.

These results highlight how core-located mutations which alter the local charge destabilize amyloid fibrils, offering mechanistic insight into amyloid resistance. The findings highlight the potential of targeted mutation or suppression of aggregation-prone variants as therapeutic strategies in amyloid diseases.

Human lysozyme is an antimicrobial protein found across various bodily fluids and tissues [31]. Mutations in lysozyme, particularly point mutations, can destabilize its native fold and promote misfolding into amyloid fibrils, which are deposited in tissues and cause hereditary systemic amyloidosis [32, 33]. One such mutation, Asp87Gly, has been identified in patients and is associated with amyloid formation [33]. In this study, the cryo-EM structure of a patient-derived lysozyme fibril containing the Asp87Gly mutation was analyzed to investigate how this point mutation affects fibril stability [34].

Using MD simulations, the stability of the patient-derived fibril was compared to a wild-type (WT) fibril model created by back-mutating the glycine at position 87 to the wild-type aspartic acid. The patient-derived fibril remained structurally stable throughout the simulations, as indicated by lower RMSD values and higher β-sheet content. In contrast, the WT fibril showed increased structural deviations, particularly in the solvent-exposed outer chains and in the region surrounding residue 87. RMSF analysis revealed higher fluctuations in these areas for the WT structure, suggesting that the charged Asp87 residue destabilizes the hydrophobic fibril core.

Further analysis of interaction energies and hydrogen bonding revealed that the WT fibril interacts more strongly with solvent, reducing the internal hydrogen bonding. These increased protein-solvent interactions are associated with destabilization and partial unfolding of the fibril, particularly at the edges. In contrast, the patient-derived fibril, with Gly87, maintains tighter packing and less solvent exposure. This suggests that the Asp87Gly mutation fits sterically into the fibril core and also promotes a thermodynamically favorable configuration by reducing unfavorable electrostatic interactions.

Overall, this study demonstrates that point mutations can significantly alter the protein–protein and protein–solvent interactions within a fibril. The results underscore the critical role of sequence composition in the fibril core for maintaining structural stability. More broadly, these findings advance our understanding of the molecular mechanisms underlying amyloid diseases and highlight the potential of stabilizing native protein structures with small molecules or ligands as a promising therapeutic strategy.

Chapter 1

Introduction

1.1 Nucleic acids structure and function

Nucleic acids store and transmit genetic information in all organisms and viruses. Deoxyribonucleic acid (DNA) carries the hereditary instructions essential for the development, functioning, and reproduction of all living organisms. In prokaryotes, DNA is located in the cytoplasm, while in eukaryotes, it is organized into chromosomes within the nucleus.

The structure of DNA was first described by Watson and Crick [35], building on the foundational work of Franklin and Wilson [36, 37]. DNA consists of two antiparallel polynucleotide strands twisted into a double helix (Figure 1.1 (A)). This structure enables accurate replication during cell division and the transmission of genetic material across generations [38]. DNA's counterpart, ribonucleic acid (RNA), is typically single-stranded and folds upon itself to form helices (Figure 1.1 (B)). Both DNA and RNA are composed of nucleotides that contain a phosphate group, a sugar moiety, and a nucleobase (Figure 1.1 (C)). Nucleotides are linked via phosphodiester bonds between the 3′ carbon of one sugar and the 5′ carbon of the next. DNA contains four types of nucleotides: adenine (A), guanine (G), cytosine (C), and thymine (T), while in the RNA structure, T is replaced by uracil (U). Complementary base-pairing via canonical Watson-Crick hydrogen bonds, like A-U (T) or G-C, stabilizes the helical structures of nucleic acids (Figure 1.1 (C)).

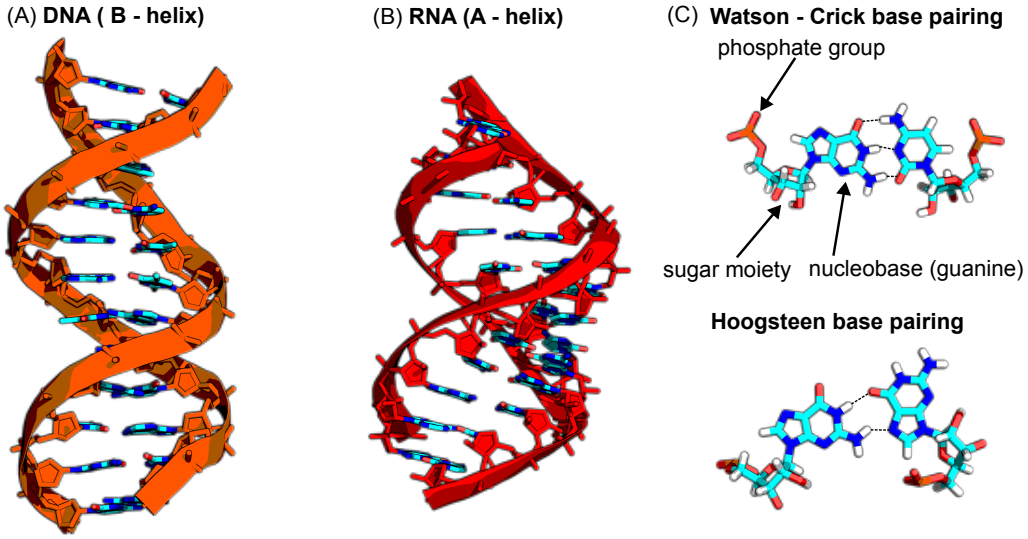

Figure 1.1: (A) Double-stranded helical representation of a DNA structure in a B-helix conformation. (B) Double-stranded helical representation of an RNA structure in an A-helix conformation. (C) Upper panel represents a canonical Watson-Crick base pair of guanine (G) and cytosine (C) bases in a DNA. Lower panel represents a non-canonical Hoogsteen base pair consisting of two guanine (G) bases in an RNA.

DNA is more chemically stable than RNA, making it better suited for long-term genetic information storage. According to the central dogma of molecular biology, DNA is transcribed into RNA, in this case messenger RNA (mRNA), which then directs the synthesis of proteins [1]. However, not all RNA molecules encode proteins. A large class of non-coding RNAs (ncRNAs) plays crucial roles in various cellular processes. Key examples include transfer RNA (tRNA) and ribosomal RNA (rRNA), which are essential for protein synthesis, microRNAs (miRNAs), which regulate gene expression post-transcriptionally, and small interfering RNAs (siRNAs), which mediate gene silencing through RNA interference [39]. Another important class, long non-coding RNAs (lncRNAs), are involved in gene regulation and have been linked to cancer mechanisms [2]. These lncRNAs are typically longer than 200 nucleotides and often possess modular structures rich in repetitive elements, which are increasingly recognized as critical for their function [40].

Due to their highly negatively charged phosphate backbone, interactions with cations are crucial for nucleic acids to overcome a significant electrostatic barrier and attain functional structures. Nucleic acid-ion interactions are governed by the ionic atmosphere, an ion distribution around nucleic acids formed by cations under the influence of a nucleic acid's electrostatic field. The ionic atmosphere consists of condensed ions, which bind directly to nucleic acids, and diffuse ions [41]. A diffusive layer is characterized by the accumulation of mobile cations and the depletion of anions. It is based on long-range electrostatic interactions between nucleic acids and ions and contributes to the thermodynamic stability of nucleic acids and their folding dynamics. Condensed ions are positioned at distinct binding sites on nucleic acids and facilitate the formation of functional nucleic acid structures.

1.1.1 RNA G-quadruplexes

G-quadruplexes (G4s) are formed by guanine-rich DNA and RNA sequences and are implicated in many important cellular processes such as transcription and translation regulation [3, 4] or pre-mRNA processing [8, 9]. Moreover, G4 structures are formed in a lncRNA called TERRA, a human telomeric RNA structure. Telomeres are specialized structures that safeguard chromosomal DNA from gradual degradation. The transcription of the C-rich strand at the telomere generates TERRA, which contains a characteristic G-rich sequence, r(UUAGGG) [42, 43], and is implicated in having an important role in telomere regulation [44]. Telomere dysfunction is associated with cell aging and cancer [45, 46], which has led to growing scientific interest in designing small drug molecules that target TERRA [47].

The structure of G4s differs from the standard helical structure, as they are formed by the non-canonical hydrogen bonding of guanines into planar quartets. These non-canonical hydrogen bonds are known as Hoogsteen bonds (Figure 1.1 (C)). Two or more guanine quartets stack on top of each other, interconnected by loops, to create the characteristic quadruplex structure (Figure 1.2 (A)). DNA G4s (dG4s) can have various loop types, depending on the length of G-tracts, base sequence, length of G4 loops, type of cations in the medium, and even the guanosine glycosidic angle. Guanine glycosidic angle is defined by atoms O4′-C1′-N9-C4. A syn-angle spans values between -90° and 90°, while the anti-

conformation defines glycosidic angles between 90° and 270° (Figure 1.2 (B)). RNA G4s (rG4s), as opposed to dG4s, only form propeller loops and are parallel, which means that the 3′ and 5′ ends of the strand are on opposite sides [48]. This is due to the additional 2′ hydroxyl group in the ribose sugar that adjusts the sugar puckering conformation of the nucleosides and forces the guanosine glycosidic angle into solely the anti-conformation [17]. It is also this hydroxyl group that makes rG4s more stable than their DNA counterparts, as it contributes to a larger number of hydrogen bonds within a molecule, as well as water coordination, which stabilizes the structure in the aqueous environments.

The quartet structure of G4s brings partially negatively charged O6 atoms within close proximity. Consequently, G4s are primarily stabilized by cations that bind to specific sites within the G4 channel located on top of a G-quartet or in-between two quartets, thereby screening the partial charge of guanine O6 atoms. G4s can be stabilized by either divalent or monovalent cations, although most studies focus on monovalent cations. This preference is due to their physiological relevance and because the higher charge density of divalent cations causes stronger electrostatic repulsion within the GQ channel, which can leave binding sites between cations unoccupied [17].

Binding affinity of cations to guanine O6 atoms follows the reverse Hofmeister series, namely $Mg^{2+} > Ca^{2+} > Sr^{2+} > Li^{+} > Ba^{2+} > Na^{+} > K^{+} = Cs^{+}$ [49]. However, the order in which cations stabilize the G4 structure is $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$ and $K^{+} > Na^{+} > Li^{+} = Cs^{+}$ [17]. This emphasizes that the ion specificity of G4 structures arises from the interplay of multiple factors, rather than being determined by a single property. Cations differ in size, hydration free energy, and binding affinity to the guanine O6 atoms. As a result, they differ in their ability to screen the guanine O6 charge and stabilize the G4 structure, which explains the ion specificity observed in G4s.

Different structural characteristics of rG4s and dG4s result in different folding mechanisms [14, 50, 51]. Due to many possible intermediate structures, the folding of dG4s is referred to as a kinetic partitioning mechanism, as opposed to the fast funnel-like folding of proteins [51, 52]. In a funneling free-energy landscape, a molecule folds toward a single free-energy minimum by gradually increasing the number of native contacts and decreasing the entropy. A kinetic partitioning folding landscape consists of multiple free energy minima separated by high barriers. Each minimum represents a local intermediate state. Therefore, dG4s remain in the local minima for longer times, making the intermediates easier to resolve by experimental methods like nuclear magnetic resonance (NMR) spectroscopy. Numerous experimental studies have been conducted to elucidate the folding intermediates of dG4s, highlighting key structural features such as hairpins, triplexes, double-hairpins, and strand-shifted conformations, all of which play critical roles in their respective folding pathways [11–13]. Information from experiments on the interactions of rG4s with ions and their folding pathways are sparse. rG4 folding mechanism resembles the funneling folding pathway, because rG4s can only form a parallel-stranded propeller structure [14]. Therefore, its intermediate states are transient and cannot be captured by high-resolution experimental techniques.

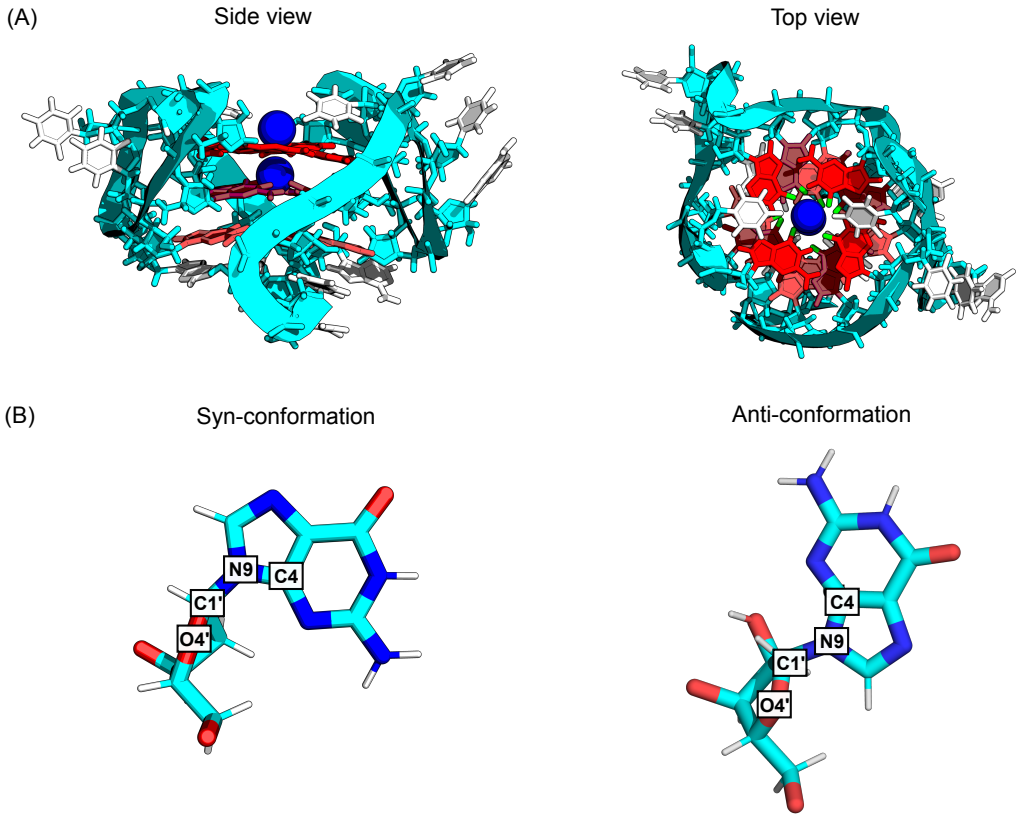

Figure 1.2: (A) RNA G-quadruplex structure resolved in an NMR solution (2KBP) [53] with two K^+ ions inside the central pore from the side and top-view. Three different guanine quartets are indicated in different shades of red. Guanine O6 atoms are indicated in green. (B) Syn- and anti-conformations represented for guanine nucleobase. Syn- and anti- are two different conformations of the glycosidic χ angle defined by atoms O4′-C1′-N9-C4, which are indicated in the picture.

The most widely used method for studying G4s is NMR spectroscopy [54–57]. It can be used to gain a detailed insight into the structure of G4s and their interactions with NMR-active ions such as NH_4^+ at an atomic level. NMR works by applying a strong magnetic field to a sample, which causes nuclei with non-zero spin (commonly ^{1}H or ^{13}C) to resonate at characteristic frequencies. By analyzing these resonances, researchers can gather detailed information about the chemical environment, bonding, and spatial arrangement of a molecule. Additionally, time-resolved NMR experiments can be used to elucidate folding pathways of dG4s [11–13]. Moreover, these experiments recently revealed that the K^+-induced folding kinetics of TERRA is an order of magnitude faster ($k_1 = 1.45/\text{min}$) compared to the dG4 of an identical nucleotide sequence ($k_1 = 0.41/\text{min}$) [14].

X-ray crystallography is another powerful technique for acquiring detailed structural information on G4s. The structures are resolved from electron densities provided by X-ray diffraction of the crystallized molecule's atoms. Crystallographic data derived from these experiments also provide information about the solvent properties around the G4s, including the ions that bind in the central channel [10]. For example, X-ray studies provided more insight into the array of hydrogen bonds that stabilize rG4s [58] as well as the interactions of dG4s with cations [59].

Circular dichroism (CD) and UV-melting analysis are used to gain insight into the

conformational changes of G4s, as well as thermodynamic properties [10, 60]. These techniques are fairly inexpensive and are usually used as complementary to NMR and X-ray diffraction studies [13, 61]. CD studies are also used in combination with cell imaging studies, in an effort to elucidate the existence of G4s *in vivo* [62].

The single-molecule Fröster resonance energy transfer (smFRET) measurement technique can resolve the distance within a molecule by measuring the energy transfer between a donor and an acceptor dye. This technique is time-sensitive and can be used to resolve conformational changes of a molecule within the changing environment. For example, smFRET technique has been used to study intermediate states of rG4 folding, due to its higher time sensitivity in comparison to other experimental techniques [15, 16]. However, the large size of the fluorophores limits the spatial resolution of the measurement.

Live cell imaging can be used to resolve rG4s in cells [63]. It is a powerful technique used to observe and study living cells in real time, providing insights into dynamic cellular processes. It allows researchers to visualize cellular structures, interactions, and functions without disrupting the natural physiological state of the cells. Various microscopic techniques are applied, one of which is Fluorescence Microscopy [64]. Here, fluorescent dyes or proteins are used to label specific cellular components. This technique was used to observe the rG4 folding and unfolding in human cervical cancer cells [7].

Finally, different experimental techniques can serve as a benchmark for parameterizing models and theoretical approaches, allowing detailed insight into the conformational dynamics of biomolecules and their interactions. For example, to gain an insight into the interaction of rG4s with ions and their folding mechanisms, molecular dynamics (MD) and Langevin dynamics (LD) simulations are typically employed in order to explore these mechanisms with high temporal and spatial resolution. These simulations make it possible to observe how ions influence the folding, stability, and dynamics of G4s [65, 66], while also revealing the role of solvent interactions and the formation of specific ion-binding sites that are crucial for the proper folding and function of these structures [67]. MD simulations, in particular, provide atomic-level resolution of the different conformational changes that rG4s can adopt, elucidating the ion-dependent structural rearrangements. On longer timescales, coarse-grained (CG) Langevin simulations can be used to explore folding pathways and identify intermediate folding states, providing a more comprehensive view of rG4 folding dynamics and their structural evolution. More elaborate discussion of these simulation methods is provided in Chapter 2 of this thesis.

1.2 Protein structure and misfolding

Proteins are ubiquitous macromolecules essential for the mechanisms of life that perform a vast array of functions crucial for the cellular processes. The vast functionality of proteins stems from their complex structures formed by a combination of many different amino acid types. The sequence of amino acids forms the primary structure of a protein. There are 20 different types of amino acids that can be combined to form protein structures, so for a protein consisting of N amino acids, there are 20^N different possible protein sequences. Amino acids form proteins via peptide bonds formed between the carboxyl group of one

amino acid and the amino group of another. Amino and carbonyl group, connected by a carbon atom called C_α, form the backbone of proteins (Figure 1.3 (A)). Amino acids differ based on their side chains, which influence their physicochemical properties, such as electrostatic charge and shape. These properties, in turn, affect non-covalent binding and protein folding. Amino acids can be positively or negatively charged, as well as neutral, and the final protein's three-dimensional structure attains the optimal distance between amino acids that repel or attract each other [68]. Order of amino acids in the protein chain also affects the pattern of mutually attractive and repulsive forces that guide the folding process. Due to differences in mutual interactions within a protein and with the surrounding molecules, small differences in protein sequence between two proteins can result in completely different functions.

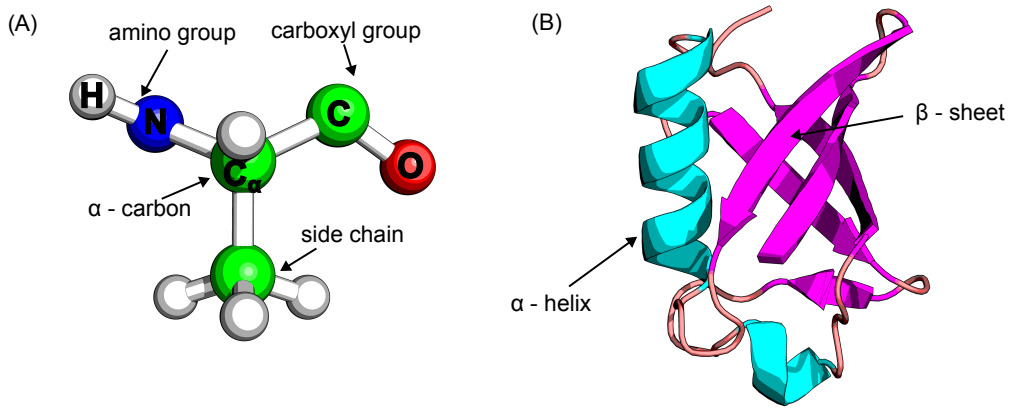

Figure 1.3: (A) Structure of the alanine (ALA) amino acid with its main building blocks labeled. (B) Example of a ubiquitin protein illustrating an α-helix and a β-sheet.

Beyond electrostatic interactions, hydrogen bonds are the most prevalent type of interaction in proteins. Hydrogen bonds are formed between two electric dipoles, more specifically a hydrogen atom attached to an electronegative donor (e.g., an amino group) and an electronegative acceptor (e.g., a carbonyl oxygen). Such hydrogen bonds are crucial for stabilizing protein structures, particularly secondary structures like α-helices and β-sheets (Figure 1.3 (B)). In α-helices, hydrogen bonds are formed by the interaction between a carbonyl oxygen of one amino acid and the amide hydrogen of another amino acid four residues apart. This regular pattern extends along the peptide chain and provides the well-known helical shape with remarkable stability. Similarly, in β-sheets, adjacent polypeptide strands are held together by the hydrogen bonds connecting their backbone atoms, forming sheet-like structures. The connected strands can extend in opposite or the same directions, making the formed sheets parallel or anti-parallel. Hydrogen bonds also play roles in stabilizing tertiary and quaternary structures.

Aside from hydrogen bonds, van der Waals (vdW) interactions play an important role in protein folding and stability. These interactions are formed between induced or temporary dipoles and are strongly dependent on the distance between the atoms and molecules. VdW interactions also stabilize molecules that cluster due to hydrophobic interactions, thereby maintaining the hydrophobic core.

Hydrophobic interactions within a protein are a result of the properties of water. Water molecules possess an electric dipole resulting from a partial negative charge of oxygen atoms and a partial positive charge of hydrogen atoms, and they bind to each other to form a network of hydrogen bonds. The network is dynamic, as individual molecules can attach and detach easily from it, maintaining a high entropy. The insertion of a non-polar solute results in a formation of ordered, cage-like water structures, resulting from water molecules binding to each other around the solute, effectively lowering the entropy of the system. Hydrophobic interactions drive the formation of a tightly packed core, shielding nonpolar residues from water and exposing the hydrophilic polar residues. The interaction energy between the polar residues and water molecules compensates for the drop in entropy.

1.2.1 Misfolding of proteins: amyloid fibrils

After synthesis, proteins fold to adopt their native conformation and fulfill their biological function. However, certain mutations or environmental conditions cause proteins to misfold, exposing the hydrophobic parts to the solvent [26]. This leads to increased propensity of proteins for aggregation and can result in the formation of ordered insoluble structures, called amyloid fibrils. Amyloid fibril formation causes a severe disease, called amyloidosis. Most dominant examples of diseases related to amyloid fibrils are neurodegenerative diseases such as Alzheimer's or Parkinson's disease as well as systemic amyloidosis, type II diabetes, and atrial amyloidosis [69].

Amyloid fibrils form ordered cross-β structures, in which β-strands stack in the direction perpendicular to the fibril axis, as a result of hydrogen bonding of backbone amide groups C=O and N-H (Figure 1.4). The resulting structures are called β-sheets which form the main building blocks of amyloid fibrils, the protofilaments. Multiple protofilaments extend on top of each other to form fibrils. The resulting fibrils can contain one or more horizontally layered protofilaments twisting around each other, forming steric zipper motifs [26]. Understanding the molecular structure and dynamics of amyloid fibril protofilaments is crucial for developing therapeutic strategies to prevent or disrupt the formation of these toxic aggregates and potentially treat or slow the progression of associated diseases.

The first amyloid fibril structures were resolved by solid state NMR spectroscopy (ssNMR) [70]. Unlike solution-state NMR, ssNMR is optimized for insoluble and non-crystalline samples and provides valuable insights into the secondary and tertiary structures of amyloid fibrils. Cryo-electron microscopy (cryo-EM) is used to provide high resolution insight into the structure of amyloid fibrils, and it is instrumental in amyloid fibril research. It is based on cooling the sample to cryogenic temperatures in order to perform transmission electron microscopy (TEM) experiments. TEM uses a focused beam of electrons that interact with the sample, producing images with a resolution up to a nanometer [27]. Recently, atomic force microscopy (AFM) has also emerged as a powerful method to study structural, morphological, and mechanical properties of amyloid fibrils at a level of individual aggregates. AFM is a single molecule measurement method that uses a cantilever probe with a microfabricated sharp tip at its end to scan the interaction forces between the tip and the sample and produce the 3D morphology of the sample [71].

MD simulations can provide meaningful information on amyloid fibril stability and interactions at the atomic level. Paired with free energy calculation methods, they can provide insights into the impact of point mutations on the stability of the fibrils [72]. Such analyses can reveal how mutations alter fibril stability, disrupt or reinforce β-sheet stacking, or influence intermolecular interactions. Additionally, MD simulations help predict aggregation propensity by assessing conformational flexibility, hydrophobicity, and the likelihood of fibril fragmentation or extension [73]. These insights are valuable for understanding disease-related amyloid formations and guiding rational drug design, where inhibitors or stabilizers can be tested computationally before experimental validation.

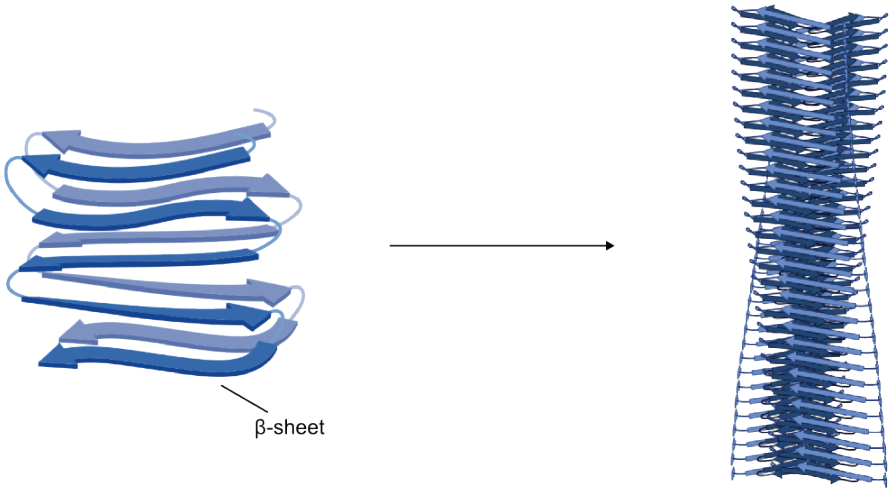

Figure 1.4: Schematic presentation of an amyloid fibril structure. Multiple protein chains form arched β-strands that stack in the direction of the fibril axis creating ordered structures called protofilaments (left). Protofilaments are the main fibril building blocks, which twist around each other to form fibril structures (right). Figure is generated with BioRender.com.

1.3 Outline of the thesis

Computational simulations provide an insight into molecular behaviors and transient states that remain challenging to characterize experimentally. In this thesis, I used MD and LD simulations to resolve the rG4 folding pathway, as well as molecular principles guiding the interactions of rG4s with ions and misfolding of proteins.

Chapter 2 introduced the methodology used troughout the thesis. It gives the theoretical overview of the simulation methodology, focusing on the MD and LD simulations.

Chapter 3 describes application of MD and LD simulations to resolve the folding pathway of rG4. It emphasizes the critical role of combining computational simulations with experimental data to uncover the physical determinants of rG4 folding. Intermediate states on the rG4 folding pathway are discussed as well as an important role of entropy in the process. Additionally, matching procedure is introduced to relate the ionic concentrations in implicit-solvent CG simulations to the local ionic environment surrounding rG4.

Chapter 4 explores rG4 as a model RNA-based ion channel using MD simulations. The study demonstrates strong K^+ selectivity, while Na^+, Li^+, Cs^+, and NH_4^+ show distinct

transport patterns influenced by ion size, hydration, and interactions with guanine O6 atoms. Free energy and pulling simulations reveal a force-dependent conduction mechanism, highlighting the importance of near-physiological conditions.

Chapter 5 investigates the structural determinants and molecular mechanisms of amyloid fibril formation using cryo-EM and MD simulations. The studies reveal how seeding, point mutations, and sequence variations influence fibril morphology, conformational stability, and amyloid resistance in Aβ, ApoA-II, and lysozyme fibrils. These findings provide mechanistic insight into amyloid-associated diseases and may serve as a guidance for rational design of therapeutic strategies targeting protein misfolding.

Chapter 2

Computational Methods

While the structural insight from experiments can provide a lot of information on the function of biomolecules, some of the most important functional properties are contained in the dynamical motion. Dynamics of biomolecules can be simulated using different types of simulations.

MD simulations allow us to observe the dynamical motion of biomolecules with atomistic spatial resolution and femtosecond temporal resolution. They are utilized to connect the impact of molecular motion to biological functionality of biomolecules and measure the equilibrium properties. However, all-atom MD simulations are computationally costly and are not able to observe long scale phenomena such as protein and nucleic acid folding or ligand binding. Enhanced sampling techniques such as temperature replica exchange (TREMD) MD simulations can be used to overcome these limitations and still be able to observe molecular motion in great detail. Alternatively, coarse-grained simulations group multiple atoms into one interacting particle, which results in saving computational time allowing for simulations of long-scale phenomena while still maintaining an appropriate temporal and spatial resolution. In my thesis, I used MD and coarse-grained (CG) LD simulations to study the thermodynamical properties of rG4s and different types of amyloid fibrils.

2.1 Molecular dynamics simulations

MD simulations can be used to gain insight into thermodynamical properties of systems. Each particle of the system is modeled using the classical approximation, and therefore the dynamics of particle over time is observed by solving the Newton's equation of motion. In this chapter, I will introduce the basic concepts of MD simulations, guided by the comprehensive outlook in the book by Frenkel and Smit [74].

2.1.1 Newton's equations of motion

MD simulations are often called computational experiments, since they follow a procedure analogous to laboratory experiments. The procedure consists of preparing the system we wish to simulate, equilibrating it, and then measuring the average of the observable. In order to initiate an MD simulation of a system, we require initial coordinates and velocities. Initial coordinates are usually 3D structures resolved by an experimental technique such as NMR, X-ray crystallography or, cryo-EM. Initial velocities are drawn from Maxwell-Boltzmann equation at a desired temperature T:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} \exp\left(-\frac{m_i v_i^2}{2k_B T}\right) \tag{2.1}$$

Trajectories are calculated over time by solving Newton's equations of motion for a system of N particles:

$$\frac{d^2\vec{r_i}}{dt^2} = \frac{\vec{F_i}}{m_i}, \qquad i = 1, ..., N. \tag{2.2}$$

Here F_i represents the force acting on particle i due to the potential energy $U(\vec{r_i})$:

$$\vec{F_i} = -\nabla U(\vec{r_i}). \tag{2.3}$$

Potential energy $U(\vec{r_i})$ is the result of bonded and non-bonded interactions of particle i with all other particles in the system.

Equations of motion for a system of N particles, which in all-atom MD simulations represent atoms constituting biomolecules, are solved numerically. Different algorithms, also known as integrators, have been created for this purpose. Verlet-like algorithms are widely used due to their reversibility, and small long time energy drift and memory requirements. The Verlet algorithm is derived by summing up forward and backward Taylor expansion of the atomic position around time t:

$$\vec{r_i}(t + \Delta t) = \vec{r_i}(t) + \vec{v_i}(t)\Delta t + \frac{\vec{F_i}(t)}{2m_i}\Delta t^2 + \frac{\Delta t^3}{3!}\dddot{\vec{r}}_i + O(\Delta t^4) \tag{2.4}$$

and

$$\vec{r_i}(t - \Delta t) = \vec{r_i}(t) - \vec{v_i}(t)\Delta t + \frac{\vec{F_i}(t)}{2m_i}\Delta t^2 - \frac{\Delta t^3}{3!}\dddot{\vec{r}}_i + O(\Delta t^4). \tag{2.5}$$

The resulting summation yields a new atomic position that depends on the previous and current position, but not on the velocity of the atom:

$$\vec{r_i}(t + \Delta t) = 2\vec{r_i}(t) - \vec{r_i}(t - \Delta t) + \frac{\vec{F_i}}{m_i}\Delta t^2. \tag{2.6}$$

Therefore, other Verlet-like algorithms are implemented that employ velocity during position updates. Such an algorithm is the Leapfrog algorithm, used as the default in the Gromacs package. Here, new positions are calculated using velocities evaluated at a half time step:

$$\vec{r_i}(t + \Delta t) = \vec{r_i}(t) + \Delta t\vec{v_i}(t + \frac{1}{2}\Delta t). \tag{2.7}$$

The time step in the MD simulation must be chosen carefully. Larger time step acceler-

ates simulations but decreases accuracy. Namely, the time step should be short enough to resolve the fastest motions in the system, which in the biological systems are typically the vibrations of the O-H bonds, with a period of 1 fs. This would restrict the size of the time step to 1 fs. However, in most cases the bonds with hydrogens are constrained to allow for larger time step and more efficient simulations, resulting in a timestep of 2 fs being the most frequently used in MD simulations.

2.1.2 Initiating an MD simulation

In order to perform an MD simulation, the biomolecule has to be placed in a box of desired shape. Since the initial coordinates are acquired from experiments, initiating the simulation directly from them may yield very high potential energy values due to atomic overlaps. Therefore, it is necessary to perform energy minimization to remove steric clashes. Subsequently, the system is equilibrated, and the simulation is performed under thermodynamic equilibrium conditions.

The use of MD simulations generates the microcanonical ensemble. However, most biological processes occur in the canonical (constant number of particles N, temperature T and volume V) or isothermal-isobaric (constant number of particles N, temperature T and pressure P) ensembles. Therefore, it is necessary to incorporate this in the simulations by coupling the system to an external temperature or pressure bath.

To prepare the system at a particular temperature, the fast, weak-coupling Berendsen thermostat is commonly employed. The system is coupled to an external heat bath at temperature T_0. Instantaneous temperature T is corrected according to the following equation:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}, \tag{2.8}$$

where τ represents the temperature coupling constant. Atomic velocities are adjusted according to the new temperature after the integration step. Berendsen thermostat suppresses fluctuations in kinetic energy and therefore does not reproduce the Boltzmann ensemble. Thus, it is not recommended for production runs aimed at measuring equilibrium properties. Nosé-Hoover thermostat samples the canonical ensemble by introducing an additional degree of freedom s associated with an external system into the Hamiltonian:

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\vec{p_i}^2}{2m_i s^2} + \mathcal{U}(\vec{r}^N) + \frac{p_s^2}{2Q} + g k_B T \ln s, \tag{2.9}$$

where Q is the effective mass related to s and g denotes the number of degrees of freedom in the system.

Pressure coupling is performed by changing the volume of the simulation box. This is done similarly to temperature coupling, by coupling the system to a pressure bath. Berendsen pressure coupling method corrects the instantaneous pressure to match the required pressure, analogously to the Berendsen thermostat in Eq. 2.8 by rescaling the

box coordinates. It can be used to equilibrate the system to the desired pressure but does not reproduce the NPT-ensemble. Therefore, the Parrinello-Rahman method is employed which adjusts the Hamiltonian similarly as Nose-Hoover temperature coupling scheme.

2.1.3 Force fields for biomolecules

Interaction potentials between atoms determine the force acting on each atom and therefore the dynamics of the system, according to Eq. 2.2. Interaction potentials are defined by functional forms with parameters depending on the biomolecule being simulated. These functions are known as biomolecular force fields and represent a classical approximation of interatomic interactions (Figure 2.1).

The potentials are defined by bonded and non-bonded terms:

$$U(\vec{r}^N) = U_{\text{bonded}}(\vec{r}^N) + U_{\text{non-bonded}}(\vec{r}^N). \tag{2.10}$$

U_{bonded} describes interactions between atoms that are connected by covalent bonds. These include bond stretching, angle bending, torsional, and improper dihedral interactions, typically approximated by harmonic and torsional potentials:

$$U_{\text{bonded}} = \frac{1}{2}\sum_{\text{bonds}} k_r(r - r_{\text{eq}})^2 + \frac{1}{2}\sum_{\text{angles}} k_\Theta(\Theta - \Theta_{\text{eq}})^2 + \frac{1}{2}\sum_{\text{dihedrals}} k_\phi(1 + cos(n\psi - \psi_0)) + \frac{1}{2}\sum_{\text{improper}} k_\Psi(\Psi - \Psi_{\text{eq}})^2.$$

The non-bonded interactions describe electrostatic interactions between charged atoms and van der Waals (vdW) interactions:

$$U_{\text{non-bonded}} = \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.11}$$

Calculation of forces is the most computationally demanding part of an MD simulation. Due to their computational cost, MD-accessible systems are much smaller than experimental ones and far below the thermodynamic limit. Therefore, the results always depend on the number of particles simulated and are susceptible to finite size effects. To overcome these limitations, MD simulations usually incorporate periodic boundary conditions (Figure 2.2). This means that the simulation box is replicated infinitely in all directions, removing the surface effects and approximating bulk behavior. Each particle can move beyond the edge of the box and re-enter through the opposite side. Since MD simulations are performed at constant particle number, a periodic image of that same particle enters the box from the other side. Particles interact with others in neighboring boxes, including their own periodic images. As a result the interaction potential becomes an infinite sum of pair interaction terms, which is not feasible computationally. However, for short-range interactions, it is common to truncate interactions beyond a cutoff distance r_c.

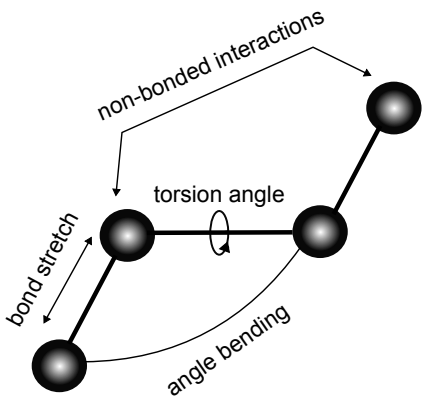

Figure 2.1: Schematic representation of the interaction potential within a molecule: non-bonded interactions and bonded interactions which include bond stretching, angle bending and torsion angle.

This method is applied to Lennard-Jones potential because it decreases rapidly with distance. The standard procedure is to truncate and shift the potential at r_c to avoid discontinuities and infinite forces, adding a correction term for the truncated tail. The neighbor list of atoms within distance r_c is updated only every few steps. To account for this, a larger distance r_v ($r_v > r_c$) is used for the neighbor list, implemented as the Verlet cutoff in the Gromacs package used in this thesis.

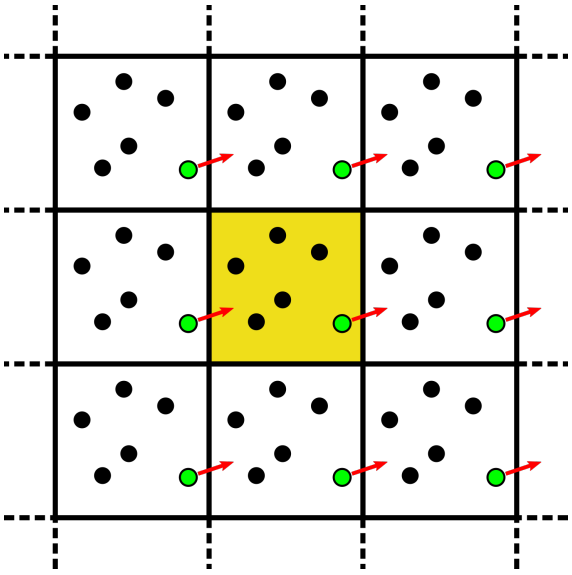

Figure 2.2: A schematic representation of periodic boundary conditions. The simulation box is shown in yellow color and its copies are in white. The green particle exits the box on the right side and at the same time reenters the box from the left side, conserving the constant number of particles in the box.

Coulomb potential, on the other hand, can't be truncated because it doesn't decrease rapidly with distance. Therefore, the Particle Mesh Ewald (PME) method is typically used, dividing the potential into a directly computed short-range part and a Fourier-based long-range part. It utilizes the periodic boundary conditions and Fourier transforms to approximate the interaction of a point charge q_i with the electrostatic potential of

surrounding charges. The distributions of surrounding charges are mapped onto the mesh points, therefore allowing the calculation of the potential at a specific position in a box without having to take into account the positions of all other particles.

The potential energy terms describing the interactions between atoms are almost always the same, however the parameters differ depending on the macromolecule type and environment. These parameters are called force fields. For example, simulating RNA in an ionic solution requires force fields for RNA, water, and ions.

In order to correctly describe RNA dynamics, an RNA force field has to capture all of its important aspects like hydrogen bonding, stacking, glycosidic torsional angle χ rotation, and backbone conformational sampling. Since RNAs are highly charged molecules, a good force field must also properly describe the electrostatic effects, as well as interactions of RNA with surrounding ions in the solution and water [75].

Polarizable force fields include the effect of electronic polarization in the description of interatomic interactions and have recently become more popular for nucleic acid simulations. Most popular polarizable force fields are AMOEBA and Drude oscillator force fields. However, they introduce additional degrees of freedom and are therefore computationally more expensive than non-polarizable force fields. Since non-polarizable force fields remain sufficiently accurate, they are still widely used and validated for diverse nucleic acid structures. Two of the most commonly used groups of non-polarizable force fields are CHARMM and AMBER. In this thesis, I used AMBER force field for RNA simulations, so further discussion will be focusing on that force field family.

AMBER force fields all stem from the parametrization by Cornell et al. (1995) [76], which defined partial charges by fitting the electrostatic potential to the quantum mechanical (QM) electrostatic potential (ESP) around the molecular compounds [77]. These charges were later optimized by the restrained ESP method, which uses hyperbolic restraints to reduce the electrostatic artifacts due to unnaturally large charges [78].

Increased availability of computational power led to longer simulations which revealed the deficiencies of the original AMBER force field ff94 and its successors, ff98 and ff99 [79]. Namely, these force fields over-stabilized the unfavorable α/γ = gauche+,trans state of the backbone torsion angles, resulting in deprivation of the B-DNA structure. Subsequent optimization, named parambsc0 optimized the α and γ torsional angles by fitting the parameters of the potential energy function to high-level quantum mechanical data [80]. Despite still maintaining deficiencies, like underestimation of the helical twist and overpopulation of γ=trans state, it has remained the state-of-the-art model for simulating B-DNA. The same bias towards γ=trans state did not affect RNAs because the transition was reversible. This suggests that force fields shouldn't be considered transferable across different nucleic acid types and highlights the need for separate parametrization of RNA and DNA force fields. [79]

Improvement upon bsc0 force field was implemented in the bsc1 force field, which optimized the sugar puckering and χ, ξ, ζ dihedrals, showing very good agreement with experimental results including for the simulations of non-canonical DNAs [81].

In case of RNA force fields, challenge lays in describing the non-canonical structures

which have a prevalent role in biological functions of RNAs. The chiOL improvement of the bsc0 force field focuses on the glycosidic torsion angle χ [79]. χ affects many structural features of nucleic acids as it describes the rotation around bonds connecting base and sugar residues. chiOL reparameterization improved the sampling of native RNA structures and complex RNA folds by removing the "ladder-like" structural artifact and correctly sampling the traditional A-RNA conformation.

AMBER force fields were also used in this work for simulation of amyloid fibril structures. Amyloid fibrils are formed by small proteins or peptides which have a β-strand secondary structure and form β-sheets. Therefore, in order to simulate amyloid fibrils with sufficient accuracy, a proper balance between α and β structure is required. This can be attained by parametrizing protein torsion angles.

The improvements of protein force fields is again based on the original ff94 AMBER force field [76], which upon increasing simulation timescales immediately revealed its deficiency in simulating α-helical region, present in most of the protein structures [82]. An improvement was made in ff96 [83] and ff99 [84] force fields by optimizing ϕ and ψ backbone dihedral angles, resulting in an improper bias towards helical conformations. ff99sb [82] version vas built upon ff94 and ff99 force fields, using quantum-level data to improve the sampling of secondary structure elements. Two further improvements were made on ff99SB force field. In the ff99SB* [85] force filed, NMR coupling data was used to attain a more accurate representation of helical and coil structures. Another approach included optimizing the side-chain torsional angles to match the ab-initio data, resulting in a ff99SB-ildn force field [62]. In this thesis, I used the combination of ff99SB* and ff99SB-ildn force fields to simulate amyloid fibrils.

80% of a simulation box consists mainly of water, therefore water models have a significant impact on accuracy of simulation results and computational costs. Water models must therefore be as simple as possible to reduce computational costs while maintaining a sufficient level of agreement with experimental properties. Rigid, non-polarizable water models are most commonly employed in classical MD simulations, in which water consists of point charges that represent hydrogen and oxygen atoms. Models differ in geometry and position of the oxygen charge, but the parameter describing the Lennard-Jones interaction is always placed only at the oxygen atom. Three groups of rigid, non-polarizable models are divided into two- three- and four-site models, based on the positions of the oxygen point charge. Three-site models place the charges at their respective atoms, while additional dummy atoms are included in four- and five- site models to describe the oxygen charge.

Water models are selected for simulations based on their ability to describe desired properties. Most common models in use are three-site models SPC [86, 87], TIP3P [88], a four-site TIP4P model [89], and their variants because they obtain the best overall agreement with the experiments regarding properties of water. In this thesis, I used TIP3P water model for simulations with RNA because the force fields for nucleic acids have been parametrized with respect to it. For simulations of amyloid fibrils, I used the TIP4P-EW water model, which is a variant of the TIP4P model for use with Ewlad techniques.

In addition to water, a solvent in the simulation box usually contains ions, which are

very important in systems with charged proteins and nucleic acids. Ions may bind to charged residues in macromolecules and therefore enhance the stability of certain folds.

In MD simulations, ions are typically depicted as point charges and their interactions with other atoms are solely determined by non-bonded interactions. Charge is determined by the valence of ions, and therefore the only way to optimize the properties of ions is by using the Lennard-Jones parameters, ϵ and σ. These have been optimized abundantly to reproduce the experimental kinetic or thermodynamic properties of aqueous salt solutions. Mamatkulov and Schwierz, in particular, developed a systematic procedure to parameterize some of the most ubiquitous mono and divalent cations as well as the chloride ions to reproduce experimental hydration free energies, activity derivatives, and characteristics of water exchange from the first hydration shell [90]. These properties are crucial to successfully capture interactions of ions with biomolecular binding sites, and are therefore used in this thesis.

2.1.4 Enhanced sampling MD simulations

Enhanced sampling methods can be incorporated to overcome the low sampling efficiency of all-atom MD simulations. Most enhanced sampling techniques either increase the energy of the system to allow for sampling of a larger conformational space, or reduce the dimensionality of the sampled space to one collective variable (CV) that can capture the process of interest. Enhanced sampling techniques used in this thesis are TREMD simulations and constant force pulling.

TREMD simulations

Temperature replica exchange is an enhanced sampling simulation method based on simulating multiple independent replicas of the system, each at a different temperature, starting from lowest to the highest. Therefore, systems at higher temperatures are able to cross the potential energy barriers, while those at lower temperatures sample around the local minima [91]. During the simulations, swaps are attempted between two replicas at different temperatures. The acceptance of the swaps is calculated according to the Metropolis acceptance criterion with the probability:

$$P(i(T1) \rightarrow j(T2)) = \min(1, \exp[(\frac{1}{k_B T_1} - \frac{1}{k_B T_2})(U_i - U_j)]), \tag{2.12}$$

where replica i is at temperature T_1 and replica j is at temperature T_2, and U_i and U_j are the potential energies of replicas at the respective temperatures.

According to this equation, exchange attempt is less likely to be accepted for two replicas that have too large temperature difference. Attempting the exchange between more than two replicas also decreases the acceptance probability because the probability depends on multiple exchanges between the systems. In order to assure a higher acceptance ratio, usually the exchange is attempted between a pair of replicas at neighboring temperatures. In this way the replicas perform a random walk along the temperature space. After the exchange is accepted, the velocities of the replicas are scaled according to the ratio of the

two temperatures:

$$\vec{v_i}(T2) = \sqrt{\frac{T_2}{T_1}} \cdot \vec{v_i}(T1)$$
$$\vec{v_j}(T1) = \sqrt{\frac{T_1}{T_2}} \cdot \vec{v_j}(T2). \tag{2.13}$$

Consequently, the canonical ensemble is conserved and the measurements of thermodynamical observables may be performed at the desired temperature. TREMD simulations are appropriate to use if the highest temperature simulated is enough to cross the activation energy.

Constant force pulling simulations

A constant force pulling simulation is a non-equilibrium simulation method in which the center of mass (COM) of one group of atoms is pulled with respect to the COM of the other group by applying a constant force. Therefore, potential applied for pulling is linear. The pulling is performed along a collective variable such as distance between the COMs of pulling groups, an angle or a dihedral angle. The distance can be defined in 3 dimensions along a vector. In this thesis, i performed constant-force pulling simulations to study the dependence of ionic conductance of an rG4 on the type of ion, more information can be found in Chapter 4.

2.2 Langevin dynamics simulations

Some relevant conformational motions cannot be observed on a microsecond timescale. In cases where the solvent does not play a crucial role in the event of interest, simulating it implicitly saves computational time and allows for timescale increase of order of magnitude and more. Therefore, there is an incentive to use Langevin equations of motion instead of Newton's equations of motion, where additional terms encompass the effect of solvent on the atoms of the simulated biomolecule. In this chapter, I will briefly describe the Langevin equations of motion and their application in simulations following the outline of the book by Daniel Zuckerman [92].

2.2.1 Langevin equations of motion

The Langevin equation describes the dynamics of a stochastic system. It is formed by adding two additional averaged terms to the Newton's equation of motion: a term representing a friction force describing the decrease of velocity due to collisions with solvent particles, and a term for a thermal force which modifies velocities of particles depending on the temperature. The Langevin equation for a particle i is:

$$m_i\ddot{\vec{r_i}} = \vec{F_i} - \gamma_i\dot{\vec{r_i}} + \vec{F_i}^{rand} \tag{2.14}$$

F_i is a conservative force on a particle i due to interactions with other particles in

the system and γ_i is a friction constant roughly describing the frequency of collision with solvent molecules. $\vec{F}_i^{rand}$ is usually represented by Gaussian distribution and describes the white noise term from collisions with solvent particles. Since F_i is a conservative force and $\vec{F}_i^{rand}$ derives from a reservoir in thermal equilibrium, the Langevin dynamics sample the Boltzmann distribution. The Langevin equations of motions are solved using the velocity form of the Verlet algorithm. In Langevin dynamics, a modified Verlet algorithm incorporating friction and random forces ensures stable integration while enabling correct canonical sampling through the fluctuation-dissipation balance [93].

2.2.2 Coarse-grained models for RNA simulations

In an attempt to maximize computational efficiency of biomolecular simulations, multiple atoms can be combined into a single interaction site, as done in coarse-grained (CG) simulations. This interaction site is modeled as a classical particle with its own unique set of parameters, representing the biomolecule it is embedded within. CG models are utilized in simulations of many different biomolecules such as membranes, large proteins and nucleic acids. They allow simulations of larger systems and longer timescales in comparison to all-atom simulations, while still describing the macromolecular features that are fundamental for the biological processes of interest. In this thesis, I used a CG model to simulate RNA folding, and I will therefore focus on CG models for nucleic acids.

CG models for RNAs are mainly used to simulate RNA folding. Different degrees of coarse-graining are used, depending on whether the research objective is a coarse representation of the folding pathway or a detailed representation of interactions between residues along the pathway. Therefore, different RNA CG models are developed, ranging in a level of detail from representing each nucleoside by one bead to a more detailed representations that are able to capture base-pair formation. Representing the potential between the beads is a challenging task because more atoms are grouped into a single interaction site and parametrization by QM calculations or experimental measurements is not trivial [94]. Nevertheless, there are several available CG RNA models which have demonstrated the ability to accurately capture specific thermodynamic properties.

HiReRNA is a high resolution model that represents each nucleotide with seven beads and effective interaction potential similar to equation 2.11 that represents physical properties of the real system [95]. Albeit it maintains a high level of detail, with the help of enhanced sampling methods it was able to predict the folded structures of smaller molecules such as pseudoknots and hairpins. Additionally, it is able to capture a detailed representation of intramolecular interactions that guide the folding process.

SimRNA is one more example of a high resolution model as it uses five beads to represent each nucleotide [96]. It uses a knowledge-based statistical potential to represent energy and Monte Carlo algorithm to sample conformational space. With the aid of secondary structure restraints, it is able to accurately predict three dimensional structures of small RNAs.

seRNA model employs ab initio calculations to approximate interaction potential of isolated RNA base pairs, and uses the approximated analytical expressions to parametrize

the CG model [97]. It represents each nucleotide with four beads and the interaction potential encompasses main physical properties which include hydrogen bonding, stacking, backbone interactions and electrostatics. It is parametrized for simulations of double-stranded RNA.

Some interesting low resolution models include oxRNA [98] and the self-organizing polymer model (SOP) [99], which use a simple representation of one bead per nucleotide. However, the potential describing the interactions between the beads is complex. oxRNA maps multiple interaction sites onto a single bead, therefore reducing the computational costs of force calculations while still maintaining key physical detail. It can fold small RNAs and characterize their thermodynamic and mechanical properties. The SOP model is a CG polymer approximation of RNA which uses potential based on interactions in the native state of the simulated RNA molecule. Due to its compactness, it has been successfully applied to reproduce unfolding trajectories of RNA molecules ranging from small hairpins to ribozymes.

Three interaction sites (TIS) coarse-grained model developed by Thirumalai and coworkers was used to simulate RNA G4 folding in this thesis [100–102]. The model represents each nucleotide with three beads which are considered the interaction sites. The beads are placed at the center of mass of phosphate group (P), sugar group (S) and base (B). The beads interact according to the following potential:

$$U_{\mathrm{TIS}} = U_{\mathrm{BL}} + U_{\mathrm{BA}} + U_{\mathrm{EV}} + U_{\mathrm{ST}} + U_{\mathrm{HB}} + U_{\mathrm{EL}}. \tag{2.15}$$

The first two terms represent the bond length and bond angle potentials. Similar to equation 2.11, they have the form of harmonic potentials: $U_{\mathrm{BL}} = k_\rho(\rho - \rho_0)^2$ and $U_{\mathrm{BA}} = k_\alpha(\alpha - \alpha_0)^2$. The equilibrium parameters ρ_0 and α_0 are parametrized to match the coarse-grained representation of an A-form RNA helix. U_{EV} represents excluded volume interactions, represented by Weeks-Chandler-Anderson potential:

$$U_{\mathrm{EV}} = \begin{cases} \epsilon_0[(\frac{D_0}{r})^2 - 2(\frac{D_0}{r})^6 + 1], & r \leq D_0 \\ 0, & r > D_0, \end{cases}$$

where r is the distance between the two beads. ϵ_0 and D_0 are parameters representing the characteristic energy and distance, and their values are assigned to characterize the stacking between the bases. The use of this same parameter across all base conformations is justified due to the predominant role of stacking and hydrogen bonding interactions in the stability of RNAs.

Stacking interactions between two bases are described by the equation:

$$U_{\mathrm{ST}} = \frac{U^0_{\mathrm{ST}}}{1 + k_r(r - r_0)^2 + k_\phi(\phi_1 - \phi_{1,0})^2 + k_\phi(\phi_2 - \phi_{2,0})^2} \tag{2.16}$$

where r represents a distance between two nucleobase beads, and ϕ_1 and ϕ_2 represent

backbone torsion angles in the model (Figure 2.4 (A)). Parameters k_ϕ and k_r are selected to minimize the deviation of r, while r_0, $\phi_{1,0}$ and $\phi_{2,0}$ are parametrized according to the NMR calculated values of standard A-form RNA helix. Parameter U^0_{ST} is the baseline energy for single-base stacking and it is parametrized using experimental thermodynamic data from RNA hairpins and pseudoknots.

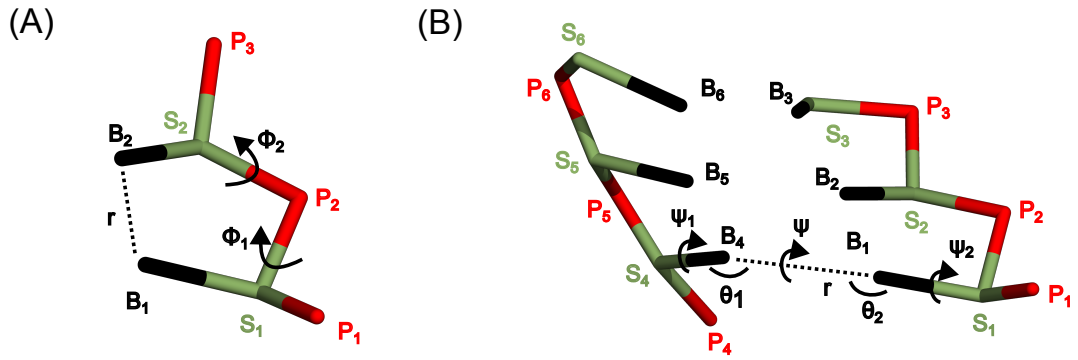

Figure 2.3: (A) Schematic representation of parameters in equation 2.16. Base (B), sugar (S) and phosphate (P) beads are indicated in black, green and red, respectively. The indices define different beads. r is the distance between the beads B_1 and B_2 in angstroms. ϕ_1(P_1,S_1,P_2,S_2) and ϕ_2(P_3,S_2,P_2,S_1) are the dihedral angles, in radians. (B) Representation of the parameters from equation 2.17. It shows a hydrogen bond between the bases B_1 and B_4. Here, r is the distance between the beads, θ_1 (S_4, B_4, B_1) and θ_2 (S_1, B_1, B_4) are the angles. ψ (S_1, B_1, B_4,S_4), ψ_1(B_1, B_4, S_4, P_5) and ψ_2(B_4, B_1, S_1, P_2) are the dihedral angles.

Hydrogen bonds are defined according to equation:

$$U_{\mathrm{HB}} = U^0_{\mathrm{HB}} \times [1 + 5(r - r_0)^2 + 1.5(\theta_1 - \theta_{1,0})^2 + 1.5(\theta_2 - \theta_{2,0})^2 + 0.15(\Psi - \Psi_0)^2 + 0.15(\Psi - \Psi_{1,0})^2 + 0.15(\Psi - \Psi_{2,0})^2]^{-1} \quad (2.17)$$

where r is the distance between two nucleobase beads, θ_1 and θ_2 are hydrogen bonding angles, Ψ a dihedral angle between the planes of the two bases, and Ψ_1 and Ψ_2 the backbone torsion angles (Figure 2.4 (B)). The equilibrium parameters are obtained from the native structure of the simulated RNA, by converting it into the coarse-grained representation. This biasing of the interaction potential to the native state is a property of the Gō model. The difference to the Gō model is that TIS includes all possible hydrogen bonds that can form secondary structure, and therefore it allows the formation of A-U, G-C, and G-U base-pairs that are at least four nucleobases apart. The equilibrium parameters for these interactions are again parametrized according to the A-form RNA helix. Due to this property, the TIS model is able to simulate non-native interactions along the RNA folding pathway.

Since TIS is an implicit solvent model, electrostatic interactions with the solvent are modeled implicitly. Stability of RNA is governed by the electrostatic screening of the negatively charged backbone. Therefore, TIS models the electrostatic screening of phosphate

groups by solvent as the linearized Debye-Hückel equation:

$$G_{DH} = \frac{Q^2 e^2}{2\mathcal{E}} \sum_{i \neq j} \frac{\exp(-|\vec{r}_i - \vec{r}_j|/\lambda)}{|\vec{r}_i - \vec{r}_j|}, \tag{2.18}$$

where $|\vec{r}_i - \vec{r}_j|$ is the distance between phosphate groups i and j, $\mathcal{E}$ is the dielectric constant of water, λ is the Debye screening length dependent on the ionic concentration c_n:

$$\lambda = \sqrt{\frac{\mathcal{E} k_B T}{8\pi e^2 N_A c_n}}. \tag{2.19}$$

The charge term Q is additionally redefined to account for the non-mobile ions that condense to the RNA backbone. It is defined according to the Manning theory of counterion condensation:

$$Q = Q^*(T) = \frac{b}{l_B(T)}. \tag{2.20}$$

In Manning theory, RNA is approximated as rod-like polyelectrolyte (Figure 2.4) and b in equation 2.20 represents unit length per unit charge while $l_B = \frac{e^2}{\mathcal{E} k_B T}$ represents the Bjerrum length. Debye-Hückel equation represents the total electrostatic energy in the model, U_{EL}.

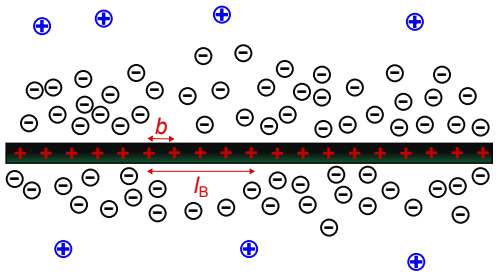

Figure 2.4: Schematic representation of the highly charged rod-like polymer with a unit length per unit charge b and the Bjerrum length l_B in a salt solution.

TIS model uses LD equation 2.14 to calculate trajectories. Therefore, the solvent is implicit which reduces the number of atoms in the calculations and additionally speeds up the simulations. This makes TIS model appropriate to simulate many RNA folding and unfolding events, especially for smaller RNAs.

2.3 Root mean square deviation

Root mean square deviation (RMSD) is a measure of the average distance between the atoms of a molecule and the corresponding atoms in a superimposed reference structure. It is calculated according to the following equation:

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\vec{r}_i - \vec{\hat{r}}_i\|^2} \tag{2.21}$$

where $\vec{r}_i$ and $\vec{\hat{r}}_i$ represent the coordinates of the i-th atom in the target and reference structures, respectively, and N is the total number of atoms. In simulations, it is typically used to calculate the deviation of the simulated molecular coordinates from the experimental reference structure.

2.4 Root mean square fluctuation

Root mean square fluctuation (RMSF) is a measure of the time-averaged mobility of individual atoms or residues in a molecule during a simulation. Unlike RMSD, which compares the structure to a reference, RMSF quantifies the intrinsic flexibility of each particle over the trajectory. It is particularly useful for identifying flexible regions, such as loops, single-stranded RNA segments, or terminal residues.

RMSF for the i-th atom is calculated as:

$$\text{RMSF}_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \|\vec{r}_i(t) - \langle\vec{r}_i\rangle\|^2}, \tag{2.22}$$

where $\vec{r}_i(t)$ is the position of atom i at time t, $\langle\vec{r}_i\rangle$ is the time-averaged position of the atom over the total simulation length T, and $\|\cdot\|$ represents the Euclidean norm.

RMSF provides insight into the local flexibility of the molecule, highlighting regions that undergo significant thermal motion or conformational rearrangements.

Chapter 3

Coarse-grained and MD simulation studies of folding pathways of RNA G-Quadruplex

This research was conducted under the supervision of Prof. Nadine Schwierz at the University of Augsburg and Max Planck Institute of Biophysics in Frankfurt am Main. The CD experiments and scientific guidance was provided by Dr. Ines Burkhart, Dr. Diana Müller and Prof. Harald Schwalbe from the Schwalbe group at the Institute for Organic Chemistry and Chemical Biology and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University, Frankfurt. It has been published in:

Ugrina M, Burkhart I, Müller D, Schwalbe H, **Schwierz N.** "RNA G-quadruplex folding is a multi-pathway process driven by conformational entropy". *Nucleic Acids Res.*, 2024, pp. 87–100.

3.1 Introduction

rG4s have emerged as key structural motifs involved in numerous cellular processes, including the regulation of transcription, translation [3, 4], and pre-mRNA processing [8, 9]. These structures form within guanine-rich RNA sequences, which are widespread in the human genome. Although the in-vivo existence of rG4s was long debated [103], recent technological advancements, particularly the development of more sensitive detection methods, have provided compelling evidence of their presence and dynamic behavior in living cells [7].

Structurally, G4s consist of planar arrangements of four guanine bases forming G-quartets, which stack upon one another to create a central channel. This channel is often flanked by loop regions. In RNA, these are typically propeller loops, whereas dG4s can display more diverse loop conformations [104] (more on the G4 structure is described in Chapter 1). The stability of both rG4s and dG4s is highly dependent on the presence of monovalent cations, with potassium (K^+) and sodium (Na^+) ions being particularly effective in maintaining the structural integrity of the G4 core [17].

An especially significant example of an rG4 is found in telomeric repeat-containing RNA (TERRA), which is transcribed from the C-rich strand of telomeric DNA [105, 106]. TERRA contains the conserved G-rich sequence motif r(UUAGGG) and plays a critical role in safeguarding genomic stability by regulating telomerase activity and protecting chromosome ends [42, 43]. Dysfunction in telomere maintenance mechanisms is closely linked to cellular aging and oncogenesis [45, 46], making TERRA an attractive target for small molecule drug development [47].

Despite their biological importance, rG4s have received comparatively less attention than dG4s [107]. This is partly due to the earlier and more straightforward in-vivo vali-

dation of dG4s through various experimental approaches [5, 6]. On the other hand, rG4s are more transient and dynamic, posing greater challenges for structural and functional characterization [7].

To unravel the molecular mechanisms underlying rG4 folding and function, detailed structural insights are required. Techniques such as NMR spectroscopy [54–57] and X-ray crystallography [108] have been instrumental in resolving the structures of fully folded G4s (more on these methods is described in Chapter 1). Time-resolved NMR has also offered insights into possible folding intermediates [11–13], revealing distinct folding kinetics between TERRA rG4s and analogous DNA sequences [50, 51]. For example, potassium-induced folding of the TERRA sequence is significantly faster ($k_1 = 1.45$/min) than its DNA counterpart ($k_1 = 0.41$/min), indicating fundamental differences in their energy landscapes and folding pathways.

Although some intermediate structures of dG4s have been resolved at atomic resolution [55, 109–111], identifying the transient rG4 intermediates remains challenging. Single-molecule fluorescence techniques, such as smFRET, have suggested the existence of hairpin and triplex intermediates during G4 folding [15, 16]. These observations highlight the complex and multi-step nature of G4 assembly, which likely includes several metastable conformations.

Computational simulations complement the experimental studies by providing atomic and coarse-grained insights into G4 folding dynamics [107, 112–115]. Atomistic MD simulations, although powerful, are limited by the long timescales associated with folding events. As a result, simulating complete folding pathways remains computationally intensive. CG models address this limitation by simplifying nucleic acid representation, reducing computational load while capturing essential folding features. For example, the HiRE-DNA model has successfully been used to map the folding landscape of human telomeric dG4s [112]. Several CG models have been developed specifically for RNA, including SimRNA [96], HiRE-RNA [95], seRNA [97] and the TIS model [100–102, 116–118].

The TIS model, developed by Thirumalai and colleagues, offers an efficient framework for studying RNA folding by representing each nucleotide with three particles and incorporating an implicit solvent. It has been validated on a variety of RNA systems, ranging from small hairpins [117] to larger ribozymes [100] and riboswitches [119].

In this work, I apply a combined approach using TIS-based CG simulations with the CD spectroscopy measurements performed by the Schwalbe Group to investigate the folding pathways and intermediate states of the monomeric 25-mer human telomeric TERRA sequence. Initially, the CG model is validated against CD experimental data, with good agreement observed across different temperatures and salt concentrations. To quantitatively match ionic effects, a concentration correction procedure incorporating all-atom MD simulations is introduced. This integrated approach enables the identification of distinct folding intermediates, including hairpins, triplexes, and double-hairpins, and reveals a diverse, branched folding pathway. Additionally, I extend this analysis to synthetic sequences with varying loop lengths and a plant-derived rG4, confirming that multi-pathway folding is a general feature of RNA G-quadruplexes.

3.2 Methods

3.2.1 Simulation Models

The primary system investigated in this study is a monomeric 25-nucleotide rG4, referred to as TERRA25. To construct an initial model, I began with the experimentally determined structure of a dimeric TERRA G4 resolved by NMR in the presence of potassium ions (PDB ID: 2KBP, see Ref. [53]). This structure corresponds to the sequence $(UAGGGUUAGGGU)_2$.

In the first modeling step, the two RNA strands were joined into a single continuous chain via nucleotides U12 and U13 using the ModeRNA homology modeling server [120]. An additional propeller loop was inserted, and a uridine (U25) was appended to the 3′ end to complete the 25-mer monomer. Next, I employed the SimRNA server [96] to perform coarse-grained energy minimization and generated ten minimum energy conformations. All-atom molecular dynamics simulations of each candidate conformation were conducted for 100 ns, and the structure exhibiting the lowest RMSD relative to the experimental reference (2KBP) was selected for further study.

The equilibrated TERRA25 structure has the sequence 5′-UAGGGUUAGGGUUAG GGUUAGGGUU-3′ and is illustrated in Figure 3.1. It consists of four consecutive GGG repeats connected by UUA loops in a propeller-like configuration. The stacked guanosines from different repeats form three distinct G-quartets, which stabilize the non-canonical G4 core.

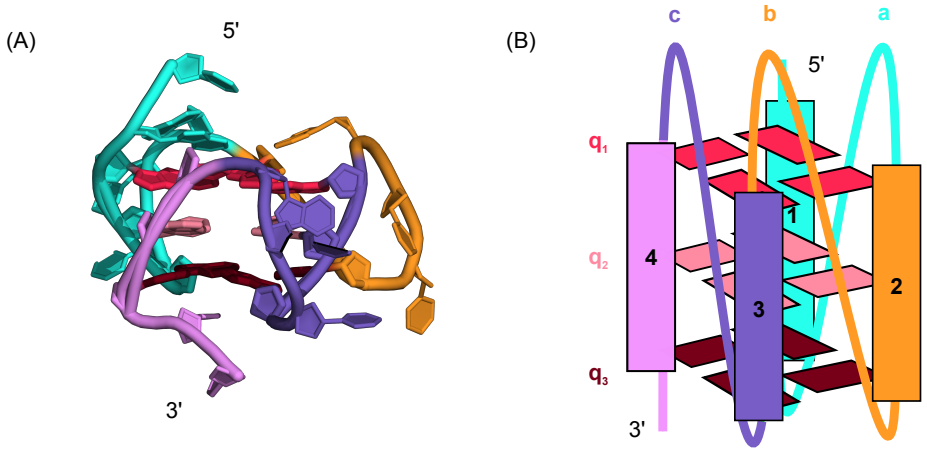

Figure 3.1: Monomeric 25-mer TERRA G4 used in simulations. **(A)** All-atom MD snapshot of the folded structure. **(B)** Schema showing the arrangement of G-rich repeats (1-4, colored blue, orange, purple, and pink from 5′ to 3′) and the three propeller loops (a, b, c) with UUA sequence. The G-quartets are labeled q_1, q_2, and q_3.

Additional four all-atom models were constructed to investigate folding intermediates of TERRA25: two full-length conformations (H1 and H2) and their corresponding minimal models ($H1_{MM}$ and $H2_{MM}$). The minimal models were generated by isolating only the relevant hairpin-forming segments from the full TERRA25 sequence. Specifically, $H1_{MM}$ comprises the sequence UAGGGUUAGGG, while $H2_{MM}$ includes the sequence GGGU-UAGGG (Figure 3.10 A and B).

To build the full-length models, we used the ModeRNA homology modeling server, employing the minimal hairpin structures as templates. In the H1 model, the hairpin is formed by the two guanosine repeats at the 5′ end, with the remainder of the RNA left unfolded (Figure 3.10 D). In contrast, the H2 model features a hairpin formed by the two central guanosine repeats, while the flanking regions remain unfolded (Figure 3.10 C).

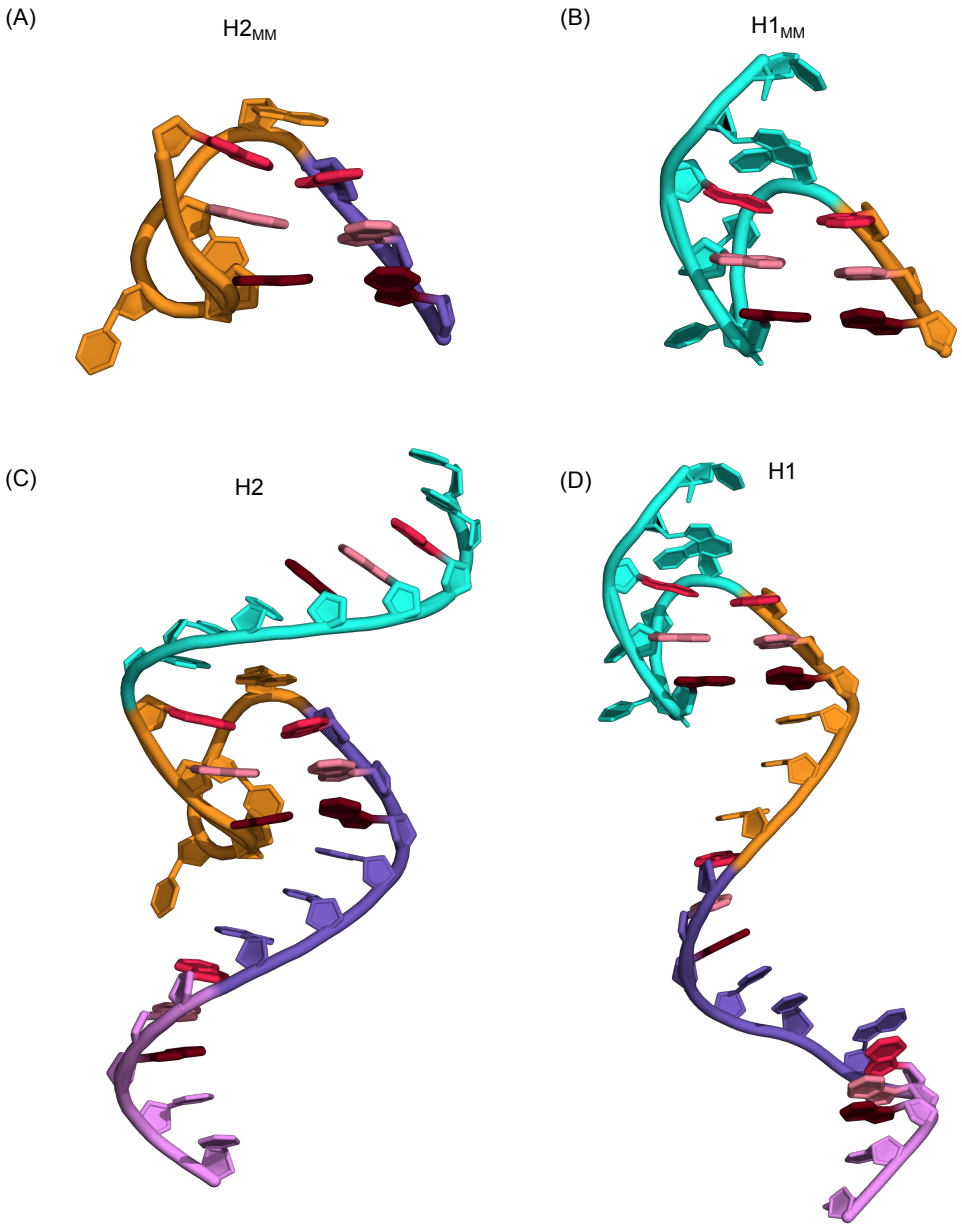

Figure 3.2: Hairpin models simulated in all-atom simulations. (A) Minimal hairpin model consisting of the two middle guanine repeats ($H2_{MM}$). (B) Minimal hairpin model consisting of the first two guanine repeats along the 5′ end ($H1_{MM}$). (C) Full-length hairpin model consisting of the two middle guanine repeats (H2). (D) Full-length hairpin model consisting of the first two guanine repeats along the 5′ end (H1).

To assess the generality of the observed folding behavior, four additional rG4 systems were also simulated (Table 3.1). The first three sequences are synthetic variants with loop lengths of one, two, and three uridines, respectively. They follow the general motif

$UAG_3U_nG_3U_nG_3U_nG_3UU$, where $n = 1, 2, 3$. Full sequences are listed in Table 3.1. These systems were constructed analogously to TERRA25 using the ModeRNA server and the TERRA25 structure as a reference.

The fourth system represents a biologically relevant G4 sequence from a plant RNA, previously studied via smFRET, which is also listed in Table 3.1 [16].Due to the presence of more than three guanosines in the final two repeats, this system required additional modeling steps. I first predicted its secondary structure using the RNAfold server from the ViennaRNA Package 2 [121], and then generated a corresponding native 3D conformation for simulations.

Table 3.1: Sequences of rG4 structures simulated in this work. Loop regions are marked in blue.

Abbreviation	Sequence
TERRA25	UAGGGUUAGGGUUAGGGUUAGGGUU
U	UAGGGUGGGUGGGUGGGUU
UU	UAGGGUUGGGUUGGGUUGGGUU
UUU	UAGGGUUUGGGUUUGGGUUUGGGUU
Plant	GGGAGGGAAGGGGAAGGGG

3.2.2 All-Atom MD Simulations

All-atom MD simulations were performed using GROMACS 2018.1 [122]. TERRA25 was modeled using the Amber99sb-ILDN* force field [76], with parmbsc0 [80] and χ_{OL3} [79] corrections for RNA backbone and glycosidic torsion angles, respectively. Solvation was performed with the TIP3P water model [123]. K^+ and Cl^- ions were parameterized using Mamatkulov-Schwierz force-field parameters developed for compatibility with TIP3P water [90], ensuring accurate thermodynamic and kinetic ion behavior. In all systems, additional K^+ ions were added to neutralize the RNA backbone, followed by equal numbers of K^+ and Cl^- to achieve target bulk concentrations. A total of 25 counterions were added for system neutralization.

In the first set of simulations, the ten lowest-energy conformations generated by SimRNA were simulated independently in 1 M KCl. Each structure was placed in a rectangular simulation box of dimensions $5.4 \times 5.0 \times 5.6$ nm^3. The trajectory with the lowest RMSD relative to the experimental structure, based on heavy atoms in the central pore, was selected as the representative equilibrated conformation for further studies.

A second set of simulations was conducted to evaluate the relationship between bulk ion concentration and local ion density around TERRA25. Simulations were carried out at eight KCl concentrations: 0.48 mM, 16 mM, 53 mM, 82 mM, 0.14 M, 0.22 M, 1.03 M, and 1.89 M. System sizes and solvent content were adjusted to maintain constant density and accommodate ionic differences. To test the generalization of the ion-matching protocol to sodium, an additional 100 ns simulation of TERRA25 was performed in 1 M NaCl.

A third set of simulations was conducted to investigate the stability of the hairpin intermediates observed in CG simulations. Each RNA structure was solvated in a cubic box such that the minimum distance between RNA and the box edge was 2.0 nm. The

simulation box sizes and number of water molecules were adjusted according to the size of the simulated hairpin structure. Each system was neutralized with 25 K^+ and 1 Cl^- ions. The low salt concentration promotes hairpin unfolding.

All systems were solvated in cubic boxes and subjected to both NVT and NPT equilibration. Temperature was maintained using a velocity rescaling thermostat with a stochastic term [124], using a coupling constant of 0.1 ps. NPT equilibration was then performed for 2 ns using the Berendsen barostat [125] with a 1 ps pressure coupling constant. Production simulations employed the same thermostat and the Parrinello-Rahman barostat [126] with a pressure coupling time constant of 2 ps.

3.2.3 CG Simulations Using the TIS Model

To investigate the long-timescale folding behavior of rG4s, I employed CG molecular dynamics simulations using the TIS RNA model. Further details on the TIS model are presented in Chapter 2.

Simulations were performed using Langevin dynamics in the low-friction limit, which accelerates sampling of conformational space while preserving thermodynamic accuracy. The time step was set to $h = 0.05\tau$, with a characteristic timescale $\tau = 50$fs. Each system was simulated in a cubic box of 70 nm with periodic boundary conditions.

To reflect experimental ionic conditions, the effective salt concentration in the CG model was calibrated using a concentration-matching protocol based on ion distribution profiles from all-atom MD simulations. This approach ensures that the local screening experienced by the RNA in the CG model corresponds to the appropriate bulk concentration in the all-atom model.

To assess the temperature dependence of G4 stability, I performed 120 independent simulations at evenly spaced temperatures from 1-120 °C, with a temperature spacing of 1 °C. Each system was simulated at a CG ionic concentration of $c_{\mathrm{CG}} = 150$ mM, corresponding to a matched experimental bulk concentration of $c_{\mathrm{bulk}} = 12$ mM. Each simulation ran for approximately 100 µs.

To investigate the effect of salt concentration on folding behavior, simulations were performed at 16 different ionic strengths ranging from 10 µM to 1 M (corresponding to $c_{\mathrm{bulk}} = 0.06$ µM to 430 mM). These simulations were conducted at two representative temperatures: 25 °C and 60 °C. Each setup involved independent trajectories of approximately 100 µs. All simulations were initiated from the fully folded rG4 structure. The initial 5 µs of each trajectory was excluded from analysis to allow for equilibration. To ensure sufficient statistical convergence at 25 °C, additional ensembles of 50 independent simulations were conducted at 0.1, 0.4, and 1 mM CG concentration. Each of these shorter trajectories was 20 µs in duration.

For direct characterization of folding and unfolding pathways, an additional simulation set was performed at 25 °C and $c_{\mathrm{CG}} = 50$ mM, corresponding to $c_{\mathrm{bulk}} = 0.3$ mM. This salt concentration was chosen to balance the populations of folded and unfolded states, thus facilitating observation of transitions in both directions. Simulations of a plant-derived rG4 sequence were conducted under similar temperature conditions at a higher CG ionic strength

($c_{\mathrm{CG}} = 400$ mM, corresponding to $c_{\mathrm{bulk}} = 80$ mM) to match experimental conditions.

To achieve high-quality statistics, 100 independent simulations were carried out for each folding pathway study. Of these, 23 trajectories were extended to 180 µs, while the remaining 77 were simulated for 52.5 µs. This produced a total simulation time of 8182.5 µ. As in previous setups, all simulations began from the folded conformation, and the first 5 µs were discarded from analysis. Notably, the number of forward (folding) and backward (unfolding) events in the trajectories was equal, indicating that the simulations had reached equilibrium and were sampling both directions of the folding pathway effectively.

3.2.4 Time Scaling and Kinetic Analysis in Coarse-Grained Simulations

In the TIS model, the solvent effect is accounted for via a combination of stochastic and friction forces in the Langevin equation. The timescale of the simulation is determined by the integration timestep and represents the coarse-grained timescale. However, establishing a direct correspondence between coarse-grained simulation time and physical (experimental) time is nontrivial. Namely, the effective solvent viscosity in the CG simulations is reduced by a factor of 100 to enhance conformational sampling, as previously described in the literature [116]. Additionally, the free energy landscape is smoothed due to the coarsening of atomistic interactions and omission of explicit solvent degrees of freedom, which accelerates barrier crossing events. To bridge the timescale gap, a time rescaling factor α can be introduced by comparing folding rates from simulation and experiment:

$$\alpha = \frac{k_{\mathrm{FU}}^{\mathrm{CG}}}{k_{\mathrm{FU}}^{\mathrm{exp}}} \tag{3.1}$$

For the TERRA25 system, the experimentally measured folding rate is $k_{\mathrm{FU}}^{\mathrm{exp}} = 2.4 \times 10^{-2}\ \mathrm{s}^{-1}$, while the corresponding simulated rate is $k_{\mathrm{FU}}^{\mathrm{CG}} = 3.75 \times 10^{5}\ \mathrm{s}_{\mathrm{CG}}^{-1}$. This yields a rescaling factor of $\alpha = 1.56 \times 10^{7}$, enabling conversion between CG time and real time.

Nevertheless, the use of a single numerical factor cannot fully account for the effects of coarse-graining and loss of microscopic detail. For this reason, all time-dependent quantities are also reported in simulation units (s_{CG}). Additionally, reduced kinetic observables were defined to allow for normalized comparisons between states:

$$\tilde{k}_{ij} = \frac{k_{ij}^{\mathrm{CG}}}{k_{\mathrm{FU}}^{\mathrm{CG}}}, \quad \tilde{\tau}_i = \frac{\tau_i^{\mathrm{CG}}}{\tau_{\mathrm{F}}^{\mathrm{CG}}}$$

Transition rate coefficients k_{ij} between stable states $j \to i$ are estimated from long equilibrium trajectories using:

$$k_{ij} = \frac{N}{2t_j}, \quad k_{ji} = \frac{N}{2t_i} \tag{3.2}$$

Here, N is the total number of observed transitions between states i and j, and t_j, t_i denote the total time spent in each state. For sufficiently long simulations at equilibrium, the number of forward and reverse transitions is approximately equal.

Assuming first-order kinetics, the mean lifetime $\langle \tau_j \rangle$ of state j is calculated as:

$$\langle \tau_j \rangle = \left(\sum_{i \neq j} k_{ij} \right)^{-1} \tag{3.3}$$

Population of state i P_i is calculated by dividing the total time spent in state i, t_i by the total simulation time t_{sim}

$$P_i = \frac{t_i}{t_{\mathrm{sim}}} \tag{3.4}$$

3.2.5 Gibbs Free Energy Calculation

To assess the thermodynamic stability of the folded states across different rG4 systems, I calculated the Gibbs free energy difference using the equilibrium folding and unfolding rates obtained from CG simulations. The free energy change was computed using the relation:

$$\Delta G = -k_B T \ln(K_{\mathrm{eq}}), \quad \text{with} \quad K_{\mathrm{eq}} = \frac{k_{\mathrm{FU}}}{k_{\mathrm{UF}}} \tag{3.5}$$

Here, K_{eq} denotes the equilibrium constant, defined as the ratio of folding to unfolding rate constants.

The simulation-derived values were compared to experimental estimates obtained from UV-melting experiments [127]. It is important to note that the experimental values represent an ensemble average over multiple loop sequences and were conducted under RNA and salt concentrations that differ from those used in simulation. To enable a fair comparison, I report relative Gibbs free energies with respect to a reference system. Specifically, the UUU loop system was used as the reference for simulations, while the L333 system was used as the experimental baseline as in [127].

3.2.6 Radius of Gyration and Native Contact Analysis

To monitor structural transitions between folded and unfolded states, I computed two global structural observables: the radius of gyration r_g and the number of native contacts n_C.

The native contact number n_C quantifies the formation of specific interactions between guanosines in the G-quartets. In the fully folded conformation, all three quartets are formed, yielding $n_C = 12$. Contact formation was evaluated using the coordination number function implemented in PLUMED [128], applied to base-bead pairs representing guanosines, and it is calculated as follows:

$$n_C = \sum_{i,j} \frac{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^m}$$

with a contact cutoff $r_0 = 0.75$ nm, a smoothing offset $d_0 = 0.08$ nm, and exponents $n = 35$ and $m = 55$. The sum runs over all pairs of guanosine bases expected to form in the native state.

The radius of gyration determines the spread of N particles around their center of mass:

$$r_g = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_{\text{COM}})^2} \tag{3.6}$$

The radius of gyration was calculated using all backbone beads to capture global chain compaction.

3.2.7 Order Parameters

To quantitatively characterize folding intermediates and transitions in the rG4 simulations, I employed a set of custom-defined order parameters. These parameters capture the formation of native contacts between G-repeats and within G-quartets.

To monitor the formation of native contacts between G-rich repeats, I define a repeat order parameter r_{ij}, which reflects the contact formation between repeats i and j (with $i \neq j$, $i, j \in \{1, 2, 3, 4\}$). This parameter is evaluated using a smoothed indicator function:

$$h(n_C) = \begin{cases} 1, & \text{if } n_C > 2c \\ \dfrac{1}{1 + \exp\left[\kappa(n_C - \gamma c)\right]}, & \text{otherwise} \end{cases} \tag{3.7}$$

Here, the parameters are set to $\kappa = -5$, $\gamma = 1.3$, and $c = 1.4$. n_C represents number of native contacts between two repeats. The repeat order parameter $r_{ij} = h$ is equal to 1 when all three native contacts between the respective repeats are formed.

To resolve and classify the twelve structurally distinct intermediate states observed during folding, I define two combined metrics, s_1 and s_2, as linear combinations of repeat order parameters:

$$s_1 = 3r_{12} + 6r_{23}, \quad s_2 = 3r_{34} + 6r_{14} \tag{3.8}$$

These composite variables help project the conformational ensemble onto a two-dimensional space.

To evaluate the formation of individual G-quartets, I define a quartet order parameter q_k for each of the three quartets $k \in \{1, 2, 3\}$. The same functional form as in Eq. (3.7) is used, but with modified parameters: $\kappa = -4$, $\gamma = 1.9$, and $c = 1.93$. Here, n_C denotes the number of native contacts among guanosines forming quartet k. When all four contacts

are present, $q_k = 1$.

The global folding status of the RNA was assessed using the fraction of folded structures, n_f. This parameter is computed based on the total number of native guanosine-guanosine contacts in the structure and uses the same switching function defined in Eq. (3.7), with parameters $\kappa = -5$, $\gamma = 2$, and $c = 4.5$. The value of n_f approaches 1 when all quartets are fully formed, corresponding to the fully folded rG4 state.

The average fraction of folded structures was computed as the fraction of total simulation time spent in the folded state. To estimate uncertainties, standard errors were calculated via block averaging with a block size of 2.5 µs.

3.2.8 Concentration Matching Between Atomistic and Coarse-Grained Simulations

In the CG simulations, the solvent is treated implicitly, and ions are uniformly distributed throughout the simulation box (Figure 3.3B). However, this simplification neglects an important physical feature of RNA systems: the formation of an electric double layer due to the high negative charge density of the RNA backbone. In the presence of monovalent salts, positively charged ions (K^+) are electrostatically attracted to the phosphodiester backbone, resulting in significant ion accumulation near the RNA surface (Figure 3.3A). Since RNA folding is strongly influenced by local ion screening, especially in the vicinity of phosphate groups, it is essential to calibrate the ion concentration used in CG simulations to reflect the local ionic environment observed in atomistic simulations. Therefore, a concentration matching protocol was developed to relate the experimental bulk ion concentration to the effective ionic strength in CG models.

The procedure consists of two main steps. First, all-atom molecular dynamics simulations were carried out at eight different bulk KCl concentrations. From these simulations, the radial ion concentration profile was computed with respect to the central pore of the folded RNA structure. Figure 3.3A shows an example profile for $c_{\text{bulk}} = 1.03$ M, where two distinct features are observed: a sharp peak at $r = 0.2$ nm, corresponding to K^+ ions located in the central pore, and a second peak at $r = 1.23$ nm, associated with K^+ adsorption at the phosphate backbone.

To quantify the effective local ion concentration, the radial profiles were averaged from the center of the rG4 channel out to a cutoff distance R_C, representing the edge of the condensed ion layer. The value of R_C was treated as a free parameter and optimized based on consistency with experimental data. A value of $R_C = 1.5$ nm provided the best agreement with CD measurements (see Figure 3.4) and was used in all subsequent analysis. The bulk concentration c_{bulk} was estimated by averaging the ion density in the final 0.5 nm region near the edge of the simulation box.

At low concentrations, the strong binding affinity of K^+ ions to RNA causes depletion of free ions in finite-sized MD boxes, making it difficult to accurately determine c_{bulk} from cation profiles. To overcome this, the bulk concentration was instead extracted from the more stable anion distribution, which remains unaffected by strong binding.

The relationship between the local ion concentration and the experimental bulk concen-

tration was obtained via linear interpolation of the computed data points. This matching curve is shown in Figure 3.3 C. The concentration-matched values were used as input in all CG simulations to ensure realistic electrostatic screening. The effectiveness of this approach was validated by comparing simulation results to experimental CD spectra, as discussed in later sections.

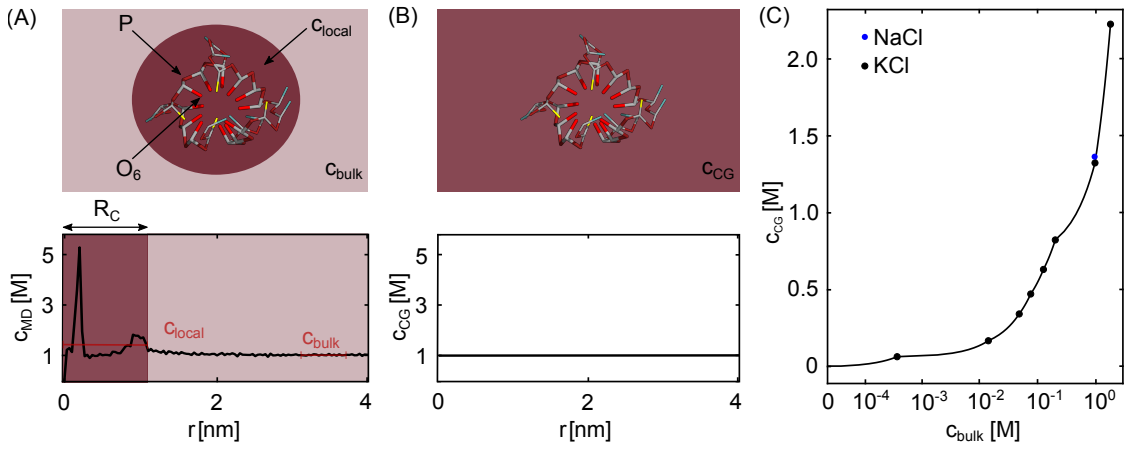

Figure 3.3: (A) Schematic representation of the difference in bulk cation concentration c_{bulk}, and the local cation concentration in the vicinity of the TERRA25 c_{CG}, with a concentration profile below for K^+ ions as a function of the distance from the TERRA25 ion channel center. The concentration profile was calculated from an all-atom MD simulation. (B) Uniform concentration profile as simulated in the TIS simulation model, along with the schematic representation to highlight the difference to the all-atom simulation. (C) Concentration matching relation of the c_{bulk} and c_{CG}. x-axis is represented on a logarithmic scale. An additional simulation was performed for 1M NaCl and the result is represented as the blue point on the figure.

3.2.9 Estimating the Fraction of Folded Structures from CD Melting Curves

The population of folded rG4s as a function of temperature was extracted from CD melting curves.

To begin, upper and lower baselines were established to represent the temperature-dependent CD signal of the fully folded and fully unfolded states, respectively. The experimental CD absorbance values were then normalized using these baselines, yielding the fraction of folded structures n_f across the temperature range.

Melting temperatures T_{m} were obtained by fitting the normalized folding curves to a sigmoidal model of the form:

$$n_f(T) = \frac{1}{1 + \exp\left(\frac{T - T_{\mathrm{m}}}{s}\right)} \tag{3.9}$$

where T_{m} denotes the melting temperature and s is the slope parameter of the melting curve. The fitting procedure was performed using the nonlinear least-squares Marquardt-Levenberg algorithm.

3.3 Results and discussion

3.3.1 Quantitative comparison between simulated and experimental fractions of folded structures

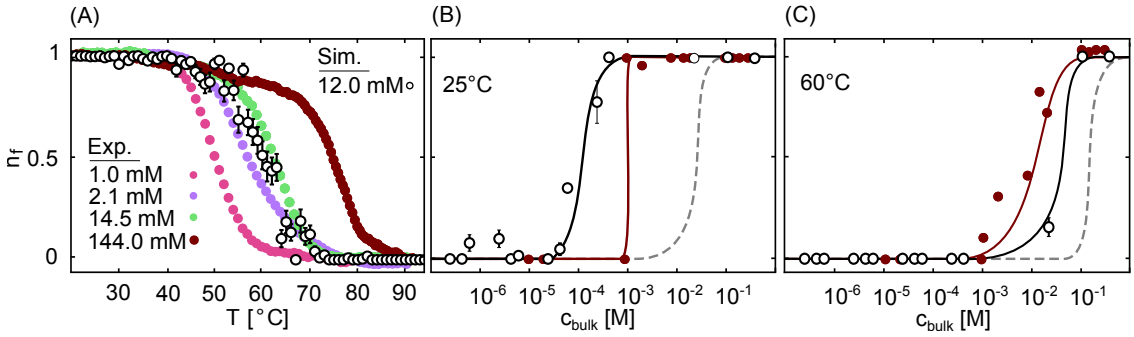

Figure 3.4: Comparison of the fraction of folded states n_f from simulations and experiments. (A) n_f as a function of temperature T from experimental CD-melting curves and TIS simulations. (B) n_f as a function of concentration from the CD experiments and TIS simulations at T = 25 °C and (C) T = 60 °C.

To validate the CG model, I compared simulation results with experimental CD melting data by analyzing the temperature and salt concentration dependence of the folded-state population n_f (Figure 3.4). Accurate comparison between the CG simulations and experiments requires proper treatment of the ion environment surrounding the RNA, particularly the electric double layer that forms around the highly charged phosphate backbone of the rG4.

In the cellular and experimental environment, positively charged ions such as K^+ are electrostatically attracted to the RNA backbone, while anions are repelled, leading to a formation of the double layer. This results in a local cation concentration that is significantly elevated compared to the bulk solution (Figure 3.3 C). To reproduce this effect in the CG simulations, which use an implicit solvent and spatially uniform ion distribution, a mapping procedure was established to relate the experimental bulk salt concentration to the effective local ion concentration in the double layer.

This mapping was derived from all-atom molecular dynamics simulations across a range of KCl concentrations. By integrating the local ion density profiles around the folded RNA structure, a matching relation was constructed that defines the CG ion concentration c_{CG} as the average local concentration within the electric double layer. The resulting relation (Figure 3.3C) was used to set the ionic conditions in all CG simulations, thereby ensuring that the screening environment accurately reflects experimental conditions.

To assess the performance of this concentration-matched model, I first examined temperature-induced unfolding of the TERRA25. Melting curves were generated from both simulations and CD measurements (Figure 3.4 A). At low temperatures, the RNA remains fully folded ($n_f \approx 1$), while increasing temperature shifts the equilibrium toward the unfolded state, causing n_f to decrease. At high temperatures, the system becomes fully unfolded ($n_f \approx 0$).

The melting temperature T_m was found to be strongly dependent on salt concentration

(see Table 3.2). Higher K^+ concentrations stabilize the folded state and increase T_m. Simulations incorporating the concentration-matching protocol correctly reproduced this trend and yielded a melting temperature of 60°C at 12 mM bulk KCl, which lies between the experimental values of 59°C and 62°C measured at 2.1 mM and 14.5 mM, respectively. In contrast, using a CG salt concentration of 150 mM without ionic matching would produce the same melting temperature. This illustrates that the CG model, when used without local concentration correction, under represents the role of ionic screening and thus misrepresents folding equilibria.

The importance of including the local ion concentration is further emphasized when examining the salt dependence of n_f at fixed temperatures of 25°C and 60°C (Figure 3.4 B,C). Without concentration matching, the CG simulations incorrectly predict the folded-to-unfolded transition at higher salt concentrations than observed experimentally. This leads to an overestimation of the destabilizing effect of salt dilution. In contrast, incorporating the matching relation derived from atomistic MD simulations enables the CG model to quantitatively reproduce the experimental data, accurately capturing both the temperature and concentration dependence of the folding transition.

Table 3.2: Melting temperatures (T_m) of TERRA25 at various bulk KCl concentrations obtained from CD experiments and coarse-grained (CG) simulations. For simulations, the effective local ion concentrations (c_CG) were derived using concentration matching based on all-atom MD simulations.

Source	$\boldsymbol{c_\mathrm{bulk}}$ [mM]	$\boldsymbol{c_\mathrm{CG}}$ [mM]	$\boldsymbol{T_\mathrm{m}}$ [°C]
Experiment	1.0	–	50.67 ± 0.04
	2.1	–	59.21 ± 0.09
	14.5	–	62.60 ± 0.20
	144.0	–	76.90 ± 0.10
Simulation	12	150	60.10 ± 0.10

3.3.2 The folding and unfolding of TERRA25 proceed sequentially through multiple pathways

Having validated the coarse-grained (CG) model, I analyzed the folding and unfolding dynamics of TERRA25 in detail using the CG simulations. Figure 3.5 B shows the time evolution of the number of native contacts n_C and the radius of gyration r_g over an 80 µs trajectory at 25 °C and a matched bulk concentration of 0.3 mM KCl. Multiple transitions are observed between the fully folded state ($n_C = 12$, $r_g \approx 15$ Å) and the unfolded state ($n_C = 0$, $r_g \approx 24$ Å).

The thermodynamic driving forces for folding were evaluated by computing the one-dimensional free energy profile $F(r_g)$, shown in Figure 3.5 C. The folded conformation is characterized by a narrow, deep free energy minimum, while the unfolded state spans a broader range of radii of gyration, consistent with a more disordered ensemble. This reflects a classical energy-entropy compensation mechanism: the folded state is stabilized by enthalpic contributions (native interactions), whereas the unfolded state is favored entropically due to its structural heterogeneity. A free energy barrier of approximately 5 $k_\mathrm{B}T$ separates the two states. The two-dimensional landscape $F(n_C, r_g)$ (inset) further

reveals the presence of metastable intermediate states, each spanning a broad distribution, indicating that each intermediate corresponds to an ensemble of structurally distinct conformations.

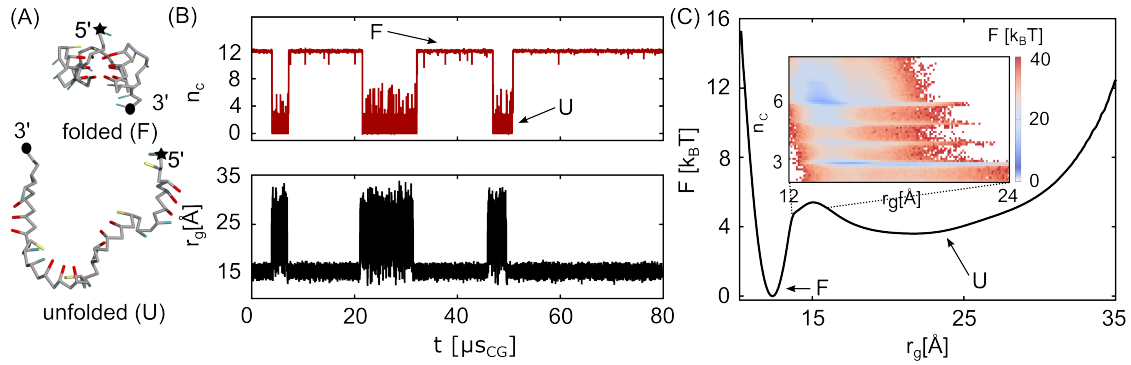

Figure 3.5: (A) Representative snapshots of folded and unfolded conformations from CG trajectories. Guanosines are shown in red, uridines in teal, and adenosines in yellow. (B) Time evolution of the number of native contacts n_C and the radius of gyration r_g during a folding trajectory. (C) One-dimensional free energy profile as a function of r_g. The inset shows a two-dimensional landscape projected onto r_g and n_C, highlighting the location and width of the folding barrier.

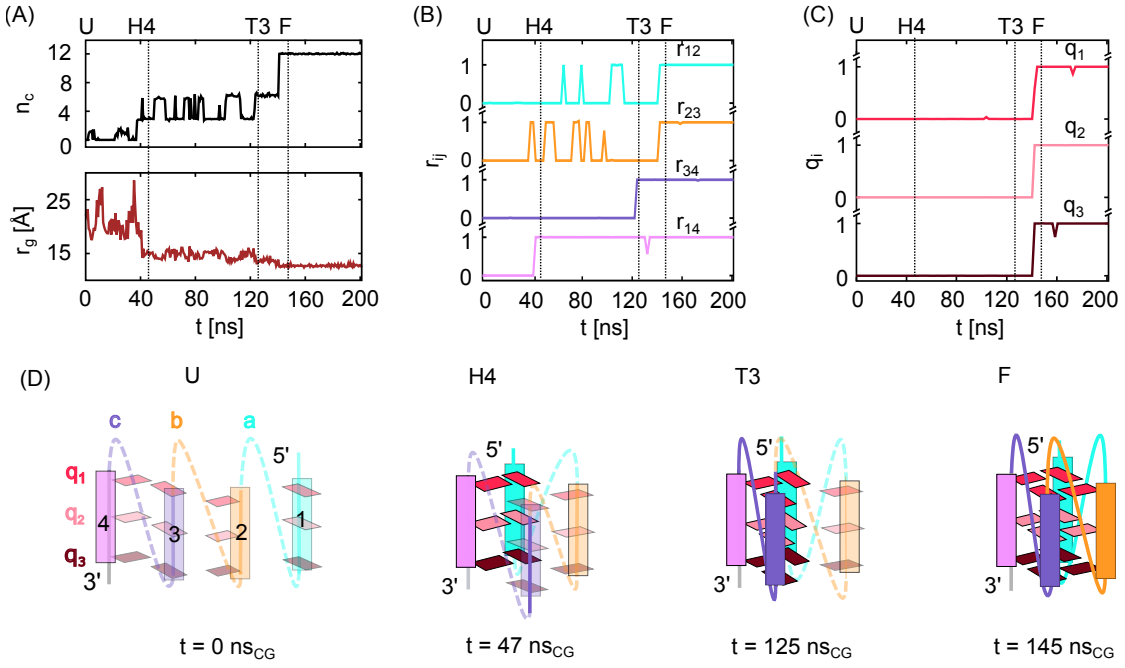

Figure 3.6: (A) Radius of gyration (r_g), (B) number of native contacts between repeats i and j (r_{ij}) and (C) number of native contacts in each quartet i (q_i) calculated from the representative folding trajectory. (D) Schematic representation of intermediate states during the folding trajectory: Unfolded state (U), hairpin intermediate (H4), triplex intermediate (T3) and the folded state (F). Bound repeats are represented in full colors and lines, and unbound repeats are represented in transparent colors or dashed lines.

To investigate the detailed folding pathway, I analyzed a representative folding trajectory shown in Figure 3.6. The number of native contacts increases in a stepwise manner, while the RNA becomes progressively more compact (Figure 3.6 A). Analysis of the repeat contact parameters r_{ij} and quartet formation parameters q_k (Figure 3.6 B, C) indicates

that folding proceeds via sequential assembly of repeats, rather than through the stepwise formation of quartets. This suggests that stacking interactions between repeats drive the initial folding steps, prior to complete quartet formation.

In this example, repeats 1 and 4 form the first stable unit-a parallel hairpin (H4) without a propeller loop (Figure 3.6 D, $t = 47$ ns$_{\mathrm{CG}}$). Subsequently, repeat 3 binds, forming a triplex structure (T3) comprised of three hydrogen-bonded guanosines per plane. At this stage, loop c at the 3′ end is established ($t = 125$ ns$_{\mathrm{CG}}$). The final folding step involves repeat 2, which completes the formation of loops a and b, and all three quartets form simultaneously ($t = 145$ ns$_{\mathrm{CG}}$), as confirmed by the rapid increase in q_k (Figure 3.6C).

Overall, the folding pathway proceeds through the sequential assembly of guanosine repeats and the associated formation of propeller loops. Two key intermediate structures are typically observed: a hairpin and a triplex. These intermediates are consistent with experimental and computational studies that proposed similar structures during the folding of RNA and DNA G-quadruplexes [11, 13, 15, 115, 129, 130].

According to simple combinatorics, four distinct hairpins can be formed by pairing any two of the four G-rich repeats, given their parallel orientation [14, 107]. Similarly, four unique triplex structures are possible. A key question is whether all of these intermediates are accessible during folding, or whether some are suppressed by kinetic or energetic barriers.

To address this, I analyzed approximately 200 folding and unfolding events and classified the intermediates observed (Figures 3.7, 3.8). Additionally, population P_i, reduced lifetime $\tilde{\tau}$, lifetime τ and rescaled τ were calculated for all intermediates and are represented in Table 3.3.

Table 3.3: Populations and lifetimes of the folded, unfolded, and intermediate states from CG simulations. The reduced lifetimes use the folded state lifetime ($\tau = 60$ μs$_{\mathrm{CG}}$) as a reference. The CG time unit can be mapped to experimental time using the scaling factor $\alpha = 1.56 \cdot 10^7$. Errors were computed by block averaging over 10 equal trajectory blocks.

State	**Population**			**Reduced Lifetime** $\tilde{\tau}$			**Lifetime** τ [s$_{\mathrm{CG}}$]			**Rescaled** τ [s]		
F	$9.51 \cdot 10^{-1}$	±	$2.1 \cdot 10^{-3}$	1.0	±	0.1	$6.14 \cdot 10^{-5}$	±	$6.1 \cdot 10^{-6}$	957.8	±	95.1
U	$4.81 \cdot 10^{-2}$	±	$2.8 \cdot 10^{-4}$	0.04	±	0.01	$2.7 \cdot 10^{-6}$	±	$6.4 \cdot 10^{-7}$	42.0	±	9.6
H1	$1.39 \cdot 10^{-5}$	±	$3.2 \cdot 10^{-7}$	$1.686 \cdot 10^{-4}$	±	$1.7 \cdot 10^{-7}$	$10.35 \cdot 10^{-9}$	±	$1.1 \cdot 10^{-11}$	0.1615	±	$1.7 \cdot 10^{-4}$
H2	$1.48 \cdot 10^{-4}$	±	$2.4 \cdot 10^{-6}$	$2.318 \cdot 10^{-4}$	±	$2.0 \cdot 10^{-7}$	$14.23 \cdot 10^{-9}$	±	$1.2 \cdot 10^{-11}$	0.2220	±	$1.9 \cdot 10^{-4}$
H3	$5.75 \cdot 10^{-5}$	±	$8.4 \cdot 10^{-7}$	$1.381 \cdot 10^{-4}$	±	$2.9 \cdot 10^{-7}$	$8.48 \cdot 10^{-9}$	±	$1.8 \cdot 10^{-11}$	0.1323	±	$2.8 \cdot 10^{-4}$
H4	$2.21 \cdot 10^{-4}$	±	$3.0 \cdot 10^{-6}$	$3.106 \cdot 10^{-4}$	±	$3.4 \cdot 10^{-7}$	$19.07 \cdot 10^{-9}$	±	$2.1 \cdot 10^{-11}$	0.2975	±	$3.3 \cdot 10^{-4}$
T1	$4.86 \cdot 10^{-5}$	±	$9.5 \cdot 10^{-7}$	$1.001 \cdot 10^{-4}$	±	$2.3 \cdot 10^{-7}$	$6.15 \cdot 10^{-9}$	±	$1.4 \cdot 10^{-11}$	0.0960	±	$2.2 \cdot 10^{-4}$
T2	$1.76 \cdot 10^{-5}$	±	$4.3 \cdot 10^{-7}$	$0.868 \cdot 10^{-4}$	±	$2.0 \cdot 10^{-7}$	$5.33 \cdot 10^{-9}$	±	$1.2 \cdot 10^{-11}$	0.0832	±	$1.9 \cdot 10^{-4}$
T3	$6.1 \cdot 10^{-4}$	±	$1.1 \cdot 10^{-5}$	$7.293 \cdot 10^{-4}$	±	$2.8 \cdot 10^{-7}$	$44.78 \cdot 10^{-9}$	±	$1.7 \cdot 10^{-11}$	0.6986	±	$2.7 \cdot 10^{-4}$
T4	$5.17 \cdot 10^{-5}$	±	$1.3 \cdot 10^{-6}$	$7.632 \cdot 10^{-4}$	±	$3.8 \cdot 10^{-7}$	$46.86 \cdot 10^{-9}$	±	$2.3 \cdot 10^{-11}$	0.7310	±	$3.6 \cdot 10^{-4}$
D1	$1.16 \cdot 10^{-4}$	±	$6.1 \cdot 10^{-6}$	$5.198 \cdot 10^{-4}$	±	$5.1 \cdot 10^{-7}$	$31.92 \cdot 10^{-9}$	±	$3.1 \cdot 10^{-11}$	0.4980	±	$4.8 \cdot 10^{-4}$
D2	$5.5 \cdot 10^{-4}$	±	$1.3 \cdot 10^{-6}$	$1.894 \cdot 10^{-4}$	±	$1.07 \cdot 10^{-7}$	$11.63 \cdot 10^{-9}$	±	$6.6 \cdot 10^{-12}$	0.1814	±	$1.0 \cdot 10^{-4}$

All four hairpin topologies were encountered in simulations, though their relative lifetimes differed. The most stable hairpin, H4, involves the terminal repeats and has a rescaled lifetime of $\tilde{\tau} = 3.106 \times 10^{-4}$, corresponding to $\tau_{\mathrm{resc}} = 0.2975$ s. Hairpins involving central repeats and propeller loops, such as H3, were shorter-lived ($\tilde{\tau} = 1.381 \times 10^{-4}$; $\tau_{\mathrm{resc}} = 0.1323$ s).

Interestingly, a hairpin analogous to H1 (formed at the 5′ end) was previously identified in smFRET experiments on plant rG4s [16], providing experimental support for its stability.

Triplex intermediates follow a similar trend: all four theoretical conformations are observed, with the longest-lived examples (T3, T4) involving the two terminal repeats and one central repeat. Their rescaled lifetimes exceed 0.7 s, whereas T1 and T2 are somewhat less stable.

A triplex structurally similar to T1 was also reported in experimental work [16], and RNA triplexes with propeller loops have been proposed for both RNA and DNA G4s with two quartets [15].

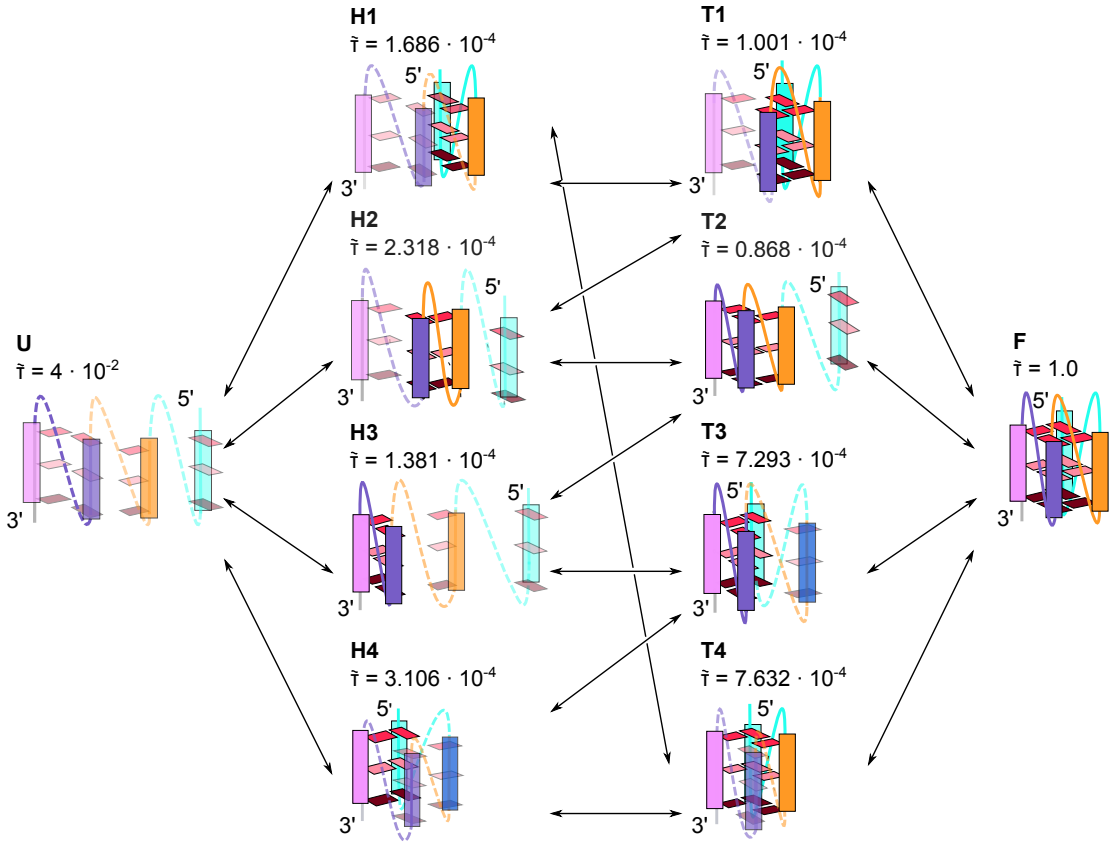

Figure 3.7: Folding and unfolding pathways of TERRA25 from the TIS simulations. Arrows represent the transitions between conformations. The pathways are involving triplex (T_i) and hairpin (H_i) intermediates. Here, i=1,2,3,4. The rescaled lifetimes $\tilde{\tau} = \tau/\tau_F$ are shown above the structures.

In addition to hairpin and triplex intermediates, a third class of intermediates was observed: double-hairpin structures (Figure 3.8). These arise when two hairpins form sequentially but without merging into a triplex. Pathways that proceed through a double-hairpin do not feature triplex intermediates. Among these, the D1 structure, which consists of a hairpin formed by terminal repeats and a hairpin formed by central repeats, was the most stable, with a lifetime of $\tilde{\tau} = 5.198 \times 10^{-4}$ ($\tau_{resc} = 0.4980$ s). A similar double-hairpin motif was also observed in NMR studies of human telomeric dG4s [13].

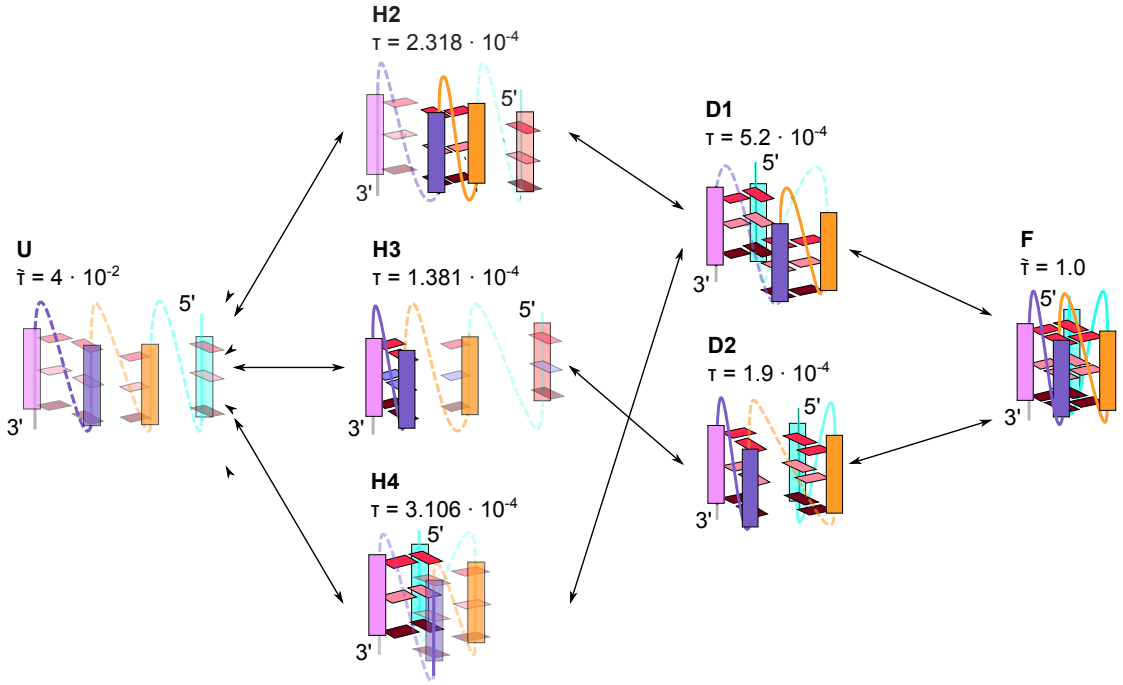

Figure 3.8: Folding and unfolding pathways of TERRA25 from the TIS simulations. Arrows represent the transitions between conformations. The pathways are involving double-hairpin (D_i) and hairpin (H_i) intermediates. The rescale lifetimes $\tilde{\tau} = \tau/\tau_F$ are shown above the structures.

In summary, the folding and unfolding of TERRA25 proceeds via sequential assembly of structural units, with the pathway passing through multiple well-defined intermediate states. The first major intermediate is typically a hairpin, followed either by a triplex or a double-hairpin structure. Each class of intermediates includes multiple conformational variants, contributing to the overall high entropy of the folding landscape. Across all simulations, four hairpins, four triplexes, and two double-hairpin structures were identified. Based on the previously established rescaling factor (Eq. 3.1), their lifetimes are estimated to lie in the range of hundreds of milliseconds (Table 3.3). The lifetimes were calculated according to Equation 3.3 because of the condition $dp/dt = 0$ in the equilibrium. Now that all the intermediate states are revealed it can be shown that for example, for hairpin H1, the rate equation calculated as: $\frac{dp_i}{dt} = \sum_{j\neq i} k_{ij}P_j - \sum_{j\neq i} k_{ji}P_i$ is:

$$\begin{aligned}\frac{dp_{\mathrm{H1}}}{dt} &= -k_{\mathrm{T1H1}}\cdot p_{\mathrm{H1}} - k_{\mathrm{D2H1}}\cdot p_{\mathrm{H1}} - k_{\mathrm{UH1}}\cdot p_{\mathrm{H1}} + k_{\mathrm{H1T1}}\cdot p_{\mathrm{T1}} + k_{\mathrm{H1D2}}\cdot p_{\mathrm{D2}} + k_{\mathrm{H1U}}\cdot p_{\mathrm{U}} \\ &= -\frac{N_{\mathrm{T1H1}}}{2t_{\mathrm{H1}}}\cdot\frac{t_{\mathrm{H1}}}{t_{\mathrm{SIM}}} - \frac{N_{\mathrm{D2H1}}}{2t_{\mathrm{H1}}}\cdot\frac{t_{\mathrm{H1}}}{t_{\mathrm{SIM}}} - \frac{N_{\mathrm{UH1}}}{2t_{\mathrm{H1}}}\cdot\frac{t_{\mathrm{H1}}}{t_{\mathrm{SIM}}} + \frac{N_{\mathrm{T1H1}}}{2t_{\mathrm{T1}}}\cdot\frac{t_{\mathrm{T1}}}{t_{\mathrm{SIM}}} \\ &+ \frac{N_{\mathrm{D2H1}}}{2t_{\mathrm{D2}}}\cdot\frac{t_{\mathrm{D2}}}{t_{\mathrm{SIM}}} + \frac{N_{\mathrm{UH1}}}{2t_{\mathrm{U}}}\cdot\frac{t_{\mathrm{U}}}{t_{\mathrm{SIM}}}\end{aligned} \tag{3.10}$$

Clearly, the values on the right side cancel out and the equilibrium is satisfied. The same is true for all other intermediate states. Rate constants for all intermediate states are presented in Table 3.4.

Table 3.4: Rates of transitions between intermediates on the TERRA25 folding pathway. Each transition is labeled by the names of the intermediates, with k_{ij} being the rate from state j to state i. CG rates are rescaled using $\alpha = 1.56 \cdot 10^7$, as given in the Methods section. The experimental folding rate for TERRA25 is $k_{\mathrm{FU}} = 2.4 \cdot 10^{-2}$ [14]. Errors were obtained via block averaging.

State i	State j	k_{ij} [10^7 s_{CG}^{-1}]			k_{ji} [10^7 s_{CG}^{-1}]			Rescaled k_{ij} [s^{-1}]			Rescaled k_{ji} [s^{-1}]		
H1	U	0.0017	±	0.0005	6.0	±	2.0	0.0011	±	0.0003	4.0	±	1.0
H2	U	0.0013	±	0.0004	4.0	±	1.0	0.008	±	0.002	2.7	±	0.8
H3	U	0.007	±	0.002	6.0	±	2.0	0.005	±	0.001	4.0	±	1.0
H4	U	0.0012	±	0.0004	2.3	±	0.8	0.008	±	0.002	1.8	±	0.5
T1	H1	3.0	±	1.0	1.0	±	0.3	2.3	±	0.7	0.7	±	0.2
T1	H2	2.3	±	0.7	7.0	±	2.0	1.6	±	0.5	4.0	±	1.0
T2	H2	0.6	±	0.2	5.0	±	2.0	0.4	±	0.1	4.0	±	1.0
T2	H3	1.2	±	0.4	4.0	±	1.0	0.8	±	0.2	2.7	±	0.8
T3	H3	5.0	±	1.0	0.4	±	0.1	3.0	±	0.9	2.8	±	0.8
T3	H4	1.7	±	0.5	0.6	±	0.2	1.1	±	0.3	0.4	±	0.1
T4	H4	0.29	±	0.09	1.2	±	0.4	0.19	±	0.06	0.8	±	0.3
F	T1	8.0	±	2.0	0.0004	±	0.0001	5.0	±	2.0	0.00028	±	0.00008
F	T2	9.0	±	3.0	0.00017	±	0.00005	6.0	±	2.0	0.00012	±	0.00003
F	T3	1.2	±	0.3	0.0008	±	0.0002	0.8	±	0.2	0.0005	±	0.0002
F	T4	1.2	±	0.4	0.00007	±	0.00002	0.8	±	0.2	0.0004	±	0.0001
D1	H4	0.7	±	0.2	1.3	±	0.4	0.5	±	0.1	0.9	±	0.3
D2	H1	1.1	±	0.3	2.9	±	0.9	0.8	±	0.2	1.9	±	0.6
D2	H3	1.4	±	0.4	1.4	±	0.4	0.9	±	0.3	1.0	±	0.3
F	D1	1.8	±	0.5	0.00022	±	0.00007	1.2	±	0.4	0.00015	±	0.00004
F	D2	4.0	±	1.0	0.000025	±	0.000007	2.9	±	0.8	0.000017	±	0.000005
F	U	0.0375	±	0.0164	0.00163	±	0.00060	0.0240	±	0.0105	0.00105	±	0.00039

In addition to native-like intermediates, transient off-pathway conformations featuring non-native contacts are occasionally observed (Figure 3.9). For example, a triplex structure with a strand shift resulting from non-native guanine interactions (Figure 3.9 A), as well as a non-native hairpin structure (Figure 3.9 B) were observed in multiple trajectories, demonstrating that the TIS model captures a broad spectrum of non-native interactions. These structures typically unfold rapidly and were not further analyzed.

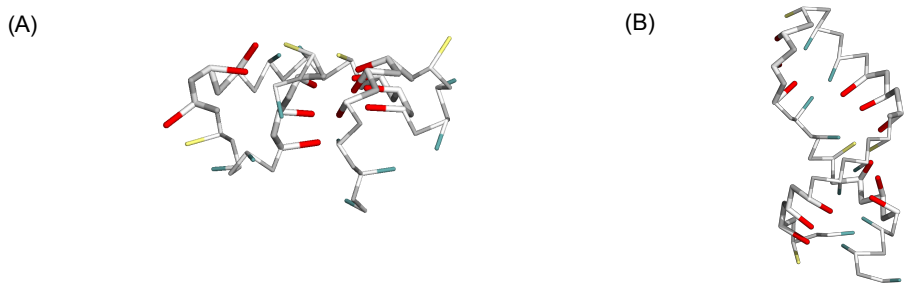

Figure 3.9: Transient non-native states observed on the TERRA25 folding pathway. (A) Strand-shifted triplex and (B) non-native helical hairpin structure.

Although the rescaling provides only an approximate conversion from CG to experimental time, additional all-atom simulations support the stability of the hairpin intermediates.

Order parameters in Figure 3.10 C and D describing the number of hydrogen bonds (N_{HB}) between repeats in full hairpin models never goes below five ($N_{HB} = 6$ in the folded state), which means that the full hairpin models remain folded throughout the simulation, despite the low ionic concentration. Minimum models, however, unfold more rapidly (Figure 3.10 A and B), consistent with earlier MD studies [113, 131].

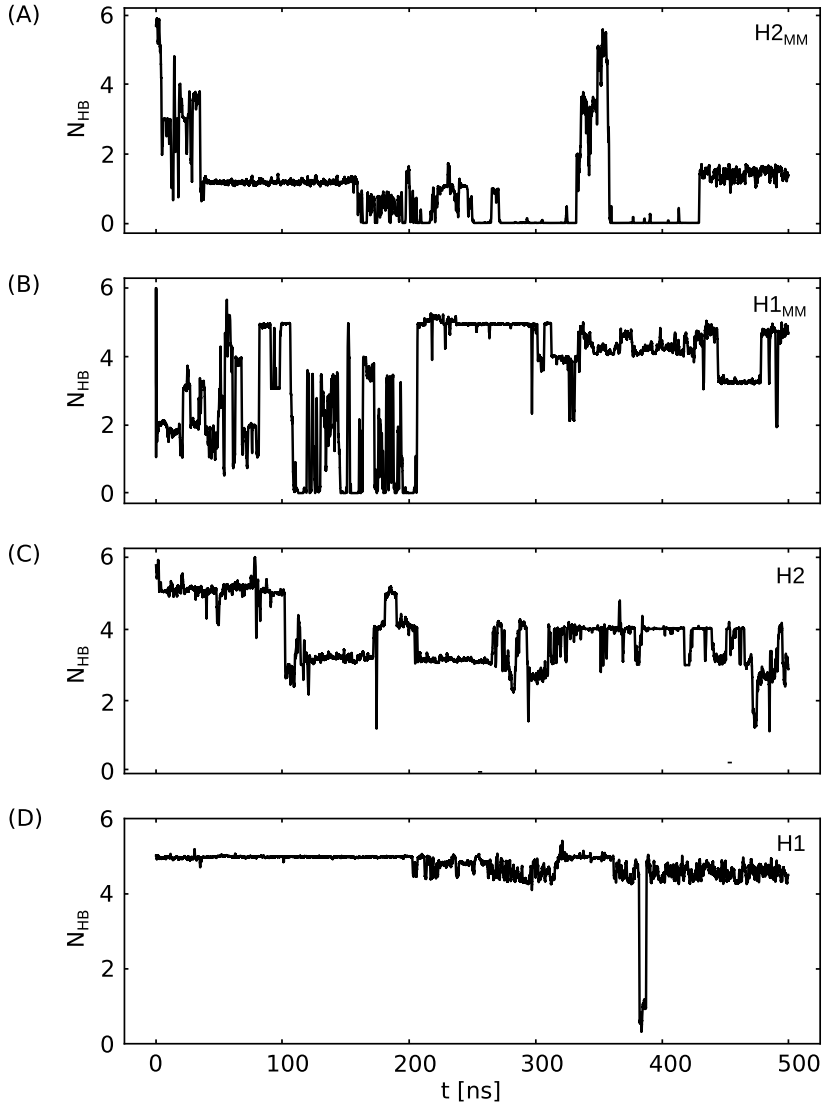

Figure 3.10: Order parameters describing the number of hydrogen bonds within hairpins simulated in all-atom simulations. (A) Order parameter calculated in a simulation involving the minimum-model hairpin formed between two middle guanosine repeats. (B) Same order parameter for minimum-model hairpin formed between two guanosine repeats along the 5′ end. (C) Order parameter for a full hairpin model formed between the two middle guanosine repeats and (D) order parameter for the full hairpin model formed between the two guanosine repeats along the 5′ end.

3.3.3 Conformational entropy is a hallmark of rG4 folding

The results from the TERRA25 system demonstrate that folding intermediates are highly degenerate, with each state comprising two to four structurally distinct conformations. This degeneracy leads to a branched, multi-pathway folding process that is driven by

conformational entropy. To explore if entropy-driven, multi-pathway folding is a general characteristic of rG4 systems, I extended the analysis to four additional rG4 systems: three synthetic sequences with varying loop lengths and a biologically relevant plant rG4 containing mixed loops (see Table 3.1). As a first step, I computed free energy profiles for each system and evaluated their relative stabilities.

The free energy landscapes projected onto the radius of gyration r_g exhibit qualitatively similar shapes across all five systems (Figure 3.11 A). As expected, the position of the first minimum shifts to larger r_g values with increasing loop length, consistent with more extended folded structures. Additionally, the height of the free energy barrier separating folded and unfolded states increases as loop length decreases, indicating that shorter loops stabilize the folded conformation and slow the kinetics of unfolding.

The most notable differences between the systems arise in the unfolded region of the free energy landscape. Specifically, short-loop sequences (e.g., U and UU) display narrower unfolded basins, reflecting a reduced number of accessible unfolded conformations and thus lower conformational entropy. In contrast, sequences with longer loops exhibit broader unfolded basins and higher free energy levels in the unfolded state.

The thermodynamic analysis reveals a general trend: rG4 stability decreases with increasing loop length (Figure 3.11 B). The results for UU, UUU, and TERRA25 show agreement with experimental data. For the U system, some deviation from experiment is observed, likely due to differences in loop sequence between the simulation model and the experimental construct.

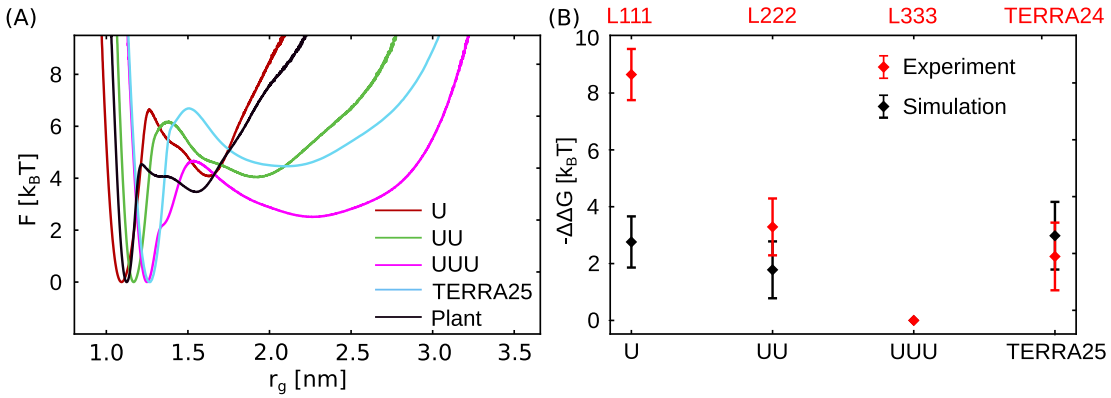

Figure 3.11: Free energy profiles and thermodynamic stabilities of different rG4 systems. (A) Free energy as a function of radius of gyration r_g. (B) Gibbs free energy changes from simulations and experiments. Experimental values are derived from van't Hoff analysis using published enthalpy and entropy data [127]. All Gibbs energies are reported relative to a reference system (UUU in simulations; L333 in experiments). Simulation errors were computed via block averaging; experimental uncertainties are from Ref. [127].

The two-dimensional free energy landscapes projected onto the composite order parameters s_1 and s_2 (Figure 3.12) provide further insight into the folding pathways of these systems. The order parameter clearly distinguishes between all possible TERRA25 intermediates, which are also depicted in the TIS-model representation in Figure 3.12. This analysis demonstrates that in order to resolve all intermediate states of rG4s, a more

complicated order parameter is needed then simply calculating end-to-end distances or values such as radius of gyration. We extended this analysis on the other simulated rG4 structures, to see if we can observe the same folding intermediates.

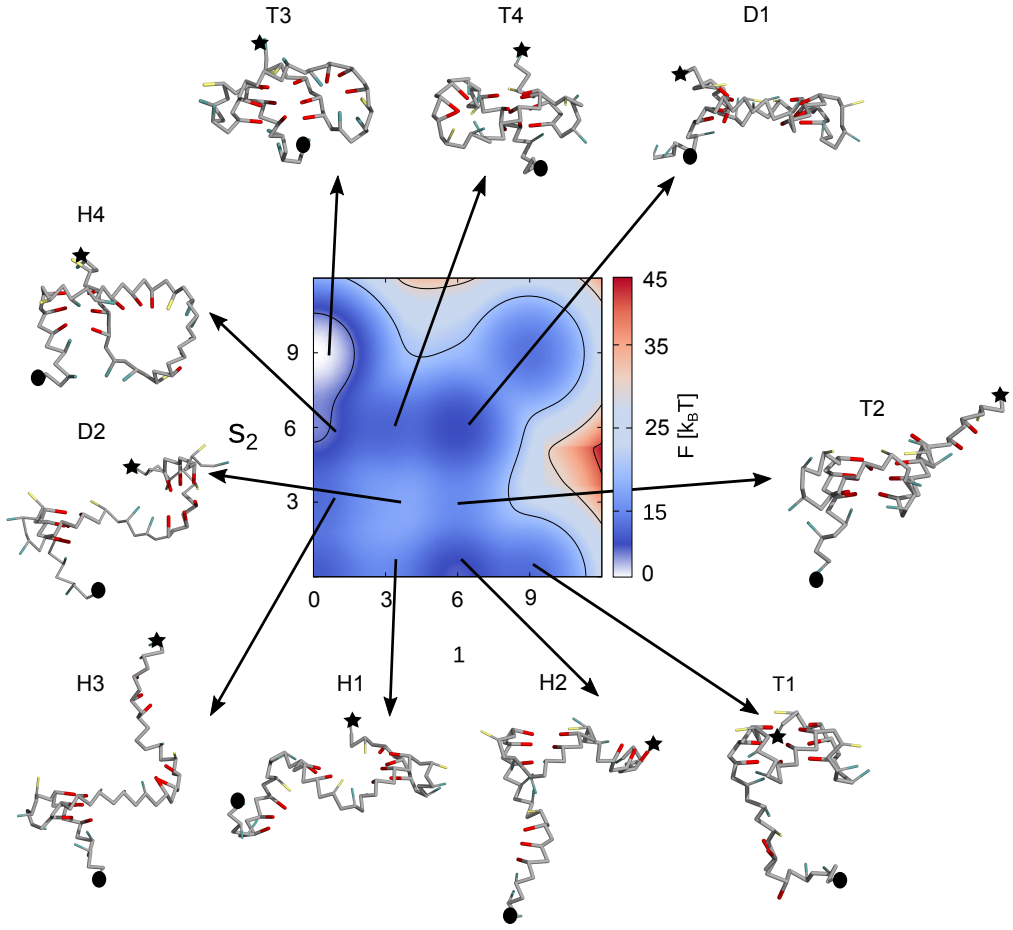

Figure 3.12: Free energy landscapes of TERRA25 folding projected onto the composite order parameters s_1 and s_2 (Eq. 3.8). Intermediate states are labeled and indicated on the landscape, as well as depicted in the TIS-model representation. Folded and unfolded free energy basins are normalized for visualization clarity.

In all five rG4 sequences, the folding landscape exhibits a branched topology with multiple parallel pathways and well-defined intermediate states (Figure 3.13). Notably, all 12 intermediate conformations identified in the TERRA25 system are also observed across the other systems. The population distribution among these intermediates, however, is not uniform and varies with the specific loop sequence and composition.

For the plant rG4, the hairpin intermediate H1 appears prominently and matches predictions from smFRET experiments [16]. While this agreement supports the relevance of the simulated pathways, experimental confirmation of other intermediates remains challenging with the smFRET due to the limited spatial resolution and label placement. Additional experimental work will be necessary to fully resolve the conformational ensemble of rG4 folding intermediates.

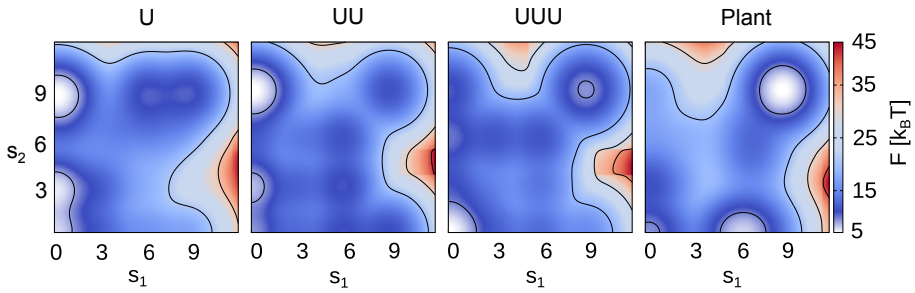

Figure 3.13: Free energy landscapes of folding projected onto the composite order parameters s_1 and s_2 (Eq. 3.8) for other simulated rG4s. The label of the simulated rG4 is written above each free energy profile. Folded and unfolded free energy basins are normalized for visualization clarity.

3.4 Conclusion

The folding kinetics of rG4s plays a pivotal role in regulating their structural function in cellular processes. In this work, I have resolved the folding pathways of a G4 structure derived from human telomeric repeat-containing RNA (TERRA25) through an integrative approach that combines all-atom MD simulations, CG simulations using the TIS model, and experimental CD measurements.

The results demonstrate that TERRA25 undergoes a branched, multi-pathway folding process, driven primarily by conformational entropy. This folding behavior is not unique to TERRA25. Analysis of four additional rG4 systems with varying loop lengths confirms that branched folding via multiple parallel intermediates is a general and intrinsic feature of rG4s. The findings highlight the central role of entropy in shaping the folding landscape of these non-canonical RNA structures.

To enable accurate CG simulations under physiological ionic conditions, we developed a concentration-matching protocol based on atomistic simulations in explicit solvent. This protocol captures the effect of the ionic double layer that forms near the highly charged RNA backbone. By matching the CG ion concentration to the local ion environment observed in atomistic simulations, the CG model reproduces the experimentally measured fraction of folded states across varying salt concentrations.

Because folding of rG4s occurs on the timescale of minutes [14], it remains inaccessible to all-atom simulations. However, the CG simulations used here reach these biologically relevant timescales, allowing for detailed resolution of the folding kinetics and intermediate states with statistically significant sampling. The simulations consistently reveal a hierarchy of on-pathway intermediates, including hairpins, triplexes, and double-hairpin states that are shared across all studied rG4 systems. Each of these intermediates encompasses multiple structural variants, contributing to their high conformational entropy.

Importantly, commonly used observables such as the end-to-end distance are insufficient to distinguish between these intermediate states. The simulation results presented here thus provide a valuable roadmap for guiding future experimental efforts, for example, by informing the design of single-molecule FRET experiments with strategically placed fluorophores [15, 16].

Looking forward, future studies may employ more detailed or hybrid models to further

explore off-pathway intermediates, including strand-shifted topologies or syn-anti guanine arrangements. Regardless of model complexity, the combined use of atomistic MD, CG simulations, and experimental validation represents a powerful framework for capturing the effects of ionic conditions and unraveling the conformational diversity and dynamics of rG4 folding.

3.5 Appendix

3.5.1 Data archive

The input files for the simulations described in this chapter are organized in the following directory tree. In the `Coarse-grained` folder are input files for the coarse-grained TIS simulations of TERRA G4 and other G4 models examined in this chapter. To run a TIS simulation e.g. of the seeded morphology at T=300 K, go to any directory, e.g. `./Coarse-grained/Plant` and run the command:

```
sbatch runfile.sh
```

In the `./Concentration_matching` folder are the input files to set up all-atom simulations for the concentration matching procedure, while in the `./Hairpin_models` folder can be found the input files for all-atom simulations of the hairpin intermediates on the TERRA G4 folding pathway. To run either simulation simply go to the desired folder, e.g. `./Hairpin_model/5prime` and run the following command:

```
gmx grompp -f mdp/run.mdp -c npt/npt.gro -p topol/topol.top
-n topol/index.ndx -o run.tpr
```

```
/hpc/gpfs2/home/g/cbio/Archive_maugrina/G4_folding
├── Coarse-grained
│   ├── RNA_cg-master
│   ├── Plant
│   │   ├── runfile.sh
│   │   ├── setup.sh
│   │   └── topol
│   ├── TERRA
│   │   ├── Concentration_25
│   │   │   ├── makescript.sh
│   │   │   └── start.sh
│   │   ├── Concentration_60
│   │   │   ├── makescript.sh
│   │   │   └── start.sh
│   │   ├── Pathway
│   │   │   ├── runfile.sh
│   │   │   └── setup.sh
│   │   ├── Temperature
│   │   │   ├── runfile.sh
│   │   │   └── setup.sh
│   │   └── topol
│   ├── U
│   │   ├── runfile.sh
│   │   ├── setup.sh
│   │   └── topol
│   ├── UU
│   │   ├── runfile.sh
```

```
│   │   ├── setup.sh
│   │   └── topol
│   └── UUU
│       ├── runfile.sh
│       ├── setup.sh
│       └── topol
├── Concentration_matching
│   ├── 0.48mM
│   │   ├── npt
│   │   │   └── npt.gro
│   │   └── topol
│   │       ├── index.ndx
│   │       └── topol.top
│   ├── 1.03M
│   │   ├── npt
│   │   │   └── npt.gro
│   │   └── topol
│   │       ├── index.ndx
│   │       └── topol.top
│   ├── 140mM
│   │   ├── npt
│   │   │   └── npt.gro
│   │   └── topol
│   │       ├── index.ndx
│   │       └── topol.top
│   ├── 16mM
│   │   ├── npt
│   │   │   └── npt.gro
│   │   └── topol
│   │       ├── index.ndx
│   │       └── topol.top
│   ├── 53mM
│   │   ├── npt
│   │   │   └── npt.gro
│   │   └── topol
│   │       ├── index.ndx
│   │       └── topol.top
│   ├── 82mM
│   │   ├── npt
│   │   │   └── npt.gro
│   │   └── topol
│   │       ├── index.ndx
│   │       └── topol.top
│   └── mdp
│       └── run.mdp
├── Hairpin_models
│   ├── 5prime
│   │   ├── mdp
│   │   │   ├── run.mdp
│   │   ├── npt
│   │   │   └── npt.gro
│   │   └── topol
│   │       ├── index.ndx
│   │       └── topol.top
│   └── middle
│       ├── mdp
│       │   └── run.mdp
│       ├── npt
│       │   └── npt.gro
│       └── topol
│           ├── index.ndx
│           └── topol.top
├── Plots
```

```
└── scripts
```

Chapter 4

Towards artificial ion channels: probing the ion specificity of TERRA G-quadruplex by enhanced sampling simulations

4.1 Introduction

Exchange of ions with the extracellular surrounding is essential for maintaining the biological functions of the cell like signaling, volume regulation, and metabolism. However, charged molecules are not able to traverse semipermeable hydrophobic bilayer cell membrane without ion transporters and ion channels. Ion channels, protein structures embedded in cell membranes, are responsible for the selective transport of charged molecules across the membrane [19, 132]. They control the flow of ions through the channel by switching between "open" conformation, which allows the ions to enter the channel, and "closed" conformation. This process is called "gating" and can be induced by trans-membrane potential, ligands or some other mechanical process [133]. They are very efficient so only a few thousand ion channel proteins per cell are enough to maintain the gradient of sodium and potassium ions across the cell membrane [132]. Faulty ion channels cause diseases like Alzheimer's, Parkinson's, cystic fibrosis etc. In turn, these diseases may be treated by incorporating functional ion channels into the cells. However, ion channels are structurally complex, have low stability and availability which is why their application in drug design is limited [19].

Scientific effort has been invested into the design of artificial ion channels that can mimic the effectiveness and selectivity of natural ion channels. Aside from being useful in treating channel dysfunction, artificial ion channels hold promise in cancer and antibacterial treatment, biosensing, and nanotechnologies [19, 20]. Two main types of ion channels are investigated: uni-molecular channels and supramolecular ion channels. The design of supramolecular ion channels is based on aggregation properties of small molecular compounds via non-covalent bonds. The goal is to design an ion channel that can be incorporated into the hydrophobic membrane (lipophilic), spans the length of the bilayer and possesses ion recognition sites. The supramolecular ion channels are designed for both cations and anions, but in this chapter the focus will be on cationic channels. The most common supramolecular cation channels are based on crown ethers, nanotubes, and self-assembled helical molecules [20]. The main cation binding sites are negatively charged groups pointing towards the central pore of the channel. The interaction between cations and channel binding sites has to include a good balance between binding affinity and permeability, leading to selectivity for a specific cation type. Ion channels with K^+ selectivity play an important role in many biological processes including controlling the action potential in neuronal and muscle cells, controlling the membrane potential, modulating hormone secretion, epithelial function etc. [134].

G4 nucleobases have emerged as a promising target for development of artificial ion channels due to the high selectivity towards cations [20, 135, 136]. Main building blocks of G4s are stacked planar quartets that consist of hydrogen-bonded guanine bases (see Chapter 1 for structural details). Inside or in between the quartets are cationic binding sites, consisting of partially negatively charged guanine O6 atoms. X-ray diffraction experiments first revealed the coordination of cations at the dG4 binding sites [21, 22]. Early NMR experiments on dG4a in NH_4^+ ion solution revealed that ions travel through the G4 channels and exchange with bulk [23, 24]. A series of enhanced sampling simulations has confirmed that the ions indeed travel through the dG4 channel pore, without exiting the walls on the side [65, 137]. The research also highlighted the dependence of G4 cationic conduction on the ionic radius and hydration free energy, as discussed in the introduction of this thesis. Interestingly, the simulations also showed the lowering of free energy barriers for NH_4^+ cations crossing the G4 channel binding site if additional ions are present in the channel, due to the electrostatic repulsion. Additionally, it confirmed the asymmetry for NH_4^+ ions exiting the channel on different ends of the dG4, as previously observed in NMR experiments [23].

First G4 ion channels were designed with hydrophobic guanosines as building blocks [138]. Further investigations revealed the ability of lipophilic G4s to self-assemble, with a central pore capable of selectively binding cations [135, 139, 140]. Elucidating the mechanism of G4 ion selectivity is a key area of current research. dG4s can adopt different conformations in dependence of cation in the solution [140]. Therefore, lipophilic dG4s are extensively used to study the ion selectivity, in particular regards to the intriguing K^+ over Na^+ selectivity. In a recent study, fluorescence-based vesicle assay in combination with voltage-clamp measurements and MD simulations were employed to explore the possibility of a dG4 to selectively transport K^+ ions [25]. An artificial ionophore was succesfully generated by inserting the G4 stacked with a lipophilic guanosine in a hydrophobic lipid membrane. Results revealed that the ionophore selectively conducts K^+ ions, induced by the K^+ concentration gradient.

Despite the majority of research being pointed at dG4s, their RNA counterparts are gaining traction due to the same functionality and higher stability than dG4s [18]. The additional hydroxide group in the ribose sugar affects guanosine χ dihedral angle, so rG4s adopt only the propeller loop parallel conformation [104]. They are cation channels with exclusively K^+ selectivity. Previous MD simulations revealed directionality of ion transport through TERRA G4, due to the non-canonical hydrogen binding of uracil bases at the 5' end of the quadruplex [141]. In this project, I used MD simulations to study the mechanism of ion selectivity of G4 TERRA. First, I performed straight-forward MD simulations to confirm the K^+ ion selectivity, after which I applied enhanced sampling methods to gain further insight into the thermodynamics of ion binding to the rG4.

4.2 Methods to analyze ion binding patterns and ion transport

The simulated structure was NMR-resolved structure of TERRA G4, with PDB code 2KBP (the same structure was used for G4-folding simulations in Chapter 3). It is a dimeric G4 structure with three guanine quartets and a sequence $(UAGGGUUAGGGU)_2$. To gain an initial insight into the mechanism of ion transport through the rG4, I first performed straightforward simulations of TERRA G4 in different ionic solutions. Then, I performed TREMD simulations, followed by constant force pulling simulations to gain insight into the ion specificity and ion channel properties of the TERRA.

Each simulation protocol was repeated for five sets of simulation systems, differing in the type of salt in the solution. Therefore, a separate simulation was performed for TERRA in KCl, NaCl, LiCl, CsCl and NH_4Cl solution. Amber99sb*-ildn force field [76] with parambsc0 [80] and χ_{OL} [79] corrections for TIP3P water was used to simulate TERRA, TIP3P model [123] for water and Mamatkulov-Schwierz parameters for ions [90]. Gromacs simulation package, version 2018.1 to 2021.5 was used for the MD simulations [122].

4.2.1 Straightforward MD simulations

In each simulation, TERRA was placed in a cubic box with an edge length of 7 nm, neutralized with 24 cations and solvated in a 1 M salt solution. Initial energy minimization was performed for each simulation run, followed by an NVT and NPT equilibration. NVT equilibration was performed using the Berendsen thermostat [124] and NPT equilibration was performed using the Berendsen barostat [125]. Each system was subjected to a production run at T=300K and for t=100 ns, using the velocity-rescaling thermostat with a stochastic term [124] and a time constant of 0.1 ps and Parrinello-Rahman pressure coupling [126] with a time constant of 0.5 ps. Exceptionally, the system containing KCl solution was simulated for t=1 µs.

In each simulation, distance along the z-axis (d_z^i) of ions from the center of mass of each quartet q_n was calculated to obtain an order parameter $\eta^i_{q_n}$. $\eta^i_{q_n}$ defines binding of the i-th ion to the n-th quartet and it was calculated using the equation 4.1.

$$\eta^i_{q_n} = \begin{cases} 1, & {d_z}^{ll} < d_z^i < {d_z}^{ul} \\ \frac{1}{1+e^{k\cdot(d_z^i - \delta d_z^i \cdot |d_z^b|)}}, & \text{otherwise} \end{cases} \tag{4.1}$$

Here, d_z^{ul} represent upper limit distance of ion from the quartet center of mass, dz^{ll} represents lower limit and $d_z^b = \frac{1}{2}({d_z}^{ul} + {d_z}^{ll})$, ${d_z}^{ul}$ =-0.025 nm, ${d_z}^{ll}$=-0.018 nm, k=5, δd_z^i=1.05.

4.2.2 TREMD simulations

TREMD simulations were performed to gain a better insight into the energetics of ion interactions with TERRA. For each system, 100 temperature replicas were simulated, ranging in temperature from T=300 K to T=400 K. Exchanges between replicas were

attempted every 500 ps.

Each system was placed in a cubic box with 7 nm edge length, neutralized and filled with 1 M salt solution. Replicas were subjected to energy minimization runs, and then to NVT equilibration using the Berendsen thermostat [124]. Harmonic position restraints with the force constant of 1000 kJ mol^{-1} nm^{-1} were applied to the heavy atoms of TERRA during the equilibration to prevent large conformational rearrangements. Production runs were performed for t=100 ns, using the velocity-rescaling thermostat [124] with a time constant of 0.1 ps for temperature coupling.

4.2.3 Pulling simulations

Constant force pulling simulations involved the following setup: in each system, a cation was placed on top of the first quartet (q_1) of TERRA G4, like in Figure 4.1. The distance between the ion and q_1 is 0.2 nm. Energy minimization run was performed for each system, followed by an NVT equilibration using the Berendsen thermostat [124] and NPT equilibration using the Berendsen barostat [125]. To ensure that atoms of TERRA and the ion sitting above q_1 remain in the same position during the equilibration process, I applied harmonic position restraints with a force constant of 200000 kJ mol^{-1} nm^{-1} to all heavy atoms of TERRA and the ion.

As mentioned previously, five systems of TERRA were simulated, differing in the type of ionic solution. 10 replicas of each system were created for the pulling simulations. The replicas were selected from the TREMD simulation runs at T=300 K, specifically selecting "open" conformations at least 2 ns away from each other. The open conformation of TERRA G4 is here considered as the state in which the uracils U12 and U24 capping the q_1 quartet point away from the quartet (Figure 4.1), so that the cation can fit on top of q_1 without steric hindrance. Additionally, 10 different starting frames were selected from the NPT equilibration simulations at distances of 500 ns from each other to perform pulling runs. This resulted in 100 pulling simulations for all of the systems, yielding a sufficient amount of data for statistical calculations.

Each system was subjected to a production "pulling" run at T=300K, using the velocity-rescaling thermostat with a stochastic term [124] and a time constant of 0.1 ps and Parrinello-Rahman pressure coupling [126] with a time constant of 0.5 ps. Pulling simulations were performed using the constant force pulling procedure, and force constants of 450 and 500 kJ mol^{-1} nm^{-1}. Additionally, to gain insight into the difference in ionic conduction mechanism of TERRA in NaCl and KCl solutions, pulling simulations were performed for these ions at force constants of 200, 300 and 350 kJ mol^{-1} nm^{-1}. The ion was pulled along the z-axis through the central pore binding sites (Figure 4.1), from the top of the tetrad q_1 (5' end) to the "exit point", which was 1.6 nm below the last tetrad (3' end). Once the ion reaches the exit point, it does not interact with the TERRA atoms anymore, reducing the amount of friction countering the pulling force. Therefore, it quickly travels a distance larger than the maximum distance in the periodic system, which is one-half of the box length, and the simulation is terminated.

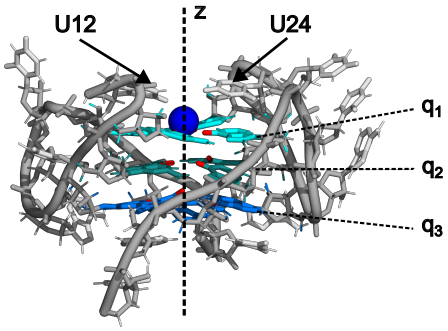

Figure 4.1: Initial set-up for constant-force pulling simulations. Three quartets are indicated as q_1, q_2 and q_3. The ions were pulled along the z-axis from the entrance of q_1 to the exit of the channel pore, below q_3.

Analysis of ion binding and transport trough rG4

From the straightforward simulations, I calculated the order parameter to indicate the binding of cations to the TERRA binding sites. The order parameter is defined by the equation 4.1.

In the TREMD simulations, the probability density function (PDF) was calculated based on the distances of the ions from the central quartet q_2 of TERRA. The free energy profiles were calculated using the Boltzmann inversion.

In the pulling simulations, the average passage time was calculated according to the following equation:

$$t_{\mathrm{avg}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} t_{\mathrm{i}}}{\mathrm{N}}, \tag{4.2}$$

where N=100 is the total number of simulations. Passage time t_{i} is defined as the time it takes an ion to travel from the entrance point of the TERRA to the exit point.

4.3 Results and discussion

4.3.1 Straightforward simulations provide an initial insight into the ionic transport

Straightforward simulations provide a qualitative insight into the binding of cations to TERRA G4 quartets. In Figure 4.2, an order parameter is calculated, representing the binding of cations to the three quartets of TERRA G4. Next to them, a snapshot of the final simulation frame is shown. In a simulation with NaCl solution, (Figure 4.2 A), two cations are found inside the channel, bound to quartets q_2 and q_3. Order parameter provides an insight into how these two ions bind and travel through the channel. First, one Na^+ cation binds to the quartet q_1, which is indicated by the order parameter η_1 having the value of one. At around t=5 ns, η_1 changes to zero, meaning that the ion is no longer bound to the quartet q_1, while at the same time η_2 becomes one, indicating the binding of Na^+ to quartet q_2. Eventually at t=20 ns, another Na^+ cation binds to q_1 and stays there until t=80 ns when it moves to q_2. Simultaneously, the first Na^+ cation moves from q_2 to

q_3. Therefore, in this simulation Na^+ cations travel collectively through the channel. On average there are N=1.8 ± 0.009 Na^+ ions in the channel during the simulation.

In Figure 4.2 B, the same is represented for the TERRA G4 simulation in KCl solution. Here, almost immediately at the beginning of the simulation, one K^+ ion binds to q_1. K^+ stays bound to the q_1 for much longer in comparison to Na^+, almost 200 ns. After that it moves to q_2, where it stays bound for the entirety of the simulation. Only at around t=900 ns, another cation binds to q_1. On average less ions are present in the channel in the KCl simulation (N=1.2 ± 0.008), and in comparison, to Na^+ cations, K^+ cations travel slower through the channel staying bound longer to a single binding site.

In a simulation in LiCl solution (Figure 4.2 C), one Li^+ ion binds to the quartet q_1 at around t=10 ns. This cation stays bound to q_1 for the rest of the simulation time and no other ion enters the channel. Therefore, there is on average N=0.9 ± 0.006 Li^+ ions in the channel. Interestingly, in a CsCl simulation (Figure 4.2 D) one Cs+ cation binds to the quartet q_1 and moves to q_2 after t=10 ns. It stays bound to q_2 for the rest of the simulation time while no other Cs^+ cations bind to the channel, resulting in the average number of ions in the channel N=1.0 ± 0.004. Finally, in NH_4Cl simulation (Figure 4.2 E), one NH_4^+ cation binds immediately to q_2. More NH_4^+ cations bind later to q_1, but do not remain bound for the entire length of the simulation. Average number of NH_4^+ cations in the channel is N=1.3 ± 0.007.

rG4s have a selectivity to K^+ cations. These results give a qualitative insight into binding of different cations to a TERRA G4. Na^+ cations travel quickly and collectively through the channel, however, the quartets must adjust to allow for the cations to go through. Due to the change of the intramolecular interactions in this process, this might affect the stability of the structure. K^+ ions bind longer to the quartets than Na^+ ions. Due to the larger size than Na^+ ions, K^+ ion is able to position itself closer to the center of the channel, which is a more favorable position to screen the partial negative charges of the O6 atoms. All of this contributes to the higher stability of the rG4 in KCl solution. Li^+ ions have a small van-der-Waals-radius, which means a higher charge density and higher hydration free energy. It is therefore possible that Li^+ ions travel through the rG4 channel with additional waters in their hydration shell which is electrostatically unfavorable. Cs^+ ion binds to the second quartet q_2 and no other ion is able to bind to the q_1. Possibly the size of the Cs^+ ion adjusts the conformation of q_1, and therefore changing the energetics of the binding site, making it unfavorable to the cation binding.

NH_4^+ and K+ ions are often compared due to their similar solvation properties [142]. Interestingly, due to their similar ionic radii, NH_4^+ cations are used as a probe for K^+ cations in ion selectivity NMR experiments [143]. These simulations show that NH_4^+ binding to the rG4 channel has similarities with K^+ binding where one ion stays permanently bound to the q_2 while other cations bind temporarily to q_1. However, comparing on the same timescale (t=100 ns), higher average number of cations is present in a simulation with KCl solution.

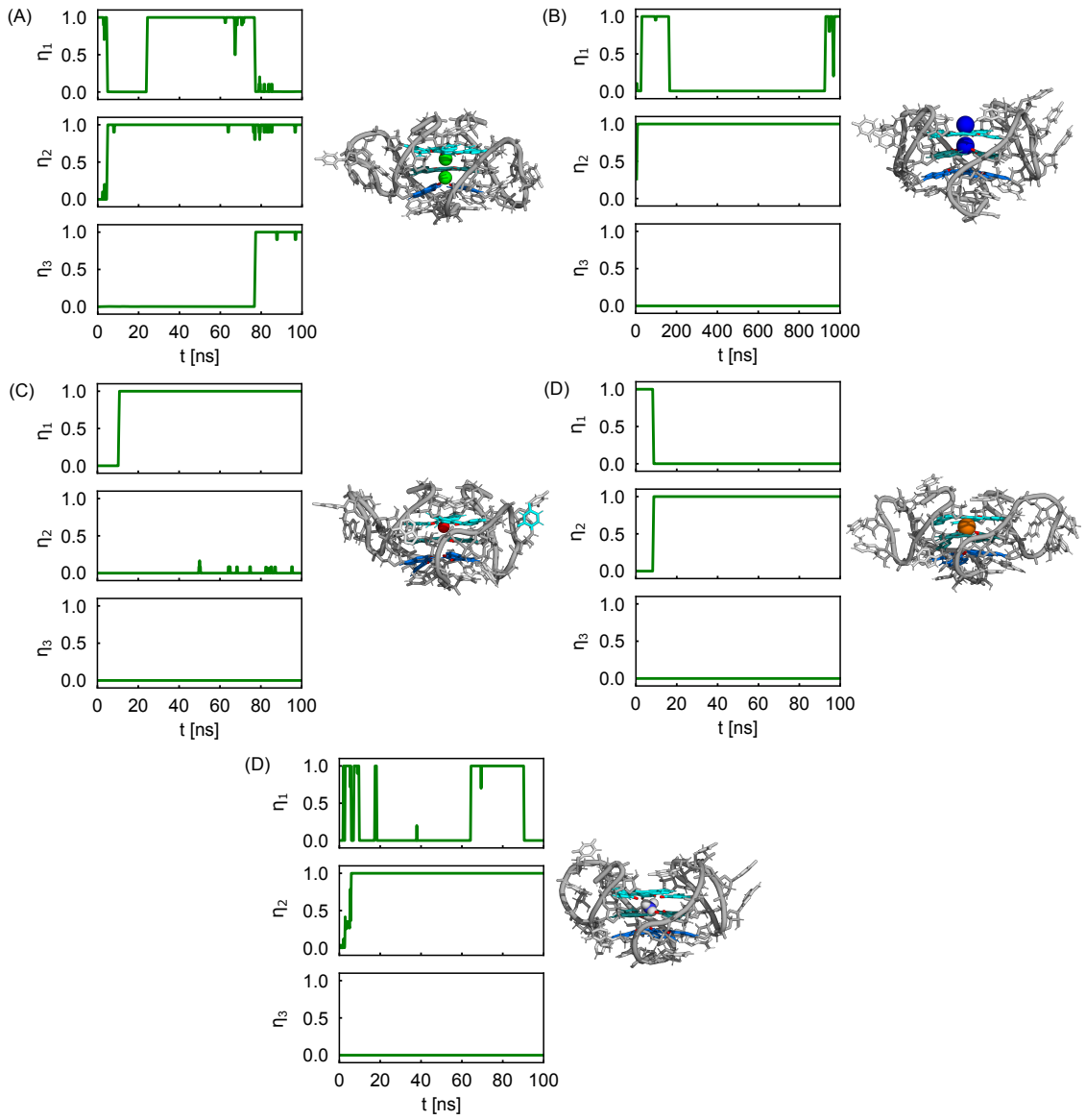

Figure 4.2: Order parameters $\eta^{i}_{q_n}$ describing the binding of (A) Na^+, (B) K^+, (C) Li^+, (D) Cs^+ and (E) NH_4^+ cations to quartets q_n, n=1,2,3. On the right side of the plots are snapshots from the individual simulations, taken at the last simulation frame.

Figure 4.3 shows RMSD of the TERRA G4 central pore for each of the simulated ionic solutions, calculated with respect to the central pore of the experimental structure. The simulation of the TERRA G4 in LiCl solution reveals a significant increase in RMSD, suggesting a lower stabilizing effect of Li^+ ions due to their relatively small ionic radius. Furthermore, potential presence of water molecules within the hydration shell of Li^+ may modulate the electrostatic interactions within the pore, which could further compromise conformational stability. RMSD of the central pore in simulations with the other ionic solutions remained similar in this set of simulations.

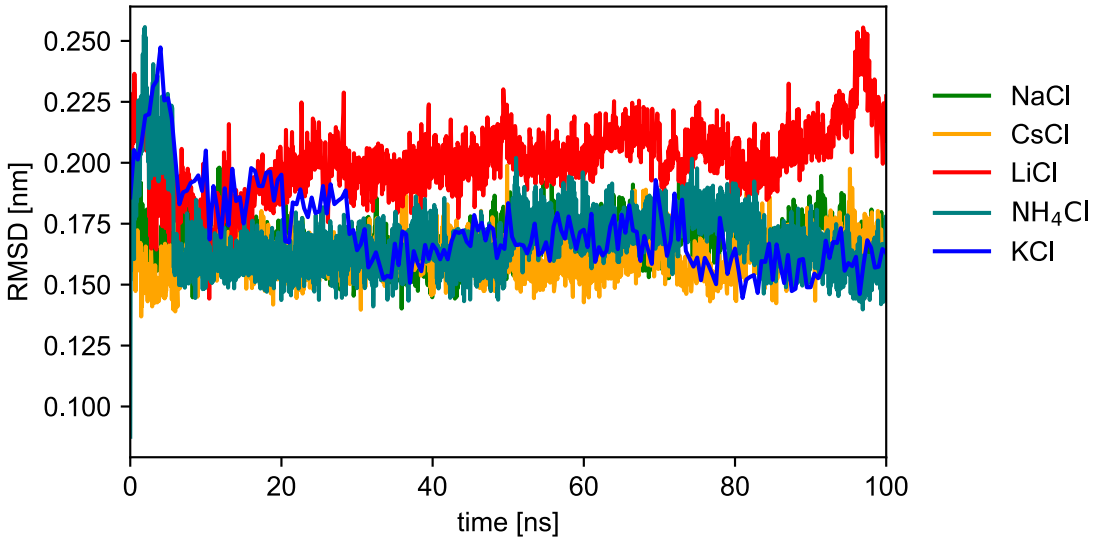

Figure 4.3: RMSD of the TERRA G4 central pore calculated from the simulations in NaCl, KCl, CsCl, LiCl and NH_4Cl solutions. In the KCl simulations, data were collected every 500 ps, whereas in simulations with other ionic solutions, data were collected every 50 ps.

4.3.2 TREMD simulations reveal energetics of cation binding to TERRA G4

I performed TREMD simulations of the TERRA G4 for each of the five ionic solutions. Figure 4.4 shows the free energy profiles for ionic transport along the z-axis through the central pore derived from the replica simulated at T=300 K. The barriers that weren't resolved extend outside of the plots.

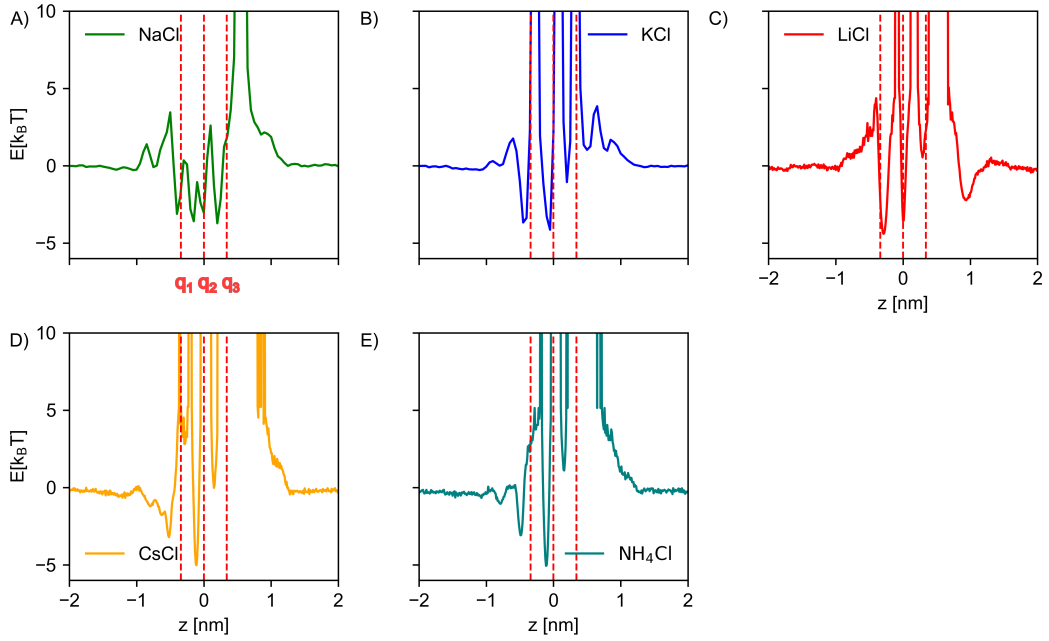

Figure 4.4: Free energy profiles of cations along the z-axis of the TERRA G4 central pore. The simulations were performed in A) NaCl, B) KCl, C) LiCl, D) CsCl and E) NH_4Cl solutions. Quartets q_1, q_2 and q_3 are indicated with red vertical lines in the plots.

Free energy profile in the NaCl solution has one unresolved barrier at the ion exit point from the central pore. The highest resolved barrier is located at the channel's entrance, and it is just above 5 k_BT. Energy barriers are elevated in the transition from q_3 to q_1, emphasizing the directional nature of the system. Overall, according to the free energy profile, Na^+ ions have the highest mobility within the rG4 channel.

Free energy profile in the simulation with KCl solution has unresolved barriers between each of the tetrads. Interestingly, K^+ ions experience lower barrier than Na^+ ions at the channel entrance, but higher barriers within the channel. According to the previous research, K^+ ions have the highest binding affinity to the guanosine O6 binding site at the quartet center in comparison to the other simulated ions [49]. Therefore, it binds faster and stays at the binding sites longer than other ions, stabilizing the rG4 structure. Lower barrier for transition from q_3 to bulk than q_1 to bulk also confirms the directional nature of the ion transport through TERRA G4, which is most likely caused by interactions with overhanging uracils at the 5' end of TERRA.

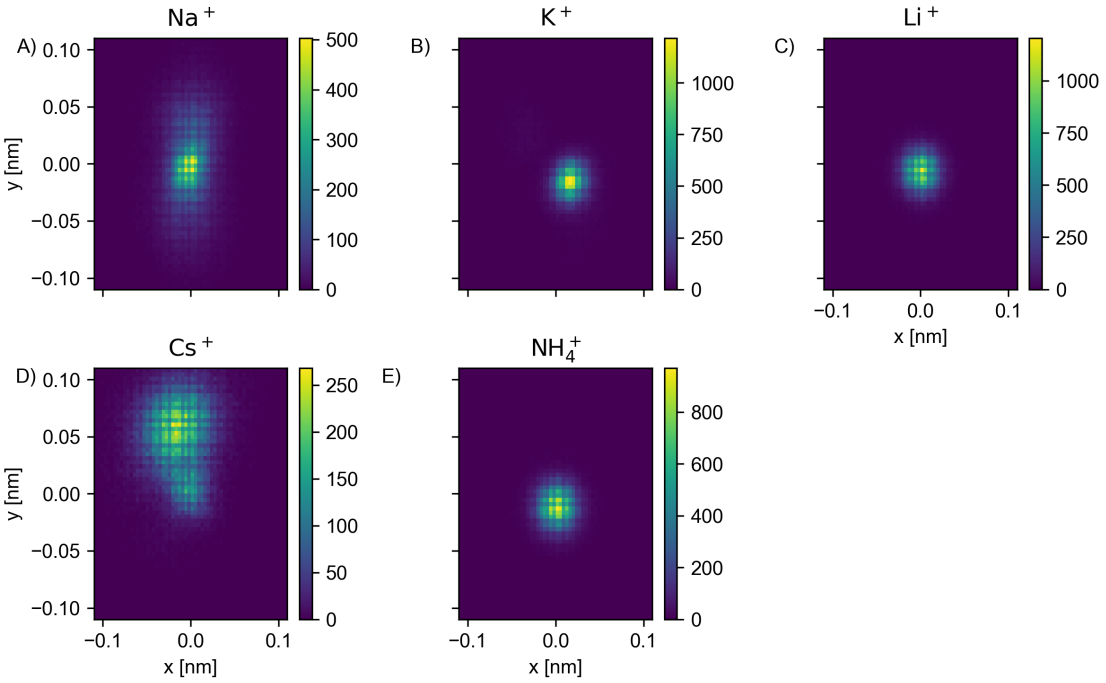

Figure 4.5: 2D histograms of cationic locations in the TERRA G4 central pore according to the xy-plane perpendicular to the direction of the cationic flow. The histograms were calculated from TERRA G4 simulations in A) NaCl, B) KCl, C) LiCl, D) CsCl and E) NH_4Cl solutions.

Free energy profiles in simulations in other ionic solutions have unresolved barriers at each transition between the quartets. Results for simulation in LiCl solution differ from the rest because all of the minima sit at the same location as the quartets. This indicates that the Li^+ binding site is in the quartet, due to its small size, and is in agreement with the results from previous simulation studies [67]. Arguably, this would make the transport of the ion through the channel more difficult as it is not able to interact with multiple quartets at the same time. Furthermore, Li^+ presents the highest barrier among the simulated ions

when transitioning from q_1 to the bulk in the direction opposite to the z-axis.

In simulations in CsCl and NH_4Cl solutions, most of the barriers remained unresolved, indicating that these ions have the lowest mobility through the central pore in comparison to other simulated ions.

From the same set of simulations, 2D histograms of ion location according to the xy-plane perpendicular to the channel direction is shown in Figure 4.5. K^+, Li^+ and NH_4^+ are positioned closest to the center of the xy-plane, while presenting the least amount of dispersion. On the other hand, Na^+ and, interestingly, Cs^+ ions present a dispersed 2D density along the x-axis, suggesting that the deviation of the ionic positions from the center of the xy-plane does not negatively affect the transport of ions through the channel.

Number of water molecules in an ion's hydration shell may significantly affect its transport through the TERRA G4 channel, due to the unfavorable interaction of partially negatively charged water oxygen with O6 atoms of the guanosines. Additionally, ions need to release waters from the hydration shell in order to interact with the O6 atoms of guanines and bind to the quartet sites. Therefore, Figure 4.5 shows number of waters in each of the ion's hydration shell at different positions along the z-axis. Unresolved parts of the z-axis extend outside of the plots. Each ion reduces number of waters in the hydration shell as it travels closer to the center of the channel so it can bind with O6 atoms of the TERRA G4 guanosines. Na^+ and K^+ ions release all the waters from the hydration shell at a few points along the z-axis. Li^+, Cs^+ and NH_4^+ ions retain the highest number of water molecules while traveling through the channel, never having less that one water in the hydration shell, which might therefore hinder their transport through the channel.

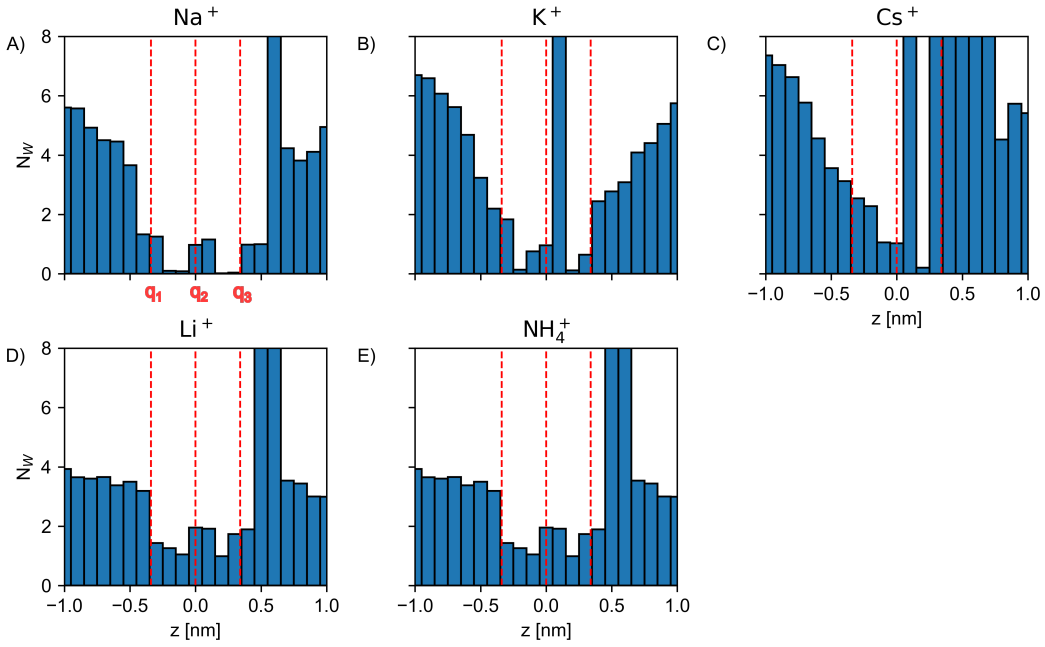

Figure 4.6: Histograms representing number of water molecules in cationic hydration shells along the z-axis of the TERRA G4 channel pore. The histograms were calculated from TREMD simulations of the TERRA G4 in A) NaCl, B) KCl, C) CsCl, D) LiCl and E) NH_4Cl solutions.

4.3.3 Force-dependent ionic conductivity and selectivity in TERRA G4 from pulling simulations

I conducted constant force pulling simulations to evaluate the ionic conductivity through the TERRA G4 central pore. In this context, the conductivity is inversely proportional to the mean first passage time (MFPT (τ)) of an ion traversing from the entrance of the TERRA G4 central pore to its exit. τ was calculated as the average over 100 constant-force pulling simulations using force constants of 450 kJ mol^{-1} nm^{-1} and 500 kJ mol^{-1} nm^{-1}. To gain a more detailed insight into the conductivity difference of K^+ and Na^+ ions, additional simulations were performed for them at 200 kJ mol^{-1} nm^{-1}, 300 kJ mol^{-1} nm^{-1} and 350 kJ mol^{-1} nm^{-1}.

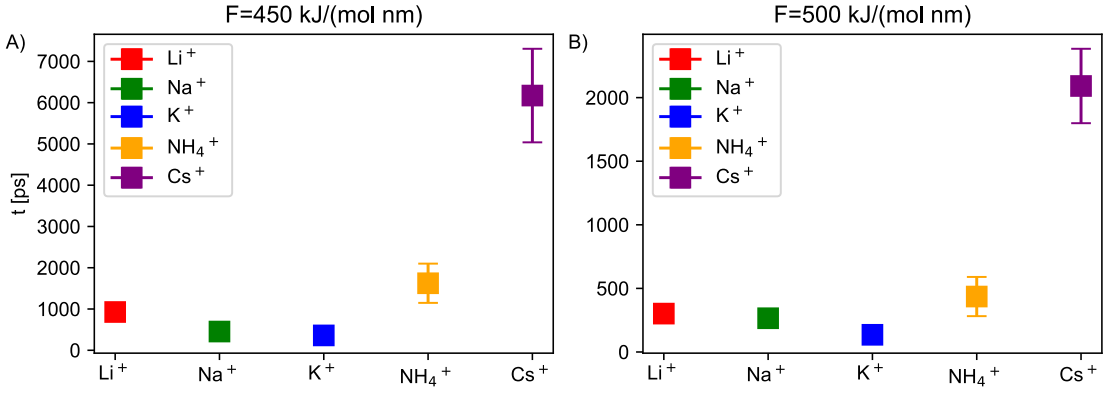

Figure 4.7: Mean first passage times (τ) of cations along the TERRA G4 central pore under the influence of the pulling force with force constants of A) 450 kJ mol^{-1} nm^{-1} and B) 500 kJ mol^{-1} nm^{-1}.

The results for the τ of all ions represented in Figure 4.7 confirm the results from previous straightforward and TREMD simulations. It takes Cs+ ions the longest time to travel through the pore. At a force of 450 kJ mol^{-1} nm^{-1}, τ=6.1 ± 1 ns, and at the force constant of 500 kJ mol^{-1} nm^{-1}, τ=2.1 ± 0.3 ns. It is followed by the NH_4^+ and Li^+ ions, as expected from the results of previous simulations. Finally, MFPTs of the the K^+ and Na^+ ions are comparable at higher force constants, down to 350 kJ mol^{-1} nm^{-1} (Figure 4.8). At 300 kJ mol^{-1} nm^{-1}, MFPT of the Na^+ ion (τ=8027 ± 1271 ps) is shorter in comparison to the K^+ ion (τ=11027 ± 1256 ps).

Recently, an electrophysiology measurement study was conducted to explore the potential of a human telomeric DNA to selectively transport ions across a lipid membrane in a stimuli-responsive manner [25]. Specifically, an endogenous human telomeric G4 DNA (h-TELO) was inserted into the lipid membrane by a synthetic guanosine-derivative MG containing a lipophilic side chain. The experiments and MD simulations from the work showed that the dG4 preferentially transports K^+ ions in comparison to Na^+, Li^+, Cs^+ and NH_4^+ ions due to the increased structural stability of the molecule in their presence. Namely, as previously mentioned, dG4 are able to adopt their structure in response to the type of ions present in the solution. Moreover, K^+ ions perpetuate the formation of parallel-stranded dG4 [140]. Since TERRA G4, like all rG4s, has a parallel-stranded conformation, the results from my simulations should align with the study, which is not the

case for higher force constants.

However, interesting results come from the simulations at lower force constants. While the MFPT of the K^+ ion is higher compared to the Na^+ ion at force constant of 300 kJ mol^{-1} nm^{-1}, this changes when the force constant is set to 200 kJ mol^{-1} nm^{-1}. Here, MFPT of the K^+ ion (τ=38997 $\pm$ 9398 ps) becomes significantly shorter than that of the Na^+ ion (τ=53026 $\pm$ 17535 ps). These results indicate that there exists an additional mechanism that promotes the transport of K^+ ions, depending on the strength of the applied external potential. However, the results at lower potentials indicate that the closer the conditions get to the in-vivo conditions (smaller external potentials resembling the trans-membrane potentials in cells), TERRA G4 preferentially transports K^+ ions.

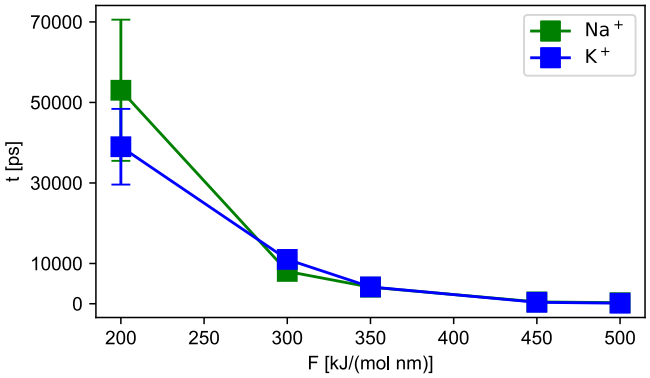

Figure 4.8: MFPT of K^+ and Na^+ ions along the TERRA G4 central pore as a function of the force constants.

To better compare the present findings with those from the h-TELO study, a current-voltage (I-V) plot for K^+ and Na^+ is presented in Figure 4.9 A. While both voltage-clamp experiments and MD simulations in the h-TELO work reveal a linear relationship between ionic conductance and voltage, the TERRA simulations in this study exhibit an apparently exponential dependence. Moreover, the voltages applied in the h-TELO research are on the order of millivolts, whereas the forces used here, when translated into voltage, are at least two orders of magnitude greater. However, for smaller voltages, the I-V dependence in figure 4.9 B appears linear.

The discrepancy of this work and the one on h-TELO can be attributed to several key differences in experimental and simulation conditions. First, dG4s can adopt various conformations depending on the ionic environment, whereas rG4s predominantly form a parallel-stranded propeller loop structure, regardless of ion type. Second, in this study, strong positional restraints (500 kJ mol^{-1} nm^{-1}) were applied to the phosphate atoms of the TERRA G4, limiting its conformational flexibility and ability to adapt according to interactions with the ions in it's channel binding sites. In contrast, h-TELO is embedded in a membrane via a lipophilic guanosine, which allows for greater mobility of the backbone atoms.

Furthermore, in the h-TELO experiments and simulations, the applied electric field influences not only the G4 itself but also surrounding molecules. This may create a cumulative effect on ion transport and facilitate structural adjustments in the overhanging

nucleosides, thereby reducing energy barriers for ion conduction.

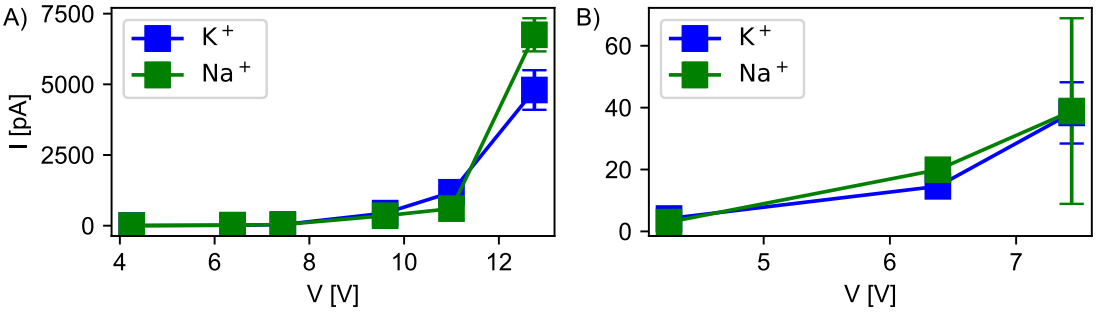

Figure 4.9: A) I-V plot for K^+ and Na^+ conductance through the TERRA G4 ion channel. B) Same plot from A), but for smaller voltages only, to highlight the differences in Na^+ and K^+ conductance.

Comparison of travel time across different points of the G4 pore at a pulling force of 350 kJ mol^{-1} nm^{-1} shows a difference in the way that Na^+ and K^+ ions interact with the atoms inside the central pore. K^+ ions spend most of the MFPT in the vicinity of q_2, after it enters the channel. Na^+ ion travels fast across q_3 and spends the largest amount of time in the vicinity of q_3. This difference is also reflected in the free energy profiles from TREMD simulations. Previous simulation studies have demonstrated that the flanking nucleotides interact with the ions upon exchange with the bulk [144]. Additionally, asymmetry of exchange with bulk at different G4 channel openings was observed in NH_4Cl simulations and experiments [23, 65, 145], as well as in TERRA MD simulations [141]. Possibly interaction with uracils at the channel exits hinders the sodium ions from exiting the channel, which is not the case for K^+ ions. This effect might be smaller at higher force constants, but might become more prominent with the decrease of the biasing potential.

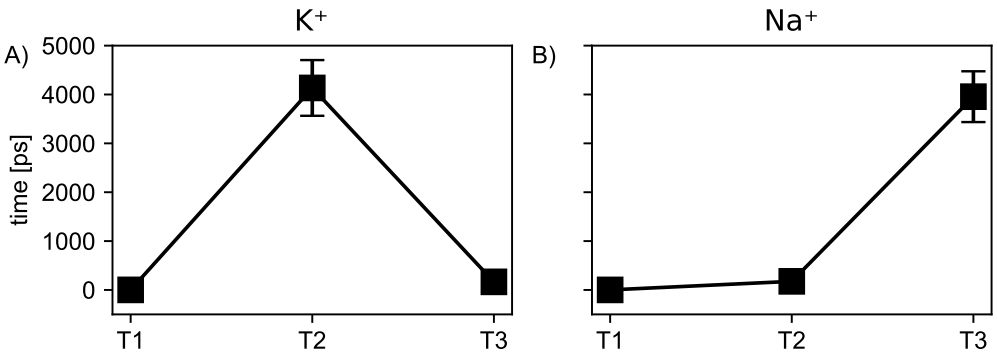

Figure 4.10: Average time that A) K^+ and B) Na^+ ions spend above quartets q_1, q_2 and q_3 indicated as T1, T2 and T3, respectively.

4.4 Conclusion

Potassium ion channels play a crucial role in many biological processes, and therefore their use in drug-design is an important research topic [134]. However, protein ion channels have limited use in this field because of their structural complexity and low stability, so the main focus is on designing artificial ion channels [19]. RNA G4s are an appropriate target for this application due to their stability and high K^+ selectivity.

I performed straightforward, TREMD and constant-force pulling MD simulations to gain insight into the mechanism of K^+ ion selectivity. Straightforward simulations yielded an initial insight into the mechanism of ion transport through TERRA G4 and ion selectivity. The results for order parameter that measured the transport of ions along the z-axis of the TERRA G4 central pore highlighted an interesting difference in transport of Na^+ ions with respect to the rest of simulated ionic solutions. While the rest of the ions traverse individually through the channel, Na^+ simulation showed that two ions traversed collectively through the TERRA G4 binding sites. RMSD calculations revealed a decreased conformational robustness of TERRA in the presence of Li^+ ions, which is likely connected to its coordination of the O6 atoms.

Free energy profiles calculated from TREMD simulations show high barriers for the transport of ions between different quartet binding sites, many of which were not resolved. Na^+ ions exhibit the least high barriers overall, however there is an unresolved barrier at the channel end, likely due to interaction of the ion with the overhanging uracils. Free energy minima that indicate the locations of the binding sites of K^+, Cs^+ and NH_4^+ ions are between the locations of quartets, but for Na^+ and Li^+ ions they are located closer to the quartet planes. Additionally, the 2D profiles along the quartet planes show that the Na^+ and Cs^+ ions are more dispersed around the center of the channel in comparison to the other ions. Finally, the results for number of waters in the ions' hydration shells indicate that Na^+ and K^+ ions contain the smallest number of waters in the hydration shell with respect to other ions, therefore making it easier for them to interact with the guanosine O6 atoms.

All together these simulations provide an important insight into the difference between simulated cations in transport mechanism through the TERRA G4. Final results from the pulling simulations indicated an interesting difference in MFPT between Na^+ and K^+ ions with respect to the applied external potential. While Na^+ ions travel faster through the channel at higher force constraints, this isn't the case at lower force constraints where K^+ ions have a shorter MFPT. This force-dependent inversion of MFPT demonstrates that high-force simulations can mask natural selectivity, highlighting the importance of probing a range of forces to infer physiological behavior. Additionally, the MFPT of Na^+ and K^+ ions differs across the different quartet binding sites. It would be interesting to resolve the barriers of the free energy profiles calculated by TREMD simulations by using a more appropriate enhanced sampling technique, such as umbrella sampling. Specifically, the difference in free energy barriers across different quartets, especially in the case of Na^+ and K^+ ions warrants further investigation. In this manner, I attempted to perform umbrella sampling simulations using K^+ ions. However, due to difficulties in retaining the ionic positions within the boundaries of the umbrella windows even at high force constants (up to 200000 kJ mol^{-1} nm^{-1}), these simulations weren't successful. Additionally, comparison of the I-V plots for TERRA and the previous work on h-TELO [25] indicates differences in the energetics of ion interactions with rG4s and dG4s at channel binding sites. These results suggest that a comprehensive understanding of the ion channel properties of rG4s requires further investigation.

4.5 Appendix

4.5.1 Data archive

Input files for the simulations described in this chapter are organized in the following directory tree. In the `Pulling_simulations` folder are input files for the constant force pulling simulations. As previously mentioned, there is 10 `run` files with 10 `npt.gro` files in each, from 10 time points of the NPT equilibration. Each of the runs can be started with the gromacs command written further below. In the `./Straightforward` folder are the input files for straightforward simulations of the system in different salt solutions. To start any of the simulations, go to any of the folders, e.g. `./Straightforward/KIO` and run the following command.

```
gmx grompp -f mdp/run.mdp -c npt/npt.gro -p topol/topol.top
-n topol/index.ndx -o run.tpr
```

In the `./TREMD` folder are the input files to set-up the TREMD simulations of the RNA G4 in the NaCl solution, serving as an example. To run the simulations, go to the `./TREMD/INPUT/NIO` folder and run:

```
sbatch run.sh
```

In the `./TREMD/RESULTS` folder are the trajectories of the replicas at T=300 K, since the simulations are computationally demanding.

```
/hpc/gpfs2/home/g/cbio/Archive_maugrina/Ion_channel
├── Pulling_simulations
│   ├── KIO
│   │   ├── mdp/
│   │   │   └── run.mdp
│   │   ├── run_1/
│   │   │   ├── npt1.gro
│   │   │   ├── npt2.gro
│   │   │   ├── npt3.gro
│   │   │   ├── npt4.gro
│   │   │   ├── npt5.gro
│   │   │   ├── npt6.gro
│   │   │   ├── npt7.gro
│   │   │   ├── npt8.gro
│   │   │   ├── npt9.gro
│   │   │   └── npt.gro
│   │   ├── ...
│   │   ├── run_9/
│   │   └── topol/
│   │       ├── index.ndx
│   │       └── topol.top
│   ├── AMO   (same structure as KIO)
│   ├── CIO   (same structure as KIO)
│   ├── LIO   (same structure as KIO)
│   └── NIO   (same structure as KIO)
├── Straightforward
│   ├── KIO
│   │   ├── mdp/
│   │   │   └── run.mdp
│   │   ├── npt/
│   │   │   └── npt.gro
│   │   └── topol/
```

```
│   │       ├── index.ndx
│   │       └── topol.top
│   ├── AMO   (same structure as KIO)
│   ├── CIO   (same structure as KIO)
│   ├── LIO   (same structure as KIO)
│   └── NIO   (same structure as KIO)
└── TREMD
    ├── INPUT
    │   └── NIO
    │       ├── mdp/
    │       ├── nvt/
    │       ├── run.sh
    │       └── topol/
    └── RESULTS
        ├── KIO
        │   ├── repl_0
        ├── AMO   (same structure as KIO)
        ├── CIO   (same structure as KIO)
        ├── LIO   (same structure as KIO)
        └── NIO   (same structure as KIO)
```

Chapter 5

Misfolding of proteins into amyloid fibrils

5.1 Introduction: Misfolding of proteins into fibrils

The following chapter address fundamental questions concerning the structural determinants and molecular mechanisms of amyloid fibril formation in proteins. Amyloid fibrils are highly ordered protein aggregates with characteristic cross-β sheet architecture, and their deposition in tissues is implicated as a cause of many neurodegenerative and systematic diseases [69]. Despite their common structural motif, fibrils can differ markedly in morphology, stability, and biological activity, and these variations are closely linked to disease onset, progression, and pathology [146]. Understanding how specific sequence features and mutations influence fibril structure is therefore essential for elucidating the molecular basis of amyloid-associated diseases.

The following research was performed under the supervision of Prof. Nadine Schwierz at the University of Augsburg, and the experimental work was performed by members of the Prof. Marcus Fändrich group at the Institute of Protein Biochemistry, Ulm University and their collaborators. In the first part, seeded and unseeded Aβ fibrils are compared, highlighting the structural differences that arise from distinct aggregation pathways and underscoring the conformational diversity of amyloid assemblies. It was published in:

Pfeiffer P B, **Ugrina M**, **Schwierz N**, Sigurdson C J, Schmidt M, Fändrich M. "Cryo-EM Analysis of the Effect of Seeding with Brain-derived Aβ Amyloid Fibrils". *J Mol Biol*, 2024, 436(4):168422.

The second part shifts focus to apolipoprotein A-II, where naturally occurring sequence variants are examined to reveal how specific amino acid substitutions can confer resistance to fibril formation and thereby modulate amyloidogenic propensity. It was published in:

Andreotti G, Baur J, **Ugrina M**, Pfeiffer P B, Hartmann M, Wiese S, Miyahara H, Higuchi K, **Schwierz N**, Schmidt M, Fändrich M. "Insights into the Structural Basis of Amyloid Resistance Provided by Cryo-EM Structures of AApoAII Amyloid Fibrils". *J Mol Biol*, 2024, 436(4):168441.

The third part investigates lysozyme-derived fibrils associated with hereditary systemic amyloidosis and demonstrates how a single point mutation can significantly affect fibril stability at the atomic level. It was published in:

Karimi-Farsijani S, Sharma K, **Ugrina M**, Kuhn, L, Pfeiffer P B, Haupt C, Wiese S, Hegenbert U, Schönland S O, **Schwierz N**, Schmidt M, Fändrich M. "Cryo-EM structure of a lysozyme-derived amyloid fibril from hereditary amyloidosis". *Nat Commun 15*, 2024,

9648.

Together, these studies integrate cryo-EM and MD simulations to establish a direct link between sequence variation, fibril morphology, and conformational stability.

5.2 Methods: MD simulations of fibrils

The simulations were performed using the GROMACS simulation package [122], version 2020.7. Periodic boundary conditions were used, and the Coulombic interactions were calculated using the Particle-mesh Ewald method [147].

Each structure was neutralized and simulated in a cubic box containing 0.15 M NaCl solution. The fibril structures were simulated using parameters from the Amber99sb-star-ildn force-field [62], while the TIP4P-EW water model [89] and Mamatkulov-Schwierz ionic parameters [90] were used for the solution.

All the systems were energy-minimized and equilibrated. First the NVT equilibrations were performed using Berendsen thermostat [124], followed by NPT equilibrations using the Berendsen barostat [125]. The production runs were performed at a constant number of particles (N), pressure (P) and temperature (T). The velocity-rescaling thermostat [124] was employed for the temperature coupling in the production runs, with a time constant τ=0.1 ps, while the Parrinello-Rahman barostat [126] with τ=5 ps was used for pressure coupling.

5.3 Results 1: Comparison of ex-vivo and in-vitro Aß amyloid fibril structures using cryo-EM and MD simulations

It is debatable what course of action takes place during the formation of amyloid fibrils, but the rate-limiting step is certain to be the nucleation process. Here, the critical amount of ordered beta strands stack to form the nucleus of the fibril. This process is energetically unfavorable since the energy gained from association of strands doesn't compensate for the entropy lost in this process. Further polymerization is faster because it is energetically more favorable [28]. However, if the nucleus is not big enough for polymerization to occur, it will dissolve. Polymerization can be accelerated by adding a preformed nucleus to the solution of monomeric precursor proteins [28]. Since the nucleus determines the properties of the fibril, the fibrils grown by seeding have the same structural properties as the seed.

Aβ peptides deposit extra cellular plaque of insoluble and highly ordered fibrillar structures in the brain tissue, and are linked to the Alzheimer's disease [28]. They are derived by a sequential process of α- or γ-secretase-mediated cleavage of amyloid precursor protein (APP) and consist mainly of 40-42 amino acids (Aβ_{1-40} and Aβ_{1-42})[148]. Multiple studies carried out on Aβ fibrils elucidated the role of seeding in fibril formation and increased fibril propagation rate [149–151]. Hence, the process of seeding could have a significant impact on the dissemination of amyloidosis within an organism and the transmission of the disease between animals [29] and may therefore be an important target for drug development [151].

Interestingly, in-vitro formed Aβ fibrils have different structural properties then ex-vivo brain-derived ones [29]. In this project, Aβ fibrils were extracted from the meningeal tissue of an Alzheimer's diseased donor and used for seeding in a solution containing containing Aβ_{1-40} peptide. The resulting structure was compared to fibrils formed in a solution without seeding. The unseeded fibrils exhibited polymorphism with two most common morphologies, Morphology I and Morphology II, differing in their cross-over distance. The two morphologies also differ in the number of peptide stacks, where Morphology I contains two peptide stacks, while Morphology II contains four peptide stacks. One morphology of seeded fibril was extracted, revealing a decrease in the lag phase due to the presence of the seed [29].

This research confirmed significant structural difference between ex-vivo and unseeded fibrils, which implies the role of environmental conditions or ligand binding in formation of the in-vivo Aβ fibrils. Namely, ex-vivo fibrils adopt a right-handed conformation while the conformation of unseeded fibrils is left-handed. Additionally, the structural properties of the ex-vivo seed were proliferated in the seeded fibril morphologies.

Moreover, cryo-EM and the calculation of interaction energies along the fibril axis highlighted the higher stability of the seeded fibrils in comparison to the unseeded fibrils. I used MD simulation to further investigate the stability difference of the seeded and unseeded fibril morphologies.

5.3.1 Methods for simulating Aß amyloid fibrils

I performed straightforward MD simulations of the seeded fibril and the Morphology I and Morphology II of the unseeded fibrils. The initial structures for the simulations were provided by cryo-EM with PDB entries 8OT1 (unseeded Morphology I), 8OT3 (unseeded Morphology II) and 8OT4 (seeded morphology). The standard protonation state was set for each of the amino acids, except for histidine. Two versions of each structure were created, each containing either only protonated or only deprotonated histidines. All structures were simulated at three different temperatures (T=300 K, T=350 K, and T=400 K) except from the structures containing only protonated histidines, which were simulated only at room temperature (T=300 K). A 200 ns production run was performed for each system.

I calculated the RMSD for each of the production runs, using the respective experimental structure as a reference.

5.3.2 Results and discussion: molecular determinants of stability in seeded and unseeded Aß fibrils

Cryo-EM structures of seeded and unseeded fibril morphologies are represented in Figure 5.1 (A). Both morphologies contain two stacked protofibrils consisting of six peptide chains. The chains in the protofibrils are stacked parallel to each-other over an intricate network of backbone hydrogen bonds. The seeded morphology is more stable, as more of the structure was resolved (Figure 5.1 (B)), and it contains five β-sheets, numbered from 1 to 5 in Figure 5.1 (A) and (B). Unseeded morphologies had more flexible N- and C-terminal ends which therefore weren't resolved (Figure 5.1 (C)). They contain three β-sheets, numbered

from 1-3. Peptide chains in the seeded morphology stack in a right-handed twist while the unseeded morphologies have a left-handed twist.

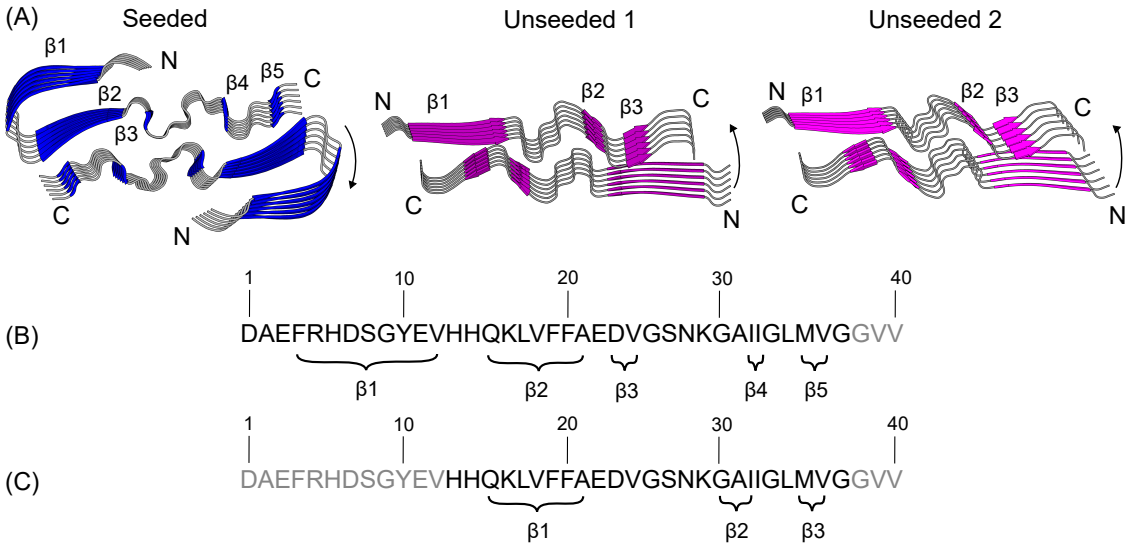

Figure 5.1: (A) Cartoon representation of the Cryo-EM-resolved structures of seeded and unseeded morphologies. The seeded morphology consists of five β-sheets per chain, indicating its higher stability. Unseeded morphologies 1 and 2 both consist of three beta sheets per chain. The right- and left- handed twists of the structures are indicated by an arrow. (B) Sequence of seeded fibril structure, unresolved terminal ends are indicated in grey. (C) Sequence of unseeded fibril structures, unresolved terminal ends are indicated in grey

Figure 5.2 shows qualitative representations of electrostatic charge surface of two stacked chains in each of the morphologies. In both seeded and unseeded morphologies, hydrophilic residues are on the outer part and exposed to solvent. In contrast, neutral hydrophobic residues are protected from the solvent, near the peptide stack. The seeded morphology contains polar contacts at the protofibril contact site between valine and asparagine residues. The contacts in the interface region of the unseeded morphologies are hydrogen bonds between glycine and leucine.

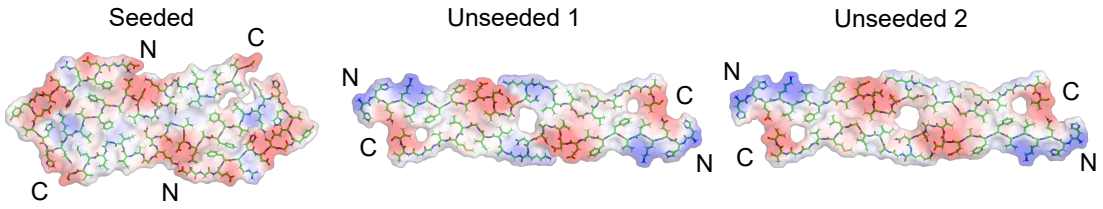

Figure 5.2: Cross sections representing two interacting chains from two protofibrils of simulated β-fibril morphologies. Average charge density is qualitatively represented to show positively charged (red), negatively charged (blue) and neutral (white) regions of the surface charge.

After 200 ns simulation time at three different temperatures, RMSDs of the final structures were calculated with respect to the appropriate experimental structures. They are shown in Figure 5.3 (B), together with snapshots of structures after 200 ns simulation at T=400 K, superimposed with respective experimental structures (Figure 5.3 (A)).

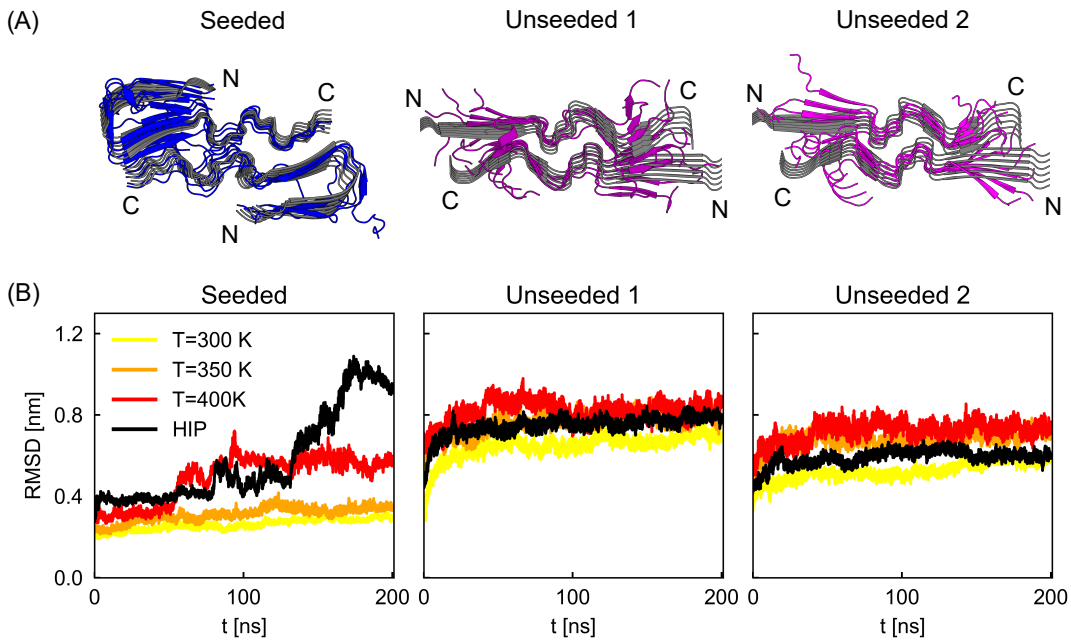

Figure 5.3: (A) Structures of simulated β-fibril morphologies: seeded morphology, unseeded Morphology I and unseeded Morphology II are represented in cartoon representation and colored blue, purple and magenta, respectively. They are superimposed with the respective structures from cryo-EM experiments that are represented in gray. (B) RMSDs of simulated structures at temperatures T=300K, T=350K and T=400 K are shown in yellow, orange and red color, respectively. RMSDs of structures simulated with protonated histidines are shown in black (HIP).

The seeded morphology has the lowest RMSD showing the highest conformational robustness. This is corroborated by the snapshots from the simulation at T=400 K in Figure 5.3 (A), as the structure remained similar to the starting structure, without larger conformational rearrangements. The RMSDs at temperatures T=300 K and T=350 K remain within 0.4 nm, which is within the error of the experimental measurement. Only at T=400 K does the RMSD increase, but is still within 0.5 nm and lower than those of unseeded morphologies. RMSDs of unseeded morphologies range from 0.5 nm for T=300 K simulation, to 0.8 nm for T=400 K simulation, indicating significant conformational rearrangements. It is evident from the superposition snapshots at T=400 K, that structures of the unseeded morphologies deviate from the initial positions. Most notable are deviations of C- and N- terminal ends, which divert from the starting structure the most, and point towards solvent (Figure 5.3 (A)). Interestingly, full N- and C- terminal ends of the unseeded morphologies weren't resolved due to their high flexibility, and the new terminal ends of the unseeded morphologies again show high flexibilities in the simulations. The seeded morphology structure, however, shows high conformational deviations when the protonation state of histidine is changed, with the rise of RMSD to 1.2 nm. Therefore, the changed charge distribution may affect the structural integrity of the entire macromolecule. In same way, the environmental conditions can have an impact on the structural stability of the fibrils. The histidines are located and rotated towards the N-terminal arch between β-sheets β1 and β2, therefore they significantly affect the hydrophobic packing in that area. Changes in the protonation state didn't affect the stability of the unseeded morphologies,

as they are anyhow located at the N-terminal ends of peptides and are solvent exposed.

5.3.3 Summary and conclusion

Fibrils formed in the presence of a seed have the same structural properties as the seed. Therefore, the difference between a fibril formed by seeding and the unseeded one formed by nucleation can be determined on the base of the structural differences. In this study our collaborators used cryo-EM microscopy and other methods to gain insight into the structural differences between Aβ amyloid fibrils formed in the seeding process and the unseeded fibrils formed by nucleation [29]. The results revealed stark structural difference in the seeded and unseeded fibril structures like different cross-over distances and number of peptide stacks forming the fibrils. Unseeded fibrils also folded into multiple morphologies, of which two were the most dominant.

I used MD simulations to simulate the seeded morphology and the two unseeded morphologies of the cryo-EM resolved structures and to compare their structural deviations from the reference cryo-EM structures. RMSDs measured after the simulations showed that unseeded morphologies adopted high conformational fluctuations, while the seeded morphology remained close to the starting structure, even at the temperature of 400 K. However, change of the protonation state of histidine that was buried in the N-terminal arch decreased the structural integrity of the seeded morphology and it showed high structural deviations. The same wasn't true for the unseeded morphologies, since the N-terminal part of the structure wasn't resolved in the experiments due to their instability. Therefore, the simulations corroborated the experimental findings and PDBePISA calculations.

The dependence of the seeded morphology on the protonation state of its histidine residues leads to the conclusion that some other environmental factors may affect the histidine protonation state, such as neutral-basic pH environment or an anionic co-factor that a diffuse density in cryo-EM experiments indicates [29]. However, no such co-factor was resolved so far even though the literature points to glycosaminoglycans that have been known to interact with the Aβ-fibril structure [152–154].

5.4 Results 2: MD simulation study into structural properties of amyloid resistance

Specific mutations in amyloid precursor proteins are related to protection against amyloid fibril formation, and this property is known as amyloid resistance [30]. Studies have shown that amyloid resistance is a property of some amyloidogenic proteins such as prion proteins, amyloid-β precursor proteins and serum amyloid A proteins [155–157]. Prion proteins are present in nervous cells and are a misfolded form of PrP proteins [158]. They are implicated in neurodegenerative diseases like kuru, Creuzfeldt-Jakob and Gerstmann-Sträussler-Scheinker diseases [30]. However, mutations at certain positions result in amyloid resistance and protection against the diseases. For example, Gly127Val mutation of the prion protein protects against kuru [155]. Serum amyloid A protein is present in the liver inflammatory mechanism and its misfolding can cause systematic AA amyloidosis [157].

There are two different variants of the protein, SAA1.1 and SAA2.2 [159]. The SAA2.2 variant encodes mutation Gln30Leu which suppresses amyloidogenesis and promotes amyloid resistance [159]. The Ala673Thr mutation in the amyloid-β precursor protein yields protection against Alzheimer's disease [156].

Our collaborators investigated the case of the amyloid resistance of apolipoprotein A-II (ApoA-II) [30]. This protein is located in the plasma and implicated in HDL cholesterol metabolism, whereas misfolded forms cause systemic AApoAII amyloidosis [160]. The amyloidosis affects humans as well as animals, in particularly mice. Studies have been made on mice that express two variants of ApoA-II protein, ApoA-IIC and ApoA-IIF. The results showed that mice which express ApoA-IIC variant develop amyloidosis in a large number of cases, and mice that express ApoA-IIF variant don't, even if the ApoA-IIC-seed is inserted [161]. Moreover, the difference in amyloid resistance is prescribed to 4 mutations between the two variants, Ser9Asn, Gln16His, Arg54Lys and Asn62Lys, but the impact of the mutations isn't specified. In this study, our collaborators used cryo-EM spectroscopy to resolve the 3-D structure of ApoA-IIC variant of ApoA-II protein, extracted from diseased mice. I used MD simulations to simulate the cryo-EM resolved ApoA-IIC fibril. To gain insight into the impact of each of the mutations on the stability of fibrils, I simulated five additional homology models of ApoA-IIC, including each of the four mutations, as well as the ApoA-IIF model.

5.4.1 Methods for simulating ApoA-II amlyoid fibrils

Cryo-EM experiments yielded three ApoA-IIC morphologies, of which Morphology I constituted 75%. I performed straightforward MD simulations of ApoA-IIC Morphology I, as well as five additional homology models based on ApoA-IIC. In the first one, I created the ApoA-IIF model by introducing all four mutations to the ApoA-IIC: Ser9Asn, Gln16His, Arg54Lys and Asn62Lys. To identify which of the four mutations present in the ApoA-IIF variant introduces the largest instability, I created four additional homology models based on ApoA-IIC, each containing only one mutation. The PDB Manipulator tool of the CHARMM-GUI web-interface was used to introduce the mutations [162]. A 200 ns production run was performed for each system.

I calculated RMSDs from each of the production runs, using the respective experimental or homology modeled structure as a reference. RMSFs per residue were also calculated for the ApoA-IIF structure, and represented as a B-factor dependent color gradient on average coordinates of specific chains. Additionally, I used the simulation rerun method implemented in Gromacs to calculate individual protein-protein and protein-solvent interaction energies. The averages are calculated using block averaging over five equal blocks for the error calculation.

5.4.2 Results and discussion: Impact of Point Mutations on the Structural and Energetic Stability of ApoA-II Amyloid Fibrils

Our collaborators resolved the 3D structure of ApoA-IIC fibril using cryo-EM spectroscopy, and the predominant morphology is Morphology 1, represented in Figure 5.4 (A). It consists

of two stacked protofibrils. Each protofibril consists of six chains stacked along the fibril axis. Each chain contains eleven β-sheets, numbered in Figure 5.4 (A) from 1 to 11 and indicated in the sequence in 5.4 (C). The arches in the chains are formed by a combination of hydrophobic packing and polar interactions between charged side-chains.

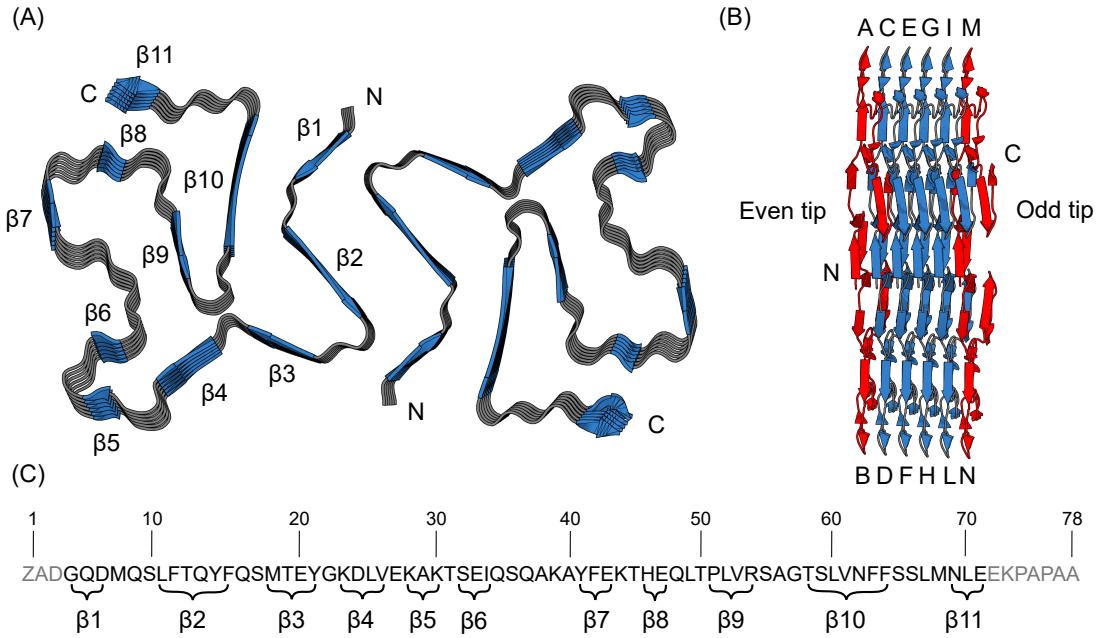

Figure 5.4: (A) Cryo-EM structure of ApoA-IIC fibril Morphology I. (B) Top view of the cryo-EM structure of ApoA-IIC fibril Morphology I with chains indicated from A to N. The outer chains are colored red, and even and odd end are indicated depending on the terminal end that is solvent exposed in the outer chains. (C) Sequence of ApoA-IIC amyloid fibril, with beta sheets indicated from 1-11. Unresolved terminal ends are indicated in gray color.

Chains that form the amyloid fibrils aren't stacked to fit into a plane along the fibril axis. Instead, C- and N- terminal ends protrude at the opposite side along the fibril axis, which is called "wrapping". Therefore, even and odd fibril ends are defined in dependence of the solvent-exposed terminal end which protrudes outside of the best-fit plane. Figure 5.4 (B) shows twelve chains of the ApoA-IIC fibril, while the outer chains are indicated in red with the indication of whether they form the even or the odd tip.

Interactions between chains from two different protofibrils come from tight packing of the residues as well as salt bridge pairs, for example Gln16 and Asp6 or Gln13 and Gln13. There are four mutations in ApoA-IIF fibril with respect to ApoA-IIC fibril, as shown in Figure 5.5 (B) and (C). Two of them change the charge of the mutated residue, Gln16His and Asn62Lys. Gln16 is located in the stack between two chains while Asn62 is located in a chain core (Figure 5.5 (A)). Gln16 and Arg54 are also bound in salt bridges.

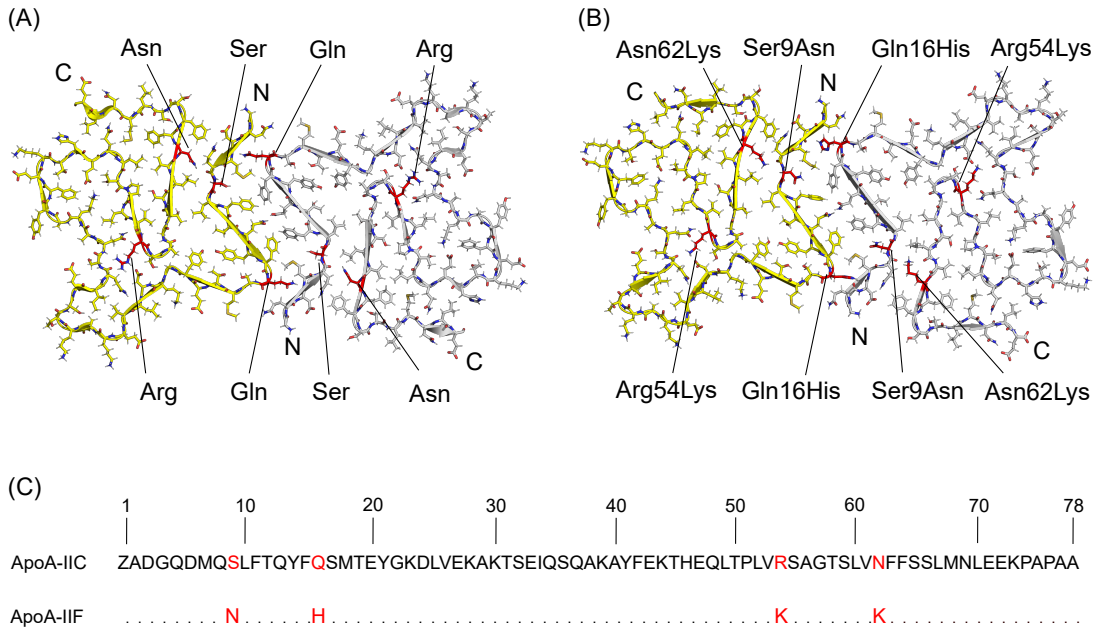

Figure 5.5: (A) Cross-section of ApoA-IIC and (B) ApoA-IIF amyloid structures, representing the mutated residues in red. (C) Aligned sequences of ApoA-IIC and ApoA-IIF fibrils, with mutations indicated.

After 200 ns simulation time, RMSD values were calculated for each simulated model, with the respective experimental or homology modeled structure as a reference. Results confirm the predictions: the ApoA-IIF fibril structure is less robust than the ApoA-IIC structure and the RMSDs reflect these results (Figure 5.6 (A)). However, the interesting takeaway arises from simulations with independent mutations. RMSDs in Figure 5.6 (A) are similar for ApoA-IIC structure and the three ApoA-IIC-based homology models which contain the mutations Ser9Asn, Gln16His, and Arg54Lys. Of these mutations, only the Gln16His changes the charge of the system. However it is located near the solvent, and presumably therefore doesn't affect the conformational stability of the structure. The other two mutations do not change the charge and by size are similar to the residue that was mutated, so they don't cause steric clashes. The Asn62Lys mutation causes the highest RMSD, which corroborates the experimental findings that this mutation is responsible for the amyloid resistance in mice [161]. This mutation replaces neutral asparagine amino acid with positively charged lysine.

Therefore, the hydrophilic lysine might be destabilizing the hydrophobic chain core. Additionally, in the previous case of Aβ-fibril simulations, replacing neutral histidine residues reduced the stability of the fibril structure, demonstrating how charge distribution largely affects the stability of the fibril fold, which largely relies on hydrophobic interactions within the core. Therefore, mutations that change the charge of residues located in the protofibril core have a significant impact on the stability of the structure. The ApoA-IIF structure shows high structural deviations from the initial ApoA-IIC model (Figure 5.6 (B)). Fluctuations can be clearly observed at the N-terminal ends of the fibril, near the fibril stack, where the residues are positioned away from the Lys62.

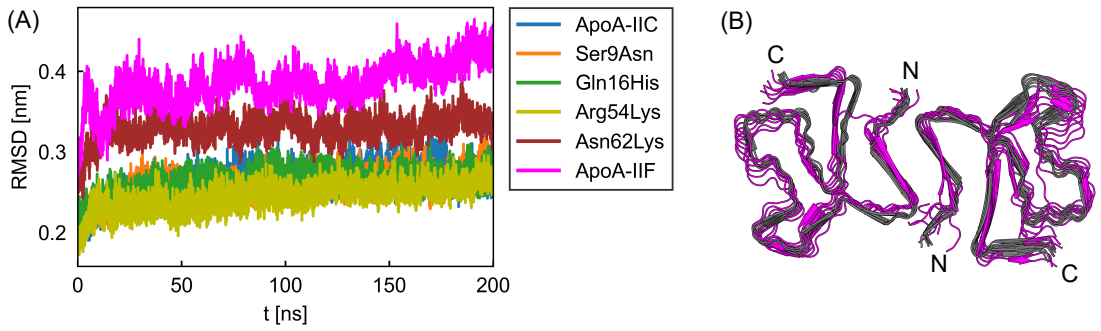

Figure 5.6: (A) RMSDs calculated from 200 ns simulations of ApoA-IIC structure and homology models. Respective experimental or homology modeled structures are used as a reference. (B) Superposition of ApoA-IIF amyloid fibril after t=200 ns simulation time (colored magenta) with the experimental structure of ApoA-IIC fibril (colored gray).

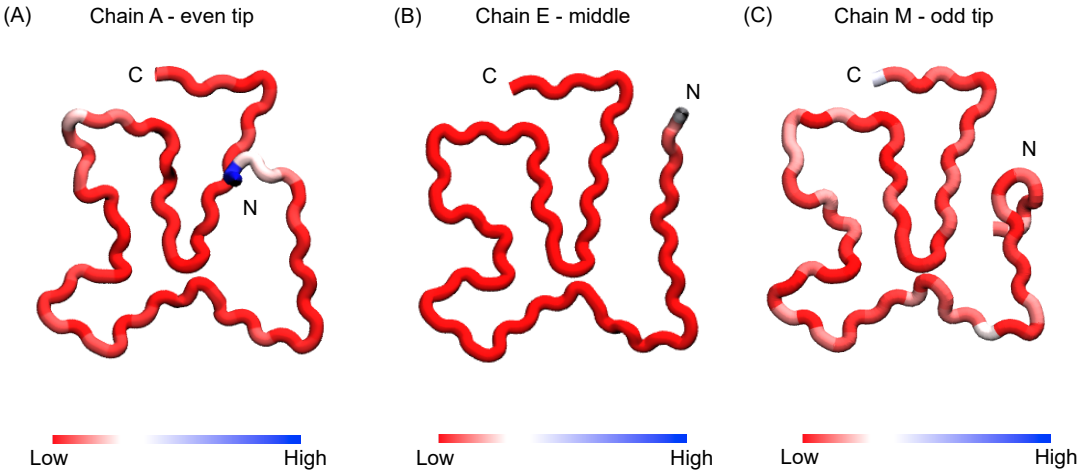

Figure 5.7: Chains from even tip (A), center (B) and odd tip (C) of the ApoA-IIF model fibril structure represented as tubes, where the color represents RMSF. Low RMSF is represented in red color, while high RMSF in blue. Color scale is represented at the bottom of each figure.

To gain further insight into the conformational changes in the ApoA-IIF model, Figure 5.7 represents average coordinates of a chain from each end of the structure with a color gradient according to RMSFs of the residues. Chains that are located at the outer ends unsurprisingly have higher RMSF values because they are more exposed to solvent than chains located in the center. All of the chains have high conformational flexibility at the terminal ends, depending on which end is solvent exposed. At the even end, N-terminal end is solvent exposed and therefore the fluctuations are higher there. At the odd end fluctuations are higher at the C-terminal end. The chain at the odd tip of the structure has higher structural fluctuations located within the arches in comparison to the one at the even tip. Therefore, if the C-terminal end is solvent exposed fluctuations are higher in the chain and upon further simulation the fibril would most likely dissociate from the odd tip.

Interaction energies of ApoA-IIC and ApoA-IIF corroborate the implications of RMSD calculations. Figure 5.8 (A) represents a difference in interaction energies of different

components of the system, defined as $\Delta E = E(\text{ApoA-IIC}) - E(\text{ApoA-IIF})$. The mutations in the ApoA-IIF model result in decreased interaction energy strength within the protein. On the other hand, ApoA-IIF fibril has a more favorable interaction energy with solvent, since more of the structure is available for interaction with the solvent. Solvent-solvent interaction has an important role in guiding the fibril folding process [163], and the results show that the solvent-solvent interaction energy is stronger in the case of ApoA-IIC fibril. All together, these interactions make the ApoA-IIC system energetically more favorable. The number of hydrogen bonds within the system reflects the results from energetic calculations. Relative to ApoA-IIC, ApoA-IIF showed fewer protein-protein hydrogen bonds and increased solvent contact. This reduces the number of solvent molecules that are able to interact, which is why the solvent interaction energy is decreased by an amount similar to the increase of protein-solvent interaction energy. Therefore, the energetic instability of the ApoA-IIF system is dominated by unfavorable interactions within the fibril, due to the change in the charge distribution within the fibril core.

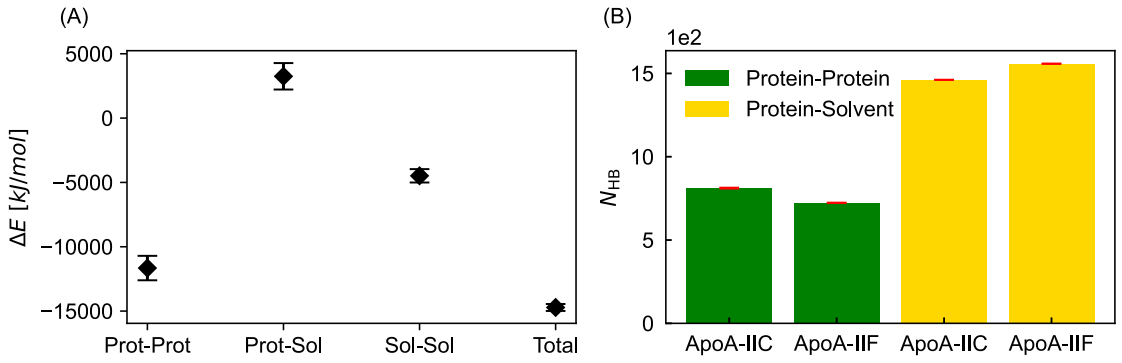

Figure 5.8: (A) Interaction energies in the protein and between protein and solvent, calculated from ApoA-IIC and ApoA-IIF simulations. (B) Average number of hydrogen bonds within protein and between protein and solvent, calculated from ApoA-IIC and ApoA-IIF simulations.

5.4.3 Summary and conclusion

In this project, the amyloid resistance property was investigated in the case of apolipoprotein A-II [30]. Our collaborators used cryo-EM microscopy to resolve the structure of ApoA-IIC protein extracted from mice suffering from systemic amyloidosis. Amyloid resistant mice express a different variant of AApoA-II protein, ApoA-IIF. This variant differs from ApoA-IIC in four mutations (Figure 5.5). I created homology models of the ApoA-IIF variant based on the ApoA-IIC, as well as four additional models, each modeling a single of the four ApoA-IIF mutations. The goal was to use MD simulations to gain an insight into the difference in stability between ApoA-IIC fibril and the homology models.

The results revealed that the ApoA-IIF variant fibril structure is less stable than ApoA-IIC, judging by the RMSDs and interaction energies. From the simulations of homology models that contained single mutations, only the one containing the Asn62Lys mutation showed lower conformational stability with higher flexibility (higher RMSF) and larger deviation from the starting structure (higher RMSD) in comparison to ApoA-IIC. Other models show similar structural stability as the ApoA-IIC model. Therefore, the mutation

that changes the charge distribution and hydrophobicity within the fibril core makes the fibril fold unfavorable. Calculations of interaction energies confirmed the conclusion drawn from the RMSD measurements that ApoA-IIF variant is energetically less favorable in a fibril fold than ApoA-IIC. The instability comes from a decrease in the interaction energies within the protein. These findings provide an important insight into the structural basis of amyloid resistance and demonstrate that point mutations in amyloid precursor proteins make the organisms susceptible to amyloidosis. Previous research has shown that eliminating the protein prone to misfolding from the organism inhibits the formation of amyloid fibrils and development of the disease [164, 165].

5.5 Results 3: Impact of point mutations on stability of lysosyme-derived amyloid fibril

Lysozyme is an ubiquitous protein in mammalian systems and it is well known for its antimicrobial activity [31]. There are three different types of lysozymes: conventional or c-type, goose or g-type and invertebrate or i-type [166]. In humans, c-type lysozyme is present in the liver, as well as articular cartilage, milk, saliva and bodily fluids [167]. Wild type lysozyme protein structure consists of α-domains and β-domains, and is additionally interconnected by four disulfide bonds. Mutations in lysozyme proteins may destabilize the native form, causing the protein to unfold and, consequently, misfold into amyloid fibrils. The fibrils are deposited in the viscera causing systemic amyloidosis. For example, human C-type lyzozyme-based systemic amyloidosis is a hereditary disease causing symptoms like renal dysfunction, gastrointestinal tract symptoms and sicca syndrome [32, 33].

Mutations that cause amyloid misfolding can be point- or double- mutations and the type of mutation determines the manifestation and progression of the disease [168]. However, even though the impact of point mutations on stability of the native protein is well researched [169, 170], it is not clear whether the mutations alter the amyloid folding pathway and stabilize specific types of amyloid structures. Insight into the structure of the ALys fibril provides more information on the location and possible impact of mutations to the fibril structure. Therefore, our collaborators have resolved the cryo-EM structure of human C-type lysozyme fibril extracted from abdominal fat tissue of a patient suffering from systemic ALys amyloidosis. The sample reveled predominance of one morphology, demonstrating the stability of this specific fibrillar structure. To gain insight into the impact of mutations on stability of the fibrillar structure, several mutations were considered. Of all the mutations, only D87G mutation positioned in the β-arc of the fibril structure does not sterically and conformationally support the fibril fold. I used MD simulations to analyze how and if this mutation affects the stability of the fibril structure.

5.5.1 Methods for simulating lysozyme fibrils

I performed MD simulations of the cryo-EM structure of the lysozyme fibril, as well as the wild-type (WT) fibril. WT structure homology model was created by introducing the D87G back-mutation into the cryo-EM structure. The PDB Manipulator tool of the

CHARMM-GUI web-interface was used to introduce the mutation [162].

To allow for better sampling six production runs were performed for each model with the starting velocities drawn randomly from a Maxwell-Boltzmann velocity distribution, resulting in a total of 2.4 μs of simulation time.

I calculated RMSD after the production runs for both simulated structures, using the respective experimental or homology modeled structures as a reference. RMSFs per residue were calculated for the WT structure and represented as a B-factor dependent color gradient on average coordinates of specific chains. To gain insight into interaction energies, I calculated the protein-protein and protein-solvent interaction energies using Gromacs simulation rerun option.

5.5.2 Results and discussion: Impact of the D87G mutation on the structural stability of lysozyme amyloid fibrils

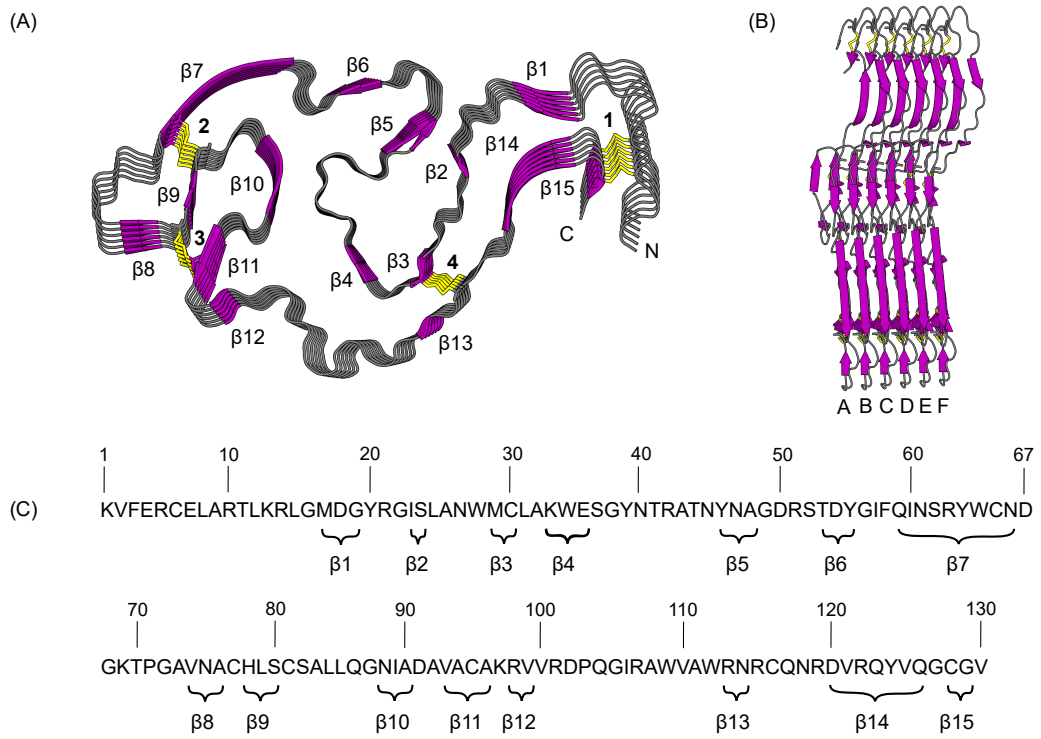

Figure 5.9: (A) Three-dimensional structure of the lysozyme fibril structure, resolved by cryo-EM spectroscopy. Loops are indicated in gray colors and β-sheets are indicated in purple and numbered from 1-15. Disulfide bonds are indicated in yellow color and numbered from 1-4. (B) Top view of the lysozyme fibril structure with chains labeled from A to F. (C) Sequence of the patient-derived lysozyme fibril represented in (A). Amino acids that form beta sheets are indicated by curly brackets.

The cryo-EM structure of the lysozyme fibril resolved by our collaborators contains a single morphology, highlighting the stability of this specific fibril structure. Consequently, all of the 130 amino acids in the fibril were successfully resolved. The Fibril contains only one stack consisting of six chains (Figure 5.9 B) with fourteen β-sheets (Figure 5.9 A and C). No α-helices are present in the fibril structure, even though they are in the native protein

fold. The fibril retained all of the four disulfide bonds, present also in the native protein, as seen in Figure 5.9 (A). These disulfide bonds additionally stabilize the fibril structure, some of them connecting the β-sheet regions of the fibril.

Figure 5.10 represents single chain of the cryo-EM resolved lysozyme fibril structure, extracted from the patient suffering from systemic amyloidosis (A) and the homology modeled wild-type structure (B). The mutated residue is found in the hydrophobic arc of the fibril. Here, neutral glycine is replaced by negatively charged aspartic acid, which changes the charge distribution and hydrophobicity in the mutated region. This, and the fact that Asp doesn't sterically fit well into the given position, is why this mutation is considered destabilizing to the fibril fold.

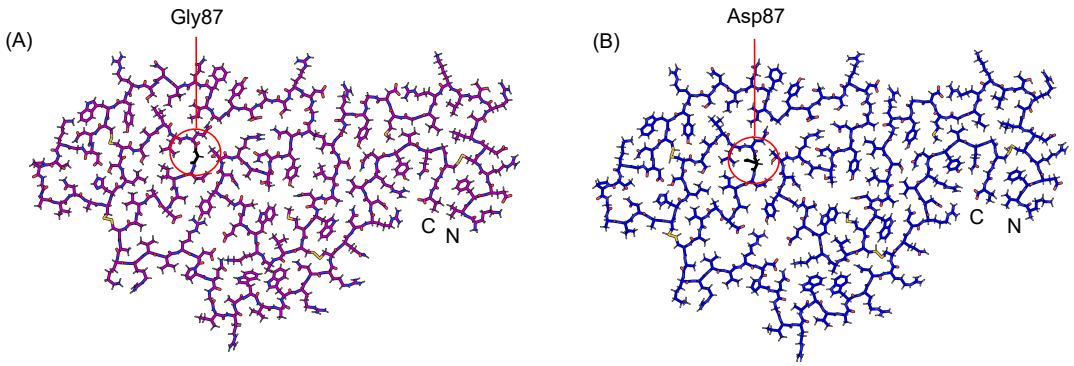

Figure 5.10: (A) Cross-section of a single chain of the patient-derived lysozyme fibril with the mutated residue Gly87 indicated in black color. (B) Cross-section of a single chain of the homology-modeled wild-type lysozyme fibril with the mutated residue Asp87 indicated in black color.

RMSD calculated after t=200 ns simulation time indicates the difference in the structural integrity between the patient and the wild-type structure (Figure 5.11 (A)). The patient-derived structure remains close to the starting structure throughout the entire simulation time, having the RMSD under 0.25 nm. Conversely, the wild-type structure has a higher RMSD in the simulation, demonstrating lower structural robustness in comparison to the patient-derived structure. The same can be observed from the snapshots of structures after the simulation, superimposed to the experimental structures (Figure 5.11 (B) and (C)). The patient derived structure remained in a position close to the starting structure. In the WT structure mostly outer parts of the fibril stack deviate significantly from the initial positions. Deviations in the structures are noticeable also in the region near the mutation, while in the patient-derived structure these regions remain near the starting structure. Observably, the D87G mutation changing the charge distribution in the hydrophobic arc of the fibril is destabilizing to the fibril fold, causing the fibril to unfold, starting from the outer chains that are closer to the solvent. The percentage of β-sheets is at 22% for the wild-type sequence after the simulation, while for the patient sequence it is at 24%, and the loss of the β-sheet secondary structure starts from the outer chains which are solvent exposed. This is expected since the dissociation of fibrils is guided by loss of secondary structure [163]. The disulfide bonds were retained throughout the simulations.

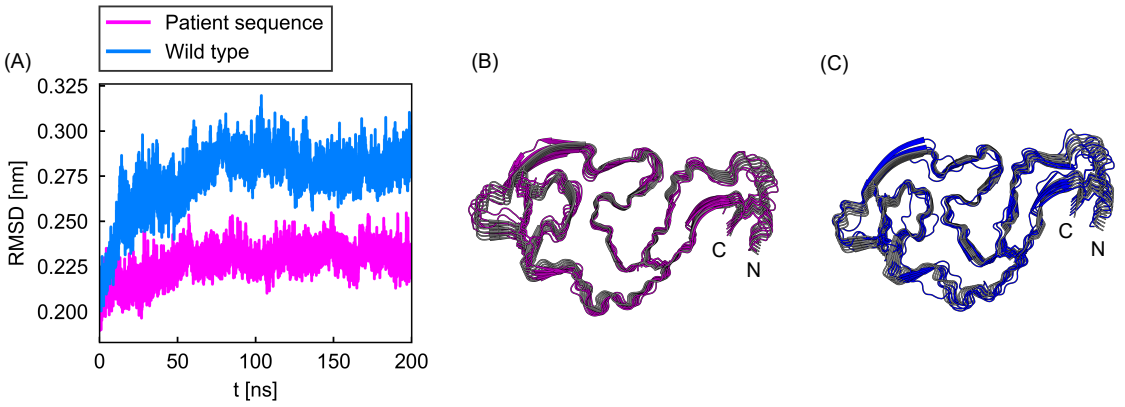

Figure 5.11: (A) RMSD of the patient-derived (pink) and wild-type structure (blue) during simulations. (B) Final structure of the patient-derived lysozyme fibril structure after the simulation (pink) superimposed to the experimentally resolved structure (gray). (C) Final structure of the wild-type lysozyme fibril structure after the simulation (blue) superimposed to the homology-modeled initial structure (gray).

To gain a more detailed insight into the regions of the WT fibril with the highest conformational stability, RMSF per residue was calculated for two outer chains and the chain in the middle and represented in a form of color gradient on the average chain coordinates (Figure 5.12). As expected, the two outer chains which are exposed to the solvent have much higher structural fluctuations than the middle chain. Interestingly, chain F shows larger fluctuations than chain A. The fluctuations in the chain F are observable in the regions around the location of the mutation, in the fibril core. Therefore, it is even more evident that the WT sequence doesn't support such fibrillar fold. Otherwise, the fluctuations are also high in the region of N terminal end in all of the chains.

Interaction energies within the protein and between protein and solvent show that the fibril unfolding is correlated to increased interactions of the fibril with the solvent. In Figure 5.13 A, difference in interaction energies of different components of the system is defined as $\Delta E = E(\text{Patient}) - E(\text{WT})$. The interaction energies within protein are within error for the patient-derived and WT structure. However, the protein-solvent interaction energy is lower for the WT structure, indicating increased contact with solvent in comparison to the patient sequence. This is reflected also in the solvent-solvent interaction energy. Since the WT structure increased interactions with solvent, it decreased number of solvent molecules available for interaction and therefore the solvent-solvent interaction energy. Therefore, solvent-solvent interaction energy is lower for the patient-derived sequence. This is expected since the mutation from WT to patient-derived sequence eliminated the charged residue from the core of the fibril and replaced it with a hydrophobic residue. Interestingly the contributions from solvent-solvent and protein solvent interaction energy differences between the WT and patient sequence cancel each-other out and the potential energy of the system is as a result similar in both systems.

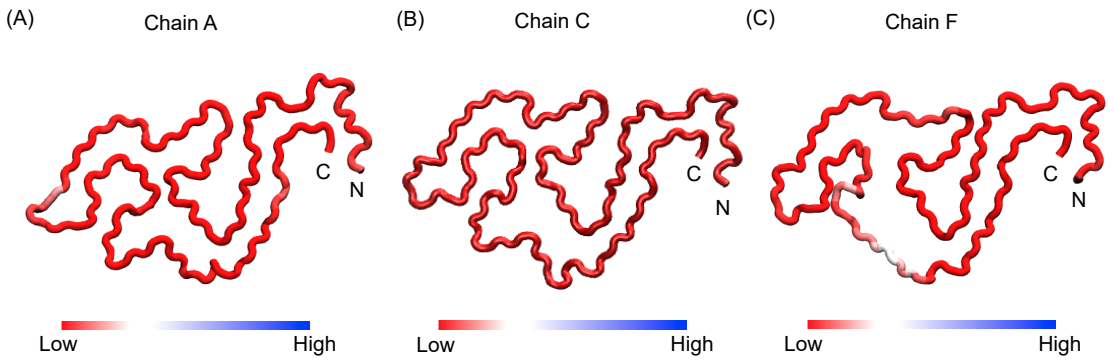

Figure 5.12: RMSF per residue represented as the color gradient on the average coordinates of outer chain A (A), middle chain E (B) and outer chain F (C) for the WT fibril structure. Color scale representing the RMSF from low to high values is indicated below the chain figures.

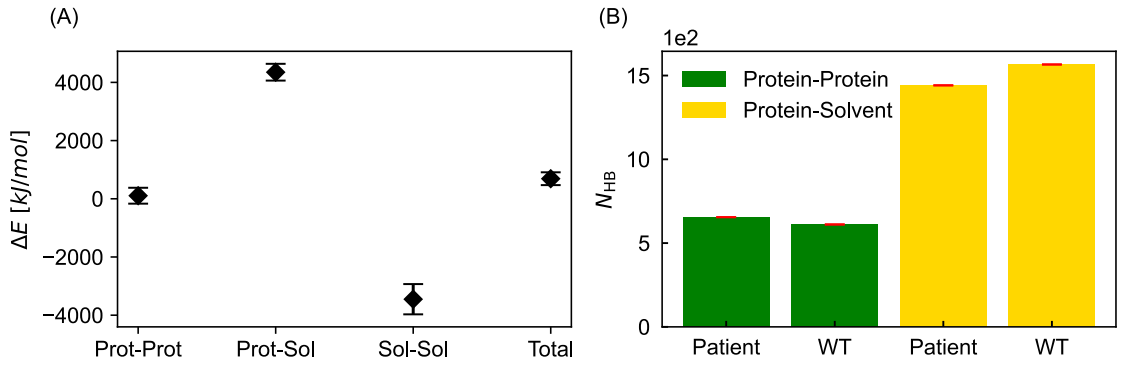

Figure 5.13: (A) Difference in protein-protein, solvent-solvent and protein-solvent interaction energies between the patient and the WT structure, defined as $\Delta E = E(\text{Patient}) - E(\text{WT})$. (B) Average number of hydrogen bonds within protein and between protein and solvent, calculated from patient and WT structure simulations. Patient-derived structure is labeled as "Patient" and wild-type structure is labeled as "WT"

Number of hydrogen bonds within the protein is smaller in case of the WT structure in comparison to the patient-derived structure. However, number of hydrogen bonds with solvent is larger in case of the WT structure. Therefore, this and the calculations of the interaction energy lead to the conclusion that mutation of the hydrophobic residue in the patient structure to the negatively charged polar residue in the WT sequence leads to increased interaction of the WT sequence with solvent, leading to conformational instability and subsequent unfolding of the fibril structure. These results are in line with previous research indicating an important role of solvent entropy in guiding the fibril folding process [163].

To confirm these findings, an average maximum distance from all six production runs between equivalent residues in neighboring chains was calculated in Figure 5.14. The chains were defined in Figure 5.9. Residues Leu84 to Arg107 of solvent exposed chains in the wild type sequence attain high distances from the rest of the fibril, while the same is not visible in case of the patient sequence. The larger distances indicate a local loosening of the structure, which in turn may lead to further unfolding of the fibril structure.

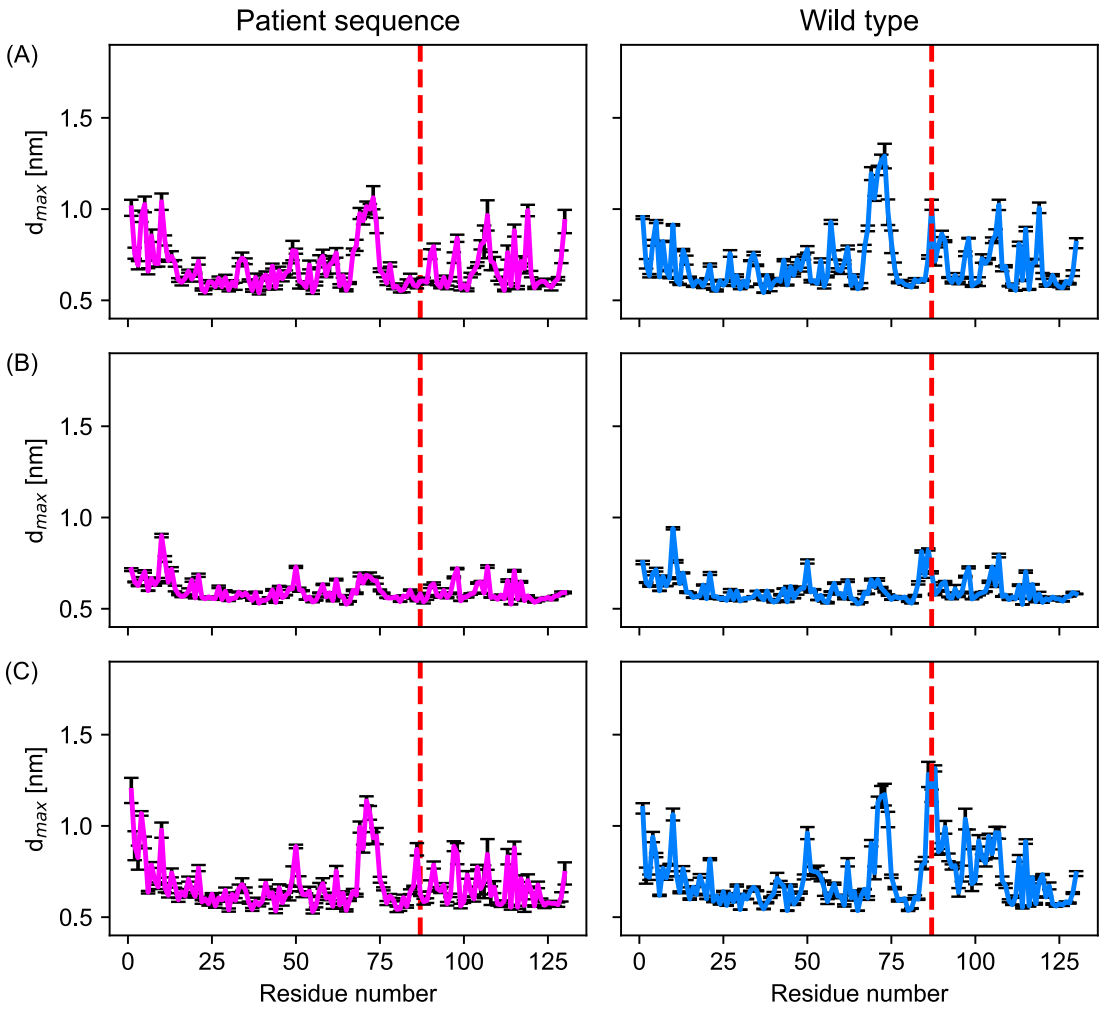

Figure 5.14: Maximal distances per reside in the patient and wild type sequence between (A) chains A and B, (B) chains C and D, and (C) chains E and F.

5.5.3 Summary and conclusion

In this project I used MD simulations to investigate the influence of mutations on the stability of the lysozyme fibril fold. Lysozyme is an important part of the mammalian immune system as it protects the organism against microbes. The mutations in the lysozyme sequence destabilize it and it unfolds to acquire aggregation prone structures, such as fibrils. Our collaborators used cryo-EM to resolve the structure of ALys amyloid fibril extracted from a patient suffering from hereditary amyloidosis caused by D87G mutation [34]. I used homology modeling to back-mutate the D87G in the cryo-EM structure, and subsequently simulated both structures using all-atom MD simulations.

WT structure was structurally less stable during the simulations in comparison to the mutated structure, as revealed by RMSD measurements. This means that the WT structure doesn't support the aspartate residue at the given position within the core of the fibril. This is in line with previous results that charged residues positioned at the fibril core destabilize the structure [29, 30]. This was confirmed by the results from RMSF calculations which showed that outer chains of the fibril have the most flexibility in the

region around the mutation site. However, calculations of the interaction energy revealed an important role that the solvent plays in the conformational stability of amyloid fibrils. Once again MD simulations are a useful tool to study the impact of point mutations on the stability of fibril structures. It is not clear how the mutation would affect the native protein, but from the results it is indicated that the structure of the fibril is closely correlated to its sequence. This is an important insight for drug design and development of ligands that can bind to specific areas of the native protein and stabilize its fold, which was previously shown as a successful strategy [171].

5.6 Appendix

5.6.1 Data archive

Part 1: Comparison of ex-vivo and in-vitro Aß amyloid fibril structures using cryoEM and MD simulations

The input files for the simulations described in the first part of this chapter, which concern the seeded and unseeded Aß fibril morphologies, are organized in the following directory tree. Simulations of the unprotonated models were carried out at temperatures of T=300, T=350, and T=400 K, and the input files for all respective simulations are provided. For example, to run a T=300 K simulation of the seeded morphology go to the directory `./Beta/seeded_np` and run the command

```
gmx grompp -f ../mdp_fibrils/run.mdp -c npt/npt.gro -p topol/topol.top
-n topol/index.ndx -o run.tpr
```

The structures after 200 ns simulation time as well as RMSDs from all simulations are provided in the `./Beta/MD_results` folder.

```
/hpc/gpfs2/home/g/cbio/Archive_maugrina/Fibrils/Beta
├── mdp_fibrils
│   ├── run350.mdp
│   ├── run400.mdp
│   └── run.mdp
├── seeded_np
│   ├── npt
│   │   └── npt.gro
│   ├── npt350
│   │   └── npt.gro
│   ├── npt400
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── seeded_p
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── unseeded1_np
│   ├── npt
│   │   └── npt.gro
│   ├── npt350
```

```
│   │   └── npt.gro
│   ├── npt400
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── unseeded1_p
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── unseeded2_np
│   ├── npt350
│   │   └── npt.gro
│   ├── npt400
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── unseeded2_p
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
└── MD_results
    ├── seeded_morphology
    │   ├── 300K
    │   ├── 350K
    │   ├── 400K
    │   └── protonated_histidine_HIP_300K
    ├── unseeded_morphology_1
    │   ├── 300K
    │   ├── 350K
    │   ├── 400K
    │   └── protonated_histidine_HIP_300K
    └── unseeded_morphology_2
        ├── 300K
        ├── 350K
        ├── 400K
        └── protonated_histidine_HIP_300K
```

Part 2: MD simulation study into structural properties of amyloid resistance

The simulation input files for the second part of this chapter, covering the AApoA-IIC fibril and AApoAII-F point mutations, are archived in the following directory tree. Go to any of the directories e.g. `./AApoAII/Native_model` and run the command

```
gmx grompp -f mdp/run.mdp -c npt/npt.gro -p topol/topol.top
-n topol/index.ndx -o run.tpr
```

The structures after 200 ns simulation time as well as RMSDs from all simulations are provided in the `./AApoAII/MD_Results` folder.

```
/hpc/gpfs2/home/g/cbio/Archive_maugrina/Fibrils/AApoAII
├── Native_model
│   ├── mdp
```

```
│   │   └── run.mdp
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── Arg54Lys
│   ├── mdp
│   │   └── run.mdp
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── Asn62Lys
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── Gln16His
│   ├── mdp
│   │   └── run.mdp
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── Ser9Asn
│   ├── mdp
│   │   └── run.mdp
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── All_mutations
│   ├── mdp
│   │   └── run.mdp
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
└── MD_Results
    ├── Rmsd
    └── structures
        ├── mutation_16
        ├── mutation_54
        ├── mutation_62
        ├── mutation_9
        ├── mutation_all
        └── native_model
```

Part 3: Impact of point mutations on stability of lysosyme-derived amyloid fibril

The simulation input files for the third part of this chapter, which pertain to the lysozyme fibril simulations, are archived in the following directory tree. Go to any of the directories e.g. `./Lysozyme_fibril/D87G` and run the command

```
gmx grompp -f mdp/run.mdp -c npt/npt.gro -p topol/topol.top
-n topol/index.ndx -o run.tpr
```

The structures after 200 ns simulation time as well as RMSDs and RMSFs from all simulations are provided in the `./Lysozyme_fibril/Results_MD` folder.

```
/hpc/gpfs2/home/g/cbio/Archive_maugrina/Fibrils/Lysozyme_fibril
├── D87G
│   ├── mdp
│   │   └── run.mdp
│   ├── npt
│   │   ├── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
├── Native
│   ├── mdp
│   │   └── run.mdp
│   ├── npt
│   │   └── npt.gro
│   └── topol
│       ├── index.ndx
│       └── topol.top
└── Results_MD
    ├── Patient_sequence
    ├── RMSFS
    │   ├── Patinent
    │   └── wt
    └── Wild_type
```

Chapter 6

Conclusions and Outlook

In my research I used MD simulations, enhanced sampling methods and Langevin dynamics simulations to investigate the folding of rG4s and their interactions with ions, as well as conformational properties of amyloid fibrils.

rG4s are non-canonical secondary structures formed by guanine-rich sequences of RNA. They consist of stacked guanine quartets stabilized by Hoogsteen hydrogen bonds and monovalent cations, particularly potassium (K^+). rG4s are increasingly recognized as key regulators of gene expression, RNA processing, and genome stability and are prevalent in lncRNAs such as TERRA [44]. They are dynamic in cells and the folding kinetics of rG4s are essential for their function in cellular processes [7]. Despite their biological importance, the structural complexity and transient nature of rG4 intermediates have made them difficult to resolve using conventional experimental methods. In this thesis, I utilized the TIS coarse-grained model to explore rG4 folding dynamics, supported by experimental data from CD spectroscopy. The simulations were performed on two biologically relevant rG4s: TERRA and a plant rG4, as well as three synthetic structures with differing loop lengths. They revealed that TERRA G4 folding follows a sequential but multi-pathway mechanism, characterized by an ensemble of intermediate states consisting of hairpins, triplexes, and double-hairpins. These states arise from local loop flexibility and base-stacking interactions and are highly dependent on temperature and salt concentration. I also demonstrated that conformational entropy plays a central role in the branched folding landscape of rG4s. I confirmed this by expanding the simulations to four other rG4s. All of the G4 structures demonstrated folding landscapes with the same intermediates as in the case of TERRA, although their populations differ across different rG4s.

A key limitation of the CG simulations used in this work is their reliance on an implicit solvent representation, which fails to fully capture the local charge screening effects and specific ion-RNA interactions critical to rG4 stability. To account for the role of ionic conditions, we developed a concentration-matching scheme using atomistic MD simulations, which allowed us to calibrate coarse-grained models to match experimental monovalent ion concentrations. This enabled accurate prediction of folded-state populations across different salt concentrations and temperatures and bridged the gap between simulations and experiments.

Due to their affinity to binding cations in the central pore, rG4s (and dG4s) are also being studied for their potential role as artificial ion channels [20, 135, 136]. Ion channels are usually protein structures that are responsible for selective transport of ions across a cellular membrane, and therefore maintaining crucial cellular processes such as signaling, cellular volume regulation and metabolism. Artificial ion channels are synthetic or biomimetic structures that replicate the key functions of natural ion channels, with a potential for

application in treating ion channel dysfunction, biosensing, and nanotechnologies [19, 20]. Nucleic acid-based ion channels, in particular those made from G4s, present a compelling target due to their properties like ion conductance and selectivity.

Using straightforward and enhanced molecular dynamics simulations including TREMD and constant-force pulling, I explored how different monovalent cations interact with and traverse the G4 channel formed by TERRA RNA. rG4s have a specificity towards K^+ ions, which bind longer to the binding site, and free energy profiles calculated from TREMD simulations indicate the highest binding affinity to G4 binding sites in comparison with other ions. Na^+ ions, on the other hand, exhibited more dynamic permeation in the straightforward simulations, with cumulative ion transport. By analyzing the MFPT of K^+ and Na^+ ions along the TERRA G4 binding sites, as well as free energy profiles from TREMD simulations, I found that the lowest quartet presents the highest barrier for Na^+ transition, most likely due to the presence of terminal overhanging nucleotides which can bind cations.

The MFPTs from the constant-force pulling simulations were the lowest for K^+ and Na^+. Interestingly, Na^+ ions exhibit a slightly lower MFPT for higher force values, while for lower force values K^+ ions have lower MFPT values. These results show that there exists an additional mechanism that slows Na^+ conductance, which can not be resolved by using high pulling force values.

MFPTs were substantially higher for the rest of the studied cations, demonstrating low conductance trough the G4 channel. Li^+ ions encountered significant energy barriers for transport, probably due to their strong hydration and small radius. Cs^+ ions also don't exhibit favorable conductance, most likely due to their size which distorts the G4 core. NH_4^+ ions present a suitable surrogate for K^+ ions in NMR experiments and show a similar binding-site pattern to the G4 core, but lower conductance, possibly due to its non-spherical shape.

In summary, these observations suggest that rG4s can act as ion-selective channels, opening promising directions for bio-inspired nanotechnology and synthetic biology applications. The work highlights how the combination of all-atom simulations with enhanced sampling enables the prediction of ion-transport properties, offering a computational path toward the rational design of RNA-based ion-conducting nanostructures. However, further work is needed to clarify the ion specificity, especially and crucially the differences in free-energy barriers of ion binding along the G4 channel. For this, an enhanced-sampling method like umbrella sampling may be used. Additionally, the influence of different structural properties like nucleotide overhangs on the ionic conductance presents an interesting research topic.

In the final chapter of the thesis, I discuss the structural and thermodynamic properties of amyloid fibrils using MD simulations, with a focus on the effects of sequence variations, protonation states, and environmental conditions. Amyloid fibrils, characterized by their cross-β sheet architecture, are implicated in the pathology of various protein misfolding diseases, including Alzheimer's, Parkinson's, and systemic amyloidosis [28, 157]. Understanding the molecular determinants of fibril stability and morphology is thus essential for both basic biophysics and therapeutic development.

The first part examined structural and stability differences between seeded and unseeded Aβ fibrils using cryo-EM and MD simulations. According the cryo-EM experiments performed by our collaborators, seeded fibrils derived from *ex vivo* brain tissue differ from the unseeded fibrils in cross-over distances and number of peptide stacks forming the fibrils.

MD simulations showed that seeded fibrils are more conformationally stable, maintaining structure even at 400 K, whereas unseeded morphologies underwent significant deviations, especially at terminal regions. Protonation of histidine residues in the N-terminal arch destabilized the seeded morphology, suggesting that local charge states and electrostatics are critical for structural integrity. Unseeded fibrils, which lack structurally resolved N-termini, were less affected by protonation changes. This work provided the insight into the structural differences between seeded and unseeded fibrils that impact their stability. Environmental conditions such as pH and presence of co-factors may further influence fibril structure and are likely to be of relevance in amyloid polymorphism and pathology.

In the second study, I explored the structural basis of amyloid resistance in ApoA-II fibrils using MD simulations. Our collaborators resolved the structure of the amyloidogenic ApoA-IIC variant using cryo-EM, while the amyloid-resistant ApoA-IIF variant, found in mice protected from systemic amyloidosis, differs by four-point mutations.

Simulations showed that ApoA-IIF is structurally less stable than ApoA-IIC, with higher RMSDs and lower intra-fibril interaction energy. Among the individual mutations tested, Asn62Lys had the strongest destabilizing effect, likely due to charge disruption in the fibril core. The remaining mutations had minimal impact. Energy analyses confirmed that ApoA-IIF's instability arises from reduced internal interactions and increased solvent exposure. These results highlight the importance of sequence-specific charge distribution and hydrophobicity in maintaining amyloid fibril stability. They offer mechanistic insight into how specific mutations protect organisms from amyloidosis, supporting the idea that modifying or suppressing aggregation-prone sequences could be a viable therapeutic approach.

In the final project, I used MD simulations to investigate the effect of the Asp87Gly mutation on the stability of a lysozyme-derived amyloid fibril, whose structure was resolved by our collaborators via cryo-EM from a patient with systemic amyloidosis. A wild-type (WT) model was created by reversing the Asp87Gly mutation in the cryo-EM structure to assess whether the native sequence supports the observed fibril fold.

Simulations showed that the WT fibril is less stable, exhibiting higher RMSD, greater flexibility (particularly in the region of the mutation), and increased solvent interactions. This instability arises from the introduction of a charged aspartate residue into a hydrophobic fibril core, disrupting packing and promoting unfolding from the solvent-exposed ends. RMSF and analysis of hydrogen bonds further confirmed these trends. Overall, the results highlight how point mutations at structurally sensitive sites, especially those involving charge changes, can influence fibril stability. These insights reinforce the importance of sequence-specific interactions in amyloid formation and may inform future therapeutic strategies targeting native-state stabilization.

This work demonstrates that MD simulations, combined with enhanced sampling meth-

ods and experimental validation, can yield detailed mechanistic insights into complex biomolecular processes such as rG4 folding, ion selectivity, and amyloid fibril stability. The rise of cutting-edge computational technologies such as artificial intelligence (AI) and machine learning (ML) are revolutionizing the field of computational biology. ML algorithms can help process the data generated by MD simulations, identify specific patterns, fine-tune simulation parameters and aid the development of new predictive force-fields. They can also guide enhanced sampling simulations by reducing the dimensionality of the simulation data and introducing adaptive algorithms that adjust sampling parameters based on the outcomes from previous simulations [172]. These approaches may help resolve folding pathways of rG4s and dG4s, especially when appropriate collective variables are identified, and our present work on rG4 folding contributes toward that goal. Additionally, free-energy barriers of ionic conductance may be quantified by umbrella sampling or adaptive force biasing techniques [173], while alanine-scanning technique incorporated into energy decomposition calculations emerged as a promising technique for studying the impact of side-chain mutations on fibril formation and stability [72]. Guided enhanced sampling techniques with optimized force fields can further improve our understanding of amyloid fibril stability and interaction energies. Therefore, it will be exciting to observe how the increase in computational power will shape the direction of computational biology in the future.

List of Publications

1. **Ugrina M**, Burkhart I, Müller D, Schwalbe H, **Schwierz N.** "RNA G-quadruplex folding is a multi-pathway process driven by conformational entropy". *Nucleic Acids Res.*, 2024, pp. 87–100

2. Pfeiffer P B, **Ugrina M**, **Schwierz N**, Sigurdson C J, Schmidt M, Fändrich M. "Cryo-EM Analysis of the Effect of Seeding with Brain-derived Aβ Amyloid Fibrils". *J Mol Biol*, 2024, 436(4):168422

3. Andreotti G, Baur J, **Ugrina M**, Pfeiffer P B, Hartmann M, Wiese S, Miyahara H, Higuchi K, **Schwierz N**, Schmidt M, Fändrich M. "Insights into the Structural Basis of Amyloid Resistance Provided by Cryo-EM Structures of AApoAII Amyloid Fibrils". *J Mol Biol*, 2024, 436(4):168441

4. Karimi-Farsijani S, Sharma K, **Ugrina M**, Kuhn, L, Pfeiffer P B, Haupt C, Wiese S, Hegenbert U, Schönland S O, **Schwierz N**, Schmidt M, Fändrich M. "Cryo-EM structure of a lysozyme-derived amyloid fibril from hereditary amyloidosis". *Nat Commun 15*, 2024, 9648

Bibliography

[1] Cirk, F. "Central Dogma of molecular biology". *Nature*, 227:561–563, 1970.

[2] Rinn, J. L. and Chang, H. Y. "Genome regulation by long noncoding RNAs". *Annual Review of Biochemistry*, 81:145–166, 2012.

[3] Wanrooij, P. H., Uhler, J. P., Simonsson, T., Falkenberg, M., and Gustafsson, C. M. "G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation". *Proc. Natl. Acad. Sci. U. S. A.*, 107(37):16072–16077, 2010.

[4] Kharel, P., Becker, G., Tsvetkov, V., and Ivanov, P. "Properties and biological impact of RNA G-quadruplexes: From order to turmoil and back". *Nucleic Acids Res.*, 48 (22):12534–12555, 2020.

[5] Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Plückthun, A. "In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei". *Proc. Natl. Acad. Sci. U. S. A.*, 98(15):8572–8577, 2001.

[6] Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D., and Lipps, H. J. "Telomerase recruitment by the telomere end binding protein-β facilitates G-quadruplex DNA unfolding in ciliates". *Nat. Struct. Mol. Biol.*, 15(6):598–604, 2008.

[7] Chen, X. C., Chen, S. B., Dai, J., Yuan, J. H., Ou, T. M., Huang, Z. S., and Tan, J. H. "Tracking the Dynamic Folding and Unfolding of RNA G-Quadruplexes in Live Cells". *Angewandte Chemie - International Edition*, 57(17):4702–4706, 2018.

[8] Dumas, L., Herviou, P., Dassi, E., Cammas, A., and Millevoi, S. "G-Quadruplexes in RNA Biology: Recent Advances and Future Directions". *Trends Biochem. Sci.*, 46(4):270–283, 2021.

[9] Zhang, J., Harvey, S. E., and Cheng, C. "A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes". *Nucleic Acids Res.*, 47(7):3667–3679, 2019.

[10] Yang, D. *G-Quadruplex Nucleic Acids, Methods and Protocols*, volume 2035. 2019.

[11] Bessi, I., Jonker, H. R., Richter, C., and Schwalbe, H. "Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex". *Angew. Chem. Int. Ed.*, 54(29):8444–8448, 2015.

[12] Grün, J. T., Blümler, A., Burkhart, I., Wirmer-Bartoschek, J., Heckel, A., and Schwalbe, H. "Unraveling the Kinetics of Spare-Tire DNA G-Quadruplex Folding". *J. Am. Chem. Soc.*, pages accepted manuscript, (published online), 2021.

[13] Frelih, T., Wang, B., Plavec, J., and Šket, P. "Pre-folded structures govern folding pathways of human telomeric G-quadruplexes". *Nucleic Acids Research*, 48(4):2189–2197, 2020.

[14] Müller, D., Bessi, I., Richter, C., and Schwalbe, H. "The folding landscapes of human telomeric RNA and DNA G-quadruplexes are markedly different". *Angew Chem. Int. Ed.*, 60:10895–10901, 2021.

[15] Zhang, A. Y. and Balasubramanian, S. "The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids". *J. Am. Chem. Soc.*, 134(46):19297–19308, 2012.

[16] Wang, L., Xu, Y.-P., Bai, D., Shan, S.-W., Xie, J., Li, Y., and Wu, W.-Q. "Insights into the structural dynamics and helicase-catalyzed unfolding of plant RNA G-quadruplexes". *J. Biol. Chem.*, 298(8):102165, 2022.

[17] Bhattacharyya, D., Arachchilage, G. M., and Basu, S. "Metal cations in G-quadruplex folding and stability". *Frontiers in Chemistry*, 4(SEP):1–14, 2016.

[18] Zaccaria, F. and Fonseca Guerra, C. "RNA versus DNA G-Quadruplex: The Origin of Increased Stability". *Chemistry - A European Journal*, 24(61):16315–16322, 2018.

[19] Yang, J., Yu, G., Sessler, J. L., Shin, I., Gale, P. A., and Huang, F. "Artificial transmembrane ion transporters as potential therapeutics". *Chem*, 7(12):3256–3291, 2021.

[20] Zheng, S. P., Huang, L. B., Sun, Z., and Barboiu, M. "Self-Assembled Artificial Ion-Channels toward Natural Selection of Functions". *Angewandte Chemie - International Edition*, 60(2):566–597, 2021.

[21] Laughlan, G., Murchie, A. I., Norman, D. G., Moore, M. H., Moody, P. C., Lilley, D. M., and Luisi, B. "The high-resolution crystal structure of a parallel-stranded guanine tetraplex". *Science*, 265(5171):520–524, 1994.

[22] Spannenberg, A., Burlakov, V. V., and Rosenthal, U. "Crystal structure of". *Zeitschrift fur Kristallographie - New Crystal Structures*, 216(4):616–618, 2001.

[23] Podbevšek, P., Hud, N. V., and Plavec, J. "NMR evaluation of ammonium ion movement within a unimolecular G-quadruplex in solution". *Nucleic Acids Research*, 35(8):2554–2563, 2007.

[24] Schultze, P., Hud, N. V., Smith, F. W., and Feigon, J. "The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G4T4G4)". *Nucleic Acids Research*, 27(15):3018–3028, 1999.

[25] Debnath, M., Chakraborty, S., Kumar, P. Y., Chaudhuri, R., and Dash, J. "Ionophore constructed from non-covalent assembly of a G-quadruplex and liponucleoside transports K + -ion across biological membranes". *Nature Communications*, (2020):1–12.

[26] Ke, P. C., Zhou, R., Serpell, L. C., Riek, R., Knowles, T. P., Lashuel, H. A., Gazit, E., Hamley, I. W., Davis, T. P., Fändrich, M., Otzen, D. E., Chapman, M. R., Dobson, C. M., Eisenberg, D. S., and Mezzenga, R. "Half a century of amyloids: Past, present and future". *Chemical Society Reviews*, 49(15):5473–5509, 2020.

[27] de Jonge, N. "Theory of the spatial resolution of (scanning) transmission electron microscopy in liquid water or ice layers". *Ultramicroscopy*, 187:113–125, 2018.

[28] Joseph T. Jarrett and Jr, P. T. L. "Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie?". *Cell*, 73(6): 1055–8, 1993.

[29] Benedikt Pfeiffer, P., Ugrina, M., Schwierz, N., Sigurdson, C. J., Schmidt, M., and Fändrich, M. "Cryo-EM analysis of the effect of seeding with brain-derived Aβ amyloid fibrils". *Journal of Molecular Biology*, 436:168422, 2023.

[30] Andreotti, G., Baur, J., Ugrina, M., Pfeiffer, P. B., Hartmann, M., Wiese, S., Miyahara, H., Higuchi, K., Schwierz, N., Schmidt, M., and Fändrich, M. "Insights into the Structural Basis of Amyloid Resistance Provided by Cryo-EM Structures of AApoAII Amyloid Fibrils". *Journal of Molecular Biology*, 436(4), 2024.

[31] Fleming, A. and Wright, A. E. "On a remarkable bacteriolytic element found in tissues and secretions". *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 93(653):306–317, 1922.

[32] Pepys, M. B., Hawkins, P. N., Booth, D. R., Vigushin, D. M., Tennent, G. A., Soutar, A. K., Totty, N., Nguyen, O., Blake, C. C., Terry, C. J., Feest, T. G., Zalin, A. M., and Hsuan, J. J. "Human lysozyme gene mutations cause hereditary systemic amyloidosis". *Nature*, 362(6420):553–557, 1993.

[33] Anker, S., Hinderhofer, K., Baur, J., Haupt, C., Röcken, C., Beimler, J., Zeier, M., Weiler, M., Wühl, E., Kimmich, C., Schönland, S., and Hegenbart, U. "Lysozyme amyloidosis—a report on a large german cohort and the characterisation of a novel amyloidogenic lysozyme gene variant". *Amyloid*, 29(4):245–254, 2022.

[34] Karimi-Farsijani, S., Sharma, K., Ugrina, M., Kuhn, L., Pfeiffer, P. B., Haupt, C., Wiese, S., Hegenbart, U., Schönland, S. O., Schwierz, N., Schmidt, M., and Fändrich, M. "Cryo-EM structure of a lysozyme-derived amyloid fibril from hereditary amyloidosis". *Nature Communications*, 15(1):1–9, 2024.

[35] Crick, F. and Watson, J. "Molecular structure of nucleic acids". *Nature*, 171:737–738, 1953.

[36] Franklin, R. E. and Gosling, R. G. "Molecular configuration in sodium thymonucleate". *Nature*, 171:740–741, 1953.

[37] Wilkins, M. H. F., Stokes, A. R., and Wilson, H. R. "Molecular Structure of Deoxypentose Nucleic Acids". *Nature*, 171:738–740, 1953.

[38] Alberts, B., Johnson, A., and Lewis, J. e. a. "Molecular biology of the cell. 4th edition". *New York: Garland Science*, 2002.

[39] Breaker, R. R. and Joyce, G. F. "The expanding view of RNA and DNA function". *Chemistry and Biology*, 21(9):1059–1065, 2014.

[40] Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L. L., Chen, R., Dean, C., Dinger, M. E., Fitzgerald, K. A., Gingeras, T. R., Guttman, M., Hirose, T., Huarte, M., Johnson, R., Kanduri, C., Kapranov, P., Lawrence, J. B., Lee, J. T., Mendell, J. T., Mercer, T. R., Moore, K. J., Nakagawa, S., Rinn, J. L., Spector, D. L., Ulitsky, I., Wan, Y., Wilusz, J. E., and Wu, M. "Long non-coding RNAs: definitions, functions, challenges and recommendations". *Nature Reviews Molecular Cell Biology*, 24(6):430–447, 2023.

[41] Draper, D. E. "A guide to ions and RNA structure". *Rna*, 10(3):335–343, 2004.

[42] Agarwala, P., Pandey, S., and Maiti, S. "The tale of RNA G-quadruplex". *Org. Biomol. Chem.*, 13(20):5570–5585, 2015.

[43] Biffi, G., Tannahill, D., and Balasubramanian, S. "An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2". *J. Am. Chem. Soc.*, 134(29):11974–11976, 2012.

[44] Luke, B. and Lingner, J. "TERRA: Telomeric repeat-containing RNA". *EMBO Journal*, 28(17):2503–2510, 2009.

[45] Gilley, D., Tanaka, H., and Herbert, B. S. "Telomere dysfunction in aging and cancer". *Int. J. Biochem. Cell Biol.*, 37(5 SPEC. ISS.):1000–1013, 2005.

[46] Marzano, F., Rapacciuolo, A., Ferrara, N., Rengo, G., Koch, W. J., and Cannavo, A. "Targeting GRK5 for treating chronic degenerative diseases". *Int. J. Mol. Sci.*, 22(4): 1–17, 2021.

[47] Zhao, C., Qin, G., Niu, J., Wang, Z., Wang, C., Ren, J., and Qu, X. "Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19?". *Angew. Chem. Int. Ed.*, 60(1):432–438, 2021.

[48] Dolinnaya, N. G., Ogloblina, A. M., and Yakubovskaya, M. G. *Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery*, volume 81. 2016.

[49] Cruz-León, S. and Schwierz, N. "Hofmeister Series for Metal-Cation-RNA Interactions: The Interplay of Binding Affinity and Exchange Kinetics". *Langmuir*, 36(21): 5979–5989, 2020.

[50] Largy, E., König, A., Ghosh, A., Ghosh, D., Benabou, S., Rosu, F., and Gabelica, V. "Mass Spectrometry of Nucleic Acid Noncovalent Complexes". *Chem. Rev.*, 122(8): 7720–7839, 2022.

[51] Marchand, A. and Gabelica, V. "Folding and misfolding pathways of G-quadruplex DNA". *Nucleic Acids Res.*, 44(22):10999–11012, 2016.

[52] Šponer, J., Bussi, G., Stadlbauer, P., Kührová, P., Banáš, P., Islam, B., Haider, S., Neidle, S., and Otyepka, M. "Folding of guanine quadruplex molecules–funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies". *Biochimica et Biophysica Acta - General Subjects*, 1861(5):1246–1263, 2017.

[53] Martadinata, H. and Phan, A. T. "Structure of propeller-type parallel-stranded rna g-quadruplexes, formed by human telomeric rna sequences in k+ solution". *J. Am. Chem. Soc.*, 131(7):2570–2578, 2009.

[54] Adrian, M., Heddi, B., and Phan, A. T. "NMR spectroscopy of G-quadruplexes". *Methods*, 57(1):11–24, 2012.

[55] Wirmer-Bartoschek, J., Bendel, L. E., Jonker, H. R., Grün, J. T., Papi, F., Bazzicalupi, C., Messori, L., Gratteri, P., and Schwalbe, H. "Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements that Affect Ligand Binding". *Angew. Chem. Int. Ed.*, 56(25):7102–7106, 2017.

[56] Binas, O., Bessi, I., and Schwalbe, H. "Structure Validation of G-Rich RNAs in Noncoding Regions of the Human Genome". *Chembiochem*, 21(11):1656–1663, 2020.

[57] Bessi, I., Bazzicalupi, C., Richter, C., Jonker, H. R., Saxena, K., Sissi, C., Chioccioli, M., Bianco, S., Bilia, A. R., Schwalbe, H., and Gratteri, P. "Spectroscopic, molecular modeling, and NMR-spectroscopic investigation of the binding mode of the natural alkaloids berberine and sanguinarine to human telomeric G-quadruplex DNA". *ACS Chem. Biol.*, 7(6):1109–1119, 2012.

[58] Collie, G. W., Haider, S. M., Neidle, S., and Parkinson, G. N. "A crystallographic and modelling study of a human telomeric rna (terra) quadruplex". *Nucleic Acids Research*, 38(16):5569–5580, 2010.

[59] Schultze, P., Hud, N. V., Smith, F. W., and Feigon, J. "The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G4T4G4)". *Nucleic Acids Research*, 27(15):3018 – 3028, 1999.

[60] Mergny, J. L. and Lacroix, L. "UV melting of G-quadruplexes". *Current Protocols in Nucleic Acid Chemistry*, (SUPPL. 37):1–15, 2009.

[61] Grün, J. T. and Schwalbe, H. "Folding dynamics of polymorphic G-quadruplex structures". *Biopolymers*, (September):1–15, 2021.

[62] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. "Improved side-chain torsion potentials for the Amber ff99SB protein force field". *Proteins: Structure, Function and Bioinformatics*, 78(8):1950–1958, 2010.

[63] Zacharogianni, M. and Rabouille, C. "Chapter 3 - Trafficking Along the Secretory Pathway in Drosophila Cell Line and Tissues: A Light and Electron Microscopy Approach". In Perez, F. and Stephens, D. J., editors, *Methods for Analysis of Golgi Complex Function*, volume 118 of *Methods in Cell Biology*, pages 35–49. Academic Press, 2013.

[64] Ettinger, A. and Wittmann, T. "Chapter 5 - Fluorescence live cell imaging". In Waters, J. C. and Wittman, T., editors, *Quantitative Imaging in Cell Biology*, volume 123 of *Methods in Cell Biology*, pages 77–94. Academic Press, 2014.

[65] Akhshi, P. and Wu, G. "Umbrella sampling molecular dynamics simulations reveal concerted ion movement through G-quadruplex DNA channels". *Physical Chemistry Chemical Physics*, 19(18):11017–11025, 2017.

[66] Fujiwara, S. I., Sawada, K., and Amisaki, T. "Molecular dynamics study on conformational differences between dGMP and 8-oxo-dGMP: Effects of metal ions". *Journal of Molecular Graphics and Modelling*, 51:158–167, 2014.

[67] Salsbury, A., Dean, T., and Lemkul, J. A. "Polarizable Molecular Dynamics Simulations of two c-kit Oncogene Promoter G-Quadruplexes: Effect of Primary and Secondary Structure on Loop and Ion Sampling". *Journal of Chemical Theory and Computation*, 2020.

[68] Kessel, A. and Ben-Tal, N. *Introduction to Proteins: Structure, Function, and Motion, SECOND EDITION (Chapman Hall/CRC Mathematical and Computational Biology)*. CRC Press, 2018.

[69] Chiti, F. and Dobson, C. M. "Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade". *Annual Review of Biochemistry*, 86:27–68, 2017.

[70] Benzinger, T. L., Gregory, D. M., Burkoth, T. S., Miller-Auer, H., Lynn, D. G., Botto, R. E., and Meredith, S. C. "Propagating structure of Alzheimer's β-amyloid(10-35) is parallel β-sheet with residues in exact register". *Proceedings of the National Academy of Sciences of the United States of America*, 95(23):13407–13412, 1998.

[71] Simone Ruggeri, F., Habchi, J., Cerreta, A., and Dietler, G. "AFM-Based Single Molecule Techniques: Unraveling the Amyloid Pathogenic Species". *Current Pharmaceutical Design*, 22(26):3950–3970, 2016.

[72] Hartmann, J. and Zacharias, M. "Analysis of amyloidogenic transthyretin mutations using continuum solvent free energy calculations". *Proteins: Structure, Function and Bioinformatics*, 90(12):2080–2090, 2022.

[73] Sasmal, S., Schwierz, N., and Head-Gordon, T. "Mechanism of nucleation and growth of Aβ40 fibrils from all-atom and coarse-grained simulations". *Journal of Physical Chemistry B*, 120(47):12088–12097, 2016.

[74] Frenkel, D. and Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*. Computational science series. Elsevier Science, 2nd edition, 2001.

[75] Salsbury, A. M. and Lemkul, J. A. "Recent developments in empirical atomistic force fields for nucleic acids and applications to studies of folding and dynamics". *Current Opinion in Structural Biology*, 67:9–17, 2021.

[76] Bayly, C. I., Merz, K. M., Ferguson, D. M., Cornell, W. D., Fox, T., Caldwell, J. W., Kollman, P. A., Cieplak, P., Gould, I. R., and Spellmeyer, D. C. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[77] Sponer, J., Bussi, G., Krepl, M., Banas, P., Bottaro, S., Cunha, R. A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurečka, P., Walter, N. G., and Otyepka, M. "RNA structural dynamics as captured by molecular simulations: A comprehensive overview". *Chemical Reviews*, 118(8):4177–4338, 2018.

[78] Woods, R. J. and Chappelle, R. "Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates". *Journal of Molecular Structure: THEOCHEM*, 527(1-3):149–156, 2000.

[79] Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham, T. E., and Jurečka, P. "Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles". *Journal of Chemical Theory and Computation*, 7(9):2886–2902, 2011.

[80] Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., and Orozco, M. "Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers". *Biophys. J.*, 92(11):3817–3829, 2007.

[81] Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., Portella, G., Battistini, F., Gelpí, J. L., González, C., Vendruscolo, M., Laughton, C. A., Harris, S. A., Case, D. A., and Orozco, M. "Parmbsc1: a refined force field for DNA simulations". *Nature Methods*, 13(1):55–58, 2016.

[82] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. "Comparison of multiple amber force fields and development of improved protein backbone parameters". *Proteins: Structure, Function and Genetics*, 65(3):712–725, 2006.

[83] Kollman, P., Dixon, R., Cornell, W., Fox, T., Chipot, C., and Pohorille, A. *The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data*, pages 83–96. Springer Netherlands, Dordrecht, 1997.

[84] Wang, J., Cieplak, P., and Kollman, P. A. "How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules?". *Journal of Computational Chemistry*, 21(12):1049–1074, 2000.

[85] Justin I. Odegaard and Chawla, A. "NIH Public Access". *Bone*, 23(1):1–7, 2008.

[86] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. *Interaction Models for Water in Relation to Protein Hydration*, pages 331–342. Springer Netherlands, Dordrecht, 1981.

[87] Berendsen, H. J., Grigera, J. R., and Straatsma, T. P. "The missing term in effective pair potentials". *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[88] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. "Comparison of simple potential functions for simulating liquid water". , 79(2):926–935, 1983.

[89] Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., and Head-Gordon, T. "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew". *The Journal of Chemical Physics*, 120(20):9665–9678, 2004.

[90] Mamatkulov, S. and Schwierz, N. "Force fields for monovalent and divalent metal cations in tip3p water based on thermodynamic and kinetic properties". *J. Chem. Phys.*, 148, 2018.

[91] Zhou, R. "Replica exchange molecular dynamics method for protein folding simulation.". *Methods in molecular biology*, 350(November):205–223, 2007.

[92] Zuckermann, D. M. *Statistical Physics of Biomolecules AN INTRODUCTION*. Taylor and Francis Group, LLC, 2010.

[93] Honeycutt, J. D. and Thirumalai, D. "The nature of folded states of globular proteins". *Biopolymers*, 32(6):695–709, 1992.

[94] Dawson, W. K., Maciejczyk, M., Jankowska, E. J., and Bujnicki, J. M. "Coarse-grained modeling of RNA 3D structure". *Methods*, 103(2016):138–156, 2016.

[95] Pasquali, S. and Derreumaux, P. "HiRE-RNA: A high resolution coarse-grained energy model for RNA". *J. Phys. Chem. B*, 114(37):11957–11966, 2010.

[96] Boniecki, M. J., Lach, G., Dawson, W. K., Tomala, K., Lukasz, P., Soltysinski, T., Rother, K. M., and Bujnicki, J. M. "SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction". *Nucleic Acids Res.*, 44(7):e63–e63, 2015.

[97] Cruz-León, S., Vázquez-Mayagoitia, A., Melchionna, S., Schwierz, N., and Fyta, M. "Coarse-Grained Double-Stranded RNA Model from Quantum-Mechanical Calculations". *J. Phys. Chem. B*, 122(32):7915–7928, 2018.

[98] Šulc, P., Romano, F., Ouldridge, T. E., Doye, J. P., and Louis, A. A. "A nucleotide-level coarse-grained model of RNA". *Journal of Chemical Physics*, 140(23), 2014.

[99] Hyeon, C. and Thirumalai, D. "Mechanical unfolding of RNA: From hairpins to structures with internal multiloops". *Biophysical Journal*, 92(3):731–743, 2007.

[100] Denesyuk, N. A. and Thirumalai, D. "How do metal ions direct ribozyme folding?". *Nat. Chem.*, 7:793–801, 2015.

[101] Denesyuk, N. A., Hori, N., and Thirumalai, D. "Molecular simulations of ion effects on the thermodynamics of rna folding". *The J. Phys. Chem. B*, 122(50):11860–11867, 2018.

[102] Nguyen, H. T., Hori, N., and Thirumalai, D. "Theory and simulations for rna folding in mixtures of monovalent and divalent cations". *Proc. Natl. Acad. Sci.*, 116(42): 21022–21030, 2019.

[103] Guo, J. U. and Bartel, D. P. "RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria". *Science*, 353(6306), 2016.

[104] Karsisiotis, A. I., O'Kane, C., and Webba da Silva, M. "DNA quadruplex folding formalism - A tutorial on quadruplex topologies". *Methods*, 64(1):28–35, 2013.

[105] Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. "Telomeric Repeat-Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends". *Science (New York, N.Y.)*, 318(November):798–801, 2007.

[106] Azzalin, C. M. and Lingner, J. "Telomeres: The silence is broken". *Cell Cycle*, 7(9): 1161–1165, 2008.

[107] Šponer, J., Bussi, G., Stadlbauer, P., Kührová, P., Banáš, P., Islam, B., Haider, S., Neidle, S., and Otyepka, M. "Folding of guanine quadruplex molecules–funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies". *Biochim. Biophys. Acta - Gen. Subj.*, 1861(5):1246–1263, 2017.

[108] Campbell, N. H. and Parkinson, G. N. "Crystallographic studies of quadruplex nucleic acids". *Methods*, 43(4):252–263, 2007.

[109] Cerofolini, L., Amato, J., Giachetti, A., Limongelli, V., Novellino, E., Parrinello, M., Fragai, M., Randazzo, A., and Luchinat, C. "G-triplex structure and formation propensity". *Nucleic Acids Res.*, 42(21):13393–13404, 2014.

[110] Dai, J., Carver, M., Punchihewa, C., Jones, R. A., and Yang, D. "Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: Insights into structure polymorphism of the human telomeric sequence". *Nucleic Acids Res.*, 35(15):4927–4940, 2007.

[111] Dai, J., Punchihewa, C., Ambrus, A., Chen, D., Jones, R. A., and Yang, D. "Structure of the intramolecular human telomeric G-quadruplex in potassium solution: A novel adenine triple formation". *Nucleic Acids Res.*, 35(7):2440–2450, 2007.

[112] Stadlbauer, P., Mazzanti, L., Cragnolini, T., Wales, D. J., Derreumaux, P., Pasquali, S., and Šponer, J. "Coarse-Grained Simulations Complemented by Atomistic Molecular Dynamics Provide New Insights into Folding and Unfolding of Human Telomeric G-Quadruplexes". *J. Chem. Theory Comput.*, 12(12):6077–6097, 2016.

[113] Stadlbauer, P., Kührová, P., Banáš, P., Koča, J., Bussi, G., Trantírek, L., Otyepka, M., and Šponer, J. "Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations". *Nucleic Acids Res.*, 43(20):9626–9644, 2015.

[114] Stadlbauer, P., Kührová, P., Vicherek, L., Banáš, P., Otyepka, M., Trantírek, L., and Šponer, J. "Parallel G-triplexes and G-hairpins as potential transitory ensembles in the folding of parallel-stranded DNA G-Quadruplexes". *Nucleic Acids Res.*, 47(14): 7276–7293, 2019.

[115] Luo, D. and Mu, Y. "Computational Insights into the Stability and Folding Pathways of Human Telomeric DNA G-Quadruplexes". *J. Phys. Chem. B*, 120(22):4912–4926, 2016.

[116] Denesyuk, N. A. and Thirumalai, D. "Coarse-grained model for predicting RNA folding thermodynamics". *J. Phys. Chem. B*, 117(17):4901–4911, 2013.

[117] Hyeon, C. and Thirumalai, D. "Mechanical unfolding of RNA hairpins". *Proc. Natl. Acad. Sci. U. S. A.*, 102(19):6789–6794, 2005.

[118] Hori, N., Denesyuk, N. A., and Thirumalai, D. "Salt Effects on the Thermodynamics of a Frameshifting RNA Pseudoknot under Tension". *J. Mol. Biol.*, 428(14):2847–2859, 2016.

[119] Fuks, C., Falkner, S., Schwierz, N., and Hengesbach, M. "Combining Coarse-Grained Simulations and Single Molecule Analysis Reveals a Three-State Folding Model of the Guanidine-II Riboswitch". *Front. Mol. Biosci.*, 9(April):1–16, 2022.

[120] Rother, M., Rother, K., Puton, T., and Bujnicki, J. M. "ModeRNA: a tool for comparative modeling of RNA 3D structure". *Nucleic Acids Res.*, 39(10):4007–4022, 2011.

[121] Hjelm, B. E., Rollins, B., Morgan, L., Sequeira, A., Mamdani, F., Pereira, F., Damas, J., Webb, M. G., Weber, M. D., Schatzberg, A. F., Barchas, J. D., Lee, F. S., Akil, H., Watson, S. J., Myers, R. M., Chao, E. C., Kimonis, V., Thompson, P. M., Bunney, W. E., and Vawter, M. P. "Splice-Break: Exploiting an RNA-seq splice junction algorithm to discover mitochondrial DNA deletion breakpoints and analyses of psychiatric disorders". *Nucleic Acids Research*, 47(10):1–14, 2019.

[122] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation". *J. Chem. Theo. Comp.*, 4(3):435–447, 2008.

[123] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. "Comparison of simple potential functions for simulating liquid water". *J. Chem. Phys.*, 79(2):926–935, 1983.

[124] Bussi, G., Donadio, D., and Parrinello, M. "Canonical sampling through velocity rescaling". *J. Chem. Phys.*, 126(1):14101, 2007.

[125] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. "Molecular dynamics with coupling to an external bath". *J. Chem. Phys.*, 81 (8):3684–3690, 1984.

[126] Parrinello, M. and Rahman, A. "Polymorphic transitions in single crystals: A new molecular dynamics method". *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[127] Zhang, A. Y., Bugaut, A., and Balasubramanian, S. "A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology". *Biochemistry*, 50(33):7251–7258, 2011.

[128] Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. "Plumed 2: New feathers for an old bird". *Comput. Phys. Commun.*, 185:604–613, 2014.

[129] Gray, R. D., Trent, J. O., and Chaires, J. B. "Folding and unfolding pathways of the human telomeric G-quadruplex". *J. Mol. Biol.*, 426(8):1629–1650, 2014.

[130] Bian, Y., Tan, C., Wang, J., Sheng, Y., Zhang, J., and Wang, W. "Atomistic Picture for the Folding Pathway of a Hybrid-1 Type Human Telomeric DNA G-quadruplex". *PLoS Comput. Biol.*, 10(4), 2014.

[131] Havrila, M., Stadlbauer, P., Kührová, P., Banáš, P., Mergny, J. L., Otyepka, M., and Šponer, J. "Structural dynamics of propeller loop: Towards folding of RNA G-quadruplex". *Nucleic Acids Res.*, 46(17):8754–8771, 2018.

[132] Acherman, M. J. and Clapham, D. E. "Ion channels — basic science and clinical disease". *Mechanisms of Disease*, pages 1575–1586, 1997.

[133] Di Resta, C. and Becchetti, A. *Introduction to Ion Channels*, pages 9–21. Springer New York, New York, NY, 2010.

[134] González, C., Baez-Nieto, D., Valencia, I., Oyarzún, I., Rojas, P., Naranjo, D., and Latorre, R. "K+ channels: Function-structural overview". *Comprehensive Physiology*, 2(3):2087–2149, 2012.

[135] Forman, S. L., Fettinger, J. C., Pieraccini, S., Gottarelli, G., and Davis, J. T. "Toward artificial ion channels: A lipophilic G-quadruplex". *Journal of the American Chemical Society*, 122(17):4060–4067, 2000.

[136] Siebenmorgen, T. and Zacharias, M. "Origin of Ion Specificity of Telomeric DNA G-Quadruplexes Investigated by Free-Energy Simulations". *Biophysical Journal*, 112 (11):2280–2290, 2017.

[137] Akhshi, P., Mosey, N. J., and Wu, G. "Free-energy landscapes of ion movement through a G-quadruplex DNA channel". *Angewandte Chemie - International Edition*, 51(12):2850–2854, 2012.

[138] Chen, L., Sakai, N., Moshiri, S. T., and Matile, S. "Toward supramolecular ion channels formed by oligonucleotide analogs: Hydrophobic guanine dimers". *Tetrahedron Letters*, 39(22):3627–3630, 1998.

[139] Kaucher, M. S., Harrell, W. A., and Davis, J. T. "A unimolecular G-quadruplex that functions as a synthetic transmembrane Na+ transporter". *Journal of the American Chemical Society*, 128(1):38–39, 2006.

[140] Li, C., Chen, H., Chen, Q., Shi, H., Yang, X., Wang, K., and Liu, J. "Lipophilic G-Quadruplex Isomers as Biomimetic Ion Channels for Conformation-Dependent Selective Transmembrane Transport". *Analytical Chemistry*, 92(14):10169–10176, 2020.

[141] Lemkul, J. A. "Same fold, different properties: Polarizable molecular dynamics simulations of telomeric and TERRA G-quadruplexes". *Nucleic Acids Research*, 48 (2):561–575, 2020.

[142] Aydin, F., Zhan, C., Ritt, C., Epsztein, R., Elimelech, M., Schwegler, E., and Pham, T. A. "Similarities and differences between potassium and ammonium ions in liquid water: A first-principles study". *Physical Chemistry Chemical Physics*, 22(4):2540–2548, 2020.

[143] Werbeck, N. D., Kirkpatrick, J., Reinstein, J., and Hansen, D. F. "Using 15N-ammonium to characterise and map potassium binding sites in proteins by NMR spectroscopy". *ChemBioChem*, 15(4):543–548, 2014.

[144] Salsbury, A. M. and Lemkul, J. A. "Molecular Dynamics Simulations of the c-kit1 Promoter G-Quadruplex: Importance of Electronic Polarization on Stability and Cooperative Ion Binding". *Journal of Physical Chemistry B*, 123(1):148–159, 2019.

[145] Podbevšek, P., Šket, P., and Plavec, J. "Stacking and not solely topology of T3 loops controls rigidity and ammonium ion movement within d(G4T3G4) 2 G-quadruplex". *Journal of the American Chemical Society*, 130(43):14287–14293, 2008.

[146] Chatani, E., Yuzu, K., Ohhashi, Y., and Goto, Y. "Current understanding of the structure, stability and dynamic properties of amyloid fibrils". *International Journal of Molecular Sciences*, 22(9), 2021.

[147] Darden, T., York, D., and Pedersen, L. "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems". *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[148] Thal, D. R., Walter, J., Saido, T. C., and Fändrich, M. "Neuropathology and biochemistry of Aβ and its aggregates in Alzheimer's disease". *Acta Neuropathologica*, 129(2):167–182, 2015.

[149] Katzmarski, N., Ziegler-Waldkirch, S., Scheffler, N., Witt, C., Abou-Ajram, C., Nuscher, B., Prinz, M., Haass, C., and Meyer-Luehmann, M. "Aβ oligomers trigger and accelerate Aβ seeding". *Brain Pathology*, 30(1):36–45, 2020.

[150] Sowade, R. F. and Jahn, T. R. "Seed-induced acceleration of amyloid -β Mediated neurotoxicity in vivo". *Nature Communications*, 8(1):1–12, 2017.

[151] Ye, L., Rasmussen, J., Kaeser, S. A., Marzesco, A., Obermüller, U., Mahler, J., Schelle, J., Odenthal, J., Krüger, C., Fritschi, S. K., Walker, L. C., Staufenbiel, M., Baumann, F., and Jucker, M. "Aβ seeding potency peaks in the early stages of cerebral β-amyloidosis". *EMBO reports*, 18(9):1536–1544, 2017.

[152] Ancsin, J. B. "Amyloidogenesis: Historical and modern observations point to heparan sulfate proteoglycans as a major culprit". *Amyloid*, 10(2):67–79, 2003.

[153] McLaurin, J. A., Franklin, T., Zhang, X., Deng, J., and Fraser, P. E. "Interactions of Alzheimer amyloid-β peptides with glycosaminoglycans: Effects on fibril nucleation and growth". *European Journal of Biochemistry*, 266(3):1101–1110, 1999.

[154] Iannuzzi, C., Irace, G., and Sirangelo, I. "The effect of glycosaminoglycans (GAGs) on amyloid aggregation and toxicity". *Molecules*, 20(2):2510–2528, 2015.

[155] Asante, E. A., Smidak, M., Grimshaw, A., Houghton, R., Tomlinson, A., Jeelani, A., Jakubcova, T., Hamdan, S., Richard-Londt, A., Linehan, J. M., Brandner, S., Alpers, M., Whitfield, J., Mead, S., Wadsworth, J. D., and Collinge, J. "A naturally occurring variant of the human prion protein completely prevents prion disease". *Nature*, 522 (7557):478–481, 2015.

[156] Jonsson, T., Atwal, J. K., Steinberg, S., Snaedal, J., Jonsson, P. V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., Hoyte, K., Gustafson, A., Liu, Y., Lu, Y., Bhangale, T., Graham, R. R., Huttenlocher, J., Bjornsdottir, G., Andreassen, O. A., Jonsson, E. G., Palotie, A., Behrens, T. W., Magnusson, O. T., Kong, A., Thorsteinsdottir, U., Watts, R. J., and Stefansson, K. "A mutation in APP protects against Alzheimer's disease and age-related cognitive decline". *Nature*, 488(7409):96, 2012.

[157] Chen, R., Chen, Q., Zheng, J., Zeng, Z., Chen, M., Li, L., and Zhang, S. "Serum amyloid protein A in inflammatory bowel disease: from bench to bedside". *Cell Death Discovery*, 9(1):1–10, 2023.

[158] Jucker, M. and Walker, L. C. "Self-propagation of pathogenic protein aggregates in neurodegenerative diseases". *Nature*, 501(7465):45–51, 2013.

[159] Mori M, H. K., Tiang G. "AA amyloidosis-resistant CE/J mice have Saa1 and Saa2 genes that encode an identical SAA isoform". *Amyloid.*, 21(1):1–8, 2014.

[160] Tailleux, A., Duriez, P., Fruchart, J. C., and Clavey, V. "Apolipoprotein A-II, HDL metabolism and atherosclerosis". *Atherosclerosis*, 164(1):1–13, 2002.

[161] Sawashita, J., Zhang, B., Hasegawa, K., Mori, M., Naiki, H., Kametani, F., and Higuchi, K. "C-terminal sequence of amyloid-resistant type F apolipoprotein A-II inhibits amyloid fibril formation of apolipoprotein A-II in mice". *Proceedings of the National Academy of Sciences of the United States of America*, 112(8):E836–E845, 2015.

[162] Park, S.-J., Kern, N., Brown, T., Lee, J., and Im, W. "CHARMM-GUI PDB Manipulator: Various PDB Structural Modifications for Biomolecular Modeling and Simulation". *Journal of Molecular Biology*, page 167995, 2023.

[163] Schwierz, N., Frost, C. V., Geissler, P. L., and Zacharias, M. "Dynamics of Seeded Aβ40-Fibril Growth from Atomistic Molecular Dynamics Simulations: Kinetic Trapping and Reduced Water Mobility in the Locking Step". *Journal of the American Chemical Society*, 138(2):527–539, 2016.

[164] Weissmann, C. and Flechsig, E. "PrP knock-out and PrP transgenic mice in prion research". *British Medical Bulletin*, 66:43–60, 2003.

[165] Vaughan, D. G., Stammerjohn, S. E., and Conkright, M. E. "Gene with Dominant-Negative Effect". *Ecological Research*, 323(March):1473–1478, 2009.

[166] Callewaert, L. and Michiels, C. W. "Lysozymes in the animal kingdom". *Journal of Biosciences*, 35(1):127–160, 2010.

[167] Morozova-Roche, L. A., Zurdo, J., Spencer, A., Noppe, W., Receveur, V., Archer, D. B., Joniau, M., and Dobson, C. M. "Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants". *Journal of Structural Biology*, 130(2-3):339–351, 2000.

[168] Dumoulin, M., Kumita, J. R., and Dobson, C. M. "Normal and aberrant biological self-assembly: Insights from studies of human lyspzyme and its amyloidogenic variants". *Accounts of Chemical Research*, 39(9):603–610, 2006.

[169] Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V., and Dobson, C. M. "Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme". *Nature Structural Biology*, 9(4):308–315, 2002.

[170] Hammarström, P., Jiang, X., Hurshman, A. R., Powers, E. T., and Kelly, J. W. "Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity". *Proceedings of the National Academy of Sciences of the United States of America*, 99(SUPPL. 4):16427–16432, 2002.

[171] Bulawa, C. E., Connelly, S., DeVit, M., Wang, L., Weigel, C., Fleming, J. A., Packman, J., Powers, E. T., Wiseman, R. L., Foss, T. R., Wilson, I. A., Kelly, J. W., and Labaudinière, R. "Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade". *Proceedings of the National Academy of Sciences of the United States of America*, 109(24):9629–9634, 2012.

[172] Lappala, A. “The next revolution in computational simulations: Harnessing AI and quantum computing in molecular dynamics”. *Current Opinion in Structural Biology*, 89(Ml):102919, 2024.

[173] Kleiman, D. E., Nadeem, H., and Shukla, D. “Adaptive Sampling Methods for Molecular Dynamics in the Era of Machine Learning”. *Journal of Physical Chemistry B*, 127(50):10669–10681, 2023.

Acknowledgments

I would first like to thank my supervisor, Prof. Nadine Schwierz. Thank you, Nadine, for giving me the opportunity to work on these interesting projects, and for your guidance and insightful advice, which have been a valuable addition to my toolkit, not only in my scientific work but also in my professional life.

I spent three years of my PhD journey at the Max Planck Institute for Biophysics in Frankfurt. I would therefore like to thank Prof. Gerhard Hummer and Dr. Jürgen Köfinger for welcoming Prof. Schwierz's group into the Theoretical Biophysics department and allowing us to take part in the weekly PhD and group meetings. I learned a lot by participating in these.

At the University of Augsburg, I would like to thank Prof. Christoph Weber and Prof. Markus Heyl for welcoming our group into their departments and making us feel included in all the group activities. I am grateful to have met and exchanged ideas with many wonderful colleagues from both groups.

I would also like to thank Prof. Harald Schwalbe from Goethe University Frankfurt and his then-PhD students Dr. Ines Burkhart and Dr. Diana Müller for a great collaboration on the RNA G4 folding project.

My sincere thanks also go to Prof. Marcus Fändrich from Ulm University, as well as members of his group Peter B. Pfeiffer, Giada Andreotti, and Sara Karimi, for the excellent collaboration on the amyloid projects.

I would like to acknowledge and thank my colleagues Ibrahim, Akhil, Christian, Adrian, and Jonas for creating a great working atmosphere and for many interesting scientific discussions. Also thank you for the feedback on my thesis.

Last but not least, I would like to thank my husband Josip, my mom Jagoda, and the rest of my family and friends, for always believing in me.