Molecular Dynamics simulations for the study of interaction between non-canonical DNA structures and biochemically relevant co-solutes

Von der Fakultät 8 - Mathematik und Physik
der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Ewa Anna Oprzeska-Zingrebe

aus Ostrów Wielkopolski, Polen

Hauptberichter: Priv.-Doz. Dr. Jens Smiatek
Mitberichter: Prof. Dr. Johannes Kästner
Mitberichterin: Prof. Dr. Maria Fyta

Tag der mündlichen Prüfung: 19 Juni 2023

Institut für Computerphysik der Universität Stuttgart

2023

The research results presented in this doctoral thesis have already been published in acknowledged peer-review international journals.

The list of scientific publications that constitute this work comprise the following titles:

1. Angulo G., Astumian R. D., et al., "New methods: general discussion.", *Faraday Discussions* vol. 195, pp. 521-556, 2016.

2. Oprzeska-Zingrebe, E. A.; Smiatek, J., "Preferential binding of urea to single-stranded DNA structures: a molecular dynamics study.", *Biophys. J.* vol. 114, no. 7, pp. 1551-1562, 2018.

3. Oprzeska-Zingrebe, E. A.; Smiatek, J., "Aqueous ionic liquids in comparison with standard co-solutes - Differences and common principles in their interaction with protein and DNA structures.", *Biophys. Rev.* vol. 10, no. 3, pp. 809-824, 2018.

4. Oprzeska-Zingrebe, E. A.; Meyer, S.; Roloff, A.; Kunte, H.-J.; Smiatek, J., "Influence of compatible solute ectoine on distinct DNA structures: thermodynamic insights into molecular binding mechanisms and destabilization effects.", *Phys. Chem. Chem. Phys.* vol. 20, no. 40, pp. 25861-25874, 2018.

5. Oprzeska-Zingrebe, E. A.; Kohagen, M.; Kästner, J.; Smiatek, J., "Unfolding of DNA by co-solutes: insights from Kirkwood–Buff integrals and transfer free energies.", *Europ. Phys. J. Spec. Top.* vol. 227, pp. 1665-1679, 2019.

6. Oprzeska-Zingrebe, E. A.; Smiatek, J., "Aqueous Mixtures of Urea and Trimethylamine-N-oxide: Evidence for Kosmotropic or Chaotropic Behavior?", *J. Phys. Chem. B* vol. 123, no. 20, pp. 4415-4424, 2019.

7. Oprzeska-Zingrebe, E. A.; Smiatek, J., "Some Notes on the Thermodynamic Accuracy of Coarse-Grained Models.", *Front. Mol. Biosci.* vol. 6, no. 87, 2019.

8. Oprzeska-Zingrebe, E. A.; Smiatek, J., "Interactions of a DNA G-quadruplex with TMAO and Urea: A Molecular Dynamics Study on Co-Solute Compensation Mechanisms.", *Phys. Chem. Chem. Phys.* vol. 23, no. 2, pp. 1254-1264, 2021.

9. Oprzeska-Zingrebe, E. A.; Smiatek, J., "Basket-type G-Quadruplex with Two Tetrads in the Presence of TMAO and Urea: A Molecular Dynamics Study", *J. Mol. Struct.* vol. 1274, no. 2, pp. 134375, 2023.

Abstract

Non-canonical nucleic acid structures, such as DNA G-quadruplexes and i-Motifs, have been proved to play an important role in key biological processes, including gene expression, replication, regulation or telomere maintenance. The presence of G-quadruplexes in promoter regions of certain oncogenes turn them into a potential target for cancer therapies. Besides their biological implications, non-canonical DNA structures are present in genomes of various organisms, who adopt certain levels of co-solutes to protect their internal structures against the harsh environment. This study presents the research on the selected non-canonical DNA structures of particular biological relevance: G-quadruplex with only two tetrads, small DNA hairpin and ssDNA strand as well as canonical double helix. The atomistic molecular dynamics (MD) simulations have been applied to elucidate the structural, configuration and solvation properties of the analyzed structures in the presence of assorted co-solutes, composing the native cellular environment in nature: urea, ectoine and trimethylamine-N-oxide (TMAO). With the application of molecular theory of solutions, one determines and exemplifies the thermodynamic properties of investigated structures in various environments close to the physiological conditions present in living cells. This study uncovers the versatile nature of DNA interaction with diverse co-solutes and water, as well as the cross-interactions between the inorganic components of the biomolecular solution. The cellular mechanisms of DNA structural stabilization and destabilization are hereby described in terms of preferential binding and preferential exclusion, with particular emphasis on the properties of solvent structure within individual solvation shells. In this regards, this work presents a comprehensive study on the intracellular interactions involving nucleic acids, thus shedding light into their microscopic properties and opening the path for further biomedical research.

Abstract

Die Wechselwirkungen zwischen Biomolekülen wie Proteinen oder Nukleinsäuren mit verschiedenen Osmolyten waren in den letzten Jahren Gegenstand intensiver Forschung. Unter ihnen gewinnen sowohl die Mittel, die die native Struktur des Biomoleküls stabilisieren, als auch diejenigen, die sie destabilisieren, besondere Bedeutung. Während die Auswirkungen dieser Kosolute auf Proteine bisher umfassend erforscht und untersucht wurden, blieben ihre Auswirkungen auf Nukleinsäuren oft im Dunkeln. Da viele dieser Kosolute nicht nur in der zellulären Umgebung der Nukleinsäuren allgegenwärtig sind, sondern auch in den Laboren für biomolekulare Studien weit verbreitet sind, ist es dringend erforderlich, ein klares Bild von den Auswirkungen der Kosolute in Nukleinsäuresystemen zu gewinnen. Neben den kanonischen helikalen DNA-Strukturen, die als Standardträger der genetischen Information gelten, enthält das Genom zahlreicher Organismen viele nicht-kanonische Strukturen, deren biologische Rolle weit über die typische Expression von Genen hinausgeht. Unter ihnen erlangen DNA-G-Quadruplexe besondere Bedeutung aufgrund ihres Vorhandenseins in transkriptionsregulatorischen Regionen zahlreicher Gene und Onkogene sowie in chromosomalen Telomeren und ihrer daraus resultierenden regulatorischen Rolle bei grundlegenden biochemischen Prozessen wie der Genexpression und der funktionellen Genomik.

Diese Arbeit befasst sich mit der Wechselwirkung von Kosoluten mit verschiedenen nichtkanonischen DNA-Strukturen, wie kurzen DNA-Hairpins oder DNA-G-Quadruplex, sowie mit der kanonischen Watson-Crick-DNA-Helix. Um einen tieferen Einblick in das Problem der DNA-Kosolute-Wechselwirkungen zu erhalten, habe ich die Berechnungsmethode der allatomaren Molekulardynamik (MD)-Simulationen in Kombination mit der Molekulartheorie der Lösungen angewendet. Der Vorteil von Computersimulationen liegt darin, dass sie Details der Co-Lösungsmitteleffekte auf atomarer Ebene liefern können, was im Experiment normalerweise nicht möglich ist. Die experimentellen Ansätze werden zwar als direkterer Weg der biomolekularen Forschung angesehen, können aber oft nicht die molekularen Details der DNA-Co-Lösungsmittel-Wechselwirkungen oder die daraus resultierenden Entfaltungsphänomene liefern. Trotz zahlreicher Vorteile haben die rechnergestützten Methoden jedoch auch einige Einschränkungen. Eine davon ist die verfügbare Simulationszeitskala, die in der Regel kürzer ist als die Zeitskala mehrerer spontan ablaufender biochemischer Prozesse, wie die Faltung von Nukleinsäuren oder die Öffnung von Basenpaaren. Daher werden bestimmte molekulare Ereignisse selbst in den längsten MD-Simulationen in der Regel nicht beobachtet.

Als ersten Ansatz zur Untersuchung der Wechselwirkung der Kosolute mit Nukleinsäurestrukturen untersuchte ich das gegenseitige Zusammenspiel zwischen verschiedenen Komponenten der reinen Lösung, wie Wasser und Kosolute, selbst. In Kapitel 3 habe ich die Eigenschaften der Wechselwirkungen von TMAO und Harnstoff in wässrigen binären und ternären Lösungen untersucht. Die erhaltenen Ergebnisse bestätigen ein eher vernachlässigbares Assoziationsverhalten zwischen den beiden Kosoluten, selbst bei höheren Konzentrationen, was durch die geringe Stabilität und Anzahl der TMAO-Harnstoff-Wasserstoffbindungen belegt wird. Es hat sich gezeigt, dass TMAO bevorzugt Wasserstoffbrückenbindungen mit Wasser bildet, und diese sind angeblich die stabilsten aller Wasserstoffbrückenbindungen, die in wässrigen TMAO-Harnstoff-Gemischen gebildet werden. Es hat sich gezeigt, dass TMAO auch die Wasserstoffbrückenbindungen von Wasser erheblich stabilisiert, während der Einfluss von Harnstoff auf die Dynamik der Wasserstoffbrückenbindung von Wasser nur gering ist. Beide Kosolute verlangsamen die Wasserdynamik, was sich in längeren dipolaren Relaxationszeiten des Wassers und den daraus resultierenden Veränderungen der Vorwärtslebensdauer der Wasserstoffbrückenbindungen zeigt. Im Allgemeinen wurde festgestellt, dass TMAO einen viel stärkeren Einfluss auf die Wassereigenschaften ausübt als Harnstoff.

Trotz der klassischen Einteilung in Chaotrope ("Strukturbrecher") und Kosmotrope ("Strukturbildner") zeigen die vorliegenden Ergebnisse, dass beide Kosolute in wässriger Lösung kosmotrope Eigenschaften aufweisen. Obwohl die Wirkungen leicht unterschiedlich sind, stärken sowohl Harnstoff als auch TMAO das Wasserstoffbrückenbindungsnetzwerk der Wassermoleküle, was durch die Zunahme der relativen Anzahl der gesamten Wasserstoffbrückenbindungen und eine Abnahme der Wasserfluktuationen verdeutlicht wird. Dies führt zu verlängerten dipolaren Relaxationszeiten und längeren Vorwärtslebensdauern der Wasserstoffbrückenbindungen. Außerdem wurde keine Kompensation des Einflusses von Harnstoff und TMAO auf die Wasserstruktur und -dynamik in

Gegenwart beider Kosolute beobachtet. Aufgrund der zwitterionischen Struktur von TMAO sind die Wasserstoffbrückenbindungen zwischen Wasser und TMAO-Molekülen im Vergleich zu Wasser und Harnstoff wesentlich stärker, was zu stabilen Hydratationsschalen um TMAO führt. Daher können die dispergierten und relativ instabilen TMAO-Harnstoff-Komplexe im Hinblick auf ihre individuellen molekularen Eigenschaften als Lösungsmittel-geteilte Kosolute-Paare betrachtet werden. Unter Berücksichtigung all dieser Ergebnisse zeigt diese Studie das Fehlen der eindeutigen chaotropen Eigenschaften von Harnstoff, was ein neues Licht auf die Bedeutung von wasservermittelten Effekten für die Stabilisierung oder Destabilisierung von Makromolekülen wirft. Da sich beide Kosolute überwiegend wie Kosmotrope verhalten, kann man daraus schließen, dass der globale Einfluss auf die Wasserstruktur und -dynamik sowie die wasservermittelten Wechselwirkungen für Stabilisierungs- oder Destabilisierungsphänomene von geringer Bedeutung sind. Dies deutet darauf hin, dass organische Kosolute die strukturellen Modifikationseffekte auf Makromoleküle über die Unterschiede in ihrem Akkumulationsmuster um den zentralen Solut ausüben können, was mit den bisherigen Erkenntnissen im Rahmen von Local/Bulk-Partitioning-Ansätzen übereinstimmt. Im Hinblick auf diese Ergebnisse sollte die Unterscheidung von organischen Kosoluten in Kosmotrope oder Chaotrope sorgfältig überdacht werden, um die Auswirkungen auf die makromolekulare Struktur zu erklären. Darüber hinaus ist zu erwarten, dass auch andere organische Kosolute ein analoges Verhalten zeigen, wie es für TMAO und Harnstoff berichtet wurde.

In Kapitel 4 habe ich die Wechselwirkung zwischen einem kurzen 7-bp-DNA-Oligonukleotid mit der Sequenz d(GCGAAGC) in seiner nativen und entfalteten Form in wässriger Lösung mit verschiedenen Konzentrationen von Harnstoff untersucht. Die vorgestellten Ergebnisse zeigen die bevorzugte Bindung von Harnstoff an die DNA unabhängig von ihrer Konformation, was in Übereinstimmung mit den jüngsten experimentellen Erkenntnissen zur Bindung von Harnstoff an DNA und RNA steht. Darüber hinaus weisen die Ergebnisse eindeutig darauf hin, dass Harnstoff den ungefalteten Zustand der DNA begünstigt, was mit den bisherigen Erkenntnissen über den Einfluss von Harnstoff auf Proteine übereinstimmt. Diese Ergebnisse zeigen den unspezifischen Bindungsmechanismus zwischen Harnstoff und Nukleinsäurekonformationen auf, der qualitativ auf verschiedene DNA-Formen übertragbar ist, auch wenn sich die Stärke der DNA-Harnstoff-Interaktion möglicherweise zwischen nicht-kanonischen DNA-Hairpins und der DNA-Doppelhelix unterscheidet. Somit gibt diese Studie Einblicke in die Natur der bevorzugten Wechselwirkungen in ternären Systemen aus DNA, Harnstoff und Wasser und liefert eine konsistente Sicht auf die energetischen Beiträge und die lokalen Wechselwirkungen zwischen DNA und Harnstoff.

In Kapitel 5 habe ich die Wechselwirkungen zwischen zwei DNA-Strukturen untersucht: einem kurzen 7-bp-DNA-Oligonukleotid mit der Sequenz d(GCGAAGC) und einer kanonischen 24-bp-B-DNA-Helix mit Ectoin. Aufgrund seiner zellschützenden Eigenschaften wird dieser Kosolute in der pharmazeutischen Industrie häufig als Bestandteil einer breiten Palette von Gesundheitsprodukten verwendet. Diese Forschungsarbeit konzentriert sich auf das Ectoin-DNA-Bindungsverhalten als Ansatz zur Untersuchung des Einflusses von Kosoluten auf die DNA-Struktur. Obwohl man festgestellt hat, dass Ectoin im Allgemeinen die Struktur von Proteinen stabilisiert, deuten die erzielten Ergebnisse auf eine starke und unspezifische Bindung von Ectoin an die DNA hin, die durch enthalpische Mechanismen mit Lennard-Jones- und Coulomb-Wechselwirkungen mit kurzer Reichweite ausgelöst wird. Somit kann Ectoin in übereinstimmung mit den jüngsten experimentellen Ergebnissen als ein DNA-Denaturierungsmittel angesehen werden. Die denaturierende Wirkung von Ectoin kann hauptsächlich auf die stark negative Ladung des DNA-Phosphodiester-Rückgrats zurückgeführt werden, die die bevorzugte Bindung zwitterionischer Ectoin-Moleküle über starke elektrostatische Wechselwirkungen in Kombination mit ausgeprägten Dispersionsenergien begünstigt. Gleichzeitig scheint die Wasserstruktur um das DNA-Molekül selbst bei hohen Ectoin-Konzentrationen nicht beeinträchtigt zu werden, was gegen einen entropischen Beitrag zum Bindungsmechanismus spricht. Auf der Grundlage statistisch-mechanischer Lösungstheorien habe ich die damit verbundene Änderung der Schmelztemperatur der DNA bestimmt und ihre qualitative übereinstimmung mit dem Experiment nachgewiesen. Die beobachtete starke Bindung von Ectoin an die DNA erklärt auch den kürzlich berichteten Schutzmechanismus, der Strahlenschäden an Nukleinsäuren verhindert. In dieser Hinsicht überwiegt das bevorzugte Bindungsverhalten von Ectoin, das die Bildung einer stabilen Schutzhülle um die DNA fördert, seine potenziell denaturierende Wirkung auf die DNA. Es kann spekuliert werden, dass diese stabile Ectoin-Schale als effiziente Barriere für Strahlenschäden interpretiert werden kann, die den Schutz der Nukleinsäuren

gewährleistet. Somit bietet diese Studie eine kohärente Sicht auf die molekularen Aspekte der Wechselwirkungen zwischen DNA und Ectoin in wässrigen Systemen und ermöglicht es, mehr Licht auf ihre Natur zu werfen.

In Kapitel 6 habe ich einen korbartigen DNA-G-Quadruplex polymorpher Natur untersucht. Dieser Typ von G-Quadruplex besteht aus zwei G-Tetrads, die für seine einzigartige Stabilität verantwortlich sind. Ich habe MD-Simulationen in biologisch relevanten Konzentrationen von Kosoluten wie TMAO und Harnstoff durchgeführt und mich dabei auf die Ergebnisse für piezophile Organismen bezogen, die TMAO-Konzentrationen von 0,6 bis 1 M annehmen, um sich an osmotischen Stress anzupassen. Unter Bezugnahme auf die Studien, die darauf hindeuten, dass TMAO in 1:2-Mischungen von TMAO und Harnstoff die denaturierende Wirkung von Harnstoff vollständig kompensieren kann, habe ich die Systeme auch in 2 M Harnstoff und 1:2-molaren Mischlösungen von TMAO und Harnstoff simuliert. Insbesondere habe ich mich auf das DNA-Bindungsverhalten mit beiden Kosoluten und mit Wasser konzentriert, was einen Einblick in deren Einfluss auf die strukturelle Stabilität von G-Quadruplexen gibt. Mit der Anwendung der Kirkwood-Buff-Theorie wurde die molekulare Natur des DNA-Kosolute-Wechselwirkungsmechanismus erforscht. Die vorgestellten Ergebnisse belegen die bevorzugte Bindung von Harnstoff und den bevorzugten Ausschluss von TMAO-Molekülen um die DNA. Die entsprechenden Ergebnisse sind vergleichbar zwischen den Einzelkomponentenlösungen und der TMAO:Harnstoff-Mischung. Auch das Akkumulationsverhalten der Kosolute, die Anzahl der Wasserstoffbrückenbindungen und die Wechselwirkungsenergien in der Mischung sind im Vergleich zu den einzelnen Kosolute-Lösungen nicht signifikant verändert. Diese Ergebnisse deuten auf eine paarweise lineare Additivität der einzelnen Beiträge der Kosolute hin. Trotz der Daten, die zeigen, dass 1 M TMAO die denaturierende Wirkung von 2 M Harnstoff auf die Proteinstruktur vollständig kompensieren kann, zeigen meine Studien an DNA G-Quadruplex, dass der Kompensationseffekt nicht vollständig ist, wenn Nukleinsäuren betrachtet werden. Obwohl das Bindungsverhalten von Harnstoff an die DNA-Struktur in Gegenwart von TMAO viel weniger stark ausgeprägt ist als in einer reinen Harnstofflösung, wird es nicht vollständig aufgehoben, was sich auf die destabilisierende Wirkung auf die G-Quadruplex-Struktur übertragen lässt. Im Einzelnen zeigen meine Ergebnisse, dass Harnstoff als starker DNA-Destabilisator angesehen werden kann, während TMAO die DNA-Struktur im Sinne der Kirkwood-Buff-Theorie von Lösungen stabilisiert, bei der Harnstoff dazu neigt, sich eng um die DNA herum anzusammeln, während TMAO in die Bulk-Phase abgestoßen wird. Tatsächlich weisen sowohl TMAO- als auch Harnstoffschichten nur schwache gegenseitige Einflüsse oder Wechselwirkungen auf. Die kombinierte Betrachtung von TMAO und Harnstoff als künstlicher Kosolute zeigt jedoch, dass der bevorzugte Bindungskoeffizient im Vergleich zu ternären Harnstofflösungen leicht herabgesetzt ist. Folglich ist die destabilisierende Wirkung von Harnstoff in Gegenwart von TMAO aufgrund des strukturstabilisierenden Kompensationsmechanismus, an dem TMAO beteiligt ist, nicht so stark ausgeprägt. Die klare Trennung der einzelnen Kosolute-Schichten und das Fehlen wesentlicher Mischeffekte zwischen Harnstoff und TMAO um die DNA herum, die für die paarweise Additivität der einzelnen bevorzugten Bindungskoeffizienten verantwortlich sind, gewährleisten das Vorhandensein der Kompensationseffekte. Wie bereits in meiner früheren Studie gezeigt wurde, übt Harnstoff seine denaturierende Wirkung auf Nukleinsäuren über einen indirekten Mechanismus aus. Die aktuellen Ergebnisse bestätigen diese Schlussfolgerung und zeigen, dass auch die Stabilisierung der DNA-Struktur durch TMAO eher über einen indirekten Mechanismus erfolgt, der eine Interferenz mit der Wasserstruktur beinhaltet, als durch die Bildung direkter Bindungen mit Nukleinsäuren. Der Mechanismus der Stabilisierung von Nukleinsäuren durch TMAO ist also ähnlich wie bei Proteinen, obwohl es auch Studien gibt, die auf eine direkte TMAO-Protein-Wechselwirkung hinweisen.

Um die in Kapitel 6 vorgestellte Forschung zu erweitern und wesentlich zu ergänzen, wird in Kapitel 7 untersucht den 2KF7-DNA-G-Quadruplex, der nur aus zwei Tetraden besteht, aus einer eher visuellen Perspektive, indem ich MD-Simulationen zusammen mit molekularer Modellierung einsetze. Die in Kapitel 6 vorgestellte Studie befasst sich mit dem Thema der Wechselwirkungen zwischen DNA und Kosoluten aus rein thermodynamischer Sicht. Die in Kapitel 7 beschriebene Arbeit gibt einen detaillierten Einblick in biomolekulare Aspekte der beschriebenen Phänomene, um das numerisch gewonnene Bild zu vervollständigen. Während die vorgenannte Arbeit hauptsächlich die Stabilisierungs- und Destabilisierungsaspekte im Sinne der Kirkwood-Buff-Theorie von Lösungen diskutiert, stelle ich in diesem Kapitel die entsprechenden molekularen Implikationen und eine weitere Begründung für die bisherigen Erkenntnisse vor. Mit Hilfe von Molekulardynamiksimula-

tionen (MD) versuche ich, die naturnahe Umgebung von Biomolekülen zu rekonstruieren, die nicht nur Wasser, sondern auch wesentliche Kosolute wie TMAO und Harnstoff einschließt. Unter Bezugnahme auf die jüngsten Erkenntnisse über die Natur piezophiler Mikroorganismen, die bestimmte Konzentrationen und Anteile von Harnstoff und TMAO annehmen, um ihre zelluläre Reaktion auf osmotischen Stress zu modulieren, untersuchte ich den DNA-G-Quadruplex in Anwesenheit von nahezu physiologischen Konzentrationen beider Kosolute allein sowie in einer 1:2 molaren TMAO:Harnstoff-Mischung. Es hat sich gezeigt, dass diese Kombination beider Kosolute die denaturierende Wirkung von Harnstoff auf Proteine vollständig ausgleicht und gleichzeitig einen Schutz gegen osmotischen Stress in Tiefsee-Mikroorganismen bietet. Diese Arbeit liefert eine detaillierte qualitative Analyse der Wechselwirkung zwischen TMAO und Harnstoff und dem G-Quadruplex der DNA im Hinblick auf seine Stamm- und Schleifenregionen zusammen mit den Nukleinbasen und dem Phosphatrückgrat. Durch die Analyse der Dichteverteilungsfunktionen erhält man einen Einblick in die Veränderungen des DNA-Solvatationsmusters bei Zugabe von Kosoluten in Verbindung mit einer erhöhten Temperatur, die über der Schmelztemperatur T_m für diese Art von DNA liegt. Bei den meisten Simulationsbedingungen kann man beobachten, dass Harnstoffmoleküle in die DNA-Furchen eindringen, während TMAO sich überwiegend von der DNA-Oberfläche fernhält. In der Mischung aus Harnstoff und TMAO zieht TMAO den Harnstoff aus dem Inneren der DNA-Struktur heraus und sorgt so für die Solvatationshülle, die zur Stabilisierung des Biomoleküls ausreicht. Nichtsdestotrotz zeigen meine Ergebnisse, dass die G-Quadruplex-Struktur der DNA im Gegensatz zu Proteinen relativ resistent gegenüber dem Einfluss von Kosoluten ist, sowohl was die Stabilisierung als auch die Destabilisierung angeht. Während die meisten Proteine einen gewissen Grad an Denaturierung erfahren, wenn sie biologisch relevanten Konzentrationen von Harnstoff ausgesetzt werden, scheint die G-Quadruplex-Struktur durch die Anwesenheit von Harnstoff eher unbeeinflusst zu bleiben. Während TMAO nachweislich die Proteinstruktur gegen die schädlichen Auswirkungen von Harnstoff und Temperatur stabilisiert, bleibt die Stabilität der G-Quadruplex-Struktur unter physiologischen Konzentrationen von TMAO weitgehend unverändert. Dies weist auf die außergewöhnliche Stabilität von DNA-G-Quadruplexen hin, die sich nicht nur in der Temperaturbeständigkeit, sondern auch in der Unverwundbarkeit gegenüber dem Einfluss von Kosoluten ausdrückt. Darüber hinaus deuten die vorgestellten Ergebnisse darauf hin, dass ein möglicher Einfluss von Kosoluten auf die strukturelle Stabilität von G-Quadruplexen weder mit der Bildung von Wasserstoffbrücken noch mit direkten Kontakten zwischen Kosoluten und DNA zusammenhängt. Weder die die DNA-Struktur stabilisierenden intramolekularen Wasserstoffbrückenbindungen scheinen durch die Temperatur noch durch die Anwesenheit von Kosoluten beeinflusst zu werden. Somit können die Veränderungen im Flexibilitätsmuster bestimmter Teile der G-Quadruplex-Struktur auf die hydratisierende Rolle des Wassers in Verbindung mit der thermischen Instabilität bei Hochtemperatursimulationen zurückgeführt werden. Unter diesem Aspekt bietet diese Studie einen neuen Einblick in zahlreiche Aspekte des Zusammenspiels zwischen TMAO und Harnstoff und nicht-kanonischen DNA-Strukturen, die - im Gegensatz zu den meisten Proteinen - eine ausgeprägte negative Ladung tragen.

Die Wechselwirkung zwischen grundlegenden Biomolekülen wie Proteinen oder Nukleinsäuren mit molekularen Wirkstoffen, die ihre native Struktur stabilisieren oder destabilisieren, war in den letzten Jahren Gegenstand intensiver Forschung. Da bestimmte Kosolute wie Harnstoff, Ectoin oder TMAO nicht nur in der nativen Umgebung der meisten Nukleinsäuren allgegenwärtig sind, sondern auch häufig in der Medizin, Industrie oder Biotechnologie eingesetzt werden, gewinnt die Notwendigkeit, sich ein klares Bild von ihrem Verhalten in Nukleinsäuresystemen zu machen, an entscheidender Bedeutung. Die in dieser Arbeit vorgestellten Arbeiten ermöglichen es, einen Einblick in die molekularen Details der Wechselwirkungen zwischen ausgewählten biologisch wichtigen Nukleinsäurestrukturen mit relevanten Kosoluten zu erhalten und damit Hinweise auf deren potentielle Anwendung für DNA-Strukturmodulationen in der modernen Biophysik, Chemie und Biotechnologie zu geben. Darüber hinaus demonstriert diese Arbeit eine neuartige Umsetzung der Kirkwood-Buff-Theorie für die Untersuchung biomolekularer Systeme. Neben der mathematischen Erweiterung und Verfeinerung der Theorie stellt diese Studie einen Ansatz zur Lösung der wesentlichen Themen der biomolekularen Physik vor, wie die molekularen Details der Teilchen-Teilchen-Wechselwirkungen in den Systemen von biologischem Interesse. Besonderer Wert wird auf die Anwendung der Kirkwood-Buff-Theorie gelegt, um die thermodynamischen Eigenschaften der simulierten Systeme zu extrahieren, die im experimentellen Aufbau nicht verfügbar sind. Darüber hinaus konzentriert sich diese Forschung auf die praktische Anwendung und Umsetzung der The-

orie zur Untersuchung ternärer und quaternärer Systeme, die der realen Umgebung der lebenden Zellen sehr nahe kommen. In dieser Hinsicht stellt diese Arbeit einen wichtigen Beitrag zum Bereich der angewandten Physik dar, insbesondere im Hinblick auf die Biophysik und die physikalische Biochemie.

Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisor, Priv.-Doz. Dr. Jens Smiatek for his valuable advice, continuous support and patience during the course of my PhD degree. His vast knowledge and extensive experience have encouraged me to cross the boundaries in the pursuit for new discoveries in science during my academic research. Additionally, I would like to extend my sincere thanks to Prof. Dr. Christian Holm, Prof. Dr. Maria Fyta and Prof. Dr. Johannes Kästner for their mentorship and treasured support which was really influential in shaping my computational methods and critiquing my results.

My special thanks are offered to Prof. Dr. Modesto Orozco for giving me the opportunity to join the research in the Molecular Modelling and Bioinformatics Group, Institute for Research in Biomedicine in Barcelona, Spain. His insightful comments and suggestions helped me to develop and significantly extend my experience and knowledge in the computational modelling methods for biophysical applications.

I would like to thank Montserrat Terrazzas and Susann Meyer for their experimental contribution to our research projects and for beneficial discussions. I would also like to thank my colleagues in the Institute for Computational Physics, Institute for Research in Biomedicine Barcelona and further research team - Francesco Colizzi, Diego Gallego, Alexandra Balaceanu, Federica Battistini, Adam Hospital, Jürgen Walther, Leonardo Darre, Johannes Zeman, Takeshi Kobayashi, Anand Narayan Krishnamoorthy, Georg Rempfer, Joost de Graaf, Frank Uhlig, Miriam Kohagen, Satyajit Patra, Seishi Shimizu and Steven Abbott for useful hints, inspiring scientific discussions and for their assistance at every stage of the research project. Your support have made it possible to conduct and finish this research project with success. I would also like to thank Frank Huber for his technical support on my study, and to the wonderful members of the ICP administration office. Your help in solving the bureaucratic difficulties and the cherished time spent together in social settings is just invaluable.

My gratitude goes to Barbara Teutsch, who encouraged me to pursue my dreams despite difficulties and inspired to cross my limits.

I would like to thank all the members of the Institute for Computational Physics and the Institute for Research in Biomedicine. It is their kind help and support that have made my study and life in Stuttgart and Barcelona a wonderful time.

My gratitude extends to the Deutsche Forschungsgemeinschaft for financial funding through the Sonderforschungsbereich 716 (SFB 716/C8), which made it for me possible to undertake my studies in the Institute for Computational Physics, University of Stuttgart. The SimTech Graduate School is gratefully acknowledged for excellent opportunities of advanced education and skill training in the field of simulation technology.

Last but not least, I would like to express my deepest gratitude to my family: my Mother, my husband Ulrich and my wonderful children for their unwavering support, encouragement and belief in me all through my PhD study. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my research project.

Contents

1 Introduction

Nucleic acids are commonly regarded as biopolymers fundamental for the existence of every form of life. Beside canonical forms such as DNA double helix [9], which are widely researched since decades, there exist also numerous non-canonical structures, which are characterized by uncommon base pairing and thus often hold unique properties. In nature, one of the most commonly present non-canonical structures are DNA i-Motifs, formed by cytosine-rich sequences [10, 11, 12, 13, 14], and G-quadruplexes, emerging from sequences rich in guanine [15, 16, 17, 13, 18, 19, 20]. To date, a wide range of biological processes in nature has been proved to be dependent on the formation of specific DNA secondary and tertiary structures. Those structures are frequently involved in fundamental processes in cells, including transcription termination, regulation of gene expression or intermolecular binding [21, 22, 23]. The importance of DNA higher-order structures lays in their involvement in numerous cellular events, including replication, transcription and translation, as well as the nucleic-acid-based mechanisms to regulate telomere shortening [17, 24]. Especially the presence of non-Watson-Crick DNA quadruplex structures in human telomeres turns them into the potential target for cancer treatment therapies [25]. Furthermore, a reversible mechanism of non-Watson-Crick DNA folding can be applied in modern nanobiotechnology. Possible applications of DNA higher-order structures on that field are nowadays the object of intensive research [26, 27].

Although often simplified in computational studies, the natural environment of most biomolecules is a complex mixture of water, ions, salts, lipids, amino acids and other components, which cross-interact and mutually influence each other. Depending on the target biomolecule, some of them may remain inert and not affect biomolecule's function or stability, whereas the others can act as destabilizers, stabilizers or even structure makers, assisting the formation of certain higher-order biomolecular structures [28, 29, 30, 31]. Due to the significance of nucleic acid structures, which act as the genetic information carrier and thus influence cell's existence, providing the cellular environment stabilizing them reaches on particular importance. Not every type of cell developed mechanical barriers to protect their DNA. Unlike most eukaryotic cells, where the genomic DNA is confined within the nuclear envelope forming cell's nucleus, prokaryotic organisms' DNA does not belong to a distinct cellular compartment. Being located freely in the cytoplasm, such DNA is exposed on interaction with other components of the cytosol, which adverts the necessity to establish the means of its preservation. Specifically, unicellular organisms exposed on severe environmental stress need to develop certain mechanisms to protect their DNA from damages. Many halophilic and halotolerant bacteria synthesize and accumulate organic osmolytes to withstand harsh environmental conditions like high salinity or temperatures and to maintain the osmotic equilibrium with the surrounding [32, 33, 34, 35, 36, 37]. These osmolytes synthesized by extremophilic microorganisms are called extremolytes [34] and chemically they are composed of sugars, amino acids, polyols and heteroside derivatives [38, 39, 34, 40]. Since they are usually able to accumulate in the cytoplasm at high concentration but without negatively influencing cell's metabolism or interfering with cellular functions, they are called *compatible solutes* or co-solutes [32, 33, 38, 39]. To date, manifold compatible solutes have been proved to stabilize [41, 42, 43, 44, 45, 46] or to destabilize [47, 48, 46, 49] proteins, both in *in vivo* and *in vitro* settings. As an example, trimethylamine-N-oxide (TMAO) and ectoine are potent stabilizers, whereas urea is known as highly effective structure destabilizer. It is worth mentioning that all of these co-solutes coexist with nucleic acid structures in their native environment, having an influence not only on the structural stability of biomolecules, but also on their function. For instance, urea has been proved to assist nucleic acids in numerous processes in cells [50, 23], whereas TMAO has been adopted by certain deep sea microorganisms to accommodate to high pressures [37, 38, 51, 36] reaching up to 1000 bar in their living environment.

Over the last decades, there were proposed distinct approaches to elucidate the mechanisms of co-solutes interactions with macromolecules and their resulting structural stabilization or destabilization. Those approaches often rely on water-mediated or direct interactions with respect to the chaotropic or kosmotropic properties. It is well-known that certain co-solutes are highly hygroscopic in terms of their water-binding behavior, which affects significantly the structure of the solvent [52, 53, 54, 55, 56, 57]. In terms of the distinction between *kosmotropes*, defined as *water-structure makers*, and *chaotropes*, specified as *water-structure breakers* [58, 59, 60], it is often assumed that kosmotropes tend to stabilize the macromolecules like proteins, DNA or RNA, whereas chaotropes

promote their unfolding [47, 48, 46]. The possible explanation of the underlaying mechanisms have been proposed with reference to the Molecular Theory of Solutions and Local/Bulk Partition Model [61, 43, 62, 63, 64, 65, 66, 67, 68, 69]. In this aspect, the destabilizing influence of *chaotropes* such as urea can be attributed to the direct preferential interaction with biomolecule, resulting in the remarkable replacement of hydrating water molecules around the macromolecule. This creates the deficit in the number of hydrating water molecules, which can be considered as a main explanation for the denaturation behavior [3, 1, 2, 43, 69, 62]. The hydrophobic mechanism has been also suggested to account for the urea-induced alterations of water structure [49]. In contrast, the preferential exclusion of kosmotropes, such as ectoine, from direct surrounding of the macromolecules results in the stabilizing preferential hydration behavior [70, 43, 71, 45]. In more detail, the co-solute molecules repelled from the close vicinity of biomolecular surface to the bulk are being successively replaced by excess water molecules, which eventually stabilize the native form [3, 4, 1, 41, 72, 73, 63, 74, 69]. Thus, water molecules form a protective layer surrounding the macromolecule.

Numerous experimental and computational findings published to date show that stabilizing and destabilizing mechanisms induced by certain co-solutes often mutually compensate each other [75, 56, 36]. It was also observed that most stabilizing or destabilizing effects are reversible [76, 43, 77, 46] and disappear in the absence of co-solutes [78, 46]. The explanation for this phenomena has been often relying on water-mediated mechanisms [79], involving high hygroscopicity of various co-solutes and their consequent potential to affect the local water structure [80, 58, 59, 55, 56, 52, 54]. The contrasting explanation for these mechanisms come from the preferential exclusion and binding mechanisms in combination with local/bulk partitioning approaches, or their combination with hydration and dehydration effects [63, 64]. In addition to their complex interactions with proteins or DNA, aqueous mixtures of stabilizers and destabilizers exhibit plenty of interesting effects even in the absence of macromolecules [81, 56, 82, 36, 83]. All respective findings pinpoint the importance of the detailed study of the individual co-solute contributions and the collective mechanisms for more comprehensive understanding of their influence on protein or DNA destabilizing and stabilizing mechanisms, also with regard to the impact on water structure and biomolecular surrounding. In this aspect, the Kirkwood-Buff (KB) theory [84, 85, 86] presents a suitable framework for the elaborate study on the co-solutes interaction with macromolecules of biological importance. It relies explicitly on molecular distribution functions called *Kirkwood–Buff integrals*, and hence can be regarded as a molecular theory for multicomponent solutions. Numerous quantities in KB theory can be perceived as the derivatives of the chemical potential, which is of crucial importance for the definition of the chemical equilibrium between product and reactant state in chemical reactions [87, 88, 69, 89, 90]. The benefits of KB approach relies also on its application for the non-ideal solutions and the absence of any prior assumptions concerning molecular size or shape. Since the KB theory does not rely on the molecular details of the interacting species, it can be applied for various molecular systems [87, 62, 91, 69, 92].

This work presents the study on the interaction of selected biologically relevant co-solutes with diverse DNA structures. The particular impact is put on the non-canonical DNA forms, such as short DNA hairpins or DNA G-quadruplex. Although the mechanisms of the influence of multiple stabilizing and destabilizing co-solutes on protein structures have been widely researched and explained to date, the details of their impact on nucleic acid structures remain to vast extent vague. With the application of thermodynamic approach in terms of Molecular Theory of Solutions, in combination with free energy calculation and molecular modelling, one can get a deeper insight into the details of molecular interaction between selected co-solutes of utmost biological importance with DNA structures. This study sheds the new light on the molecular details of the phenomena in biological systems and can serve as the first approach for the application of the underlying mechanisms in modern biophysics and biochemistry.

2 Theoretical background

2.1 Selected biologically relevant nucleic acids structures

As carriers of genetic information, nucleic acids - RNA and DNA - are considered as one of the most important biomolecules in the cells of the living organisms. One can differentiate between ribonucleic acid (RNA), containing ribose in the sugar phosphate backbone, and deoxyribonucleic acid (DNA), having in its backbone 2-deoxyribose instead. A classical DNA molecule (so-called Watson-Crick structure [9] or B-DNA) is composed of two polynucleotide chains coiled around each other, whereas RNA occurs typically as single-stranded molecule (ssRNA). Nucleic acid chains carry the genetic information about the synthesis of proteins, which are relevant for the functioning, reproduction, growth and development of all know living organisms as well as certain viruses.

Beside canonical DNA double helix [9, 93] (Fig. 1a), being the most common structure existing in nature, there exist numerous other inter- and intramolecular secondary structures [94, 19] such as DNA i-Motifs, formed by the sequences rich in cytosine [13, 14, 10, 11, 12], DNA hairpins and G-quadruplexes [17, 13, 16], originating from guanine-rich sequences [15, 16, 17, 18, 19, 20]. Those structures often emerge as a result of formation of non-canonical hydrogen bonds between two purine or two pirimidine nucleobases. The formation of specific DNA secondary and tertiary structures has been frequently involved in fundamental processes in cells, including replication, transcription termination, regulation of gene expression or intermolecular binding [21, 22, 23]. Furthermore, a reversible mechanism of non-Watson-Crick DNA folding can be applied in modern nanobiotechnology. Possible applications of DNA higher-order structures on that field are nowadays the object of intensive research [26, 27].

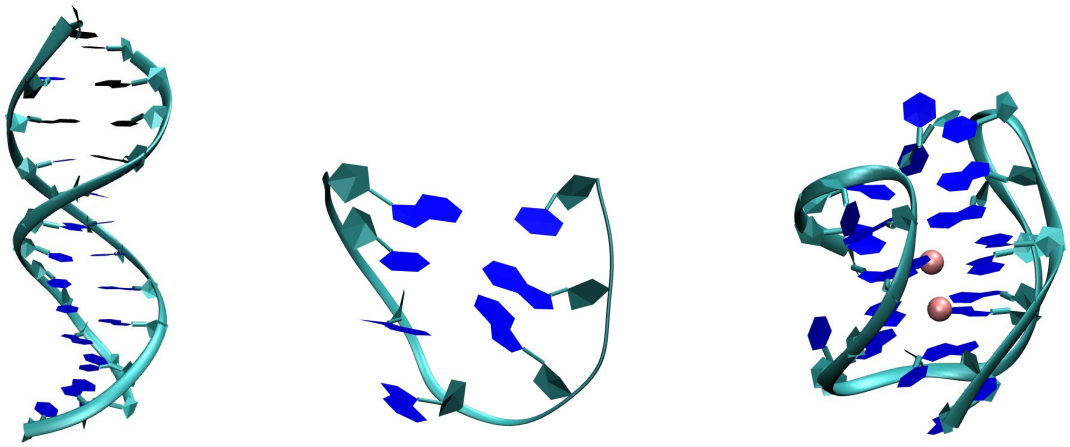
Due to their involvement in fundamental cellular processes and pathways [95, 96, 97, 98, 99, 100, 101, 19] as well as their regulatory role in certain human genetic diseases [97], particularly DNA G-quadruplexes reached on increased scientific interest in the recent years. G-quadruplexes (Fig. 1c) are non-canonical DNA structures formed by two or more guanine tetrads (G-tetrads) stacked on top of each other [99, 20, 100, 102, 103, 1, 2], whose O6 carbonyls are complexed with monovalent cations like K^+ or Na^+ necessary for quadruplex stability [100, 102, 103, 20, 101]. Individual G-tetrads are assembled via cyclic arrangement of four guanines paired through Hoogsten hydrogen bonding [20], with total of eight hydrogen bonds and four grooves characterizing each G-quartet [100, 102, 103]. G-quadruplexes can adopt variety of structures, depending on their sequence and the environmental factors [99, 100, 104, 105, 101, 19, 20, 103]. They are characterized by extraordinary thermal stability [19, 20] related to their structure. G-quadruplexes are located in functionally important regions of the genome [20] such as telomeric parts of chromosomes [106, 24, 96, 97, 99, 104, 101, 20, 19, 17, 16] or non-telomeric promoter regions[16, 19], which translates directly into their role in numerous biological processes *in vivo*. G-quadruplexes are involved in such pivotal cellular processes like gene expression, transcription, translation, DNA replication and maintenance of telomeres [24, 19, 95, 20]. In particular, formation of intramolecular G-quadruplexes in telomeric regions inhibits telomere elongation by sending a 'stop' signal to responsible polymerase [19, 107], turning those structures into a potential target for anti-cancer drug design [15, 25, 95].

In vivo nucleic acid hairpin structures (Fig. 1b) occur as a result of the base pairs formation within a single-stranded DNA or RNA molecule [108]. The typical features of a DNA hairpin involve the presence of a base-paired helix, terminated with a short unpaired loop [3, 4]. In particular short DNA minihairpins are frequently found in biologically important regions [109], such as the promoter regions of certain genes [110, 111] or replication origins of viruses [112]. The studies on the unfolding mechanism of DNA i-Motifs show that the unfolding pathway in 300 K in the absence of protonated cytosines goes through a stable hairpin conformation, which is restrained from further unfolding by a high energy barrier [113, 114, 115]. In this way, hairpin structures prove to be one of the most stable equilibrium intermediates on the formation pathways of DNA higher-order structures [113].

Small DNA hairpins with the sequence d(GCGNAGC), where N corresponds to any of the nucleobases adenine (A), guanine (G), cytosine (C) or thymine (T), have been recently applied as model systems in computational and experimental studies [117, 3] due to their extraordinary

stability in terms of high melting temperatures and resistance against nucleases [116, 111]. With regard to the structural properties of hairpins, the sequence d(GCGAAGC) appears to form one of the most thermodynamically stable structures [109, 117] due to the complex interplay between intramolecular hydrogen bonding and aromatic base pair stacking [118].

DNA hairpins are particularly suited to study the interaction with co-solutes in more detail [119]. Due to their well-defined native structure and short sequence length, those structures are ideally adapted for simulation approaches [117] in order to study the underlying thermodynamic properties. The importance of DNA hairpins and other higher order DNA structures relies on their outstanding relation between structure and function [109, 21, 13, 14, 16, 22, 17, 24, 25, 26, 27, 110, 111, 112]. Those structures are also characterized by a strong tendency to unfold under slight changes of the environmental conditions, with important consequences for biological functionality [113, 114, 120].



(a) B-DNA helix (PDB: 1BNA)

(b) DNA hairpin (PDB: 1KR8)

(c) DNA G-quadruplex (PDB: 2KF8)

Figure 1: Examples of DNA structures (from the left): canonical Watson-Crick helix [121], DNA hairpin [122] and G-quadruplex composed of 2 tetrads, complexed with two stabilizing K^+ ions in the central channel [123].

2.2 Selected biologically relevant co-solutes

Although often idealized in most computational and *in vitro* studies, the physiological environment in cells can be regarded as dense aqueous mixtures consisting of ions and macromolecular compounds. Beside lipids, amino acids or nucleotides, further important components are low weight organic co-solutes, also called co-solvents, depending on the concentration, whose presence can influence the molecular structure of DNA and proteins considerably [46]. Along with the stabilization by crowding mechanisms [124], co-solutes or ions can exert both stabilizing or destabilizing effects on macromolecular conformations [125]. Over the last years, there were numerous studies on protein-related co-solute effects [49, 46] and consequently on the properties of typical co-solutes like trimethylamine N-oxide (TMAO), guanidinium, ectoine, urea or complex ions, and their impact on peptides, proteins and homopolymers. Those effects were studied in numerous experiments and simulations [47, 126, 49, 46, 127, 128, 55, 129, 57, 56, 130, 131, 132, 69, 133, 134], with particular emphasis on the stabilization and destabilization mechanisms, closely related to the accumulation behavior of the co-solute species [135, 136, 57, 137, 47, 126, 125, 49, 138, 139, 127, 46, 140, 141, 129, 142, 143, 134].

To date, a vast range of naturally occurring compatible solutes have been shown to stabilize proteins, both in *in vitro* and *in vivo* settings [41, 42, 43, 44, 63, 74, 67, 144, 45, 46]. Although the molecular details of the underlying mechanism are still not completely understood [145, 69], numerous studies attribute the resulting protein stabilization mechanism to a preferential exclusion of co-solutes from the vicinity of the macromolecule [70, 43, 71, 45], such that the co-solute

molecules repelled from the direct vicinity of the protein surface are successively replaced by excess water molecules. This leads consequently to the stabilization of the native form in terms of a preferential hydration mechanism [41, 63, 74, 72, 73, 69, 5]. On the other hand, preferential binding of the co-solute combined with protein dehydration results in the destabilization of the structure [43, 62, 69, 5]. Basing on the refined molecular theories of solution [87, 69, 62, 146, 63, 74, 5] it can be shown that the interaction of co-solutes with water molecules exerts an influence on the resulting binding behavior with macromolecules [147, 67, 144, 114, 3, 69, 72, 73]. In this regard, the influence of water molecules on certain osmolytes can be of pronounced importance [147, 67, 144, 115, 114, 148, 55, 56].

The Fig. 2 presents selected important co-solutes for biomolecular research.

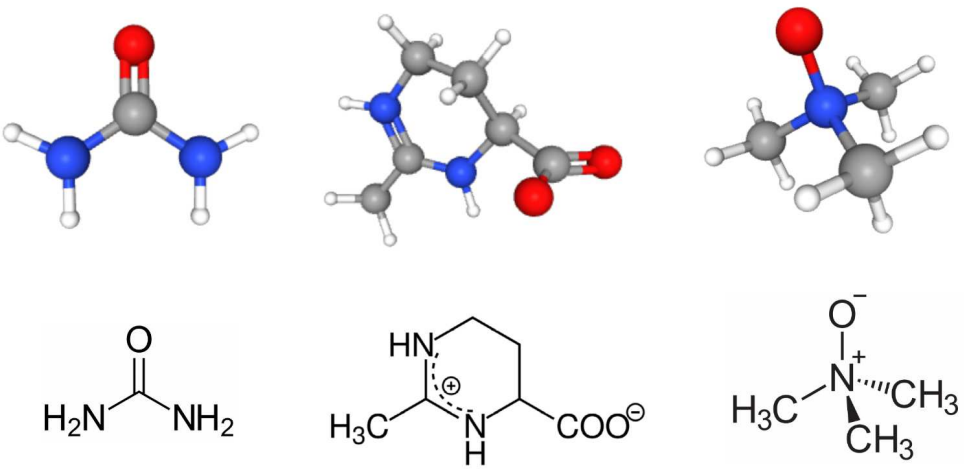


Figure 2: Molecular (top) and chemical (bottom) structures of urea (left side), ectoine (middle) and TMAO (right side).

One of the most broadly investigated and vastly applied in modern biotechnology co-solutes is urea (carbonyldiamide), shown in Fig. 2 on the left. It is an organic compound, known for its destabilizing impact on protein structure [149, 150, 151, 152, 153, 46, 38, 154, 155, 156, 157, 158]. The intracellular presence of urea is of particular importance due the fact that it assists the functionality of nucleic acids in numerous cellular processes [50, 23]. Several studies report a direct and unspecific binding of urea to certain macromolecules prompted by dispersion interactions, which induces the replacement of water molecules in the first hydration shell, and as a consequence a denaturation effect [127, 129, 3, 6]. Further research results suggest also a specific binding of urea to aromatic groups, which was also reported for urea-nucleobase interactions [142, 159]. Hitherto experimental results reveal a destabilizing effect of urea on canonical DNA conformations [160, 161, 162]. However, certain studies show that particular DNA structures may remain intact even in very highly concentrated urea solutions [161]. Certain experimental studies published to date [162] highlight a preferential binding of urea to model compounds such as nucleic acid base analogues, nucleosides and nucleotide monophosphates [162], but also a destabilizing effect of urea on higher order structures like DNA hairpins [119]. The complementary thermodynamic analysis pointed to a combined interaction in terms of enthalpic and entropic contributions, which allows to speculate that urea shows a stronger preferential binding to the unfolded DNA structure.

Beside destabilizing agents, trimethylamine-N-oxide (TMAO) (Fig. 2, right side) is one of the most important stabilizing co-solutes present in the cytosol of various cells, which recently attracted emerging attention [37, 36, 163, 164, 137, 57]. The importance of TMAO in modern biotechnology and biochemistry lays in its unique combined hydrophilic/hydrophobic properties [145], its influence on the structural and dynamical properties of water [165, 166, 167, 168, 169, 82], its combined interactions with urea [81, 170, 171, 82, 83, 36, 172] and its protecting role against pressure-induced destabilization mechanisms [173, 174, 164, 175]. The latter property concerns not only the protection of biomolecules [176, 177, 178], but also the thermo-responsive polymers

like poly(N-isopropylacrylamide) (PNIPAM) [179, 180, 135, 136, 57, 181, 182, 183]. Contrary to urea, the binding behavior observed for TMAO [177, 136, 57] turned out to differ significantly, also from other stabilizing agents like ectoine [140]. Although TMAO was found to bind preferentially to proteins [136, 177], contrary to urea it was not supposed to induce a dehydration effect. Thus, the corresponding new mechanism was named 'preferential attraction' [57]. In the last years, the combined urea-TMAO effects on proteins have reached considerable attention [155, 166, 149, 184, 185, 75, 156, 186, 182]. Numerous studies report that TMAO has the ability to counteracts the denaturing effect of urea *in vivo* [187, 153, 46, 154, 155, 156, 157, 158, 36], although the molecular details of this compensation mechanism are still not entirely understood. Some works suggest either a direct interaction as well as water-mediated effects as potential reasons for the occurrence of the compensation behavior [188, 170]. Notwithstanding, recent observations of a complete compensation mechanism for proteins in TMAO/urea solutions with particular molar ratios [155, 156] raise the question whether a similar effect can be observed also for higher-order structures of polynucleic acids.

Another important stabilizing agent, which attracted particular interest with regard to its broad industrial use in pharmaceutical products and cosmetics [189, 190, 191], is ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid) (Fig. 2, middle). It has been found to be produced *in vivo* in the cells of numerous extremophilic microorganisms such as halophilic and halotolerant bacteria [32, 33, 191]. Due to its hygroscopicity [72, 115, 148, 55, 56] even in the presence of high salt concentrations [115, 148, 56], ectoine was found to facilitate the survival of halophilic bacteria in extreme saline environments. The strong water-binding behavior of ectoine is primarily attributed to the zwitterionic properties of the molecule in aqueous environments [115, 148, 55, 56], which is represented by a negatively charged carboxygroup and a positively charged and protonated nitrogen atom as shown in Fig. 2. The protein stabilizing effect of ectoine via preferential exclusion from protein surfaces has been often related to enthalpy-entropy compensation mechanisms, where the favorable enthalpy change is induced by an increased ordering of water molecules via the formation of water-ectoine hydrogen bonds [115, 192]. Additionally, a favorable entropy change [192] occurs as a result of the weakening of water-water hydrogen bonds in the vicinity of the solute. With relation to hitherto studies, it was assumed that ectoine stabilizes proteins [72, 193, 192] both by direct and indirect mechanisms [73], involving the modifications of the hydration shell [73, 55] and the repulsion from uncharged surfaces [148]. Since ectoine holds charged groups, the study on the behavior and interactions between zwitterionic ectoine and highly charged biomolecules, such as DNA, RNA or certain proteins, is of particular importance. Although commonly considered as a protectant against UV or ionizing irradiation [194, 195, 196], in experimental studies it was found that ectoine actually enhances strand breaks in oligonucleotides [197]. These results indicate that preferential exclusion of ectoine from positive or neutral groups, which was suggested to be responsible for its stabilizing properties [73, 69, 87], may in fact not entirely apply to negatively charged moieties. Consequently, it was also reported that hydroxyectoine, a closely related ectoine-derivative, preferentially binds to negatively charged spheres [114]. This mechanism may be also valid for the negatively charged DNA backbone. Ectoine and its close analogues have been also observed to decrease significantly the melting temperature of long DNA structures, which suggests their potential use as destabilizing PCR enhancers [198].

Beside the complex interactions with proteins or nucleic acids, aqueous mixtures of stabilizing and destabilizing co-solutes were proved to show a variety of interesting effects even in the absence of macromolecules [81, 56, 82, 36, 83].

Despite the vast amount of literature on co-solute effects on proteins and the influence of salt ions on nucleobase conformations [199], the interaction of co-solutes with DNA was only sparsely investigated. [119, 160, 161, 162, 159, 142]. An important difference between nucleic acids structures and the majority of proteins is the pronounced negative charge of DNA backbone, which can influence significantly the electrostatics of the biomolecular solution and hence also the interaction properties between the solution components.

2.3 Physiological mechanisms of nucleic acids stabilization

Although non-canonical DNA structures are often characterized by an enhanced stability, numerous external factors such as changes in temperature, pH values, hydration conditions or even the exchange of ions may lead to structural destabilization as observed for other macromolecules like proteins [37]. For that reason, specifically for microorganisms living under harsh environmental conditions the presence of certain biologically relevant co-solutes is of particular importance [38], as they can selectively stabilize or destabilize protein and oligonucleotide conformations without affecting their chemical structure [46]. Unlike most eukaryotic cells, where the genomic DNA is confined within the nuclear envelope forming cell's nucleus, prokaryotic organisms' DNA does not belong to a distinct cellular compartment. Being located freely in the cytoplasm, such DNA is exposed to interaction with other components of the cytosol, which adverts the necessity to establish the means of its preservation. In particular, unicellular organisms exposed on severe environmental stress need to develop certain mechanisms to protect their DNA from damages. For such organisms, a common survival strategy involves the production of co-solutes that maintain the macromolecular conformations *in vivo* against the detrimental effects of environmental factors such as osmotic stress, high temperature or salinity [36, 37], as well as against the destabilizing impact of other molecules [46, 41, 43, 63, 74, 45, 42, 44, 67, 144].

It has been found that the cytoplasm of many extremophilic organisms can be compared to a highly concentrated aqueous solutions of stabilizing and destabilizing co-solutes [200, 38, 67, 144]. Since they are typically accumulated in the cytoplasm at molar concentrations without exerting a negative influence on cell metabolism, some of them are also called *compatible solutes* [32, 33, 38, 39]. Numerous halophilic and halotolerant bacteria were proved to synthesize such compounds to withstand high saline environments by maintaining the osmotic equilibrium of the cell with the extracellular environment [32, 33, 34, 35]. Another remarkable example comes from the studies on the adaptation mechanisms of deep sea organisms, who acquire certain osmolytes, such as trimethylamin-N-oxide (TMAO), to adapt to high pressures [37, 38, 51, 36] reaching up to 1000 bar in their living environment. In the organisms of marine animals, TMAO is generated by microbial metabolism and accumulates in their tissues to protect their cellular structures against thermal, osmotic and biomolecular destabilization [152, 151, 36]. Despite being known as a common protein denaturant [3, 6], numerous deep sea organisms accumulate also urea as an osmolyte and buoyancy factor. In those organisms, TMAO helps to restore proteins native structure and functionality that has been lost due to the presence of urea [149, 150, 151, 152, 153, 46, 38, 154, 155, 156, 157, 158], along with counteracting the osmotic pressure coming from urea itself [36]. Thus, adapting high concentrations of both TMAO and urea allows the marine piezophilic organisms to preserve the level of their osmotic pressure comparable to that of the seawater [37]. In the mixture of both osmolytes, protein secondary structure is protected through a subtle interplay between the impact of different concentrations of TMAO and urea in the cytosol [201, 6]. It has been found that the concentration of TMAO in the cells of piezophilic microorganisms oscillate around 0.6 M, and the concentration of 1 M can already provide a complete compensation of the destabilizing effect on proteins. At the same time, the proportions of 1:2 of TMAO and urea result in entire compensation of urea-caused protein denaturation [155, 156]. However, although the protecting influence of TMAO on protein structure is indisputable [163, 164, 137], the molecular details of the stabilization mechanisms are still not completely understood [145, 69, 152, 164]. It is assumed that certain co-solutes are preferentially excluded from biomolecular surface [1, 2, 43, 45, 70, 71, 164], which leads to an enhanced hydration [1, 2, 41, 63, 74, 69, 72, 73, 5, 164]. As far as TMAO is concerned, its amphiphilic nature facilitates the formation of hydrogen bonds with water and hence preferential exculsion from certain protein functional groups [164]. Consequently, preferential binding of the co-solute accompanied by elevated dehydration could be responsible for protein structure destabilization [43, 69, 62, 5]. On the other hand, DNA structures usually have a pronounced negative charge in contrast to typically uncharged proteins. Since this can influence the interaction of DNA with other components of the solution, it raises the question whether also DNA higher order structures can be stabilized by certain amounts of TMAO.

2.4 Molecular Theory of Solutions

The understanding of the fundamental properties of solutions, such as the biomolecular equilibria, effects of co-solutes or local composition in terms of preferential solvation has been based on the Fluctuation Theory (FT) of solutions [202, 203, 204, 85, 66]. The FT Theory associates the thermodynamic properties of the solution with the local fluctuations under relevant thermodynamic constraints. The general concept of the FT theory links the fluctuation characteristics of an open system in grand canonical ensemble, where the chemical potential μ, volume V and temperature T of the system are constant, to the properties of the system defined under isothermal-isobaric Gibbs ensemble (constant temperature T, pressure p and the number of particles N) [204]. Hence, it helps the understanding of solution behavior in closed systems, which are are among the most common experimental setup of the system. In other words, fluctuations in an open system can be transformed via FT to describe the quantities in an equivalent closed system. The relation between the open and closed systems is shown schematically in Fig. 3. Here, an open system can be approximated by the local region of defined volume within the bulk solution, which represents a closed system. The temperature of both systems is constant. However, the local domain enables the exchange of particles in order to maintain the same chemical potentials as defined in the bulk solution.

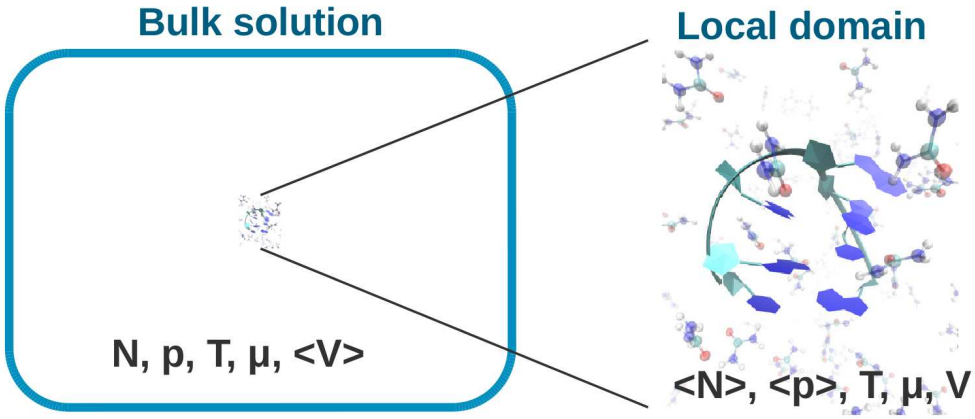


Figure 3: The schematic representation of the bulk versus local properties of the solution at constant average temperature T. The bulk region is characterized by constant number of particles N, and the chemical potentials are derived as the ensemble averages. In order to preserve a set of constant chemical potentials as determined by the bulk, the molecules are allowed to leave and enter a small local region. Thus, the local region represents an open system, whereas the bulk solution approximates a closed system.

The transformation between the expressions for the fluctuations in grand canonical ensemble and the relevant expressions for the thermodynamic properties of a closed system is provided by the statistical thermodynamics [204]. It defines the characteristic functions of the system in terms of the corresponding partition functions, which encode the partitioning of the probabilities among various microstates of the system. The partition functions involve summing over the available microstates, as expressed in the Eqn. 10 in Ref. [204].

The above mentioned expressions involve typically the fluctuations of energy and particle number. Equivalently, to replace the number fluctuations one can apply the distribution functions, where the distributions are related to an open system [204]. In that case, the results of the integration over the system volume are not as trivial as for the closed systems. Let us consider the probability $\rho^{(n)}(\{r\}\{dr\})$ of N_1 molecules of species 1 and N_2 molecules of species 2 etc. being located within the distance $d\{r\}$ at $\{r\}$ in the grand canonical ensemble [204]:

$$\int \rho^{(n)}(\{r\}\{dr\}) = \langle \prod_i \frac{N_i!}{(N_i - n_i)!} \rangle \tag{1}$$

where Π_i is the product over various species i in the mixture and n_i is the number of thermodynamically independent molecules of the species i in the system. For binary systems containing

the molecular species i, this yields [204]

$$\int \rho_i^{(1)}(r_1)dr_1 = \langle N_i \rangle \tag{2}$$

Consequently, for ternary and higher order systems, the following integrals can be evaluated over all the available particle positions [204]:

$$\iint \rho_{ij}^{(2)}(r_1, r_2)dr_1 dr_2 = \langle N_i N_j \rangle - \delta_{ij}\langle N_i \rangle \tag{3}$$

$$\iiint \rho_{ijk}^{(3)}(r_1, r_2, r_3)dr_1 dr_2 dr_3 = \langle N_i N_j N_k \rangle - \delta_{ik}\langle N_i N_j \rangle - \delta_{ij}\langle N_i N_k \rangle - \delta_{jk}\langle N_j N_i \rangle + 2\delta_{ij}\delta_{jk}\langle N_i \rangle \tag{4}$$

where the Kronecker delta δ_{ij} is equal 1 for i = j, and 0 otherwise. The equations above link the continuous functions expressed as integrals with the fluctuations within the system. The expressions for those fluctuations by corresponding distribution functions $g^{(n)}$ for species i, j, k are as follows [204]:

$$g_i^{(1)}(r_1) = \frac{\rho_i^{(1)}(r_1)}{\rho_i} = 1 \tag{5}$$

$$g_{ij}^{(2)}(r_1, r_2) = \frac{\rho_{ij}^{(2)}(r_1, r_2)}{\rho_i \rho_j} \tag{6}$$

$$g_{ijk}^{(3)}(r_1, r_2, r_3) = \frac{\rho_{ijk}^{(3)}(r_1, r_2, r_3)}{\rho_i \rho_j \rho_k} \tag{7}$$

Hence, the integrals over the position of the central particle, with reference to the inter-particle distance $r_{12} = |r_2 - r_1|$, can be related to the distribution functions, which are henceforward referred to as the radial distribution functions (RDFs). With the above formulations, the integrals, further referred to as the Kirkwood-Buff integrals (KBIs), are related to the fluctuations of the particle number over RDFs in the grand canonical ensemble [204].

The KBIs are the essential components of the Kirkwood-Buff theory, which has been developed by J. G. Kirkwood and F. P. Buff in 1951 and further developed in the subsequent years [84, 202, 87, 85]. The theory, enabling the determination of the microscopic characteristics of the system from its macroscopic thermodynamic parameters, evolved from the Fluctuation Theory, linking the experimentally available system properties in Gibbs ensemble with the corresponding properties expressed in grand canonical ensemble. The Kirkwood-Buff theory bases on the concept of the specific interaction between the solute and the co-solvent being a consequence of preferential solvation or preferential binding [202, 203, 85]. The original Kirkwood-Buff (KB) theory is confined to the application of fluctuations of the number of particles within the molecular system, and it relates the fluctuations and integrals over particle distribution functions to the properties of the closed systems [204] and to the changes in the osmotic pressure in semi-open systems [84, 204]. These involve specifically the isothermal compressibility, partial molar volumes or chemical potentials [204]. Further works extend the traditional KB formulation to particle-energy and energy-energy fluctuations [205]. Traditionally, the KB theory has been formulated under the isothermal conditions [84, 204].

The KB inversion procedure enables to obtain the expressions for the particle number fluctuations in terms of the isothermal-isobaric (Gibbs) data, which are available in experiment [204]. The most straightforward approach refers to the application of the pseudo chemical potential [202, 85] and partial molar volumes [204]. Thus, the inversion procedure enables to obtain the Kirkwood-Buff integrals (KBIs) from the experimental data in order to provide the accurate description of the preferential solvation characteristics [202, 204].

The versatility of the KB theory makes it applicable to any kind of molecules, regardless of their shape or size. Since its introduction in 1951, the original framework has been significantly extended and modified by numerous contributions [85, 206, 88, 207, 146, 208, 66, 69], and made

applicable for computer simulations [209, 87]. As the input data, the KB theory requires radial distribution functions (RDFs) [204], expressed by the Eqn. 5, Eqn. 6, Eqn. 7 and following. Let us define the radial distribution function $g_{\alpha\beta}(r)$ of the co-solute β around the solute α over the distance r with the following formula:

$$g_{\alpha\beta}(r) = \frac{\rho_\beta}{\rho_{\beta,\infty}} \tag{8}$$

The RDFs refer to the distributions of the components of the solution upon averaging over the remaining molecular degrees of freedom [204]. Hence, the cumulative particle number radial distribution function (CN-RDF) confined within a limited distance R can be described as:

$$n_{\alpha\beta}(r) = 4\pi\rho_\beta \int_0^R r'^2 g_{\alpha\beta}(r')\,\mathrm{d}r' \tag{9}$$

where ρ_β is a number density of the co-solute β.

The Kirkwood-Buff integral (KBI) [85, 203] is defined as follows:

$$G_{\alpha\beta} = 4\pi \int_0^\infty r^2[g_{\alpha\beta}(r) - 1]\,\mathrm{d}r \tag{10}$$

with α being the solute (DNA molecule in this case), β the co-solute and γ the solvent.

The approximation for finite distances, with a cut-off distance r_c fulfilling the condition $g_{\alpha\beta}(r) \approx 1$ for $r \geq r_c$ can be expressed by the formula

$$G_{\alpha\beta} \approx G_{\alpha\beta}(r_c) = 4\pi \int_0^{r_c} r^2[g_{\alpha\beta}(r) - 1]\,\mathrm{d}r \tag{11}$$

The KBI can also be expressed accordingly in the NpT or NVT ensemble [85, 203], and hence evaluated straightforwardly for well-converged RDFs.

KBIs enable to assess the affinity of two different components of a solution, with the positive value of KB integral indicating the preferential binding of the central solute molecule by the co-solute [202], and the negative value - preferential exclusion. In other words, they give the information of the excess or deficit of solvent and co-solute molecules around the solute [210, 85, 202, 203]. Because of quantifying the deviations from the random distribution in the close environment of the central molecule, the KBIs are more informative than the pure particle fluctuation definition [204].

For a binary solution with two components, where the main component has index $i, j = 1$ and co-solute or co-solvent has index $i, j = 3$, there is a strict connection between the KB integrals and the osmotic pressure π ccording to the formula

$$\frac{\pi}{k_B T} = \rho_3 + B_2\rho_3^2 + B_3\rho_3^3 \ldots \tag{12}$$

where ρ_j stands for the total number density of species j, B_2 defines the osmotic second virial coefficient and B_3 its higher terms [211], and k_B and T denote the Boltzmann constant and the absolute temperature, respectively.

In conjunction with KB integrals, B_2 is defined as [211]

$$B_2 = -2\pi \int_0^\infty r^2\,[g_{33}(r) - 1]\,\mathrm{d}r \tag{13}$$

which hence can be written as

$$B_2 = -\frac{1}{2}G_{33} \tag{14}$$

under the constraint of negligible co-solute density.

Furthermore, one can define the partial molar volumes of the species as [85]

$$V_1 = \frac{1 + \rho_2(G_{33} - G_{13})}{\rho_1 + \rho_3 + \rho_1\rho_3(G_{11}G_{33} - 2G_{13})} \tag{15}$$

and

$$V_3 = \frac{1 + \rho_1(G_{11} - G_{13})}{\rho_1 + \rho_3 + \rho_1\rho_2(G_{11}G_{33} - 2G_{13})} \tag{16}$$

The above relation points out the close relation of KB theory with fundamental thermodynamic quantities.

To investigate qualitatively the local distribution of the co-solute around the DNA oligonucleotide, the concept of preferential interactions in the light of Kirkwood-Buff theory [210, 146, 87, 69] has been employed. According to a convention for ternary systems, the solvent (water) is usually indexed by '1', the solute (different DNA structures) by '2' and the co-solute by '3'. Consequently, one can define the preferential binding coefficient ν_{23} [62, 146, 87], which can be described by the formula:

$$\nu_{23} = -\left(\frac{\partial\mu_2}{\partial\mu_3}\right)_{p,T,\rho_2\to 0} = \rho_3(G_{23} - G_{21}) \tag{17}$$

with the chemical potentials μ_2 for solutes like DNA and μ_3 for co-solute at constant pressure p, constant temperature T and vanishing DNA number densities $\rho_2 \to 0$. Osmolyte is preferentially bound if $\nu_{23} > 0$ and preferentially excluded if $\nu_{23} < 0$.

Preferential binding coefficient as described by the Eqn. 17 is defined for the infinite DNA dilution, where the DNA-DNA interactions can be neglected. In computer simulation, this can be approximated by placing a single DNA molecule in the solvation box.

In computer simulations with fixed number of water and co-solvent molecules, one has to assume that the molecules in the vicinity of the DNA surface are no longer a part of the bulk solution, but rather contribute to the local distribution. Hence, the preferential binding coefficient should be refined to account for the changes in the local composition as well as in the bulk solution distribution (see: Eqn. 56 in [87]). Under this assumption, the refined preferential binding coefficient takes the form:

$$\nu_{23}(r) = n_{23}(r) - \frac{n_3 - n_{23}(r)}{n_1 - n_{21}(r)} n_{21}(r) \tag{18}$$

with n_3 and n_1 defining the finite number of co-solute and water molecules in the simulation box, in conjunction with the corresponding values for the cumulative number of particles according to Eqn. 9.

An useful expression to study the stability of DNA conformations [87, 69] is the difference between the preferential binding coefficients. In accordance with Eqn. 17, one can define

$$\Delta\nu_{23} = -\left(\frac{\partial\Delta\mu_2}{\partial\mu_3}\right)_{p,T,\rho_2\to 0} \tag{19}$$

which leads to

$$\Delta\nu_{23}(r) = \nu_{23}^{\mathrm{D}}(r) - \nu_{23}^{\mathrm{N}}(r) = \Delta n_{23}(r) - \frac{\rho_3}{\rho_1}\Delta n_{21}(r) \tag{20}$$

where $\Delta n_{2\beta} = n_{2\beta}^{\mathrm{D}} - n_{2\beta}^{\mathrm{N}}$ and $\beta = 1, 3$ define the difference in the number of water or urea molecules around the denatured (as denoted by the superscript D) and the native (superscript N) DNA conformation.

The chemical equilibrium between both DNA structures in terms of chemical stability is related to the differences in the preferential binding coefficients according to

$$\Delta\mu_2 = \mu_2^{\mathrm{D}} - \mu_2^{\mathrm{N}} = -RT\ln K \tag{21}$$

where R stands for the molar gas constant, and the chemical equilibrium constant $K = [\mathrm{D}]_0/[\mathrm{N}]_0$ represents the ratio between the concentration of native (N) and denatured DNA structures (D) in infinite dilution [87, 69]. Upon inserting this equation into Eqn. 17, one receives

$$\Delta\nu_{23} = RT\left(\frac{\partial \ln K}{\partial \mu_3}\right)_{T,p,\rho_2\to 0} \tag{22}$$

which indicates that positive values of $\Delta\nu_{23}$ induce a shift of the chemical equilibrium to the denatured state, whereas a negative value imposes a stabilization of the native state [69]. According to those definitions, the stabilizing co-solutes typically involve the values of $\Delta\nu_{23} < 0$, whereas the destabilizing ones reveal positive values of $\Delta\nu_{23} > 0$. However, it has to be noted that one assumes a simple two-step DNA unfolding process with two well-defined local and global free energy minima in terms of the thermodynamic analysis presented above.

Assuming an ideal solution around the main solute and an arbitrarily chosen equilibrium configuration 'E' without any molecular interactions (denoted by the subscript '0'), the preferential binding coefficient for such system would be $\nu_{23}^0 = 0$ as a result of the absence of any conservative interactions. According to the formula

$$\Delta\nu_{23}^0 = \nu_{23} - \nu_{23}^0 \tag{23}$$

it follows that $\Delta\nu_{23}^0 < 0$ for $\nu_{23} < 0$ and $\Delta\nu_{23}^0 > 0$ for $\nu_{23} > 0$, respectively. Relating it to the hitherto studies on chemical equilibrium constants between native and unfolded conformations [4, 7], one can introduce the chemical equilibrium constant between the reference 'E' and any further arbitrarily chosen stable equilibrium conformation 'M' in an ideal solution according to

$$K^0 = [\mathrm{M}]^0/[\mathrm{E}]^0 \tag{24}$$

where $[\mathrm{E}]^0$ and $[\mathrm{M}]^0$ represent the corresponding concentrations of the main solute in the equilibrium reference and a further stable conformation. Knowing that

$$K_{\mathrm{cs}} \approx K^0 \exp(\Delta\nu_{23}) \tag{25}$$

one can combine this relation with Eqn. 23 to obtain

$$K_{\mathrm{cs}} \approx \frac{[\mathrm{M}]^0}{[\mathrm{E}]^0}\exp(\Delta\nu_{23}^0) \approx \frac{[\mathrm{M}]^{\mathrm{cs}}}{[\mathrm{E}]^0} \tag{26}$$

where M^{cs} describes the state M in presence of co-solutes.

In agreement with Eqn. 23, a positive value for ν_{23} results in $\Delta\nu_{23}^0 > 0$ and hence a preferred change with respect to $[\mathrm{M}^{\mathrm{cs}}] > [\mathrm{E}^0]$, whereas a negative preferential binding coefficient refers to a stabilization of the structure E^0 with respect to $[\mathrm{M}^{\mathrm{cs}}] < [\mathrm{E}^0]$. The state 'M' can be chosen arbitrarily if one is only interested in general stabilization variations in terms of the equilibrium conformation 'E^0'. Thus, it is sufficient for the general study of stabilization or destabilization influences of co-solutes to define fixed reference conformations for 'M' that do not vary over time. Consequently, co-solutes with $\Delta\nu_{23}^0 > 0$ can be considered as destabilizing agents, whereas values of $\Delta\nu_{23}^0 < 0$ refer to stabilizers.

A simplified equilibrium reaction for two-step folding/unfolding process reads then

$$N \rightleftharpoons D$$

with the chemical equilibrium constant [212]

$$K = \prod_j a_j^{\nu_j} \tag{27}$$

where a_j defines the average chemical activity of the DNA in the corresponding state combined with the stoichiometric cofficient ν_j. This coefficient is set to $\nu_D = 1$ for the denatured state, and $\nu_N = -1$ for the native state.

The chemical potential of an individual state x is defined as

$$\mu_x = \mu_x^0 + RT \ln a_x \tag{28}$$

where μ_x^0 stands for the standard chemical potential, R the universal gas constant and T the absolute temperature. With reference to the chemical equilibrium ($\Delta\mu = \sum_j \nu_j \mu_j = 0$) [212], it thus follows

$$\Delta\mu^0 = \sum_x \nu_x \mu_j^0 = -RT \ln K \tag{29}$$

The above expression is equivalent to

$$\Delta\mu^0 = -RT \lim_{\rho_2 \to 0} \ln \left(\frac{N_D^0}{N_N^0} \right) \tag{30}$$

with the number N_x^0 of molecules in the corresponding state. This relation is valid for an ideal and infinitely dilute DNA solutions ($a_x \approx \rho_x$), as denoted by the superscript '0'.

The estimate for the stabilization/destabilization tendency of co-solute species [87, 46, 69] is further expressed by the m-value, which can be evaluated by the difference in the preferential binding coefficients (Eqn. 22). The m-value approach bases on the assumption of a linear decrease or increase of the unfolding free energy for different co-solute concentrations.

In order to include co-solute induced effects, let us assume that a finite number density of co-solute species ρ_3 in the solution induces a modified change in the chemical potential of DNA structures

$$\Delta\mu_{\mathrm{cs}} = \Delta\mu^0 - mRT\rho_3 \tag{31}$$

with the $m-$value of the co-solute involved [147, 67, 144, 46]. Additionally,

$$\Delta\mu_{\mathrm{cs}} = -RT \lim_{\rho_2 \to 0} \ln K_{\mathrm{cs}} = -RT \lim_{\rho_2 \to 0} \left[\ln \left(\frac{N_D^0}{N_N^0} \right) + \ln \left(\frac{N_D^{\mathrm{cs}}}{N_N^{\mathrm{cs}}} \right) \right] \tag{32}$$

with N_x^{cs} defining the number of DNA structures in the corresponding microstate affected by co-solutes [69, 5]. Combining the Eqn. 32 with Eqn. 20 give an estimate of the stability of DNA forms according to

$$\Delta\nu_{23} = -\left(\frac{\partial \Delta\mu_{\mathrm{cs}}}{\partial \mu_3} \right)_{p,T,\rho_2 \to 0} = RT \left(\frac{\partial \ln K_{\mathrm{cs}}}{\partial \mu_3} \right)_{p,T,\rho_2 \to 0} \tag{33}$$

which emphasizes the fact that $\Delta\nu_{23} > 0$ involves a shift of the chemical equilibrium to the denatured state, whereas $\Delta\nu_{23} < 0$ indicates the stabilization of the native DNA state by co-solute.

Consequently, the actual unfolding free energy in presence of co-solutes can be expressed by the formula

$$\Delta G = \Delta G^0 - mRT\rho_3 \tag{34}$$

where ΔG^0 stands for the unfolding free energy without co-solutes and ρ_3 for the actual concentration of co-solute species. The m-value can be calculated [87, 69] as

$$m = \frac{a_{33} \Delta\nu_{23}}{\rho_3} \tag{35}$$

with the derivative of the chemical activity in binary solution (or $\rho_2 \to 0$) defined as

$$a_{33} = \left(\frac{\partial \ln a_3}{\partial \ln \rho_3} \right)_{T,p} = \frac{1}{1 + \rho_3 (G_{33} - G_{31})} \tag{36}$$

In the formula presented above, the corresponding KBIs for co-solute–co-solute and co-solute–solvent RDFs are defined in absence of the main solute [87, 69, 213]. For nearly ideal conditions at low co-solute concentrations, the chemical activity $a_{33} \approx 1$. Consequently,

$$(K_{\mathrm{cs}})_{T,p,\rho_3\to 0} \propto \exp(\Delta\nu_{23}) \tag{37}$$

Under the assumption of an ideal behavior with $a_{33} \approx 1$ for all co-solute concentrations, the m-value is often calculated in terms of molar concentrations [46, 162]. This yields the apparent m^{app}-value according to

$$m^{\mathrm{app}} = RT\frac{\Delta\nu_{23}}{\rho_3} \tag{38}$$

This expression is frequently applied to quantify co-solute effects [46].

For low densities of the co-solute, which simulates a nearly ideal behavior, the excess volumes in terms of KB integrals correspond to the partial molar volumes [85]

$$\Delta\nu_{23} = \rho_3(\Delta V_3 - \Delta V_1), \tag{39}$$

where V_j with $j = 1, 3$ define the partial molar volumes of co-solute and solvent species. The difference in the partial molar volume of the macromolecule upon unfolding can be described as [64]

$$\Delta V_2 = -V_1\rho_1\Delta G_{21} - V_3\rho_3\Delta G_{23} \tag{40}$$

which leads to

$$(\Delta V_2)_{T,p,\rho_3\to 0} \approx -V_1\rho_1\Delta G_{21} \tag{41}$$

upon the constraint of low co-solute concentrations. Assuming a small change of the partial molar volume upon denaturation in terms of $(\Delta V_2)_{T,p,\rho_3\to 0} \approx 0$ so that $\Delta G_{21} \approx 0$, for Eqn. 20 the relation is as follows

$$\Delta\nu_{23} \approx \rho_3\Delta G_{23} \tag{42}$$

Combining it with Eqn. 22 one obtains

$$RT\left(\frac{\partial \ln K}{\partial \ln a_3}\right)_{T,p,\rho_3\to 0} = \rho_3\Delta G_{23}, \tag{43}$$

which leads to

$$K_{\mathrm{cs}} = K\exp(\rho_3\Delta G_{23}), \tag{44}$$

Those relations show that any shift of the chemical equilibrium relies to high extent on the co-solute-macromolecular binding or exclusion behavior.

For low co-solute densities, other important parameters to study the shifts in chemical equilibrium are the transfer free energy $F_{23}^{\dagger}$ and the difference in the transfer free energy $\Delta F_{23}^{\dagger}$, defined by [214]

$$F_{23}^{\dagger} = -RT\nu_{23} \tag{45}$$

and thus

$$\Delta F_{23}^{\dagger} = -RT\Delta\nu_{23} \tag{46}$$

Those parameters are the estimates for the energy and the energy difference needed to transfer a co-solute molecule from infinity to close vicinity around a particular central biomolecule. Insertion of the difference in the transfer free energy into Eqn. 22 leads to the relation

$$-RT\left(\frac{\partial \ln K}{\partial \ln a_3}\right)_{T,p,\rho_3\to 0} = \Delta F_{23}^{\dagger}. \tag{47}$$

which can be applied to study any shift of the chemical equilibrium.

Hitherto studies [210] show that in the analysis of solute–co-solvent interactions the solvent effects can be disregarded only when the preferential interaction parameter ν_{23} is confined within the range $-1 < \nu_{23} < 1$, which in most situations is not the case. Especially for destabilizing co-solutes, it was shown that the contribution from hydration shells to the overall preferential interaction mechanism cannot be neglected [63]. To take the contributions of the solvent into account, the preferential hydration parameter [43, 62, 74, 63, 210] is defined as

$$\Gamma_{21} = -\frac{\rho_1}{\rho_3}\nu_{23} \tag{48}$$

This parameter reveals a close relationship with the preferential binding coefficient ν_{23}. In this aspect, a strong binding of co-solutes promotes a pronounced replacement of water molecules in terms of $\Gamma_{21} < 0$, whereas a preferential exclusion of co-solutes induces a preferential hydration according to $\Gamma_{21} > 0$.

To investigate the distribution of the solvent around a solute in terms of the local orientation of the hydration shell, one can study the angular distribution of solvent molecules around DNA, described by the solvent orientation parameters f_1 and f_2 [215], defined by

$$f_1 = \langle \cos\, \theta_1 \rangle \tag{49}$$

and

$$f_2 = \langle 3\cos^2\, \theta_2 - 1 \rangle \tag{50}$$

with the angle θ_1 being spanned between the vector from a water hydrogen atom to the mid-point between the oxygen and a further hydrogen atom of a different water molecule, and the angle θ_2 extended between the surface of the solute and the normal of the water molecule plane. The values of f_1 and f_2 give the insight into the local arrangement of water molecules around DNA.

2.5 Local/Bulk Partitioning Model

The analysis of the preferential accumulation or exclusion of solutes near the oligonucleotide surface as well as the interpretation of the impact those solutes exert on the DNA structure according to the Kirkwood-Buff theory can be complemented by the application of the local/bulk partitioning model of Record *et al.* [216]. The idea of the model lays in the separation between the region near the oligonucleotide surface, which is forming the local domain, and the bulk solution. In this aspect, solvent (water) is being treated as the partitioning medium between the local and the bulk regions around an inserted solute.

Local/bulk partition coefficient, quantifying the accumulation of the co-solute molecules around the DNA conformations, is defined by the formula [61, 216, 141, 143]:

$$K_p(r) = \frac{\langle n_{23}(r)\rangle / \langle n_{21}(r)\rangle}{n_3/n_1} \tag{51}$$

where $\langle n_{23}(r)\rangle / \langle n_{21}(r)\rangle$ is the ratio of the cumulative number RDFs for co-solute and solvent respectively, and the brackets $\langle \dots \rangle$ denote a mean number of molecules at the corresponding distance in terms of Eqn. 9.

With regard to different distances from the main solute, one has to assume the deviations in local solution concentrations in comparison to the bulk solution. The insertion of Eqn. 51 into Eqn. 20 after neglecting the brackets [143] gives

$$\Delta\nu_{23}(r) = \Delta n_{23}(r) - \Delta\left(\frac{n_{23}(r)}{K_p(r)}\right) \tag{52}$$

which points a denaturation effect for $\Delta K_p(r) = K_p^{\mathrm{D}}(r) - K_p^{\mathrm{N}}(r) > 0$ and $\Delta n_{23}(r) > 0$.

2.6 Hydrogen bonds and electrostatic properties of solution

The study of hydrogen bond properties for water and co-solute species gives the insight into the complex interplay between the components of the molecular solution. According to the Luzar-Chandler criteria for the definition of hydrogen bonds [217, 218], the maximum distance between hydrogen bond donor and acceptor pairs should be limited to 0.35 nm and a maximum angle to 30°. The stability of hydrogen bonds is estimated by the calculation of hydrogen bond forward lifetimes τ_{F} from the average autocorrelation function of the existence criterion in accordance with

$$C(t) = \frac{\langle s_i(t_0) s_i(t) \rangle}{\langle s_i^2(t_0) \rangle} \tag{53}$$

with $s_i(t) = (0, 1)$ for an intact ($s_i(t) = 1$) or a broken ($s_i(t) = 0$) hydrogen bond i at time t [217, 218]. The hydrogen bond forward lifetime τ_{F} is then expressed by

$$\tau_{\mathrm{F}} = \int_0^\infty C(t)\, dt \tag{54}$$

which allows to estimate the average time interval of an existing hydrogen bond. The forward lifetime τ_{F} of a hydrogen bond is associated with the activation free energy via the relation

$$\tau_{\mathrm{F}} = \frac{h}{k_B T} \exp\left(\frac{\Delta F^*}{k_B T}\right), \tag{55}$$

where $k_B T$ denotes the thermal energy, h the Planck constant and ΔF^* the activation energy necessary to break an arbitrarily chosen hydrogen bond [219]. Consequently, larger lifetimes correspond to larger activation energies, which is associated with the strengthening of the hydrogen bond network according to the dynamical retardation effect. Thus, long hydrogen bond forward lifetimes correspond to large activation free energies in accordance with Eqn. 55.

Analysis of co-solute effects on water dynamics can be complemented by the study of temporal molecular orientations. The autocorrelation function for the dipolar orientation vector $\boldsymbol{\mu}$ at times t and t_0 is related to the dipolar relaxation time τ_{D} according to the formula

$$\langle \boldsymbol{\mu}(t) \boldsymbol{\mu}(t_0) \rangle \propto \exp(-t/\tau_{\mathrm{D}})^{\beta}. \tag{56}$$

It gives a measure for the timescale of molecular fluctuations with the stretching exponent β. For simplicity, the stretching exponent for water usually shows values $\beta < 1$ which points out a complex decoupling mechanism for individual molecular configurations, including distinct rotational and translational contributions [220, 221].

In order to get insight into ordering effects, one can calculate the relative permittivity ϵ_{r} (dielectric constant) according to the dipole moment fluctuation formula [222, 223, 224]

$$\epsilon_{\mathrm{r}} = 1 + \frac{4\pi}{3} \frac{\langle \boldsymbol{M}_{\mathrm{tot}}^2 \rangle}{\langle V \rangle k_{\mathrm{B}} T} \tag{57}$$

where $\langle \boldsymbol{M}_{\mathrm{tot}}^2 \rangle$ stands for the average squared net total molecular dipole moment in a simulation box of average volume $\langle V \rangle$. If the calculations are restricted to the dielectric constant of water, the corresponding value of the water phase does not typically coincide with the dielectric constant of the solution. Such approach would provide a reliable comparison between pure water behavior and the properties of water molecules in binary or ternary solutions, thus giving a feasible estimate for the influence of co-solutes on the water dynamics and structure. Increasing or decreasing value of ϵ_{r} hence suggests a more or less ordered water phase, respectively, in comparison to pure water.

2.7 Molecular force fields

In molecular modelling and chemistry, the molecular force field is the potential energy function of a system of atoms or coarse-grained particles defining molecules and its parameters. Force fields can be interpreted in terms of intra- and intermolecular interactions present in the system [225]. Mathematically, the molecular force field energy function is given by the Eqn. 58 [225]:

$$V(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{v_n}{2}(1 + cos(n\omega - \gamma)) + \\ + \sum_{i=1}^{N} \sum_{i=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \tag{58}$$

where $V(r^N)$ defines the potential energy as a function of positions r of N atoms, l_i and $l_{i,0}$ are actual and reference bond lengths, θ_i and $\theta_{i,0}$ are actual and reference valence angles, ω is the torsion angle, γ angle phase, k_i and v_n force constants, $\epsilon_{i,j}$ is the minimum of van der Waals force for a pair of atoms i and j, $r_{i,j}$ distance between atoms i and j, $\sigma_{i,j}$ the distance between i and j atoms where van der Waals force is equal to zero, q_i and q_j the charges of atoms i and j, and ϵ_0 is the dielectric constant.

The force field can be thus defined by the combination of the following terms: term corresponding to the interaction between a pair of bonded atoms, summated over all the bond lengths (the first term in Eqn. 58), term describing the valence angles in the molecules (the second term in Eqn. 58), term for the torsion potential (the third term in Eqn. 58) and the term for non-bonded interactions, typically modeled with the application of Coulomb potential for electrostatic interactions and Lennard-Jones potential for van der Waals interactions (the last term in Eqn. 58).

The classical force fields applied most commonly in computer simulations of biomolecules are GROMOS, AMBER, CHARMM and OPLS [226, 227, 228]. GROMOS (GROningen MOlecular Simulation) force field is historically one of the first force fields for molecular dynamics simulation [229]. The optimization of the united atom force field was performed with respect to the properties of alkanes, with van der Waals parameters derived from hydrocarbon crystal structure calculations. The force field version released in 1987 used 0.8 nm nonbonded cut-off radii on amino acids, whereas the reparametrized version from 1996 applied 1.4 nm cut-off radii [229, 230]. CHARMM (Chemistry at HARvard Molecular Mechanics) force fields, initially released in 1983, are widely applied to date for molecular dynamics simulations of wide variety of systems, including nucleic acids, lipids or proteins [231]. For protein simulations, the CHARMM force field include the united-atom CHARMM19, all-atom CHARMM22 [232], CHARMM27, CHARMM36 and their modifications [233]. The CHARMM22 is parametrized for TIP3P water model. CHARMM27 force field is often applied for DNA, RNA and lipids simulations [234]. The OPLS (Optimized Potentials for Liquid Simulations) force field is optimized to fit the experimental properties of liquids, including heat of vaporization and density [235, 236]. Additionally, the OPLS parameters fit also the torsional properties of the gas phase. United atom OPLS-UA force fields incorporate hydrogen atoms to the nearest carbon and the parameters of both atoms are treated as common, whereas in all-atom OPLS-AA force field every atom is treated explicitly. In aqueous solutions, OPLS force fields are typically parametrized to conform with TIP3P or TIP4P water models.

The AMBER (Assisted Model Building with Energy Refinement) force field, whose versions are employed in this study, has been optimized to fit the properties of various biomolecules [237, 226]. The functional form of the AMBER force field is expressed by the Eqn. 59:

$$V(r^N) = \sum_{bonds} k_i(l_i - l_{i,0})^2 + \sum_{angles} k_i(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \sum_{n} \frac{v_{i,n}}{2}(1 + cos(n\omega_i - \gamma_i)) + \\ + \sum_{i=1}^{N-1} \sum_{i=i+1}^{N} f_{ij} \left(\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \tag{59}$$

with the parameters analogous to those in Eqn. 58. For molecular dynamics simulations of nucleic acids, the AMBER parm99 force field has been refined in order to improve the representation

of the α/γ torsional term in nucleic acids [227]. The resulting AMBER ParmBSC0 force field retained the standard AMBER parm94-99 parametrizations with the exception of the torsional parameters α and γ, which have been refined. Together with the progresses in the performance of computational methods, certain deviations from experimental data have been discovered as the lengths of the MD simulations of nucleic acids reached the microsecond regime. This resulted in further refinement of the ParmBSC0 force field, in particular concerning the proper estimation of twist, sugar puckering, terminal fraying and biases in ϵ and ζ torsions [228]. The refined ParmBSC1 force field has been parameterized from quantum mechanical data to create a general-purpose force field for DNA simulations [228].

2.8 Optimization of the molecular system

The potential energy of a system can be described as the multidimensional and complicated function of atomic coordinates. The minimum of the potential energy corresponds to those arrangements of atoms for which the states of the system are stable. Deviations from the minimum result in higher energy, which is translated into less stable configuration of atoms. The purpose of the optimization of the system's molecular structure is to search for a local energetic minimum with the application of an appropriate algorithm that acts on the energy function [238]. In more detail, the minimization procedure corresponds to the search for a point in configuration space where the first partial derivative of the energy function f with respect to each of the Cartesian or internal coordinates of the atoms x_i is equal to zero, and where the second derivatives are all positive. Mathematically [238]:

$$\frac{\partial f}{\partial x_i} = 0 \; ; \; \frac{\partial^2 f}{\partial {x_i}^2} > 0 \tag{60}$$

The Steepest Descent (Steep) algorithm [238], applied for the optimization purposes in this study, belongs to the group of the first-order minimization algorithms, basing on the first derivatives (gradients, g) of the function. This iteration method assumes a gradual movement of the atoms along the net force acting on them as long as the deviations in their coordinates bring the system closer to the minimum. The initial configuration of the system described by the vector $\mathbf{x_1}$, defines the starting point for the first iteration. For each subsequent of k iterations, the molecular configuration resulting from the previous step, corresponding to the vector $\mathbf{x_{k-1}}$, defines the new starting point. Thus, for 3N Cartesian coordinates, the direction of the movement is represented by 3N-dimensional unit vector $\mathbf{s_k}$

$$\mathbf{s_k} = -g_k/|g_k| \tag{61}$$

Consequently, taking a step of arbitrary length λ_k along the direction of the unit vector $\mathbf{s_k}$ determine the location of the minimum. The set of the newly defined coordinates after performing k iterations is then described by the Eqn. 62 [238]:

$$\mathbf{x_{k+1}} = \mathbf{x_k} + \lambda_k \mathbf{s_k} \tag{62}$$

2.9 Molecular Dynamics simulation

The principle of molecular dynamics (MD) depends on the computer simulation of trajectories of atoms and molecules by solving numerically Newton equation of motion for each of the interacting particles in the system. The trajectory thus defines the fluctuations of the positions and velocities of the particles with time. The function of the potential energy defines forces acting on atoms. Consequently, every time step Δt for the specified simulation time t, the equation of motion defined as

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{63}$$

where m_i is the mass of the particle i, x_i the coordinate along which the motion of particle i occurs, and F_{x_i} the force on particle i in the direction x_i, is being solved via Verlet integration algorithm [239, 225, 240]. This algorithm applies the coordinates of atoms $x_i(t)$ and the accelerations $a_i(t)$ at time t together with the coordinates from the previous step $x_i(t - \Delta t)$ in order to calculate the new positions $x_i(t + \Delta t)$ at time $t + \Delta t$ [239, 225]:

$$x_i(t + \Delta t) = 2x_i(t) - x_i(t - \Delta t) + \Delta t^2 a_i(t) \tag{64}$$

The velocities in the system are thus calculated as follows:

$$v_i(t) = \frac{x_i(t + \Delta t) - x_i(t - \Delta t)}{2\Delta t} \tag{65}$$

The Verlet algorithm comes in several variants, the most popular being the basic Störmer–Verlet algorithm, velocity Verlet and the leap-frog algorithm [225]. The leap–frog method, which is one of the most commonly applied integrator in computer simulations, uses the positions x_i at time t and the velocities v_i at time $t - 1/2\Delta t$. The positions and velocities in the system are then updated with reference to the forces $F_i(t)$ according to the following equations [225]:

$$v_i(t + \frac{1}{2}\Delta t) = v_i(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m_i} F_i(t) \tag{66}$$

$$x_i(t + \Delta t) = x_i(t) + \Delta t v_i(t + \frac{1}{2}\Delta t) \tag{67}$$

In these aspects, MD simulations provide the information about the state of an individual molecule in a specified moment along the trajectory, which can be beneficial for the applications where the experimental approaches are intricate, such as NMR or X-ray crystallography, which provide the results averaged over all the molecules present in the system or over specified time frame [241, 242]. The limitations of the MD simulation approach involves the allowed simulation time and the molecular system size, which requires a compromise between the time of the production run and the complexity of the simulated system [240].

As a rule, solving the Newtonian equations of motion during a molecular dynamics simulation follows the principle of energy conservation, which conforms with the microcanonical ensemble (NVE). Simulating the system at the specified constant temperature T requires however the implementation of a temperature–controlling algorithm called *thermostat* [225, 243]. One of the most commonly applied thermostats has been introduced by S. Nosé and further developed by W. G. Hoover [244, 245]. The Nosé–Hoover thermostat couples the system to the target temperature by introducing an artificial variable s, which plays the role of an exteral system for the physical system described by the particle coordinates x_i, masses m_i and potential energy $E(x)$. This variable serves as a scaling parameter, such that the time increment in the scaled system dt' is related to the real system like $dt' = sdt$, whereas the atomic coordinates $x'_i = x_i$ remain unchanged [244, 245]. Hence, this method is called the Extended System (ES) method. The Nosé expression for the real velocity in the extended system reads then

$$\frac{dx_i}{dt} = s\frac{dx_i}{dt'} = s\frac{dx'_i}{dt'} \tag{68}$$

and the Nosé formulation of the equation of motion in the Lagrangian form [244, 245]:

$$\frac{d^2x'_i}{dt'^2} = -\frac{1}{m_i s^2}\frac{\partial E}{\partial x'_i} - \frac{2}{s}\frac{ds}{dt'}\frac{dx'_i}{dt'} \tag{69}$$

Besides the simulations at constant temperature, certain simulations require the control of the system's pressure in order to keep it constant or to exert an external stress on the simulated system. The algorithms called *barostats* are regulating the pressure in the system by scaling the atomic coordinates and hence adjusting the system's volume. The classical method inroduced by H. C. Andersen employs the concept of an extended system variable, which serves as an additional degree of freedom to fine-tune the volume of the simulation cell in response to the changes of pressure [246]. This method has been further developed and extended in subsequent years [247, 248, 249]. Due to the similarity of the equations of motion, the Nosé–Hoover thermostat is usually combined with Parrinello–Rahman barostat [247]. The Parrinello–Rahman pressure coupling method allows not only the volume scaling, but also adjusting the shape of the simulation box under external stress [247].

2.10 Exploring free energy via Metadynamics

Since its origin, metadynamics algorithm [250, 251, 252] has been widely applied for studying the system's potenial of mean force (PMF) and for reconstructing its free energy surface [253]. The principle of the method lays in enhancing the rate of transitions between long-living metastable states by biasing iteratively the system's potential along the reduced reaction coordinate $s = s(x)$ called the Collective Variable (CV) [251, 253]. Collective variables typically correspond to certain parameters of the simulated structure that could be measured throughout the simulation. The examples are the distance between stacked base pairs in nucleic acids, distance to the binding site or other active sites, dihedrals, angles or similar. In this aspect, metadynamics allows acceleration of rare events, at the same time enabling the estimation of the free energy in complex molecular systems [254].

The free energy landscape of the system along s can be thus reconstructed by the deposition in the regular time intervals τ of the biasing potential V_B, which is defined by the Eqn. 70 [255]:

$$V_B(s,t) = \sum_{t=0,\tau,2\tau,\ldots} h e^{\frac{-(s-s_t)^2}{2\sigma^2}} \tag{70}$$

with h and σ being respectively Gaussian height and width, s_t defining the value of the CV at the time step τ and s the subsequent values of the CV along the grid bins at the time step τ. Thus, the total potential of the system during a metadynamics simulation is expressed by the sum of the original potential $V(x)$, enhanced by a biasing potential term V_B.

To date, there have been developed numerous extensions and variants of the metadynamics method, adapted to study diverse systems under the conditions of various complexity. The effectivity of the classical metadynamics and its alternative derivations relies strongly on the choice of the CVs. Neglecting an important CV can result in the wrong estimate of the free energy and erroneous prediction of the transition mechanisms [254]. The ideal collective variable covers the phenomena both in the metastable states and in the transition states between them [254]. There are various methods for finding the optimal CVs, each of them associated with certain advantages and disadvantages. One of the most broadly applied is the committor parametrization, which bases on the likelihood maximization of being on a transition path [256]. Although this method can provide the most optimal CV, its determination requires the knowledge of the initial and final states of the system and can be computationally very costly [256, 254]. The knowledge about initial and final states is required also in the spectral gap optimization method [257]. This approach is computationally cheap and allows to find the best variable in a set. However, the parametrization of the new variable is not allowed [257]. The methods of computing the path collective variable allows the description of the reaction pathways with high complexity, at the same time enabling iterative optimization [258, 259, 260]. The disadvantage however lays in the requirement of the knowledge about the initial and final state of the system [258, 259, 260]. This disadvantage can be overcome by applying the machine learning methods. These procedures typically require only ensemble averages [261, 262, 263] or focus on the kinetics of the system [264, 265, 266]. The limitations however involve the necessity to apply re-weighting methods [264, 265, 266], and the defined barriers may be described only suboptimally by the resulting CVs [261, 262, 263].

Beside the ordinary metadynamics, which allows the induction of transitions between the metastable states even with non-ideal CV, another commonly applied method is the well-tempered meradynamics (WT-MetaD) [267]. The advantage of the well-tempered scheme relies on its ability to provide an exact estimator of the free energy, which is difficult to achieve otherwise [267]. The principle of the well-tempering involves the construction of the history-dependent potential as the sum of Gaussian functions deposited in the CV space along the trajectory. This approach introduces ω as the rate of bias deposition as well as the tunable simulation parameters to promote exploring the physically relevant regions of the free energy surface (FES). The latter involves tuning the temperature ΔT to increase the barrier crossing. The exploration of the FES can be confined to the energy range of the $T + \Delta T$ by applying the finite ΔT. This facilitates the exploration of the CVs space and limits it to the CV regions of interest [267]. Thus, the height of the added Gaussians h is modified in the well-tempered ensemble according to the Eqn. 71 [267]:

$$h = \omega\tau e^{-\frac{V_B(s,t)}{\Delta T}} \tag{71}$$

where ω is the bias deposition rate in the units of energy over the units of time, $V_B(s,t)$ is the free energy estimate at the current time step and current position of the CV, τ the time interval of Gaussian deposition and ΔT is the tunable temperature parameter controlling the rate of Gaussian height h reduction with the progress of filling the potential wells. It is worth mentioning that at the beginning of the metadynamics run the biasing potential is equal to zero, and hence $h = \omega$. The well-tempeted metadynamics is usually used to change the Gaussian size and width adaptively, which involves the algorithm where the simulation history is applied to enhance sampling speed [252].

To improve the sampling and take advantage of the parallel performance, individual metadynamics simulations, called *replicas*, can be coupled together in the multi-replica approach. The metadynamics simulations involving replica exchange approach have been shown to improve the convergence of free energy estimates [268]. Methods involving the application of several replicas include the bias-exchange MetaD [269], collective-variable tempering MetaD [270], multiple walker MetaD [271] and the parallel tempering MetaD (PT-MetaD) [272]. The latter method implicates performing multiple metadynamics simulations at different temperatures. For all these temperatures, the free energy profiles are calculated parallely, which prompts the system to diffuse in the CV landscape. The simulated configurations are then being exchanged periodically according to the replica exchange principle [272]. This scheme prevents the colder replicas from being trapped in the local minima due to the exchange with the ones having higher temperature. The combination of metadynamics algorithm with parallel tempering improves the overall performance of the metadynamics method by promoting the exploration of the low probability regions and sampling the degrees of freedom that are hidden otherwise. An alternative approach combines the parallel-tempered metadynamics (PT-MetaD) with well-tempered ensemble (WTE) [273]. When the energy is used as a CV in the WT-MetaD scheme, the sampled biased ensemble is called WTE. The merge between PT-MetaD and the WTE increases the efficiency of either of the methods alone in exploring the FES [273].

3 TMAO and urea in pure water

In this section, the properties of aqueous urea–TMAO solutions at physiologically relevant concentrations are discussed [6]. The atomistic molecular dynamics (MD) simulations have been applied to analyze the individual and combined influences of both co-solutes on water structure and dynamics in terms of binary and ternary solutions at various mixing ratios. As key indicators, the radial distribution functions, the number of hydrogen bonds, the dipolar relaxation times and the resulting dielectric constants for different mixing ratios of urea and TMAO up to the concentrations of 3 M are studied. Contrary to commonly applied assumptions on water-related differences between TMAO and urea, the findings presented here imply hat both co-solutes have to be regarded as water-structure makers. Therefore, these results are in clear contrast to the experimentally observed compensating effects of combined urea and TMAO influences in solutions with regard to protein structures. Here, it is also demonstrated that both co-solutes show clearly kosmotropic properties, which challenges the validity of water-mediated effects for biomolecular structure stabilization or destabilization mechanisms.

3.1 Methodology

I carried out atomistic MD simulations of aqueous solutions with TMAO and urea at physiological concentrations up to 3 M. The mixing ratios (TMAO:urea = 1:1, 1:2, 1:3, 2:1, 2:2, 2:3, 3:1, 3:2 and 3:3) M have been applied. In addition, I performed 6 control simulations of binary aqueous solutions containing 1-3 M urea or 1-3 M TMAO molecules as well as one simulation of pure water. The co-solutes have been immersed in a a cubic simulation box with an initial box length of 5 nm in presence of three-dimensional periodic boundary conditions. The systems were simulated with the use of GROMACS 4.6.5 software package [274, 275, 276], with the Kirkwood-Buff (KB) force field for urea [277], a refined Kast force field for TMAO [175] and the SPC/E water model [278]. The choice of this force field combination was dictated by its validity for the study of aqueous solutions [277, 279]. The Particle Mesh Ewald (PME) method [280, 281] has been applied to study long-range electrostatic interactions, with the short-range cutoff distance of 0.9 nm. The same cutoff distance was also used for the evaluation of Lennard-Jones interactions with Lorentz-Berthelot mixing rules. The LINCS algorithm [282] was employed to constrain all covalent bonds between atoms. Prior to the production MD run, the energy of the systems was minimized with a steepest descent algorithm. A constant temperature of 300 K has been provided by the use of the Nose-Hoover thermostat [244, 245] in the equilibration runs. First, the systems were equilibrated for 3 ns (time step Δt = 2 fs) in the constant temperature–constant volume NVT ensemble, followed by a 3 ns equilibration run in the constant temperature–constant pressure NpT ensemble. A constant pressure was maintained with the use of the Parrinello-Rahman barostat [247] with a reference pressure of 1 bar. Consequently, each solution was simulated for 150 ns (time step Δt = 2 fs) in the NpT ensemble at 300 K.

3.2 Results and discussion

3.2.1 Radial distribution functions

The radial distribution functions (RDFs) for water molecules around urea and around TMAO for various mixing ratios are shown in Fig. 4.

The results show a pronounced accumulation of water molecules around TMAO, which is manifested by several peaks in the TMAO-water RDFs at short distances $r \leq 0.75$ nm (Fig. 4, top). The corresponding positions of the peak at $r \leq 0.5$ nm can be related to the first hydration shell and highlight different locations of water molecules around TMAO. It can be noted that the peak heights increase for higher TMAO concentrations, which can be rationalized by the strong binding of water molecules and consequently a decreasing amount of bulk water in the simulation box. In relation to these observations, the potential of mean force (PMF) between water and the respective TMAO molecules, defined by $\Delta_{\mathrm{PMF}}(r) = -k_B T \ln g_{\gamma\delta}(r)$, shows a slight decrease in binding energy when TMAO concentration increases ($\Delta_{\mathrm{PMF}}^{\mathrm{max}} \approx -0.1\, k_B T$ to $-0.5\, k_B T$ from 1 M TMAO to 3 M TMAO) at $r_c = 0.45$ nm. A minor strengthening of these interactions is provided also by the presence of urea. This is manifested by the most favorable binding energy between TMAO and urea observed in a 3:3 M mixture, followed by 3:2 M and 3:1 M TMAO–urea aqueous solutions.

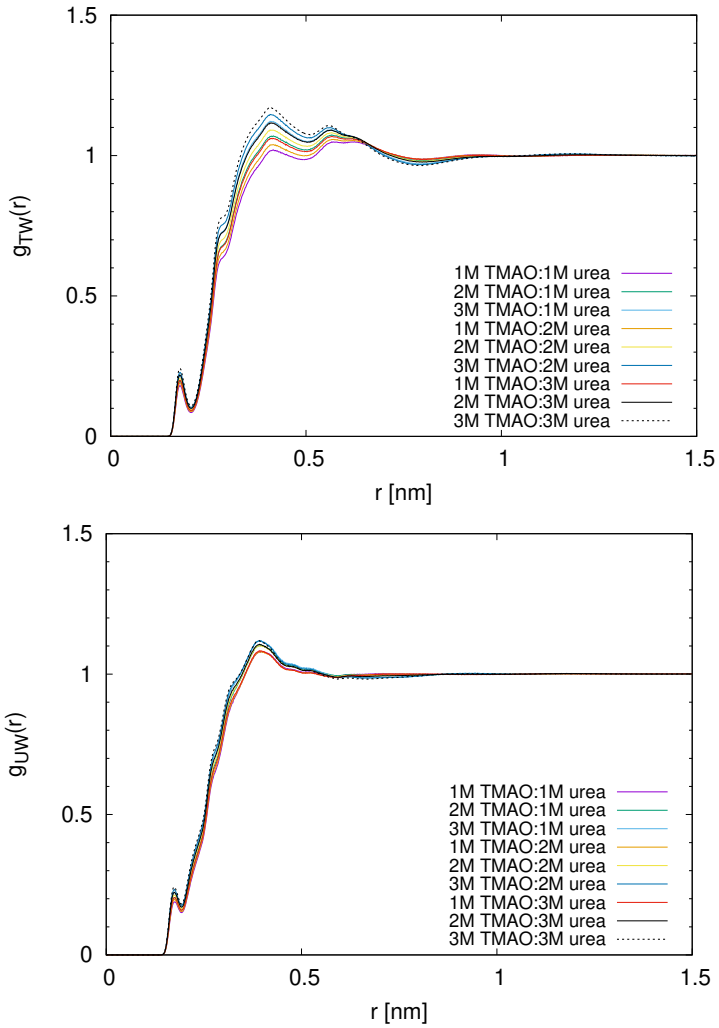


Figure 4: Center-of-mass radial distribution functions $g_{\mathrm{TW}}(r)$ and $g_{\mathrm{UW}}(r)$ for water molecules around TMAO (top) and around urea (bottom) at different mixing ratios as denoted in the legend. Reprinted with permission from [6]. Copyright 2019 American Chemical Society.

On the other hand, the water binding tendency in 1:1 M, 1:2 M and 1:3 M TMAO–urea mixtures is significantly lower. This shows that larger concentrations of TMAO molecules are necessary for the same effect to occur.

Preferential interaction between urea and water molecules are evident also from the RDFs between both species (Fig. 4, bottom). The heights of the corresponding RDF peaks are comparable to those observed for TMAO–water interactions, which highlights also the analogy of the binding strengths between both sets of molecules. In mixtures, the binding strength between urea and water molecules does not depend significantly on the concentration of TMAO as it was observed for the impact of urea on TMAO-water interactions and *vice versa*. Hence, the observed variations of urea–water RDFs with increasing TMAO concentrations are only minor. The highest urea–water RDF peaks are thus observed for 3:3 M TMAO–urea mixtures, followed by 3:2 M and 3:1 M mixing ratios.

The RDFs calculated for TMAO and urea aqueous binary solutions at concentrations 1-3 M are shown in Fig. 5. Similarly like in ternary solutions (Fig. 4), also in binary mixtures there is only negligible variation in RDFs for increasing urea concentration. In contrast, increase in TMAO concentration is associated with stronger binding of water molecules. This shows that the properties of binary solutions resemble strongly those of ternary mixtures, meaning the absence of the mutual compensation effects from individual TMAO and urea contributions.

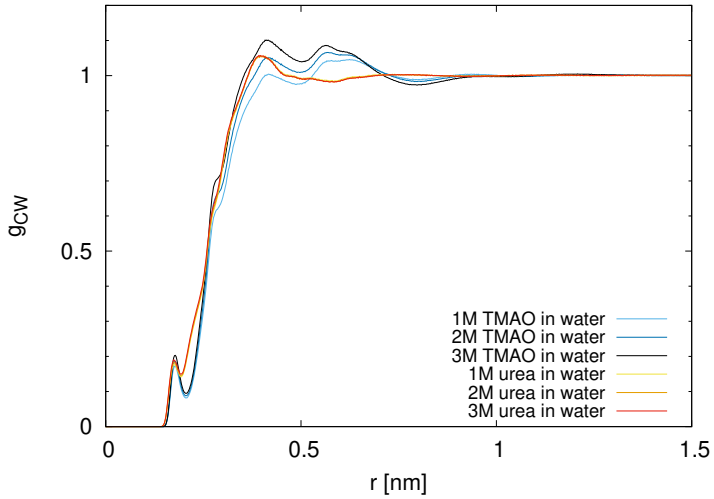


Figure 5: Center-of-mass radial distribution functions for water around urea or TMAO molecules, respectively, in binary mixtures at different molar concentrations as denoted in the legend. Reprinted with permission from [6]. Copyright 2019 American Chemical Society.

Taking into consideration both global and local length scales, the observed RDFs show that the long-range decay of structured water shells induced by TMAO ($r_c \approx 1$ nm) is broader than that caused by urea ($r_c \approx 0.7$ nm). This can be rationalized by the slightly larger molecular volume of TMAO in comparison to urea. This hypothesis is further elucidated by the analysis of the values of molecular volumes V_m, octanol-water partition coefficients and total polar solvent accessible surface areas (TPSAs) for TMAO and urea, which are presented in Tab. 1. The parameters were calculated by numerical approach from Ref. [283].

Table 1: Molecular volumes V_m, total polar solvent-accessible surface areas (TPSAs) and octanol–water partition coefficients $\log_{10} \mathrm{P}$ for TMAO and urea molecules.

Species	V_m [nm^3]	TPSA [nm^2]	$\log_{10} P$
Urea	0.54	0.69	-1.00
TMAO	0.83	0.17	-3.46

It can be noted that the molecular volume is higher for TMAO in comparison to urea, whereas a significantly higher TPSA values are observed for urea. However, the values for the octanol-water partition coefficients show that TMAO exhibits a more hydrophilic behavior. This can be explained by TMAO zwitterionic nature and the corresponding higher number of polar groups than urea [145].

Fig. 6 presents the center-of-mass RDFs between TMAO and urea molecules at different molar concentrations. Here, only a minor tendency of co-solute accumulation can be observed, which is in agreement with hitherto experimental findings [167, 82, 36].

Despite negligible PMF and binding energy values, a slight increase in the peak height for growing TMAO concentrations can be observed. This growth is not influenced by urea concentration in the solution. Interestingly, the first RDF peak is located around 0.6 nm, which corresponds to the position of the first hydration shell around TMAO (Fig. 4). Taking into consideration the hydrophilic properties of both co-solutes (Tab. 1), one can consider TMAO–urea aggregates as well-hydrated complexes, or solvent-shared co-solute pairs. Hence, the results presented in this chapter confirm the experimental findings [167, 172, 82, 36], attributing the reported weak association behavior between TMAO and urea molecules to a minor hydrogen bond mismatch.

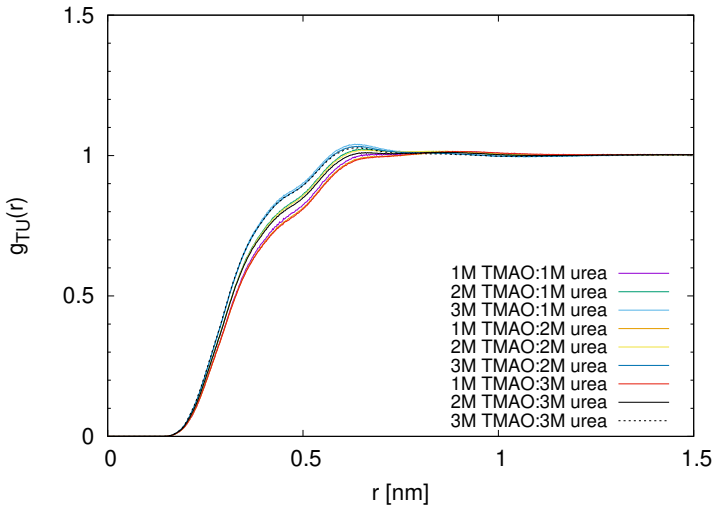


Figure 6: Center-of-mass radial distribution functions between TMAO and urea molecules at different molar concentrations as denoted in the legend. Reprinted with permission from [6]. Copyright 2019 American Chemical Society.

3.2.2 Properties of hydrogen bonds

I analyzed the number N_{HB} of hydrogen bonds (H-Bonds) between particular components: TMAO and urea, urea and water, TMAO and water and between water molecules, for both ternary nad binary solutions. The results for the normalized number of hydrogen bonds $N_{\mathrm{HB}}/N_{\mathrm{H_2O}}$, divided by the number of water molecules $N_{\mathrm{H_2O}}$ in the simulation box are ilustrated in Fig. 7. A short-hand notation for the individual mixtures is explained in Tab. 2.

Table 2: Short-hand notation for binary and ternary aqueous solutions of urea and TMAO.

Notation	Solution	Concentration urea [M]	Concentration TMAO [M]
W	Pure water	–	–
1T	Binary	–	1
2T	Binary	–	2
3T	Binary	–	3
1U	Binary	1	–
2U	Binary	2	–
3U	Binary	3	–
1U1T	Ternary	1	1
1U2T	Ternary	1	2
1U3T	Ternary	1	3
2U1T	Ternary	2	1
2U2T	Ternary	2	2
2U3T	Ternary	2	3
3U1T	Ternary	3	1
3U2T	Ternary	3	2
3U3T	Ternary	3	3

It can be observed that the number of H-bonds formed between TMAO and urea is negligibly small in comparison to the H-bond number between either of the co-solutes and water, in agreement with recent experimental findings [172, 36] and the previous observations based on the analysis of the TMAO-urea RDFs (Fig. 6). The growing number of TMAO-water H-bonds independently of the urea concentration confirms the tendency of higher water accumulation for increasing TMAO concentrations. Furthermore, also the number of urea-water H-bonds increases with increasing urea molarity regardless of TMAO concentration. Interestingly, in equimolar TMAO and urea

solutions (1U1T, 2U2T, 3U3T), the number of H-bonds formed with water is slightly higher for urea when compared to TMAO, although the PMF values would suggest a stronger binding of water molecules to TMAO. The slightly higher amount of H-bonds between water and urea can be rationalized by the higher TPSA values, in reference to Tab. 1.

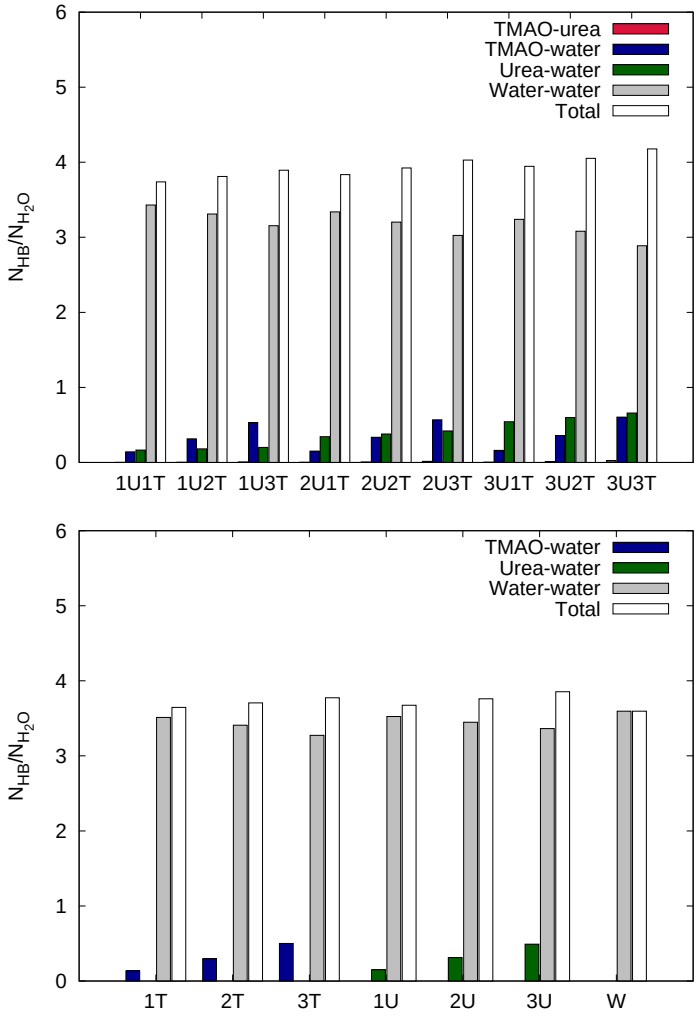


Figure 7: Top: Normalized number of hydrogen bonds $N_{\mathrm{HB}}/H_{\mathrm{H_2O}}$ between TMAO-urea, TMAO-water, urea-water and water-water molecules for different ternary solutions. The sum of all normalized hydrogen bonds corresponds to the total hydrogen bond number. Bottom: Normalized number of hydrogen bonds $N_{\mathrm{HB}}/H_{\mathrm{H_2O}}$ between TMAO-water, urea-water and water-water molecules in binary aqueous solutions of TMAO and urea as well as in pure water. The sum of all normalized hydrogen bonds corresponds to the total hydrogen bond number. Definitions of short-hand notations are presented in Tab. 2. Reprinted with permission from [6]. Copyright 2019 American Chemical Society.

It can be seen that the number of intermolecular hydrogen bonds between water molecules decreases for larger concentrations of co-solutes, which is a consequence of an increasing number of H-bonds between water molecules and co-solutes in such mixtures. This effect is more pronounced for increasing TMAO concentrations when compared to increasing urea molarity. At the same time, the total number of normalized hydrogen bonds increases with growing TMAO and urea concentrations, which can be explained by the overall decreased amount of water molecules in the solution and the multiple number of H-bonds formed between the molecules of water and the corresponding co-solute. These results suggest that the hydrogen-bond structure of water remains intact in the presence of both urea and TMAO.

Similar conclusions can be drawn basing on the results for binary mixtures of urea and TMAO (Fig. 7, bottom). For certain concentration of co-solute, the number of H-bonds formed between the co-solute and water is comparable to that reported for ternary solution and remarkably lower than the number of water-water H-bonds in either mixture. Although the number of water-water H-bonds decreases in the presence of co-solutes in comparison to pure water, the normalized total number of hydrogen bonds in the solution increases. At the same time, the number of H-bonds formed between co-solutes of the same type (TMAO-TMAO and urea-urea) is insignificantly small, pointing out that the clustering effects can be neglected. Thus, one can conclude that the formation of more complex and rigid hydrogen bond network in the solution is favored in the presence of co-solutes, and that the individual components in ternary solutions behave analogous to their corresponding counterparts in binary solutions, which implies that the mutual compensating influences on the H-bond properties are negligible.

I analyzed also the energetic stability of the hydrogen bonds, expressed as the average hydrogen bond forward lifetimes. The results for binary and ternary solutions are presented in Fig. 8.

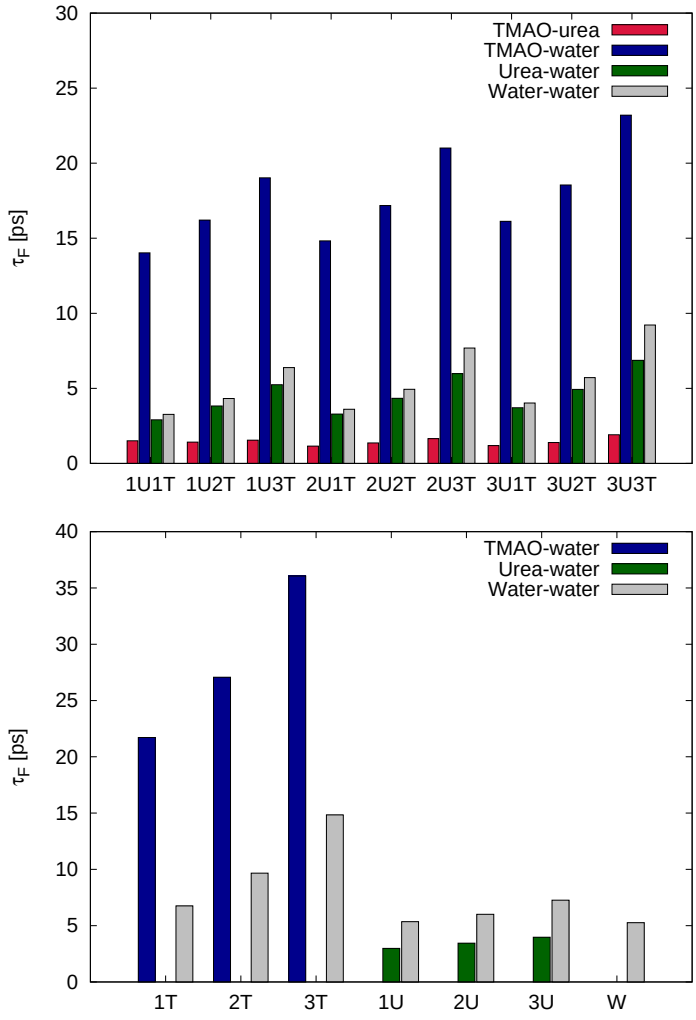


Figure 8: Top: Average forward lifetime τ_F of hydrogen bonds formed between each component of ternary solutions for different co-solute mixing concentrations as denoted in the legend. Bottom: Average forward lifetime τ_F of hydrogen bonds formed between each component of aqueous binary solutions with urea or TMAO, respectively, and in pure water for different co-solute concentrations. Detailed definitions of short-hand notations for solutions are presented in Tab. 2. Reprinted with permission from [6]. Copyright 2019 American Chemical Society.

It can be noted that the average lifetime of hydrogen bonds between TMAO and urea molecules in ternary solutions is rather low, independently on the concentration of either co-solute. This suggests that the stability of such hydrogen bonds is rather meager. On the other hand, the stability of hydrogen bonds formed between water and any of the co-solutes is remarkably higher. In particular, water–TMAO H-bonds appear to be the most stable of all the hydrogen bonds formed within the solutions, which confirms the experimental findings published to date, reporting weak urea–water and strong TMAO–water hydrogen bonds formed in aqueous ternary mixtures of urea and TMAO [36]. When water–water H-bonds are considered, in binary and ternary solutions their stability is enhanced by increasing TMAO concentrations regardless of urea concentrations. However, for a given TMAO concentration the H-bonds are significantly more stable in binary TMAO–water solutions in comparison to ternary mixtures of TMAO, urea and water. These results suggest that urea makes the hydrogen bond network of water around TMAO in a ternary solution weaker in comparison to the hydrogen bond network around TMAO in a binary solution. This phenomenon can be explained by the water binding property of urea.

Due to collective effects [56], both in binary and ternary mixtures the hydrogen bond relaxation times increase with increasing co-solute concentrations, with 3:3 TMAO:urea mixtures exhibiting the longest forward lifetime for water–TMAO hydrogen bonds in ternary solutions. This can be rationalized by the combined contributions from both co-solutes. Analogously, for given TMAO concentration the increase in the lifetime of TMAO–water hydrogen bonds was observed for growing urea concentration. However, in binary solutions an increase in urea molarity was not associated with any significant change in hydrogen bond forward lifetime between urea and water molecules. When these outcomes are compared with those for pure water, it can be noticed that an increase of co-solute concentrations enhances the forward lifetimes of hydrogen bonds between water molecules significantly, which implies a pronounced strengthening of the hydrogen bond network in the solution. Noteworthy, the characteristics of hydrogen bonds between urea and water appears to be improved in ternary solutions in comparison to binary urea-water mixtures. These results indicate that both co-solutes strengthen not only the water–water hydrogen bond network but also the hydrogen bonds formed with other solutes or biomolecules, such as DNA or proteins. It can be concluded that TMAO exerts the greater impact on the variation of hydrogen bond dynamics for water molecules than urea, which agrees with hitherto computational and experimental outcomes [284]. Hence, in the presence of TMAO it comes to the significant increase in water–water hydrogen bond lifetimes, and as a consequence to a pronounced stabilization of the hydrogen bond structure of water [285, 170]. Since the most pronounced enhancement of the average forward lifetimes of water-water H-bonds is observed in the ternary TMAO–urea mixtures with highest concentrations of either TMAO or urea in comparison to pure water, one can conclude that the combination of urea and TMAO strengthens cooperatively the structure of hydrogen bonds, in agreement with previous findings [172]. To summarize, the lifetimes of hydrogen bonds in ternary mixtures increase in the following order: urea–water < water–water < TMAO–water, with the most stable hydrogen bonds are formed between TMAO and water molecules.

3.2.3 Dielectric constant and dipolar relaxation times of water molecules

I analyzed the average dipolar relaxation times τ_{D} of water molecules in the presence of different concentrations of urea and TMAO in order to get insight into the influence of co-solutes on the local water dynamics. The results for binary and ternary solutions are shown in Fig. 9.

In Fig. 9, the concentrations of urea are fixed, whereas the number of TMAO molecules varies. The excess concentration of urea, introduced for the reasons of comparability between the mixtures, is defined as $\Delta c_{\mathrm{urea}} = c_{\mathrm{urea}} - c_{\mathrm{TMAO}}$, with c_{urea} and c_{TMAO} being the actual concentration of urea and TMAO molecules in solution. The positive values of Δc_{urea} denote the molar excess of urea over TMAO, whereas the negative values stand for a deficit of urea molecules over TMAO. The results show that the dipolar relaxation times of water molecules in pure water are lower than in ternary aqueous mixtures. This suggests that water dynamics is slowed down by the presence of co-solutes. The relaxation times are increasing only weakly with growing urea concentration and strongly with TMAO concentration, such that the largest values for the dipolar relaxation times in all solutions are observed for the highest TMAO molarities. For 1 M urea combined with 3 M TMAO at $\Delta c_{\mathrm{urea}} = -2$ M, the dipolar relaxation time is $\tau_{\mathrm{D}} \approx 18$ ps, and is growing through

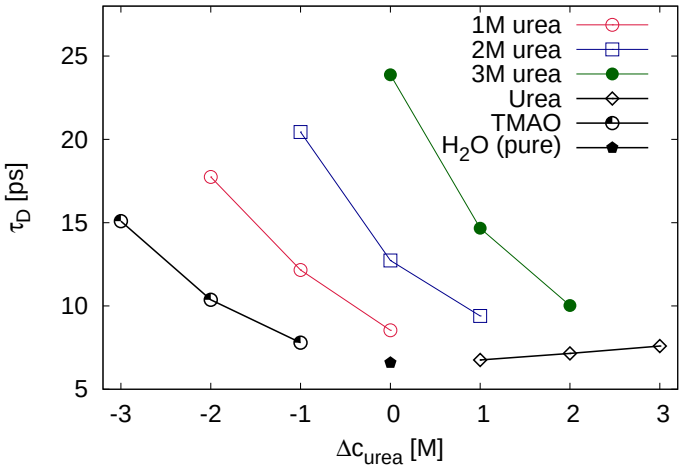


Figure 9: Dipolar relaxation times of water molecules in pure water (denoted as H_2O (pure)), in binary solutions of urea (denoted as Urea) and TMAO (denoted as TMAO) and ternary aqueous solutions with TMAO and urea molecules for fixed urea concentrations (denoted as 1M urea, 2M urea and 3M urea). The excess concentration of urea $\Delta c_{\mathrm{urea}} = c_{\mathrm{urea}} - c_{\mathrm{TMAO}}$ is explained in more detail in the main text. All lines are guides for the eyes. Reprinted with permission from [6]. Copyright 2019 American Chemical Society.

$\tau_{\mathrm{D}} \approx 21$ ps for 2 M urea at $\Delta c_{\mathrm{urea}} = -1$ M to $\tau_{\mathrm{D}} \approx 24$ ps for 3 M urea at $\Delta c_{\mathrm{urea}} = 0$ M. On the contrary, the increase of dipolar relaxation times for 1 M, 2 M and 3 M urea, combined with fixed 1 M TMAO concentration, which reads $\tau_{\mathrm{D}} \approx 8$ ps ($\Delta c_{\mathrm{urea}} = 0$ M), $\tau_{\mathrm{D}} \approx 9$ ps ($\Delta c_{\mathrm{urea}} = 1$ M) and $\tau_{\mathrm{D}} \approx 10$ ps ($\Delta c_{\mathrm{urea}} = 2$ M) respectively, is less significant. When the equimolar TMAO and urea concentrations (1:1, 2:2, and 3:3 M) at $\Delta c_{\mathrm{urea}} = 0$ M are considered, it can be seen that the dipolar relaxation times grow with growing co-solute concentrations. In agreement with the observations for hydrogen bond forward lifetimes, these results also show that the enhancing effects are stimulated mainly by the molecules of TMAO.

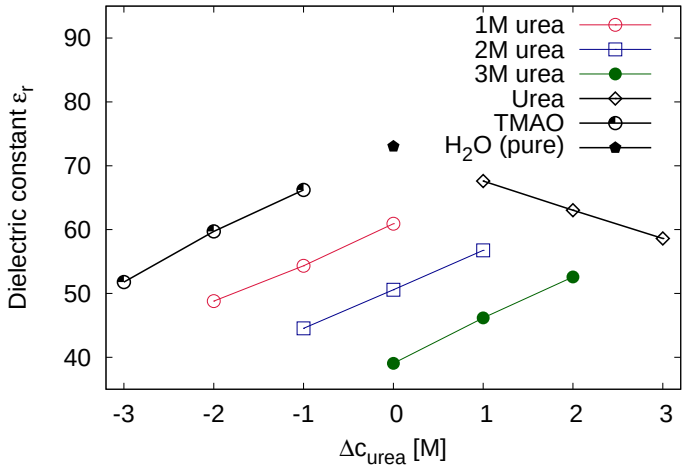


Figure 10: Water dielectric constant ϵ_r for pure water (denoted as H_2O (pure)), for binary solutions of urea (denoted as Urea) and TMAO (denoted as TMAO) and ternary aqueous solutions with TMAO and urea molecules for fixed urea concentrations (denoted as 1M urea, 2M urea and 3M urea). The excess concentration of urea $\Delta c_{\mathrm{urea}} = c_{\mathrm{urea}} - c_{\mathrm{TMAO}}$ is explained in more detail in the main text. All lines are guides for the eyes. Reprinted with permission from [6]. Copyright 2019 American Chemical Society.

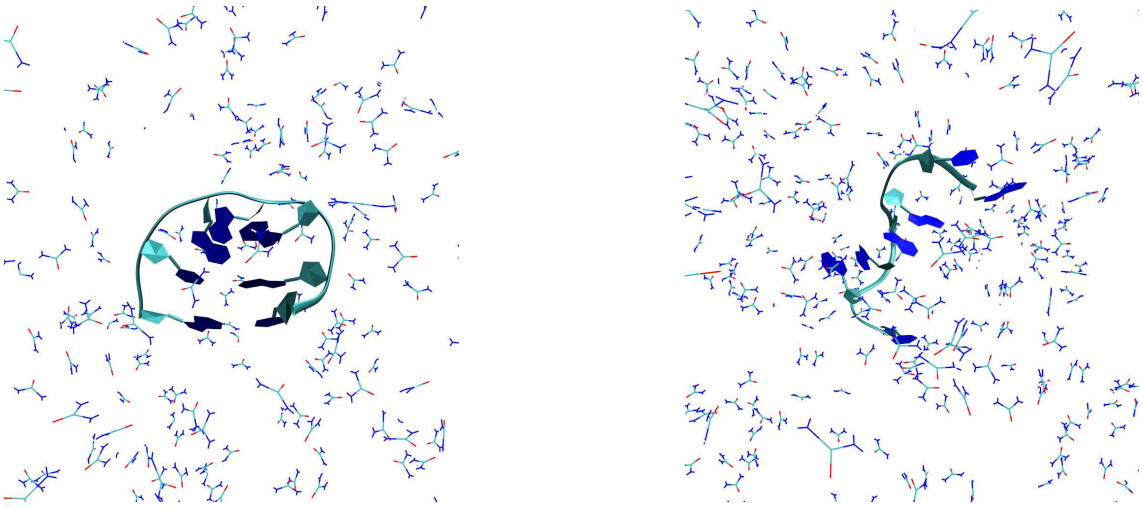
In Fig. 10 there are illustrated the values of water dielectric constant. It can be observed that the water dielectric constant ϵ_r decreases noticeably in the presence of co-solutes in comparison to pure water. In agreement with my previous results, this effect is stronger for increasing TMAO concentrations. However, a significant reduction of the water dielectric constant in binary solution happens even in the presence of urea. Thus, it can be concluded that the properties of water in the bulk phase are influenced by the local binding of water molecules around co-solutes. In consequence, this leads to the perturbation of the hydrogen bond network structure. Since the increase in co-solute molarity implies the increase in the fraction of locally bound water molecules, this leads consequently to the decrease in dielectric constant. As one can see, the decrease in the dielectric constant value are very similar in binary and ternary solutions of both co-solutes, which suggests that this effect is not influenced by the specific molecular features of TMAO or urea. Since the water phase appears to be less ordered and less mobile in the presence of both co-solutes comparing to pure water, this outcome would question the validity of the chaotropic and kosmotropic distinction between urea and TMAO.

4 Binding properties of urea to small DNA hairpins and its postulated impact on DNA stability

In this section, I study a short 7-bp DNA hairpin with the sequence 5'-D(*GP*CP*GP*AP*AP*GP*C)-3' (RCSB Protein Data Bank entry: 1KR8) [118] in the presence of varying concentrations of urea [3]. In nature, this oligonucleotide takes a form of a hairpin, composed of two Watson-Crick d(GC) base pairs forming the stem, and the nucleobase sequence d(GAA) constituting the loop region. The molecule consists of 7 nucleic base pairs, which corresponds to 222 atoms included in the DNA structure. Despite relatively small size, the oligonucleotides with a general sequence d(GCGNAGC), where N stands for any of the nucleobases A, G, C or T, are proved to be extraordinarily stable, which is manifested by a high melting temperature and resistance of such structures against nucleases [111, 116]. Among them, the fragment d(GCGAAGC) corresponds to the highest thermodynamic stability [109]. This stability is proved to be the result of a complex interplay between the formation of hydrogen bonds and aromatic base stacking [118]. These particular features turns the d(GCGAAGC) DNA hairpin into a perfect simple model system to investigate the stability and the intermolecular interaction of more complex DNA structures.

With this research, one gets a deeper insight into the nature and the mechanism of the DNA-urea interactions. The Kirkwood-Buff theory [84, 87] has been applied to approach the distribution of the molecules in the ternary solution of the oligonucleotide, co-solute urea and the solvent in more details. In order to address the problem of possible influence of urea on the free energy of DNA unfolding, I employed the Metadynamics method [250, 251, 252], biasing the normal dynamics of the system with a history-dependent potential. This study allows to get a more comprehensive understanding of the stability of the DNA structures in the presence of urea as well as the insight into the interplay between both co-solvents.

The snapshots of unfolded (11b) and native (11a) form of 1KR8 DNA surrounded by urea molecules in the concentration of 2 M is shown in Fig. 11.



(a) Native form of DNA hairpin

(b) Unfolded form of DNA hairpin

Figure 11: Native and unfolded 1KR8 DNA form in the presence of 2 M urea. Reprinted with permission from [3].

4.1 Methodology

4.1.1 High temperature molecular dynamics simulation of DNA in pure solution

The initial input structure of 1KR8 DNA oligonucleotide has been obtained from the RCSB Protein Data Bank [122]. This structure represents a native folded DNA form.

DNA unfolding can be perceived as a rare event. Thus, in order to access the unfolded form of the 1KR8 DNA structure, I performed high temperature molecular dynamics (MD) simulations

of the native DNA form [118]. The method of high temperature simulations have the unique advantage of accelerating rare events and hence enabling a more comprehensive exploration of the conformational space within the limit of the simulation time [115].

Prior to the MD run, the system was solvated in the cubic water box of the size 10.25 nm x 10.25 nm x 10.25 nm. The net charge was neutralized by the addition of 6 Na^+ ions. The energy of the system was minimized with the Steepest Descent algorithm. The system was equilibrated for 100 ns (time step $\Delta t = 2$ fs) in the NVT ensemble with position restraints on DNA molecule and Nose-Hoover thermostat [286, 244, 245] to control the temperature of the system. Then the system was equilibrated for 100 ns (time step $\Delta t = 2$ fs) in the NpT ensemble, with position restraints on DNA, Parrinello-Rahman thermostat [247] to control system's pressure and the reference pressure of 1 bar. The molecular dynamics (MD) simulations were performed in Amber99SB-ParmBSC0 force field [227] with the TIP3P water model (Transferable Interatomic Potential - 3 Point) [287] and no position restraints applied on DNA, using the Gromacs 4.6.2. software package [275]. The advantage of ParmBSC0 force field for DNA simulations lays in its ability to produce stable trajectories in multi-microsecond time scale, at the same time providing a correct and accurate description of the details of oligonucleotide structure. Particle Mesh Ewald (PME) method [288, 280, 281] has been employed to treat long-range electrostatic interaction, with 0.9 nm of short-range cutoffs for electrostatic and van der Waals interactions. The three dimensional (3D) periodic boundary conditions were applied on the system. All bonds have been constrained with the LINCS algorithm [282]. During the MD run, the temperature was controlled by the Nose-Hoover thermostat and the pressure by Parrinello-Rahman barostat. I performed four MD simulations in 400 K and one simulation in 500 K, each of them with the duration of 50 ns and the time step $\Delta t = 2$ fs. After the MD simulation, the Principal Component Analysis (PCA) [289, 290] has been performed to extract the variables, which describe the system optimally and hence can be used as a collective variable for the metadynamics run.

4.1.2 Molecular Dynamics simulation of 1KR8 DNA oligonucleotide in the presence of urea

In order to approach the details of the interaction of urea with the DNA oligonucleotide depending on its conformation, two systems have been constructed: a native 1KR8 DNA form, obtaind directly from the RCSB Protein Data Bank [122], immersed into a cubic water box of the size 6.25 nm x 6.26 nm x 6.26 nm, and an unfolded DNA structure extracted from the high temperature MD trajectory and immersed into a cubic water box of the size 7.59 nm x 7.59 nm x 7.59 nm. The net charge of both systems was neutralized with the addition of 6 Na^+ ions.

Urea is known to be a strong protein denaturant [47]. Although in laboratories urea can be used in the concentrations as high as 10 M, such concentrations are reaching far beyond the biological level, which we are interested in. Moreover, my previous simulations of DNA in the presence of urea (unpublished) have shown that the system becomes unstable for urea concentrations exceeding 4 M. Therefore in this study the systems are restricted to the lower concentrations of urea.

For both native and unfolded DNA forms, I prepared a simulation box containing 0 M, 1 M, 2 M and 3 M urea. Different sizes of the simulation boxes were applied to preserve the same urea concentrations for both DNA conformations. Prior to the MD simulation, the energy of all the systems was minimized using the Steepest Descent algorithm and the systems were equilibrated in the NVT ensemble (for the parameters of the equilibration, see the subsection 4.1.1). The 200 ns MD simulations (time step $\Delta t = 2$ fs) in explicit solvent have been performed in Amber99SB-ParmBSC0 force field [227] in the temperature 300 K, with TIP3P water model [287] and position restraints applied on DNA and 3D periodic boundary conditions. The PME method [288, 280, 281] has been applied for the treatment of long-range electrostatics, with the 0.9 nm cut-off for the short-range van der Waals and electrostatic interactions. All bonds have been constrained with the LINCS algorithm [282]. During the MD run, the temperature of the system was controlled by the Nose-Hoover thermostat [286, 244, 245].

The calculations of the Kirkwood-Buff integrals and the derived parameters along with the preferential interaction parameters and the local/bulk partitioning coefficients have been performed using a self-written code. The calculations of RDF and CN-RDF between DNA–urea and DNA–

water have been performed for the whole DNA molecule as well as for each particular nucleotide separately. The calculations of water orientation parameters were performed with the use of the Gromacs 4.6.2 software package [275].

4.1.3 Reconstructing free energy surface of 1KR8 DNA in the presence of urea

The reconstruction of the free energy landscape of 1KR8 DNA oligonucleotide in the presence of 0-3 M urea in Amber99SB-ParmBSC0 force field [227] has been performed in the temperature 300 K and with the application of TIP3P water model [287]. The native input DNA structure has been obtained from the RCSB Protein Data Bank [122]. Prior to the metadynamics run, the system has been solvated in a water cubic box with the dimensions 6.26 nm x 6.26 nm x 6.26 nm and its energy has been minimized with the use of Steepest Descent algorithm. Each of the systems was equilibrated for 2 ns (time step Δt = 2 fs) in the NVT ensemble, with position restraints put on DNA, and Nose-Hoover thermostat [286, 244, 245] to control the temperature.

The results of the Principal Component Analysis (PCA) show that the most appropriate reaction coordinate to describe the development of this system is end-to-end distance R_e between the center of mass of the terminal residues of the DNA oligonucleotide, 5'-G and 3'-C. Therefore the end-to-end distance has been chosen as the collective variable for the metadynamics run. Consequently, I performed 70 ns (time step Δt = 2 fs) metadynamics simulations of the DNA in the presence of 0-3 M urea with the use of Gromacs 5.0.4 [275, 276] and PLUMED 2.1.1 [291] software package without position restraints on DNA. The PME method has been applied for treatment of long-range electrostatics and 0.9 nm cut-off for short-range van der Waals and electrostatic interactions. All bonds have been constrained with the LINCS algorithm [282]. The 300 K temperature of the system was kept constant and controlled by Nose-Hoover thermostat [286, 244, 245]. The bias potential V_B with the Gaussian height h of 0.024 kcal/mol and Gaussian width σ of 0.05 nm has been deposited to the system along the CV, which was end-to-end distance between the center of mass of terminal residues. The results of free energy calculations have been analyzed using self-written code.

4.2 Results and discussion

4.2.1 Radial distribution functions

I evaluated the radial distribution function (RDF) of water and urea molecules around the inserted DNA oligonucleotide. The resulted RDF between DNA and urea is shown in Fig. 12 and RDF between DNA and water in Fig. 13.

The analysis of the radial distribution function shows that the local density of urea in a close vicinity of DNA surface is slightly higher around the unfolded than the native DNA form (see: Fig. 12). However, even though urea exhibits a slight binding preference to the unfolded DNA, this effect is not very pronounced. It can be noticed that increasing the concentration of urea results in subtly stronger repulsion of the osmolyte from DNA surface. This concentration-dependent phenomenon is more pronounced for the DNA in native form. The RDF of water molecules around DNA indicates no presence of clearly distinguishable solvation shells (Fig. 13). It can be noted that higher concentrations of urea result in slightly lower accumulation of water molecules in the close vicinity of DNA surface, both for unfolded and native DNA form. This suggests that growing urea concentration leads to a minor repulsion of water from the surface of the biomolecule. However, this effect is not very remarkable.

In the work of Nordstrom *et al.* [292] it is suggested that the accumulation of urea around the DNA surface depends on the nucleotide composition. Therefore I calculated RDF between DNA and urea for each particular nucleotide. The results are presented in Fig. 14.

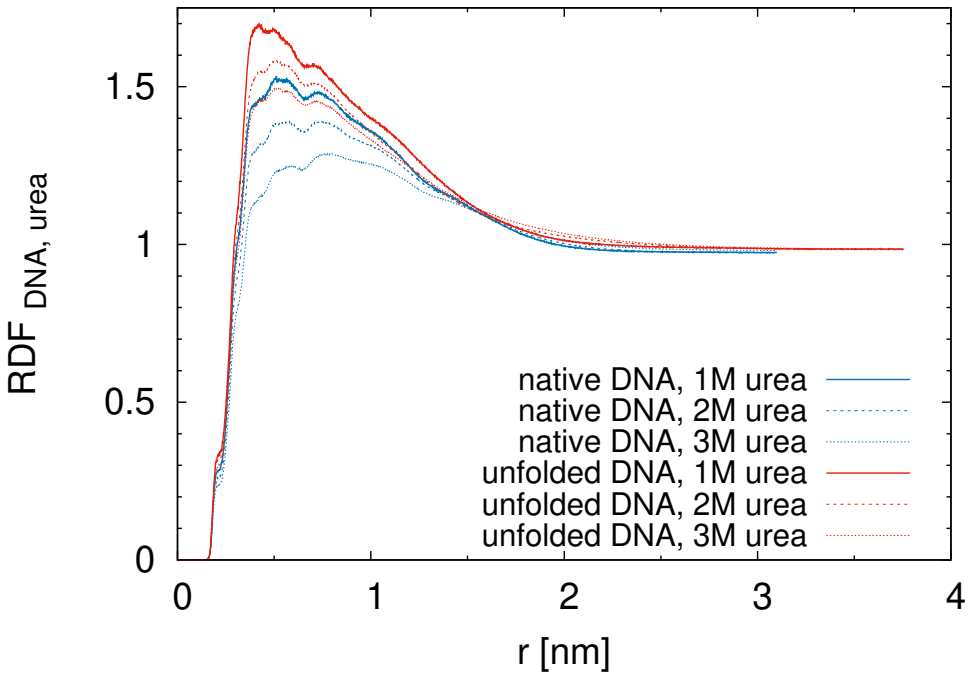


Figure 12: Radial distribution function of urea molecules around the native and the unfolded form of the DNA hairpin. Local density of urea in a close distance to the DNA surface is slightly higher for the unfolded than for the native DNA form.

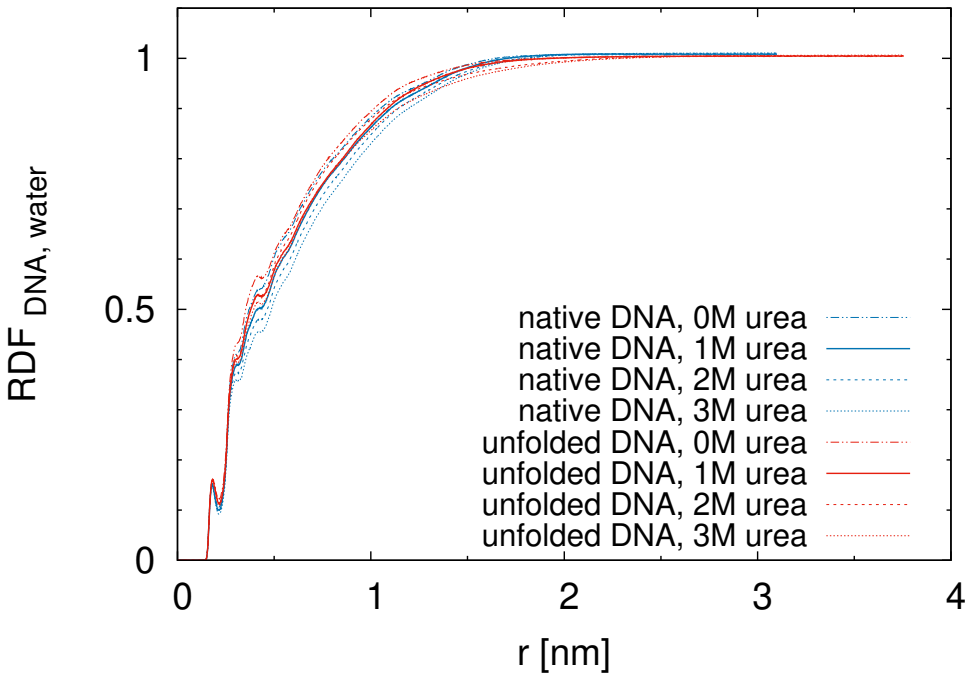


Figure 13: Radial distribution function of water molecules around the native and the unfolded form of the DNA hairpin. The distribution exhibits the lack of distinct water shells. The distribution of water around the DNA molecule appears not to be influenced by growing urea concentration, which serves as the counter-evidence for the indirect mechanism of urea–DNA interaction.

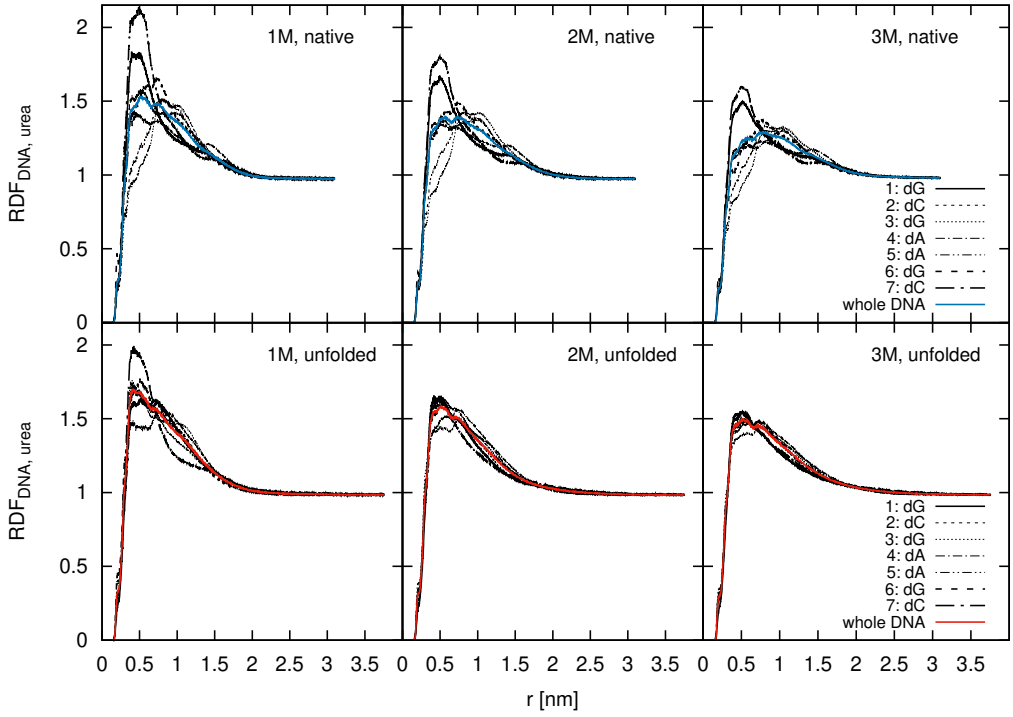


Figure 14: Radial distribution function of urea around 1KR8 DNA, defined for each nucleic residue. The solid blue and red line correspond to the RDF DNA–urea calculated for the whole 1KR8 oligonucleotide in the native and ufolded form respectively. Reprinted with permission from [3].

These findings do not correlate with the experimental findings of Nordstrom *et al.* [292], who proposed that urea shows stronger preference for adenine (A) and thymine (T) bases at the single stranded DNA surface exposed as a consequence of melting the double helix. The calculations of the radial distribution function (RDF) and cumulative radial distribution function (CN-RDF) between DNA and urea for each particular DNA residue for the extended unfolded form of the analyzed oligonucleotide (see: Fig. 14, lower panel) shows no apparent differences, which suggests that urea exhibits no preferences in binding towards particular nucleobases. The calculations of water order parameters for each DNA residues confirm these findings. The observation is quite different if one considers radial distribution function around particular DNA residues for the native folded DNA form. Here, the accumulation of urea molecules is apparently higher around the first (5'-dG) and the last (dC-3') nucleotide for 1-3 M urea. However, in case of the hairpin-formed structures, such behavior can be explained in terms of steric hindrance. The terminal 5'-dG and dC-3' residues forming the stem are exposed to the environment more than the d(GAA), which are forming the loop and hence buried inside the structure. In this regard, preferential accumulation of urea around terminal G and C nucleotides can be attributed to the greater availability of those residues for the interaction with urea. These results show that the nature of urea binding is unspecific and independent on the nucleotide composition. These findings support the thermodynamic approach in terms of the KB theory, such that the average and general binding behavior of urea can be studied regardless of individual contributions.

4.2.2 Preferential interaction coefficients

The preferential binding coefficient $\nu_{23}(r)$ has been calculated for each urea concentration around the native and unfolded DNA structure. The results are shown in Fig. 15.

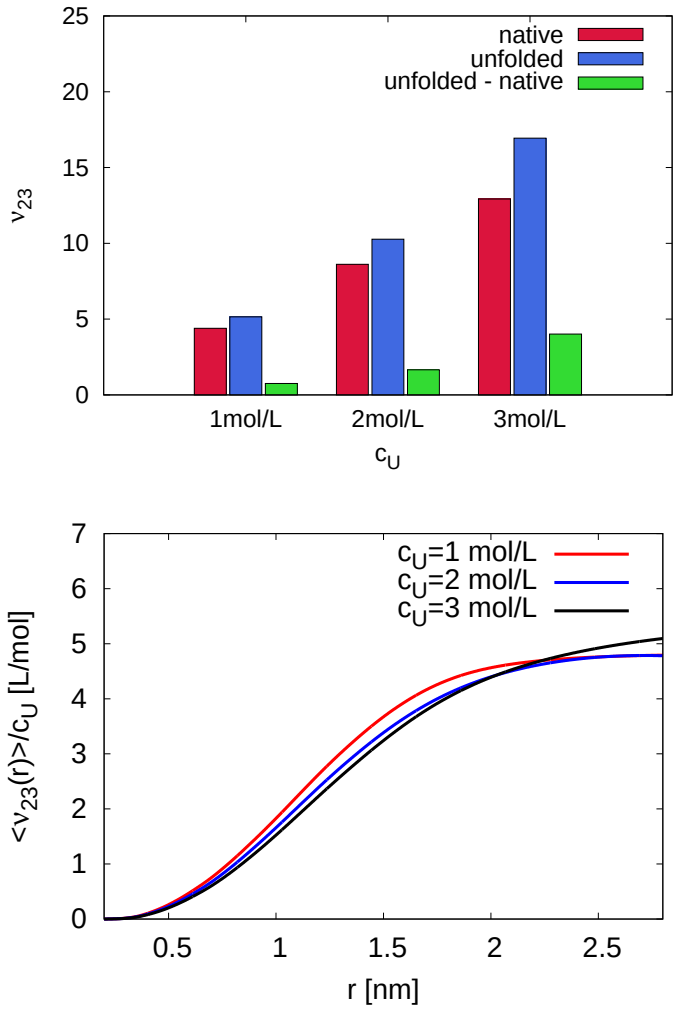


Figure 15: Top: Preferential binding coefficients ν_{23} for urea (center of mass) at different concentrations to both DNA conformations with phosphorus atoms as reference. The green bars express the differences of the preferential binding coefficients $\Delta\nu_{23}$. Bottom: Preferential binding coefficient $\langle\nu_{23}(r)/c_U\rangle$ averaged over both DNA conformations and divided by the actual urea concentration. Reprinted with permission from [3].

For all the concentrations of urea, it can be seen that the preferential binding coefficient takes positive values for both native and unfolded DNA form. This indicates that urea is preferentially bound to DNA. Since the values of $\nu_{23}(r)$ are higher for the unfolded than for native DNA conformation and $\Delta\nu_{23} > 0$, it can be suggested that urea induces DNA denaturation. According to Eqn. 20 one can thus conclude that there is a shift of the chemical equilibrium towards the unfolded state, which allows to interpret urea as a DNA structure denaturant.

The bottom panel of Fig. 15 shows the values of the preferential binding coefficient averaged over the native and unfolded DNA conformation $\langle\nu_{23}(r)\rangle/c_U$ and divided by the actual urea concentration c_U. For all urea concentrations, the preferential binding coefficients reveal positive values $\nu_{23}(r_c)/c_U \approx 5$ L/mol at $r > r_c = 2.5$ nm, which unequivocally indicates the strong and linear concentration-dependent binding behavior of urea to DNA structures, with a transfer free energy [69] defined as $G^* = -RT\nu_{23}$, where $\nu_{23} \approx \nu_{23}(r_c)$, around $G^*/c_U \approx -3$ kcal·L/mol^2. The observed constant decrease of the transfer free energy in terms of urea concentration can be explained by the nearly ideal behavior of urea in water [293]. In this regard, in the ideal ternary

solutions the values for the corresponding KBI differences $\nu_{23}(r)/\rho_3 = G_{21}(r) - G_{23}(r)$ are constant (See: Eqn. 17). Thus, also the differences in the excess volumes expressed by the KBIs, remain the same despite different urea molarities [87, 69]. This approach is valid for low and intermediate urea concentrations [293].

For low values of the preferential binding coefficient, $\nu_{23} < 1$, the effects of co-solvent association with the solute can be attributed nearly completely to the co-solute binding [210]. However, for higher values of ν_{23}, as observed in this study, the effects of water distribution take on significance with regard to possible dehydration behavior. To account for the solvent effects in more details, I investigated the preferential hydration parameter [43], defined by the Eqn. 48. In contrary to the preferential binding, the hydration coefficient gives the information about the exclusion of co-solvents from the surface of biomolecule. Hence, I evaluated the values for the preferential hydration parameter according to Eqn. 48. The obtained values, divided by the actual urea concentration, are $\Gamma_{21}/c_u \approx -275$ L/mol. This indicates clearly a strong dehydration behavior, implying that urea replaces water molecules in the local solvation shell around DNA. Since it is known that the helical DNA unwinds upon dehydration, up till the entire denaturation in severe cases [294], this result suggests a pronounced denaturation due to the dehydration of the oligonucleotide structures.

4.2.3 Local/Bulk Partitioning model for DNA - urea interaction

The effect of urea on DNA oligonucleotide stability has been interpreted according to the local/bulk partitioning model (LBPM) of Record *et al.* [61, 216] (Fig. 16). The definition of LBPM as reported by Record *et al.* [216] describes the solute (water) as the medium separating the distribution of the co-solvents between the bulk solution and the region located near the surface of a biomolecule, called the local domain.

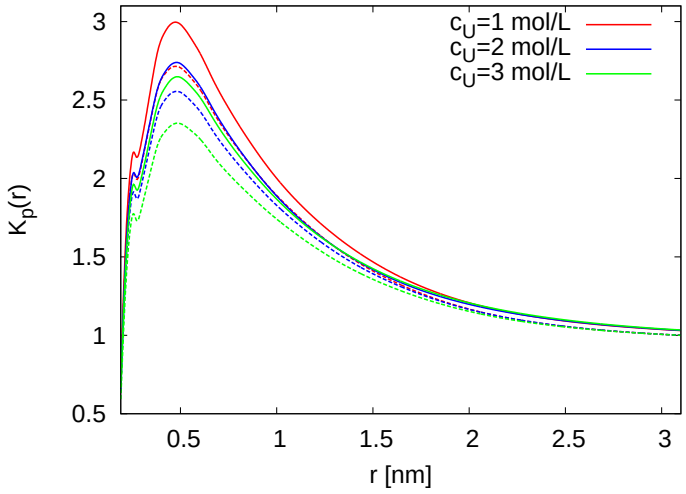


Figure 16: Local/bulk partition coefficients $K_p(r)$ for different urea concentrations (as defined by different colors in the legend) around the individual DNA conformations. Solid lines represent the values of $K_p(r)$ for the unfolded conformation, whereas dashed lines represent the corresponding results for the native DNA hairpin structure. Reprinted with permission from [3].

The analysis of local/bulk partition coefficient K_p (Fig. 16) indicates a strong accumulation of urea at short distances $r \leq 1.5$ nm around both native and unfolded oligonucleotide conformations, particularly for the lower urea concentrations. This concentration effect is more pronounced for the unfolded DNA form. This effect can be attributed to the fact that the local/bulk partition coefficient quantifies the amount of attracted urea molecules in reference to the total number of molecules in the solution. The maximum of the peak in Fig. 16 is located at $r_{\max} = 0.5$ nm, which clearly indicates a strong accumulation behavior of urea around the DNA. This accumulation effect is higher for the unfolded structures. At larger distances $r \geq 1.5$ nm, there is only negligible difference between the local urea concentration and the bulk concentration, presumably resulting

from the presence of coulombic and dispersion interactions between DNA and urea. Taking all these results into consideration one can assume that urea induces a denaturation impact on hairpin DNA, which is in agreement with Ref. [119].

4.2.4 Interaction energies and the properties of hydrogen bonds

In order to have a deeper insight into the interactions between DNA and urea in the studied system, I performed a detailed analysis of the energetic contributions to the solute–co-solvent interactions. To answer the question, whether DNA–urea attraction is driven dominantly by the electrostatic forces, or is it rather a purely Lennard-Jones process, I calculated the energetic contributions from the general Lennard-Jones short-range (LJ-SR) and electrostatic Coulomb short-range (Coulomb-SR) interactions to the total energy of the system. The results are shown in Fig. 17.

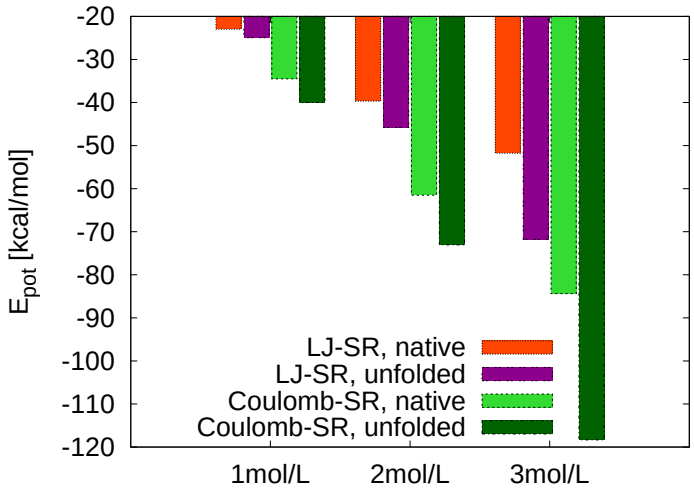


Figure 17: Lennard-Jones (LJ-SR) and electrostatic short-range (Coulomb-SR) energies between urea and DNA for both structures and for different urea concentrations. Reprinted with permission from [3].

It can be observed that all the energies take a negative value, which indicates that both LJ and Coulomb interactions between DNA conformations and urea affect the accumulation behavior. Negative Coulomb energies corresponding to the electrostatic attractive forces can be attributed to the highly negative charge of the DNA molecule, which is not to be neglected despite the overall electroneutrality of the system, and the molecular properties of urea. Since urea is bearing a pronounced static dipole moment [295], it can be considered as a highly polar molecule. This leads to asymmetric charge distribution over the individual atoms and consequently to the emergence of electron-poor and electron-rich groups of the molecule. This rationalizes the occurrence of strong electrostatic interactions between DNA and urea in presence of the highly charged oligonucleotide.

When both interaction energies are considered, their net values increase linearly with increasing urea concentrations, which points out the concentration-dependent binding energies of approximately $E_{\mathrm{LJ}}/c_U \approx -21$ kcal$\cdot$ L/mol^2 for Lennard-Jones and $E_{\mathrm{C}}/c_U \approx -37$ kcal$\cdot$ L/mol^2 for Coulomb interactions when averaged over both native and unfolded conformations. Hence, the electrostatic effects appear to dominate the conservative interactions between DNA and urea with 64% prevalence. This can be considered as a relatively high value for uncharged co-solute molecules. Moreover, the binding energy is more negative for unfolded than hairpin DNA regardless of urea concentration. This can be rationalized by the higher local urea concentrations around the unfolded state in comparison to the native conformation (see: Fig. 16). Taking into consideration the values of preferential binding coefficients (see: Fig. 15) and their relation to the m-value, one can conclude that DNA unfolding is promoted predominantly by the enthalpic contributions in terms of conservative interactions, in agreement with previous studies [162]. Since not only electrostatic, but also the LJ energies contribute significantly to the binding of urea to DNA, it shows that both

energetic contributions favor a strong and unspecific binding of urea to DNA.

To make a distinction between the preferential attraction and preferential binding mechanism [57], I calculated the number of hydrogen bonds between DNA and water and between DNA and urea for both DNA conformations. The results are shown in Fig. 18.

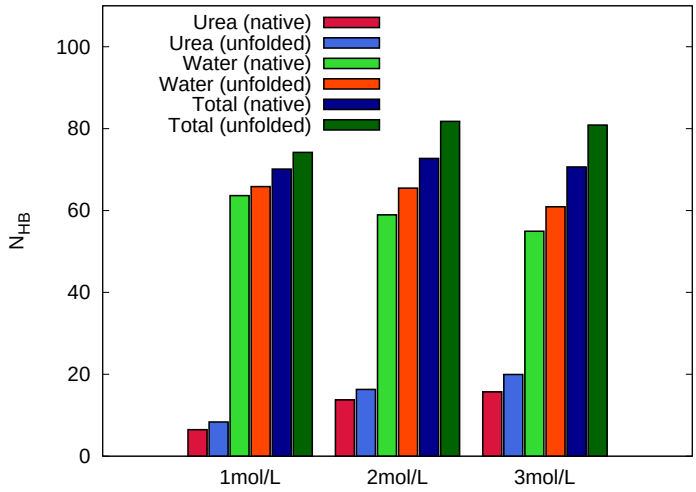


Figure 18: Number of hydrogen bonds between both conformations of DNA and urea or water, respectively, for different urea concentrations. The total number of hydrogen bonds is the sum of water–DNA and urea–DNA hydrogen bonds. Reprinted with permission from [3].

For both native and unfolded DNA conformations it can be observed that the number of DNA–urea hydrogen bonds increases with increasing urea concentration, whereas the number of DNA–water hydrogen bonds decreases. This finding stands in agreement with my analysis of the preferential hydration parameter Γ_{21}. Thus, it can be concluded that urea replaces water molecules in the first hydration shell around DNA. One can see that the total number of hydrogen bonds (water–DNA and urea–DNA) is higher for unfolded than for native DNA conformation. This suggests that the unfolded state is thermodynamically favored in agreement with $\Delta\nu_{23} > 0$. All these findings indicate that increasing concentrations of urea provoke the successive replacement of water molecules in the first solvation shell around DNA, which can be considered as a straightforward outcome of the preferential binding mechanism. Hence, one can conclude that the interaction of DNA with urea occurs rather via a direct mechanism than due to solvent-mediated effects [57].

In Fig. 19, there is presented the number of urea–DNA hydrogen bonds N_{U} divided by the total number of hydrogen bonds N_{HB} in the analyzed systems. The results show a remarkable increase of urea–DNA hydrogen bonds with increasing urea molarity, with the maximum value observed for $c_U = 3$ mol/L. For this concentration of urea nearly 25% of all hydrogen bonds are represented by urea–DNA H-bonds, which indicates that the hydrogen bonds between DNA and water loose in their importance and urea-related hydrogen bonds gain in their contribution when the molarity of urea increases. This stands in agreement with a preferential binding mechanism.

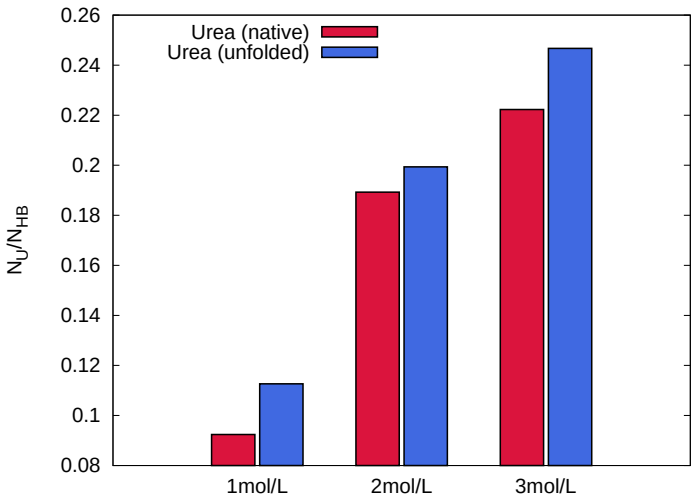


Figure 19: Number of DNA–urea hydrogen bonds N_U divided by the total number of hydrogen bonds N_{HB} with regard to different DNA conformations and increasing urea concentrations. Reprinted with permission from [3].

4.2.5 Solvent orientation parameters

The addition of co-solutes can influence the environment of the biomolecule. Osmolytes such as urea can perturb the water structure near the DNA surface, leading to the alterations in the intramolecular interactions as well as the interactions with the other components of the system [140, 58]. As a result, one can observe different entropic contributions to the solvation free energy, which are typically attributed to particular angular water conformations [140]. In order to get insight into the angular distribution of water molecules around native and unfolded DNA conformations, I calculated water orientation parameters in pure DNA solution and in the presence of 1-3 M urea according to the equations Eqns. 49 and 50. The results for f_1 are shown in Fig. 20 and for f_2 in Fig. 21.

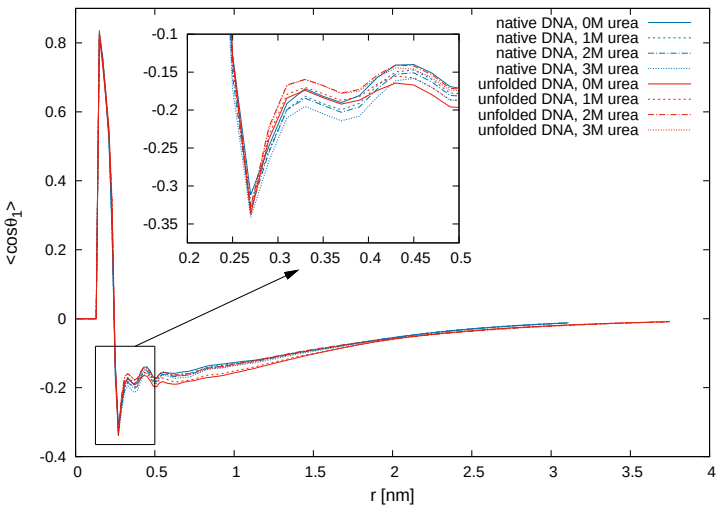


Figure 20: Water order parameter f_1 according to Eqn. 49 for water molecules around DNA in presence of varying urea concentrations as denoted in the legend. The inset represents the values for the corresponding distances as highlighted by the arrow. Reprinted with permission from [3].

The results shown in Fig. 20 demonstrate that the addition of urea exerts nearly no influence on the orientation of water shell at short length scales around DNA when compared to pure binary

solution of DNA and water. Slight deviations are observed only at length scales $r = 0.3$ - 1.5 nm, which corresponds to the peak values of the local/bulk partition coefficients (see: Fig. 16). This indicates that urea accumulation in close vicinity to DNA surface exerts only marginal influence on the orientation of water molecules, which highlights the ideal behavior of urea in aqueous solution [293].

The results for f_2 (see: Fig. 21) exhibit comparable features.

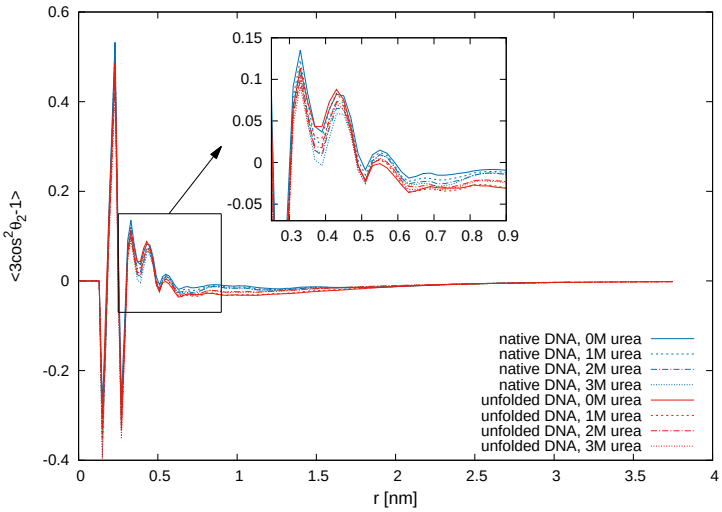


Figure 21: Water order parameter f_2 according to Eqn. 50 for water molecules around DNA in presence of varying urea concentrations as denoted in the legend. The inset represents the values for the corresponding distances as highlighted by the arrow. Reprinted with permission from [3].

The influence of urea on the water orientation parameters is only marginal, with slight deviations to DNA in pure water observed for $r > 0.45$. This corresponds to water molecules in the second and higher order hydration shells. Thus, it can be concluded that even the concentrations of urea as high as $c_U = 3$ mol/L affect the local order of hydrating water molecules only slightly.

The above analyses lead to the question whether urea binding happens according to direct or indirect mechanism. According to Zhang *et al.* [49], indirect effects can be attributed to the influence the osmolyte exerts on the structure of the bulk water or the solvation shells near the biomolecule. As the obtained results show (see: Fig. 13, 20 and 21), water structure around DNA oligonucleotide seems to be undisturbed by the addition of co-solute. Therefore, binding of urea could be ascribed presumably to the direct mechanism, potentially involving the formation of hydrogen bonds between osmolyte and DNA molecules as it happens for proteins, rather than to the indirect hydrophobic mechanism.

4.2.6 Free energy landscapes

The reconstruction of the free energy landscape of the native form of 1KR8 DNA hairpin in the presence of 0-3 M urea has been performed with the use of metadynamics method. Having identified the energetic minima, each of them has been associated with the corresponding DNA structure of relative stability. Fig. 22 shows the graphs of the free energy landscapes of 1KR8 DNA hairpin at various urea concentrations.

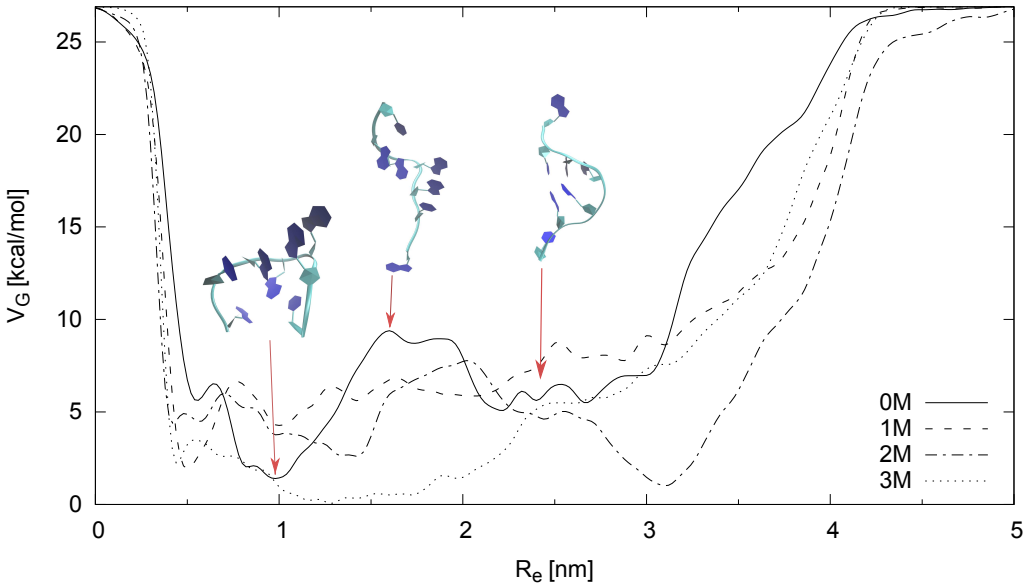


Figure 22: Free energy landscapes of 1KR8 DNA hairpin alone and in the presence of 1 M, 2 M and 3 M urea. Snapshots represent the most energetically favorable and hence the most stable DNA conformations in pure water solution, as well as the structure corresponding to the intermediate state between those two conformations of relative stability.

Fig. 22 presents the snapshots of DNA structures in a pure water solution. Two terminal conformations correspond to the most stable and hence energetically most preferred ones. The middle conformation represents an intermediate state. Both stable conformations are separated by a relative high free energy barrier of ca. 7.5 kcal/mol. As it can be observed on the subsequent graphs, the addition of growing concentrations of urea leads to slight lowering of the energy barrier between two conformational states of relative stability. This effect is more pronounced for higher urea concentrations. For 3 M urea, the energy barrier does not exist anymore and the energy minimum corresponds to the native folded state. According to my previous findings from the unbiased MD simulations, urea is favoring the unfolded conformation, which is especially visible for the higher co-solute concentrations. These findings are further confirmed by the metadynamics analysis. For 3 M urea, escaping the free energy minimum of the native state and shifting to the unfolded conformation occurs practically without any energetic hindrance. Even for lower urea concentrations (1 M and 2 M), although some residual energetic barrier exists, its height is significantly lower than for the systems without an osmolyte. Therefore, the results of the metadynamics calculations give a qualitative confirmation of the previous observations on the nature of DNA–urea interaction.

5 Binding properties of ectoine to small DNA hairpin and canonical DNA helix

Ectoine has been commonly known as an agent stabilizing protein structure [72, 193, 192] by direct and indirect mechanism [73], involving the repelling from uncharged surfaces [148] and hydration alteration [73, 55]. However, due to the significant structural difference between typical proteins and nucleic acids conformations, with the latter bearing a pronounced negative charge at the backbone, it has been speculated whether the same interaction mechanisms apply also for DNA. Recent experimental findings by Meyer *et al.* demonstrate that a prolonged DNA incubation with ectoine in a cell- and ion-free aqueous solution results in much more pronounced structural changes to DNA as compared to incubation in pure water [197]. The authors point out that ectoine, even though commonly considered as protectant also for DNA eg. against UV-related damages [194], actually enhances strand breaks in DNA in the experimental setting. These results suggest that ectoine preferential exclusion from positive or neutral surfaces, which is responsible for its stabilizing properties, may in fact not completely apply to negetively charged surfaces. It has been postulated that ectoine-derivative hydroxyectoine binds preferentially to negatively charged biomolecular spheres [114], and it is known that DNA holds a prononuced negative charge resulting from the phosphate backbone. Thus, there is a need to elucidate the molecular details of ectoine interactions with DNA in more details.

In this study, I investigate a small 7-base pair (bp) DNA hairpin [118] (PDB: 1KR8 [122]) in its native and unfolded form, as well as 24-bp duplex B-DNA structure [296] (PDB: 1D29 [297]) in the presence of various ectoine concentrations up to its solubility limits in water [298, 299] via atomistic molecular dynamics (MD) simulations [4]. Beside a model double helix of regular stability defind by 50% GC content, which is a typical Watson-Crick DNA form found in chromosomal DNA [93], I consider also the atypical DNA hairpin structure composed of two Watson-Crick d(GC) base pairs forming the stem and the triplet d(GAA) composing the loop. Small DNA hairpins with the general sequence d(GCGNAGC), where N corresponds to any of the nucleobases A, G, C or T, were recently applied as model systems in experimental and computational studies [117, 3] due to their extraordinary stability manifested by high melting temperatures and resistance against nucleases [116, 111]. Especially the fragment d(GCGAAGC) appears to be the most thermodynamically stable [109, 117] as a result of the complex interplay between aromatic base pair stacking and intramolecular hydrogen bonding [118].

This research focuses explicitly on the thermodynamic aspects associated with DNA–ectoine interactions, in terms of the binding mechanism and the analysis of enthalpic and entropic contributions. The preferential binding model has been applied with regard to molecular theory of solutions, so that any dynamic behavior in terms of unfolding pathways and the occurrence of metastable conformations can be disregarded. The obtained results reveal a strong accumulation of ectoine in close vicinity of all DNA structures, with higher deposition being observed around more negatively charged B-DNA form. With reference to the corresponding findings it can be assumed that ectoine exerts a destabilizing effect on double-stranded DNA (dsDNA) in good agreement with last experimental results [198, 197]. The analysis of energetic contributions reveals that the interaction between DNA and ectoine is dominated by electrostatic and dispersion energies, with prevalence of either of them depending on ectoine concentration and DNA form. Although *in vivo* conditions often assume the presence of excess salt, the studied systems are approximated with excess salt-free solutions due to the recent results showing that ectoine retains its properties in saline environment [115, 148, 56].

5.1 Methodology

The input native DNA hairpin structure (PDB ID: 1KR8) [118, 122] with the sequence 5'-GCGAAGC-3' consists of two G-C Watson-Crick base pairs forming the stem, and the GAA unpaired nucleotides composing the loop. The most stable unfolded form of 1KR8 oligonucleotide has been constructed according to the protocol described in Ref. [3]. In this study, I performed molecular dynamics (MD) simulations of 5 initial native and uncorrelated and 5 initial unfolded and uncorrelated conformations of 1KR8 DNA obtained from the position-restrained simulations in pure water as described in Ref. [3]. Each of the states was simulated for 150 ns. The use of distinct conformations

enables to avoid the problem of conformation-dependent sampling effects due to the averaging of the results over the corresponding position-restrained structures. In this respect, transitions into metastable DNA states can be avoided. Moreover, this protocol allows the proper sampling of the ectoine distribution around the folded and unfolded DNA state, where each state including 5 conformations was in total simulated for 750 ns. The representative snapshot of 1KR8 DNA is shown in Fig. 1b, and the classical B-DNA in Fig. 1a. The double-stranded B-DNA dodecamer (PDB ID: 1D29) [296, 297] with the sequence 5'-CGTGAATTCACG-3' is composed of 12 canonical nucleic base pairs. The snapshots of the native and unfolded hairpin conformation in the presence of 1 M ectoine is shown in Fig. 23, and the simulated B-DNA form in 1 M ectoine solution in Fig. 24.

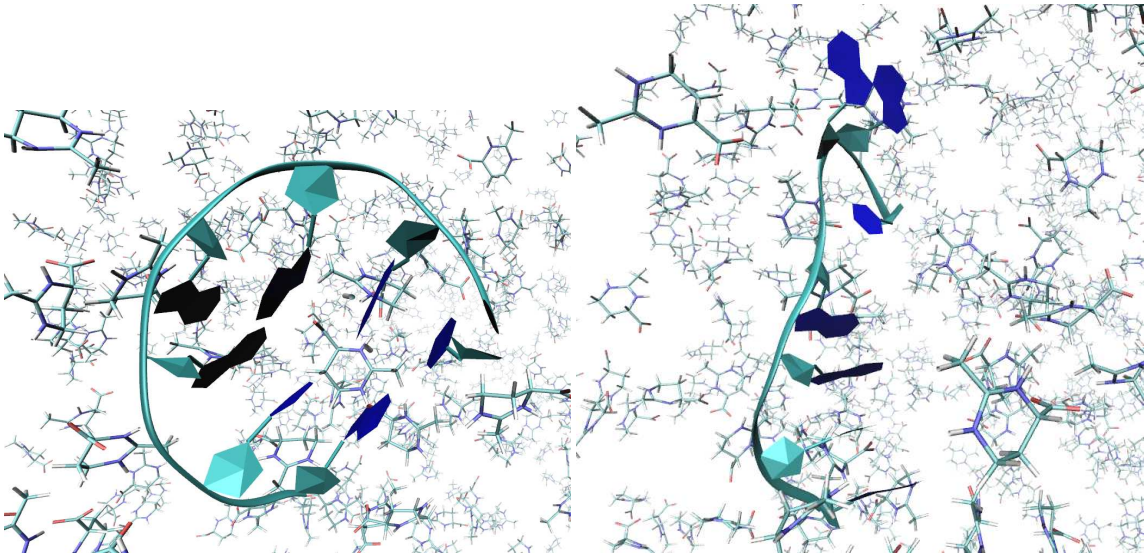


Figure 23: Snapshots of the native (left) and unfolded (right) DNA hairpin forms in the presence of 1 M ectoine concentration.

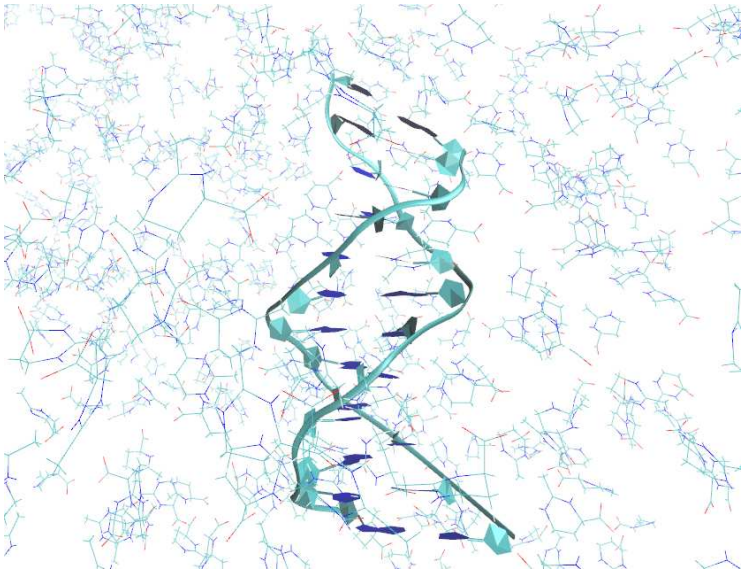


Figure 24: Snapshot of the B-DNA structure in presence of 1 M ectoine concentration.

I performed MD simulations of each of the DNA conformations in the presence of 0 M, 0.5 M, 1 M, 2 M and 3 M ectoine, which is below the ectoine solubility limit of 4 M [300]. It has been shown that in the aqueous solutions the zwitterionic ectoine form is 10.81 kcal/mol [115] more stable than the neutral one [115, 56, 72]. Therefore in this simulation setting the zwitterionic ectoine form has been applied.

All atomistic MD simulations were performed with the use of Gromacs 4.6.2 software package [275]. Amber99SB-ParmBSC1 force field [228] with TIP3P water model [287] has been applied together with the Generalized Amber Force Field (GAFF) [301, 302] for ectoine. The ectoine force field has been derived with the use of AMBERTOOLS 14 software suite [303] in combination with ACPYPE [304] and refined according to the model force field presented in Ref. [115]. The choice of Amber99SB-ParmBSC1 is rationalized in more detail in Ref. [305], and GAFF models have been successfully applied for other co-solutes before [129, 57]. All the systems were immersed in a cubic water box fulfilling the condition of the minimum distance of 4 nm from the outermost DNA atom to the box edge, in order to provide suficient bulk phase. The Particle Mesh Ewald (PME) method [280, 281] has been employed to treat long-range electrostatics and the short-range cutoff of 0.9 nm for Lennard-Jones and Coulomb interactions. The LINCS algorithm [282] has been applied to constrain all covalent bonds. The DNA net charge was neutralized by the addition of 6 Na^+ ions in the simulations of all the 1KR8 oligonucleotide forms, and the addition of 22 Na^+ ions in the B-DNA simulations. Prior to MD run, the systems were energy minimized with the Steepest Descent algorithm. The 3D periodic boundary conditions and position restraints on heavy atoms of DNA molecule have been implemented. The constant temperature of 300 K has been preserved by the Nose-Hoover thermostat [244, 245]. The systems were then equilibrated for 3 ns (time step Δt = 2 fs) in the NVT ensemble, followed by 3 ns equilibration in the NpT ensemble. Constant pressure has been maintained by the Parrinello-Rahman barostat [247] with a reference pressure of 1 bar. The 150 ns (time step Δt = 2 fs) production run of the equilibrated systems has been performed in the NpT ensemble at temperature of 300 K, with position restraints on heavy DNA atoms. The large number of ectoine molecules and the large sizes of the simulation box in combination with long simulation times provides the conditions, where the Kirkwood-Buff integrals and related expressions can be reliably applied.

5.2 Results and discussion

5.2.1 Preferential binding and local/bulk partition coefficients

I calculatd radial distribution functions $g_{23}(r)$ between ectoine and short DNA oligonucleotide in its native and unfolded form, as well as between ectoine and B-DNA. As reference positions, the phosphorus atoms of the DNA backbone and the ectoine center-of-mass have been chosen. The averaged results for individual conformations are shown in Fig. 25.

The results show that ectoine at low concentrations of $c_{\mathrm{E}} = 1$ M experiences strong attraction to DNA at short and intermediate distances ($r \leq 2$ nm), especially towards the unfolded single-stranded DNA form, which is indicated by the values of $g_{23}(r) > 1$. The magnitude of ectoine accumulation in the vicinity of the unfolded 1KR8 DNA is comparable to that observed for B-DNA, indicating a comparable potential of mean force in terms of $\Delta F = -k_B T \ln g_{23}(r)$. This effect is especially pronounced for lower ectoine concentrations of 0.5 M and 1 M. For higher ectoine molarity, one can observe the saturation effects, which obstruct further accumulation of the co-solute close to the DNA surface. It can be observed that ectoine concentrations higher than $c_{\mathrm{E}} \geq 1.5$ M result in the values of $g_{23}(r) < 1$ at short distances $r \leq 1.2$ nm. This indicates the lower local ectoine density around DNA in comparison to the bulk phase. These findings can be attributed to the saturation effects around DNA, where the local ectoine solvation shell around DNA is completely occupied at specific concentrations. As a consequence, further additional or excess ectoine molecules will be repelled into bulk solution. Moreover, it can be noticed, that the attraction of ectoine at low molar concentrations towards B-DNA structure follows a two step process, which is manifested as a shoulder between $r = 1\,\mathrm{nm} - 3$ nm. In the first step, ectoine molecules at low concentrations are filling the first hydration shell within the distances of $r \leq 1$ nm, which corresponds to the first steep RDF peak. In the second step, at higher concentrations the second hydration shell between $r = 1\,\mathrm{nm} - 3$ is being filled. For higher ectoine molarity, with both hydration shells being fully occupied, no more ectoine molecules can be attracted directly to the DNA surface and hence they remain in the bulk solution. As a consequence, no further RDF peak is observed. The distribution of ectoine molecules around DNA looks quite different if one considers small 7-bp DNA conformations. Here one can observe only one broad single RDF peak between $r = 0.5$ nm -2 nm (top panel of Fig. 25), corresponding to the ectoines occupying the first hydration shell. The observed differences to the distribution of ectoine around B-DNA can be

rationalized by averaging over distinct 1KR8 conformations and their enhanced flexibilty in terms of varying solvent accessible surface areas. This prohibits the formation of spatially organized ectoine solvation shells. Despite structural deviations between single-stranded (ssDNA) hairpin and double-stranded (dsDNA) helix DNA, the conclusions concerning the attraction of ectoine are valid for all considered DNA conformations.

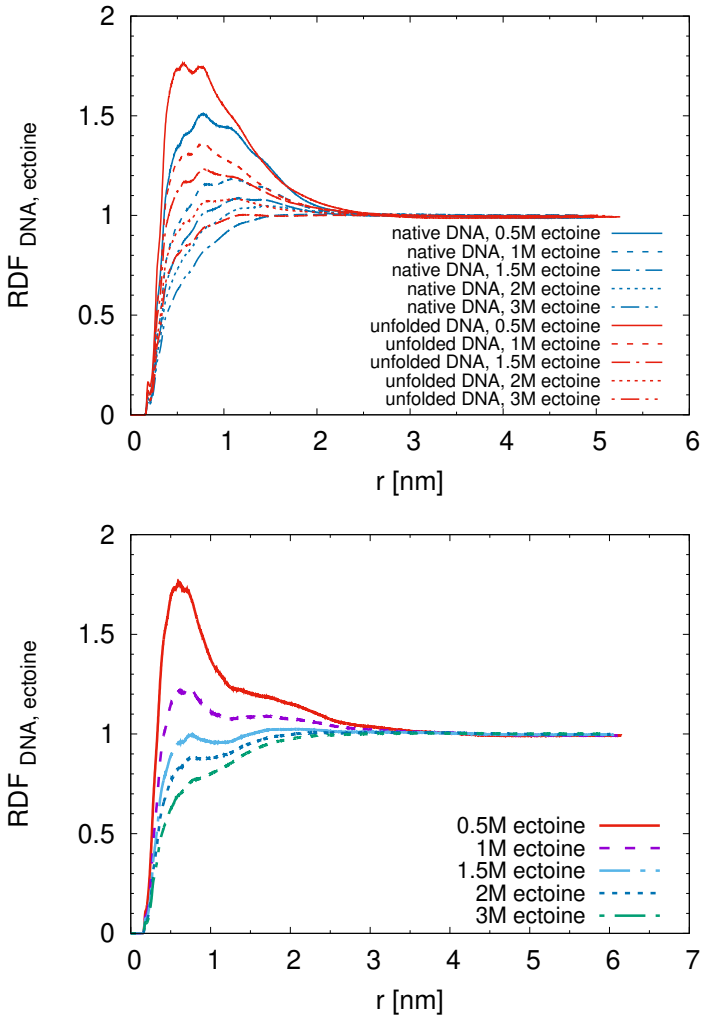


Figure 25: Top: Radial distribution functions (RDFs) of ectoine molecules at various concentrations around folded and unfolded forms of 1KR8 DNA as denoted in the legend. The RDFs around the native folded form are shown in blue color, while all RDFs around the unfolded states are labeled with red color. Bottom: RDFs of ectoine molecules at various concentrations around stable structure of double-stranded helical B-DNA as denoted in the legend. Phosphorus atoms of the DNA backbone and the center-of-mass of ectoine were chosen as reference positions for all RDFs. Reprinted with permission from [4].

To get a deeper insight into the preferential binding or exclusions mechanisms underlying the observed pattern of ectoine–DNA interactions, I calculated the preferential binding coefficients ν_{23} as defined by Eqn. 18 with $r \geq r_c = 4\,\text{nm}$. The results for all ectoine concentrations and all considered DNA structures are presented in Fig. 26. The large cutoff distance value has been applied according to the definitions given in Eqn. 11 and Eqn. 17, which involves the condition of converged KB integrals and preferential binding coefficients values. Thus, the value of $r_c = 4$ nm ensures the proper convergence of all the values of interest.

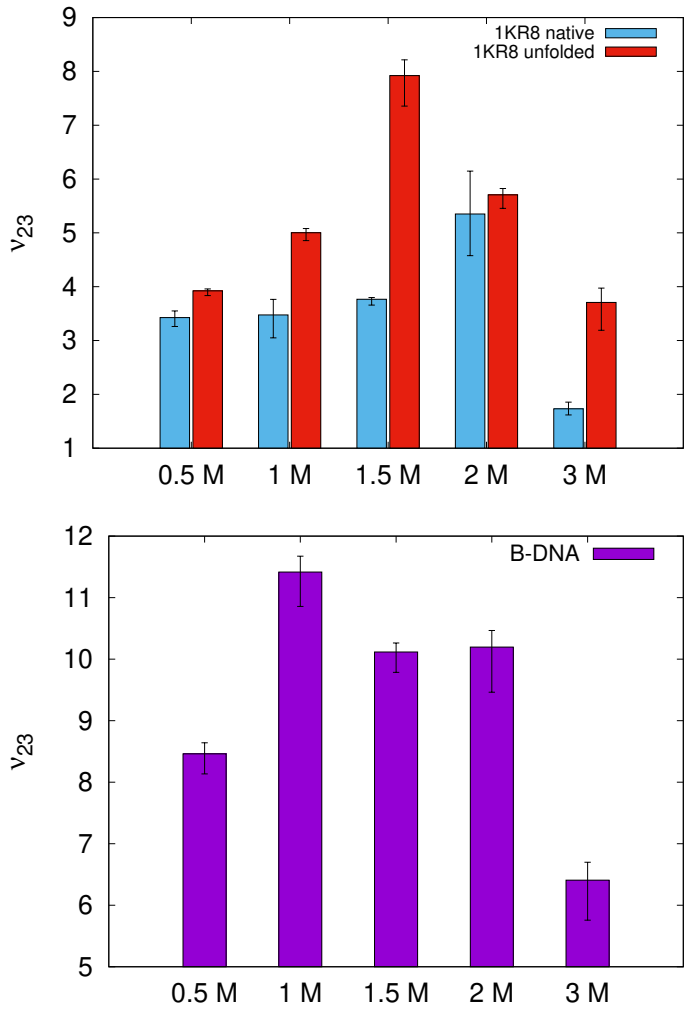


Figure 26: Top: Preferential binding coefficients ν_{23} for ectoine at different concentrations around 1KR8 DNA in its native and unfolded form as denoted in the legend. All values for the native 1KR8 DNA are marked with blue color, whereas the corresponding results for the unfolded 1KR8 DNA are colored in red. Bottom: Preferential binding coefficients ν_{23} for ectoine at different concentrations around the native structure of B-DNA. As reference positions, the phosphorus atoms of the DNA backbone and the center-of-mass of ectoine and water molecules have been chosen. The cut-off distance r_c for the evaluation of the finite length KBIs was chosen as $r_c = 4$ nm in order to ensure the presence of converged KBI values in bulk region around the DNA. Reprinted with permission from [4].

It can be seen that for all the DNA forms and all ectoine concentrations the preferential binding coefficients reveal positive values, which indicates that ectoine is preferentially bound to DNA regardless of its conformation. This finding is rather surprising for osmoprotectant agents like ectoine, which often exhibit the preferential exclusion behavior [53, 114]. On the contrary, destabilizers like urea are typically characterized by a preferential binding mechanism [69, 3]. With reference to the small 1KR8 DNA, one can observe higher values of ν_{23} around the unfolded when compared to the native conformation. As a result, the values of the differences of the preferential binding coefficients defined in Eqn. 20 show $\Delta\nu_{23} \gg 0$. This implies the surprising destabilization effect on negatively charged DNA conformations induced by ectoine, which is in good agreement with recent experimental findings [198, 197]. These results suggest the shift in chemical equilibrium towards the unfolded state according to the Eq. 20, which allows to interpret ectoine as the destabilizer of the negatively charged DNA. However, this conclusion may not apply to the neutral molecules. As it was shown experimentally [197], ectoine does not seem to exert a denaturing impact on DNA in pH 7.5, in which DNA charge can already be neutralized [197]. However, in

pH 6.6, enabling to preserve the negative DNA charge, ectoine turns out to destabilize DNA in agreement with my computational findings. Hence, it can be speculated that the osmoprotectant co-solute ectoine exhibits a comparable DNA binding behavior like urea, which is considered to be a strong denaturant [3].

Further hint on the destabilization effect of ectoine on DNA comes from the analysis of the local/bulk partition coefficients as defined in Eqn. 51. The results presented in Fig. 27 reveal a pronounced accumulation of ectoine at short distances of $r \leq 1.5$ nm around all DNA forms, with a maximum at $r \approx 0.5\,\text{nm}$ being located within the first solvation shell.

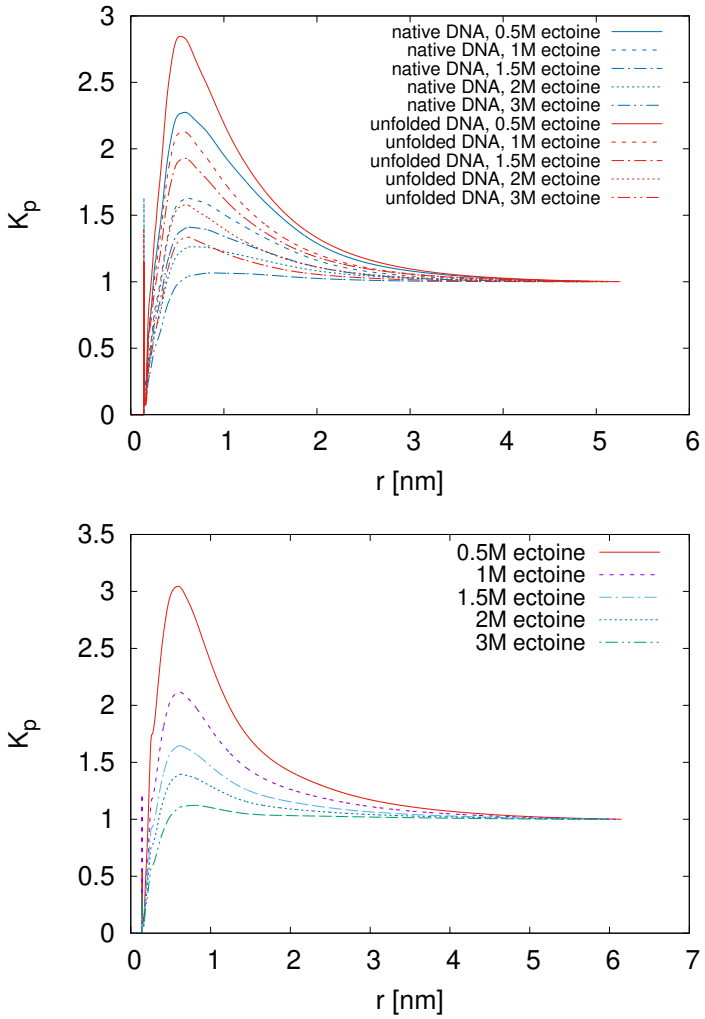


Figure 27: Top: Local/bulk partition coefficient $K_p(r)$ for different ectoine concentrations around native and unfolded 1KR8 DNA conformations as denoted in the legend. The values of $K_p(r)$ around the native folded form are shown in blue color, while all values for the unfolded states are labeled with red color. Bottom: Local/bulk partition coefficient $K_p(r)$ for different ectoine concentrations around the well-defined B-DNA form as denoted in the legend. As reference positions, the phosphorus atoms of the DNA backbone and the center-of-mass of ectoine and water molecules have been chosen. Reprinted with permission from [4].

This effect is especially pronounced for lower ectoine concentration as a consequence of the local/bulk partition coefficient quantifying the amount of attracted ectoine molecules with reference to the total number of ectoines in the bulk solution. The steep increase in $K_p(r)$ values within $r \leq 0.5$ nm also indicate the presence of water molecules around DNA. Hence, at these distances the RDF values show $g_{21}(r) > g_{23}(r)$. This shows that although ectoine directly binds to DNA, which is reflected in high $K_p(r)$ values at short distances, there is also a non-negligible amount of

DNA-hydrating water molecules. These molecules accompany the ectoine solvation shell to avoid 'dry' DNA structures. Regarding the hairpin DNA structures, it can be noticed that the values of $K_p(r)$ are higher for unfolded when compared to the native DNA form for particular ectoine concentrations, which additionally supports the assumption of the unfolding impact of ectoine on DNA [143]. Moreover, for each ectoine concentration the observed $K_p(r)$ values are higher for B-DNA as compared to DNA hairpins. This can be attributed to the larger solvent accessible surface area of 24-bp B-DNA, which leaves signifiantly more space for ectoine binding when compared to short 7-bp oligonucleotides.

5.2.2 Interaction energies and intramolecular hydrogen bonding

To get the insight into the energetic aspects of the interaction between ectoine molecules and DNA, I analyzed in details the energetic contributions to the binding mechanism, with particular emphasis on the electrostatic and dispersion interactions in terms of short-range Lennard-Jones (LJ) and Coulomb energies. The results for short 7-bp DNA oligonucleotides and canonical B-DNA helix are shown in Fig. 28.

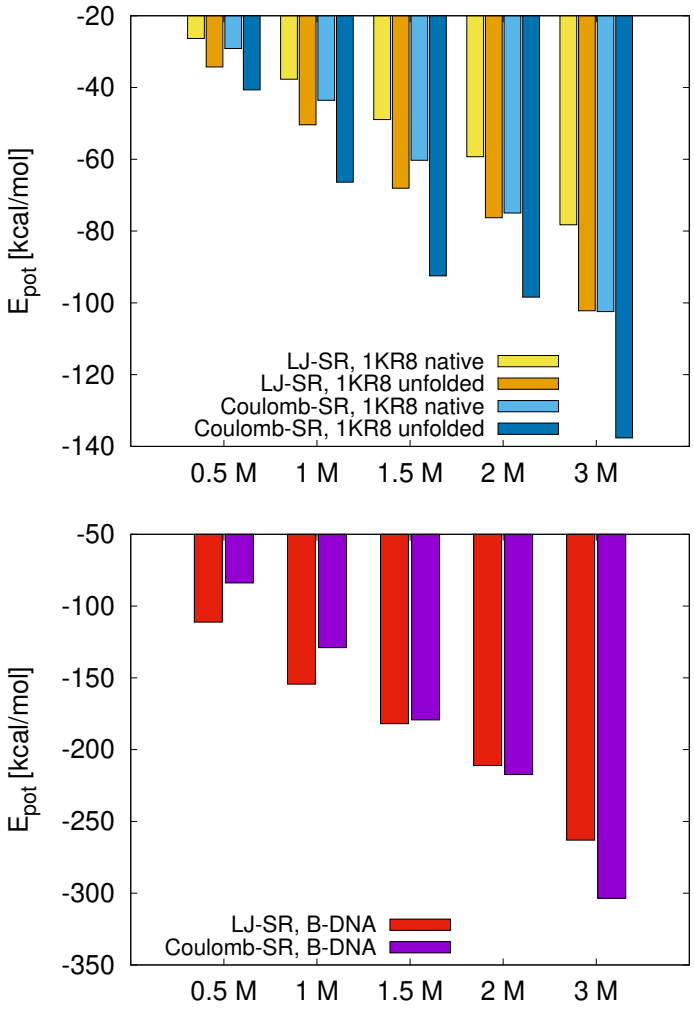


Figure 28: Top: Lennard-Jones (LJ-SR) and electrostatic short-range (Coulomb-SR) energies between ectoine at different concentrations and native and unfolded 1KR8 DNA form as denoted in the legend. Bottom: LJ-SR (red bars) and Coulomb-SR (violet bars) energies between ectoine at different concentrations and B-DNA form. Reprinted with permission from [4].

It can be noted that both Lennard-Jones and Coulomb energies take negative values regardless of ectoine concentration and DNA conformation. Thus, it can be concluded that both types

of DNA–ectoine conservative interactions strongly affect the ectoine accumulation characetristics. This effect can be attributed to the relatively large size of ectoine molecules, which can lead to strong dispersion interactions, and its zwitterionic form resulting in pronounced electrostatic interactions. Considering 1KR8 DNA oligonucleotides, one can see a nearly linear increase in the values of both interaction energies for growing ectoine concentrations, with generally more negative values observed for the unfolded conformation when compared to the native hairpin (Fig. 28). This can be associated with higher local ectoine concentrations around the unfolded DNA form than the native one in agreement with my previous results concerning preferenial binding and local/bulk partition coefficients.

The results obtained for 1KR8 DNA structures show that Coulomb short-range interactions dominate the overall DNA-ectoine interactions, although the contribution from Lennard-Jones short-range energies cannot be neglected. The prevalence of electrostatic energies can be rationalized by the molecular properties of DNA and ectoine themselves. As it was mentioned before, in aqueous solution ectoine takes a zwitterionic form, with carboxylic group bearing the negative charge and a partially mesomeric structure for the location of the positive charge in the pyrimidine ring. Since DNA itself has a pronounced negative charge, the occurence of strong electrostatic interactions between DNA phosphate backbone and ectoine is favored. However, when considering the B-DNA structure it can be observed that the contribution from dispersion interaction dominates the binding for lower ectoine concentrations up till 1.5 M (Fig. 28, bottom panel). This phenomenon can be explained by the nature of both electrostatic and LJ energies. Weak dispersion interactions are much more short-ranged that the strong electrostatic forces, thus being responsible for the binding at short distances within the first ectoine solvation shell around DNA. With increasing ectoine concentration and consequent filling of each particular solvation shell, the excess ectoine molecules are no longer in direct contact with DNA atoms and the distance between them and DNA increases in consequence of their spatial repulsion to larger distances. These repelled ectoine molecules exert stronger electrostatic forces at intermediate distances of $r \approx 1.0$ nm in comparison to LJ interactions. Thus, for higher ectoine concentrations the Coulomb energies dominate over LJ interactions. In contrast to B-DNA helix, the results averaged over different conformations of the more flexible short DNA oligonucleotides reveal that the formation of well-defined ectoine solvation shells is obstructed, so that the coulombic interactions always dominate over LJ ones. Taking all the above findings into account it can be concluded that both Lennard-Jones and Coulomb interactions contribute to strong and unspecific ectoine binding to DNA, which eventually favors DNA unfolding. Furthermore, taking into consideration the large negative values of combined LJ and Coulomb energies, it can be postulated that the binding of ectoine is enthalpy-driven, which stands in good agreement with previous findings for urea [3].

Co-solute effects can occur via indirect mechanism, where the co-solute molecules exert influence on the structure of bulk water or the solvation shells near the biomolecule [49], or by direct formation of the hydrogen bonds with the biomolecule. To approach this problem in more detail, I calculated the number of hydrogen bonds between particular DNA conformations and ectoine and between DNA and water for all ectoine concentrations. Thus, also a possible replacement of hydrating water molecules by an increasing concentration of ectoine molecules around the DNA can be studied. The results for all DNA conformations are shown in Fig. 29.

It can be observed that increasing ectoine concentrations are associated with increasing number of DNA–ectoine hydrogen bonds and consequent decreasing the DNA–water hydrogen bond number. It suggests that ectoine replaces water molecules at short distances to the DNA as a consequence of the preferential binding mechanism. Moreover, the unfolded 1KR8 DNA oligonucleotide attracts higher number of DNA–ectoine and DNA–water hydrogen bonds than the folded conformation, which is manifested by the total number of hydrogen bonds. This points out that the unfolded conformation is thermodynamically preferred. The notably larger size of ectoine molecules when compared to the molecules of water results in the decrease of the total number of hydrogen bonds with increasing concentrations of ectoine.

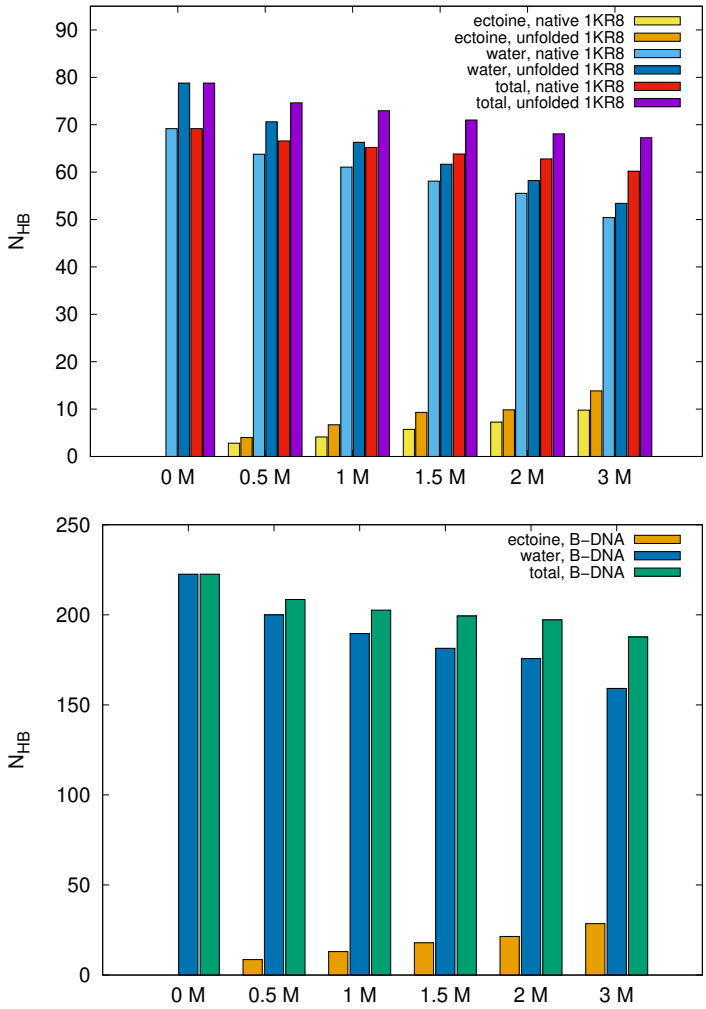


Figure 29: Top: Number of hydrogen bonds N_{HB} between 1KR8 DNA in native and unfolded form and ectoine or water for different ectoine concentrations as denoted in the legend. The sum of DNA–ectoine and DNA–water hydrogen bonds corresponds to the total number of hydrogen bonds. Bottom: Number of hydrogen bonds between B-DNA and ectoine or water for different ectoine concentrations. The sum of DNA–ectoine and DNA–water hydrogen bonds corresponds to the total number of hydrogen bonds. Reprinted with permission from [4].

Further confirmation of these results comes from the calculation of the ratio of ectoine–DNA hydrogen bonds N_{ectoine} to the total number of hydrogen bonds N_{total} formed by DNA (Fig. 30).

Here one can observe a strong linear increase of the contribution from DNA–ectoine hydrogen bonding to the total hydrogen bonds formed by the biomolecule with increasing ectoine concentration. These findings are in agreement with preferential binding model and with the previous outcomes for DNA-urea interactions [3], where the interactions between DNA and ectoine occur via a direct mechanism involving the formation of hydrogen bonds in contrast to solvent-mediated effects [49].

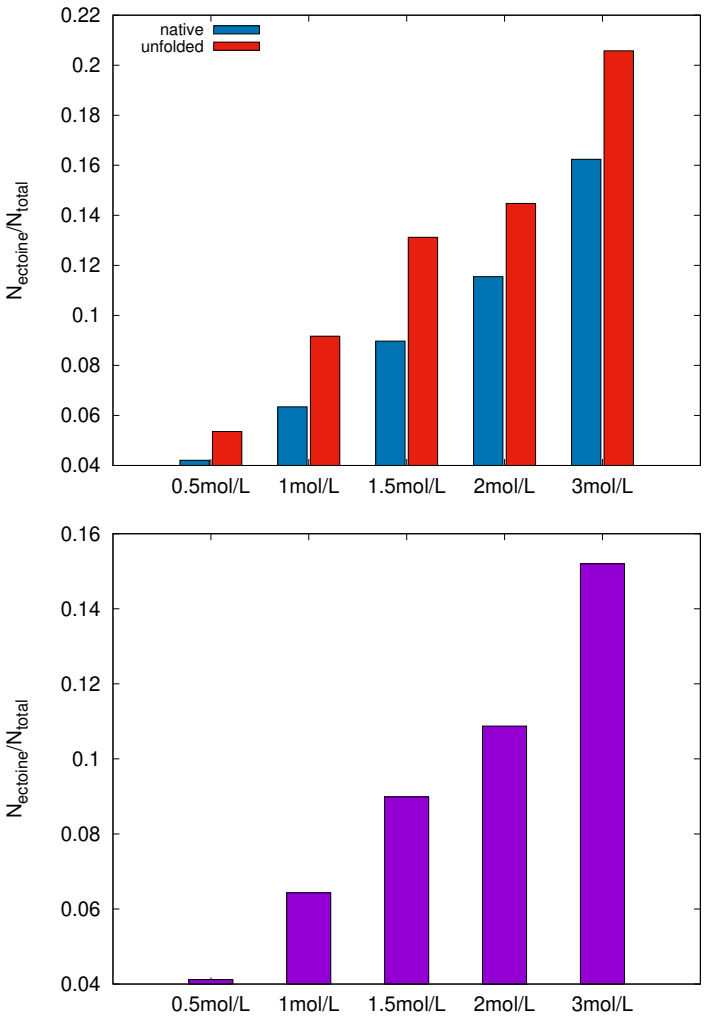


Figure 30: Top: Number of DNA–ectoine hydrogen bonds for native and unfolded 1KR8 DNA form N_{ectoine} divided by the total number of hydrogen bonds N_{total} for different ectoine concentrations. Bottom: Number of DNA–ectoine hydrogen bonds for B-DNA N_{ectoine} divided by the total number of hydrogen bonds N_{total} for different ectoine concentrations. Reprinted with permission from [4]

5.2.3 Solvent orientation parameters

The presence of ectoine around biomolecules like proteins has been often regarded as a factor influencing the structure of hydrating water [114, 55, 56], resulting in distinct entropic contributions to the solvation free energy [114]. To get a deeper insight into this phenomenon, I studied the angular distribution of water molecules around different DNA forms expressed as water orientation parameters defined by Eqn. 49 and 50. The results for f_1 for all ectoine concentrations as well as for the pure water solution are shown in Fig. 31 and 32, and for the f_2 in Fig. 33 and 34.

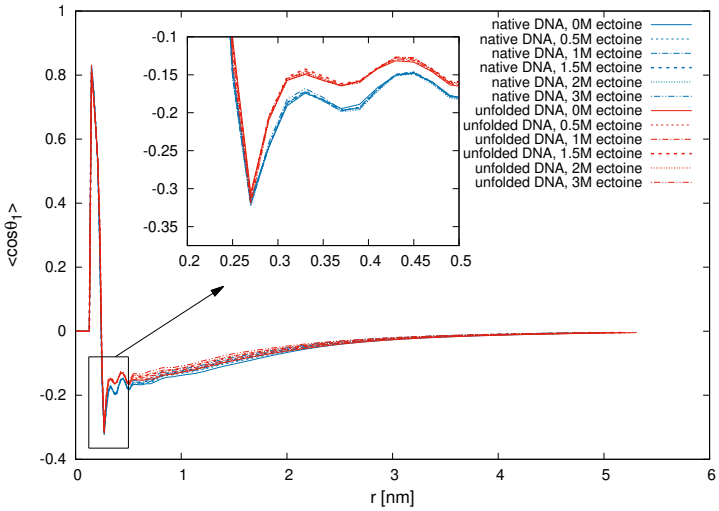


Figure 31: Water order parameter f_1 according to Eqn. 49 for water molecules around 1KR8 7-bp DNA oligonucleotide in its native and unfolded form for varying ectoine concentrations. The inset represents the close-up of the selected part of the graph as pointed by the arrow. Reprinted with permission from [4]

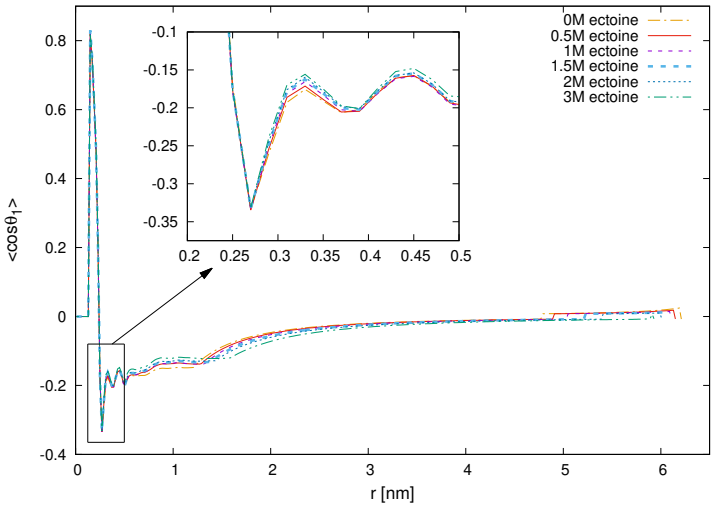


Figure 32: Water order parameter f_1 according to Eqn. 49 for water molecules around B-DNA structure for varying ectoine concentrations. The inset represents the close-up of the selected part of the graph as pointed by the arrow. Reprinted with permission from [4]

The presented results show that the orientation of water molecules at short distances around DNA remains almost unaffected in the presence of ectoine as compared to the pure water and DNA solution. Slight deviations can be observed only at the distances of 0.5-3 nm for B-DNA (Fig. 32) and 0.5-2.5 nm for 1KR8 oligonucleotides (Fig. 31), which corresponds to the higher order solvation shells. Since ectoine, which is preferentially bound within the first solvation shell around DNA, appears to exert only marginal influence on the orientation of water molecules, the indirect mechanism of DNA–ectoine interaction involving the pronounced changes in water structure can be with good approximation disregarded.

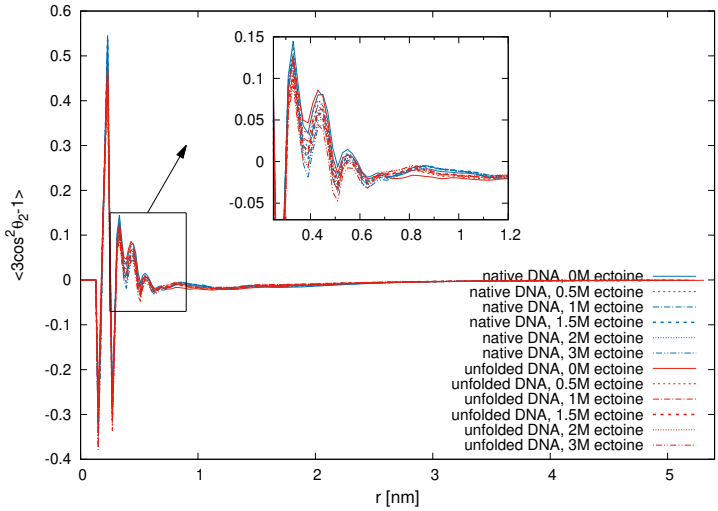


Figure 33: Water order parameter f_2 according to Eqn. 50 for water molecules around 1KR8 7-bp DNA oligonucleotide in its native and unfolded form for varying ectoine concentrations. The inset represents the close-up of the selected part of the graph as pointed by the arrow. Reprinted with permission from [4]

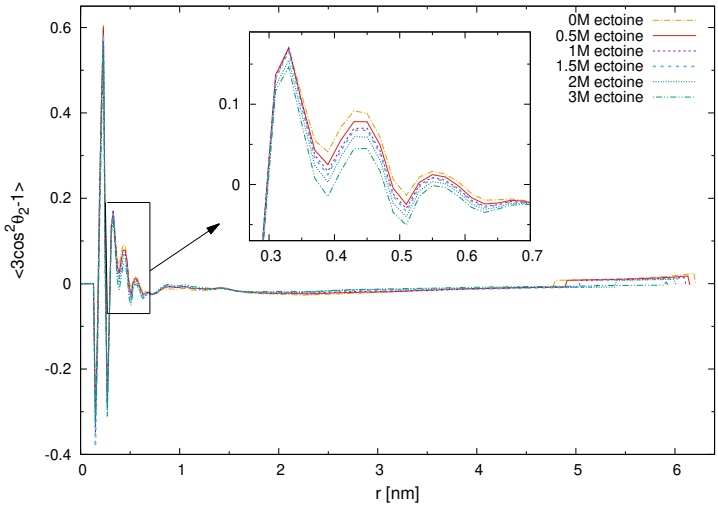


Figure 34: Water order parameter f_2 according to Eqn. 50 for water molecules around B-DNA structure for varying ectoine concentrations. The inset represents the close-up of the selected part of the graph as pointed by the arrow. Reprinted with permission from [4]

Additional support to the direct mechanism of DNA–ectoine interaction comes from the analysis of water order parameter f_2. Similarly to f_1, also here the influence of ectoine on local water structure is negligible regardless of ectoine concentration and concerns only the distances $r > 45$ nm, which are corresponding to the second and higher order hydration shells. Thus, the pronounced solvent-mediated effects can be ignored. Taking into consideration all the presented findings it can be concluded that DNA–ectoine binding occurs via direct mechanism involving the formation of hydrogen bonds and not by indirect altering the orientation of water molecules in the vicinity of biomolecular surface.

5.2.4 Change of DNA melting temperature estimated via m-values

The change of the DNA melting temperature can be approached by the calculation of the individual m-values according to Eqn. 35. Hence, the m-values have been estimated for single-stranded 1KR8 DNA for each ectoine concentration. The individual m-values have been obtained by combining the values of a_{33}, obtained from different ectoine–water simulations (without DNA), with the corresponding values of the differences in the ectoine preferential binding coefficients to folded and unfolded DNA conformations from the ectoine-DNA simulations. The calculation of the derivative of chemical activity a_{33} for ectoine in water, defined by the Eqn. 36, where G_{33} and G_{31} stand for the ectoine-ectoine (EE) and the ectoine-water (EW) Kirkwood-Buff integrals, respectively, allows to get an insight into the non-ideal effects of the simulated solution. Hence, the ideal mixture is characterized by $a_{EE} = 1$, whereas the values $a_{EE} \neq 1$ indicate the deviation from the ideal behavior as discussed in Ref. [213]. To estimate the activity coefficients, I performed 150 ns atomistic MD simulations for 0.5-3 M ectoine in pure water using the same simulation protocol and setup as described hereby in the "Methodology" section for 1KR8 DNA hairpins. The space occupied by DNA in hairpin simulations was replaced by water molecules in order to fill the free volume. The obtained values for a_{EE} are shown in Fig. 35.

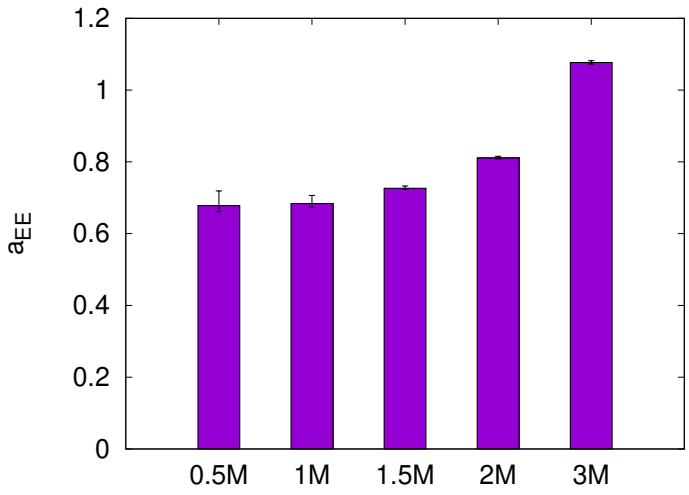


Figure 35: Derivative of the chemical activity a_{EE} in 0.5-3 M ectoine solutions. Reprinted with permission from [4]

The calculated a_{EE} values follow the rule $a_{EE} \neq 1$, which indicates the slightly non-ideal behavior of the simulated solutions. For lower ectoine molarity, the non-ideality of distribution is more pronounced. Since $a_{EE} < 1$, the Eqn. 36 yields $G_{EW} > G_{EE}$, which indicates that the hydration properties of zwitterionic ectoine molecules in the studied systems are well pronounced and that the tendency of ectoine molecules to form clusters is relatively weak.

The results for m-values, calculated according to Eqn. 35, are shown in Fig. 36. These findings show that the m_c-values differ for all concentrations in the range between $m = (0.1 - 2.0)$ L/mol. The non-monotonous variations of the m-values can be associated with the significant deviations in the differences in preferential binding coefficients $\Delta\nu_{23}$ [87] (see: Fig. 26). The values of the derivatives of the chemical activity a_{33} (a_{EE} in a binary solution) are confined in the range $0.65 < a_{\mathrm{EE}} < 1.10$, which points out a weak ectoine-ectoine accumulation behavior in aqueous solutions for low and intermediate concentrations [213, 92].

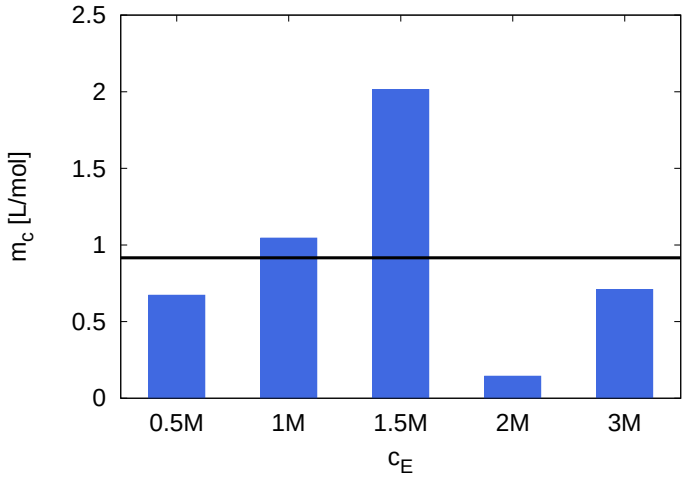


Figure 36: Corresponding m_c-values with regard to folded and unfolded 1KR8 7-bp DNA oligonucleotides for different concentrations of ectoine. The black line denotes the average m_c-value with $\langle m_c \rangle = (0.92 \pm 0.31)$ L/mol. Reprinted with permission from [4]

Since a linear influence of ectoine on the melting temperature can be assumed, one can consider a constant m_c-value by averaging over all concentrations, as explained in Ref. [92]. This assumption is rationalized by previous literature values for protein structure-stabilizing co-solutes like TMAO [46]. The average m_c-value shown in Fig. 36 for all concentrations yields $\langle m_c \rangle = 0.92 \pm 0.31$ L/mol and has a positive value, which indicates the decrease of the DNA melting temperature in presence of ectoine according to Eqn. 35. This finding stands in good agreement with previous experimental findings, where the presence of particular ectoine derivatives was associated with the lowering of the melting temperature of long dsDNA [198]. Moreover, the effect of ectoine on melting temperature has been shown to be higher for GC-rich nucleotides than for GC-poor ones. This correlates well with the properties of 1KR8 oligonucleotide, where the GC content amount to 71%. Since a proper evaluation of m-values requires low concentrations of the co-solute [87], the average of the m_c-values for the two lowest ectoine concentrations gives $\langle m_c \rangle_0 = 0.86 \pm 0.19$ L/mol. This value differs only slightly from the average value $\langle m_c \rangle$ over the full set of concentrations.

It has been shown in previous studies that most of the reported m- or m_c-values depend only slightly on the DNA conformation [3]. Hence, it can be assumed that comparable m- or m_c-values can also be obtained for other DNA forms, enabling the comparison of these findings with experimental outcomes. Recent experimental study reports the value of the entropy change from closed to open forms of a long DNA hairpin to be $\Delta S^0 = 0.17 \pm 0.01$ kJ/(mol· K) [119]. Taking this value as a rough estimate for the studied oligonucleotide system and its approximated value of $\langle m_c \rangle$, the melting temperature decrease can be evaluated as $\Delta T/K = -(13.5 \pm 4.5) \cdot c_E$ (L/mol). This value stands in agreement with previous experimental findings [198] for other dsDNA in distinct buffer, which yields $\Delta T/K \approx -4 \cdot c_E$ (L/mol). These findings provide an explanation for the significant decrease of the DNA melting temperature in presence of ectoine, suggesting a potential destabilizing effect of this co-solute on DNA structure. In summary, all presented findings indicate that in contrast to the general stabilizing impact of ectoine on proteins, for nucleic acid structures ectoine acts as a destabilizer comparable to the co-solute urea.

6 Binding properties of TMAO and urea to DNA G-quadruplex

In this study, I investigate the effect of zwitterionic TMAO and urea in the near-physiological concentrations of 1 M and 2 M accordingly, as well as the relevant 1:2 TMAO:urea mixture [155, 156], on the structural and dynamical characteristics of a human telomeric DNA G-quadruplex (PDB: 2KF7 [123, 306]) [1]. This type of G-quadruplex has been shown to exhibit a polymorphic nature [306, 18, 105], where its conformational state depends on the conditions in the solution. Under physiological conditions the basket form prevails, and upon the addition of TMAO, 2KF7 G-quadruplex acquires predominantly the antiparallel basket form.

In this research, a specific focus is set on the accumulation behavior of both TMAO and urea around the DNA, approached by the means of atomistic MD simulations. This allows to get the insights into the stabilizing or destabilizing mechanisms in terms of Kirkwood-Buff theory. In this chapter, I focus explicitly on thermodynamic aspects of DNA interaction with both co-solutes with regard to the binding mechanism and entropic and enthalpic contributions. Additionally, a theoretical approach allowing to study stabilization and destabilization effects of equilibrium conformations by means of a single arbitrarily chosen reference DNA state is introduced. Complementary to the previous studies on the interactions between single co-solutes like TMAO or urea and DNA [3, 4], the combined consideration of stabilizers and destabilizers via statistical thermodynamics sheds new light onto the underlying compensation mechanisms. Hereby presented findings highlight a destabilizing effect of urea and a stabilizing influence of TMAO in good agreement with previous experimental results. Furthermore, the compensating co-solute effects of TMAO and urea can be rationalized by the local arrangement of the species around the DNA molecule, and the cross-correlation effects between the co-solutes in the mixture can be almost entirely ignored. This points out the linear additivity of stabilizing and destabilizing interactions. These analyses also underpin the importance of water-related effects for the corresponding observations, which indicates that water-related effects like kosmotropic or chaotropic concepts are not applicable.

6.1 Methodology

The input DNA structure (PDB entry code: 2KF7 [307], structure obtained by NMR) is the human telomeric basket-type antiparallel G-quadruplex composed of two G-tetrads [306]. It consists of 22 nucleotides with the sequence 5'-GGGTTAGGGTTAGGGTTAGGGT-3'. The original structure included bromoguanine in the position DG7. For the purpose of this study, bromoguanine was replaced with guanine with the use of VMD (version 1.9.3) software [308, 309]. The tetrads are formed by the residues: dG1, dG8, dG20, dG14 (tetrad 1) and dG2, dG7, dG19, dG15 (tetrad 2), whereas deoxyguanosines dG3, dG9, dG13 and dG21 belong to the non-tetrad region. Despite relatively small size, this G-quadruplex containing only 2 tetrads has been found to be characterized by enhanced stability over its three-G-tetrad counterparts [306]. In particular, those structures exhibit extraordinary thermal stability [19, 20] related to the presence of stabilizing monovalent ion in the central channel of G-quadruplex. The exceptional stability of this DNA type makes it particularly suitable for studying the local distribution of co-solutes around the DNA, as well as their influence on thermal fluctuations.

A snapshot of 2KF7 basket-type G-quadruplex complexed with two stabilizing K^+ ions is shown in Fig. 37.

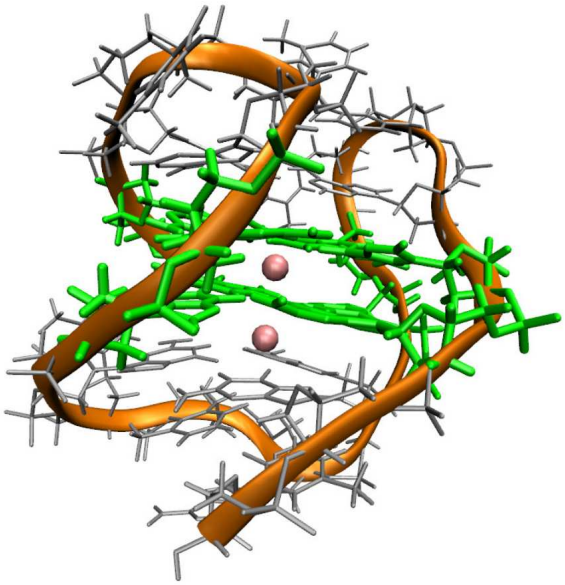


Figure 37: Basket-type human telomeric G-quadruplex complex (PDB ID: 2KF7) with two K^+ ions (pink). Stabilizing G-tetrad regions are marked with green color, the backbone is marked in orange. Reprinted with permission from [1].

All-atomistic molecular dynamics (MD) simulations were performed with the Gromacs 4.6.5 software package [275]. The Kirkwood-Buff force field for urea [277] and Garcia force field for TMAO [310] have been applied in combination with TIP3P water model [287]. Recent studies show that those force fields reflect the biophysical properties of both co-solutes in terms of their interactions with biomolecules the most accurately [279, 311]. For the G-quadruplex, I applied two force fields: ParmBSC1 [228] and its modified version, where the radii for van der Waals interactions were adjusted according to the procedure described in Ref. [312]. The referenced reparametrization of the Lennard-Jones parameters in terms of the van-der-Waals (vdW) attractive interactions was shown to provide a more accurate representation of the free energy profiles for certain nucleic acid structures [312]. Negative DNA backbone charge was neutralized by the addition of 21 Na^+ ions. Particle Mesh Ewald (PME) method [280, 281] was employed to treat long-range electrostatics. Short-range electrostatic as well as Lennard-Jones interactions were truncated with a cutoff of 0.9 nm. LINCS algorithm [282] was applied to constrain all bonds. The systems containing single DNA molecule have been simulated in pure water and in the presence of 1 M TMAO, 2 M urea and a mixture of 1 M TMAO and 2 M urea in agreement with the experimental reference conditions [155, 156]. The short-hand notation for the individual systems is shown in Table 3.

Table 3: Short hand notation for aqueous solutions of DNA (index '2') in presence of distinct urea c_{Urea} and TMAO c_{TMAO} concentrations. The notations 1TM and 2UM are introduced in order to distinguish explicitly between the individual results for TMAO and urea molecules in the mixtures. Reprinted with permission from [1].

Notation	c_{TMAO} [M]	c_{Urea} [M]
W	0	0
1T	1	0
2U	0	2
1T2U	1	2
1TM (TMAO)	1	2
2UM (Urea)	1	2

Prior to MD simulations, each of the systems was energy minimized with the Steepest Descent algorithm. Consequently, the systems were equilibrated for 3 ns (time step Δt = 2 fs) in the NVT ensemble with position restraints on DNA heavy atoms. The temperature of 300 K was maintained with the use of Nose-Hoover thermostat [244, 245]. Subsequently, the systems were equilibrated for 3 ns (time step Δt = 2 fs) in the NpT ensemble with position restrains on DNA

heavy atoms, Nose-Hoover thermostat for temperature coupling and Parrinello-Rahman barostat [247] with reference pressure of 1 bar to control the pressure. The lengths of the cubic boxes in terms of periodic boundary conditions were approximately equal 11 nm after equilibration. This enables the applicability of the finite-size approximations according to Eqns. 11 and 18. I performed 150 ns (time step $\Delta t = 2$ fs) production run for each of the equilibrated systems at the temperature of 300 K in the NpT ensemble, with position restraints on DNA in order to sample minimally fluctuating DNA conformations.

6.2 Results and discussion

6.2.1 Radial and spatial distribution functions

I calculated radial distribution functions (RDFs) around DNA for urea and TMAO molecules, respectively, with the molecular centers-of-masses for each species as reference positions. The center of mass of the G-quadruplex molecule was calculated with reference to the molecular masses and positions of the nucleobases, sugars and backbone atoms. The corresponding results are shown in Fig. 38.

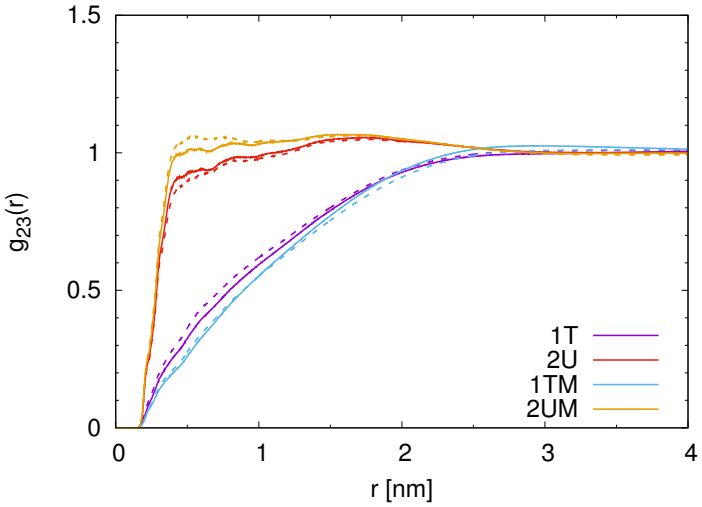


Figure 38: RDFs for TMAO and urea molecules, respectively, around the DNA for different solutions (Table 3). Solid lines refer to the simulations with the ParmBSC1 force field and dotted lines to the simulations with the vdW-modified force field [312]. The molecular center-of-masses for each species were used as reference positions. Reprinted with permission from [1].

The results show an evident accumulation of urea molecules in the close vicinity to DNA surface at $r \approx 0.4$ nm, whereas TMAO is clearly repelled to the bulk phase at distances $r > 2$ nm. However, the results for the urea–TMAO mixture (1TM and 2UM) differ only slightly from the results for the single co-solute solutions (1T and 2U), with the largest differences observed for urea at $r < 1.5$ nm. At that distances, urea molecules in the mixture exhibit a slightly smaller accumulation effect in comparison to the individual urea solution. At the same time, the positions of the main accumulation peaks remain unchanged. On the contrary, for TMAO in the mixture with urea the repulsion from DNA surface is somewhat more pronounced when compared to TMAO solution alone. Taking into consideration the clearly distinguishable urea and TMAO mean positions, long-range ordering effects become evident. Moreover, the convergence of all RDFs to unity at distances $r > 3$ nm indicates a distinction between local and bulk regions. Thus, urea is accumulated in the local region around the DNA whereas TMAO is excluded to the bulk phase. Although some discrepancies between the results obtained for ParmBSC1 and ParmBSC1_vdW fields exists, with larger accumulation effects observed for ParmBSC1_vdW, the general trend is maintained for all simulated systems.

The spatial distribution functions for TMAO and urea around the DNA are shown in Fig. 39. It can be observed that the average positions of urea molecules in the individual solution

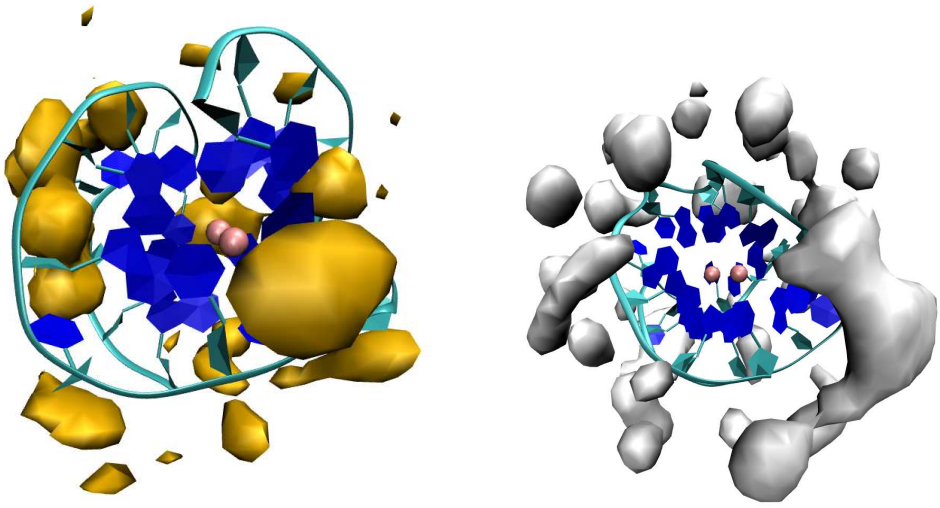


Figure 39: Spatial distribution functions of urea (left) and TMAO (right) around the DNA G-quadruplex. The DNA is shown in licorice representation from a top view with blue nucleobases and two potassium ions (pink spheres). The yellow regions mark average position of urea molecules and the white shaded regions highlight average positions of TMAO. All results are from single component solutions of urea (2U) and TMAO (1T). Reprinted with permission from [1].

(2U) are in close contact with particular parts of the DNA G-quadruplex structure, which is in agreement with the results from the RDF calculations. Urea exhibits a close accumulation around the backbone and even some insertion between the stacked nucleobases. These findings are in good agreement with the previous study [3], showing that urea interacts preferably with all components of DNA molecules. Contrary to urea, TMAO exhibits a repulsive behavior from the G-quadruplex. One can see that the internal structure of the DNA is not exposed to the solvent, which is manifested by the interaction of all nucleobases with each other or the ions. The limited accessibility of the nucleobases rationalizes the lack of interaction between TMAO and the G-quadruplex.

6.2.2 Preferential binding and local/bulk partition coefficients

In order to get insight into the accumulation of co-solutes around DNA G-quadruplex, I calculated the preferential binding coefficient ν_{23}, refined with regard to differences in the local composition of the solution in close vicinity to the solute and in bulk solution. For the systems containing a mixture of TMAO and urea, the combined co-solute approach, where TMAO and urea are treated collectively as a mixed shell of co-solutes, has been implemented. The details of such approach has already been published in Ref. [6]. The same approach is applied for further analysis involving local/bulk partitioning model. The results are shown in Fig. 40.

For all the systems, the preferential binding coefficient for single urea solutions (2U) reveals positive value with ($\nu_{23} \approx 11$) for ParmBSC1 force field, whereas for TMAO the preferential binding coefficient is always negative ($\nu_{23} \approx -4$). This observation is valid for both applied force fields. According to the preferential binding model in the frame of Kirkwood-Buff theory [69, 146, 87, 210], this indicates a strong binding behavior of urea to DNA G-quadruplex, whereas TMAO is preferentially excluded. This can be rationalized by the molecular structure of TMAO, which in the simulation setting, corresponding to pH 7.0 in nature, adopts a zwitterionic form. This leads to the repulsion between its negatively charged oxygen atoms and highly negative DNA backbone. Moreover, spatial restriction associated with TMAO size hinders the co-solute molecules from entering DNA grooves effectively.

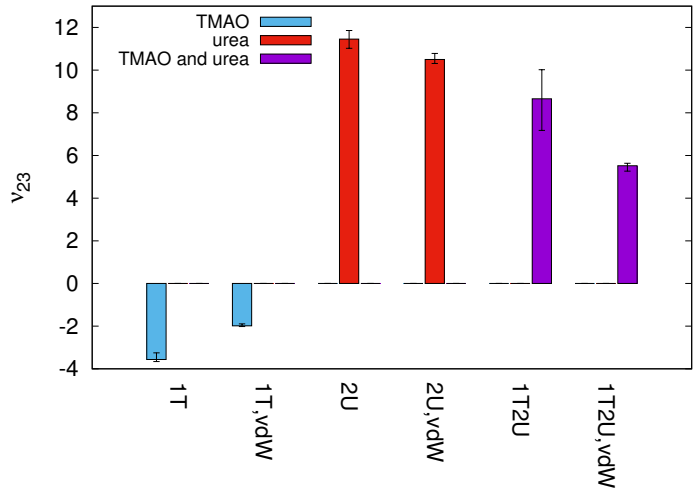


Figure 40: Preferential binding coefficients ν_{23} for single TMAO (1T) and urea solutions (2U) and the mixture (1T2U) around the DNA G-quadruplex. A combined co-solute definition consisting of all TMAO and urea molecules is considered for 1T2U. The index 'vdW' refers to the van-der-Waals-modified force field. Definitions of short-hand notations are given in Table 3. Reprinted with permission from [1].

Numerous studies published to date attribute preferential exclusion of co-solutes from the vicinity of macromolecules to their structural stabilization [70, 43, 71, 45]. In this aspect, the repulsion of the co-solute molecules from the closest environment of the biomolecule proceeds with successive filling of the resulting free space by excess water molecules, which stabilizes the native conformation via preferential hydration mechanism [41, 63, 74, 72, 73, 69, 5]. On the other hand, preferential binding of the co-solute with consequent dehydration of the biomolecule results in destabilization of its structure [43, 62, 69, 5]. Taking this into consideration, one can assume that urea acts as a DNA destabilizing agent in agreement with previous studies on small DNA hairpins [3]. Consequently, TMAO can be considered as the stabilizer of DNA native structure. The protecting properties of TMAO have been already well explored and established for proteins [155, 163, 164, 137]. Despite the differences between both types of biomolecules, such as pronounced negative DNA charge which most of proteins are lacking, TMAO shows the features indicating the protecting role also for DNA structure.

Preferential binding coefficient for the artificial combined TMAO-urea co-solute (1T2U), similarly to urea, is also positive, which suggests the possible destabilization of DNA native conformation in such mixtures. However, the value of $\nu_{23} \approx 9$ is remarkably lower in the solution where TMAO is present in comparison to the systems containing urea only. This indicates that the presence of TMAO compensates to some degree the destabilizing effect of urea, although the compensation is not complete. Furthermore, it can be concluded that the resulting value of the preferential binding coefficient for the mixture can be attributed to the sum of the individual urea and TMAO contributions.

Further evidence of the destabilizing effect of urea on G-quadruplex structure comes from the analysis of the local/bulk partitioning coefficient K_p, which is presented in Fig. 41. Analogously to ν_{23}, for the mixture of TMAO and urea molecules (1T2U) I employed a combined co-solute approach, where TMAO and urea are treated collectively as a single co-solute species [6].

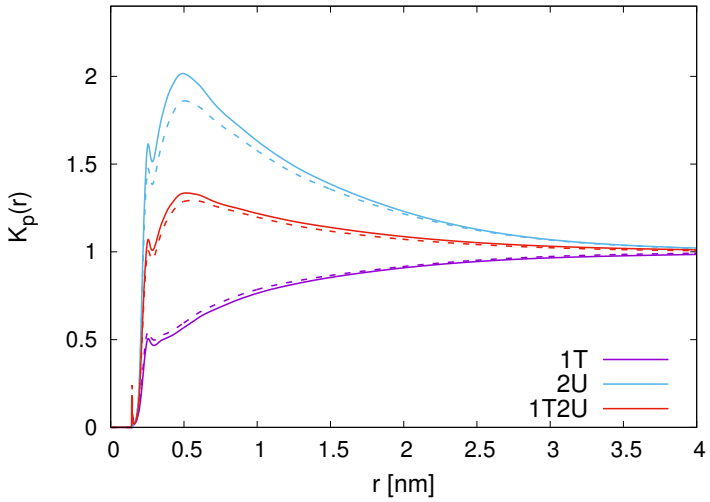


Figure 41: Local/bulk partition coefficients $K_p(r)$ for single TMAO (1T) and urea (2U) solutions in combination with the mixture (1T2U) around the DNA G-quadruplex. A combined co-solute definition consisting of all TMAO and urea molecules is considered for 1T2U. Solid lines refer to the simulations with the ParmBSC1 force field and dotted lines to the simulations with the vdW-modified force field. Definitions of short-hand notations are given in Table 3. Reprinted with permission from [1].

The values of $K_p(r) > 1$ indicate a pronounced accumulation of urea molecules (2U) at short distances of $r < 3$ nm around DNA G-quadruplex, with the maximum at $r_{\max} \approx 0.5$ nm. Together with the positive values of ν_{23}, these results confirm that urea is preferentially attracted to the DNA, which is translated into its destabilizing effect on G-quadruplex structure. On the contrary, local/bulk partitioning coefficient K_p for TMAO (1T) is lower than unity without a clearly defined maximum value. This highlights a preferential exclusion of TMAO to the bulk phase. Concerning the contributions from single components in the 1:2 TMAO:urea mixtures, the local/bulk partition coefficient $K_p(r) > 1$ with the maximum at $r_{\max} \approx 0.5$ similarly to those for single component urea solutions (2U). However, it can be noticed that the $K_p(r)$ value at $r = r_{\max} = 0.5$ nm for 1T2U is nearly four times lower than that for urea (2U) and approximately two times higher than for single TMAO solution (1T). This indicates the binding behavior of urea to the biomolecule in the mixtures, which is indeed much less pronounced than in pure urea solutions. These results are a sign of the screening impact of TMAO on the destabilizing influence of urea. By counteracting the action of urea, TMAO contributes to stabilization of DNA native conformation even in the presence of such effective denaturant as urea. It is a point of discussion whether this counteracting effect occurs via direct TMAO–urea interaction or other mechanisms like the alterations in water structure or changes in biomolecule hydration pattern if both TMAO and urea are involved. Recent studies evidence against the direct formation of hydrogen bonds between TMAO and urea [36, 6, 172, 82], although it is suggested that both co-solutes interact with each other via hydrophobic association [36]. Stabilization of biomolecule by TMAO in the presence of urea would be thus the result of the changes in biomolecular solvation pattern [158, 172, 6]. It is known that TMAO preferetially forms hydrogen bonds with water than with urea [36, 6], and the lifetime of TMAO–water H-bonds as well as their strength is significantly higher than water–water or urea–water H-bonds [36, 6, 158]. This can be associated with the disparity between the strength of weakly donating urea hydrogen atom and strongly accepting oxygen atom of TMAO, which as a consequence hampers the direct TMAO–urea H-bonding [36]. A more detailed approach to this question will be presented in the next chapter. The stabilization of biomolecule by TMAO in the presence of urea would be thus associated with strengthening the water network around biomolecule, hence inhibiting its unfolding, as suggested in Refs. [158, 82, 6].

In pure urea solutions, as well as in urea/TMAO mixtures, one can observe a steep increase of $K_p(r)$ values within the distance $r \leq 0.5$ nm, which indicates the presence of water molecules around G-quadruplex. This produces a stable solvation shell to avoid DNA dehydration even when

the strong denaturant urea is present. There are no pronounced differences observed between the simulations in ParmBSC1 and ParmBSC1_vdW force fields, which highlights the universality of these observations.

In Fig. 42 there are shown the KBIs for the co-solutes around the DNA G-quadruplex.

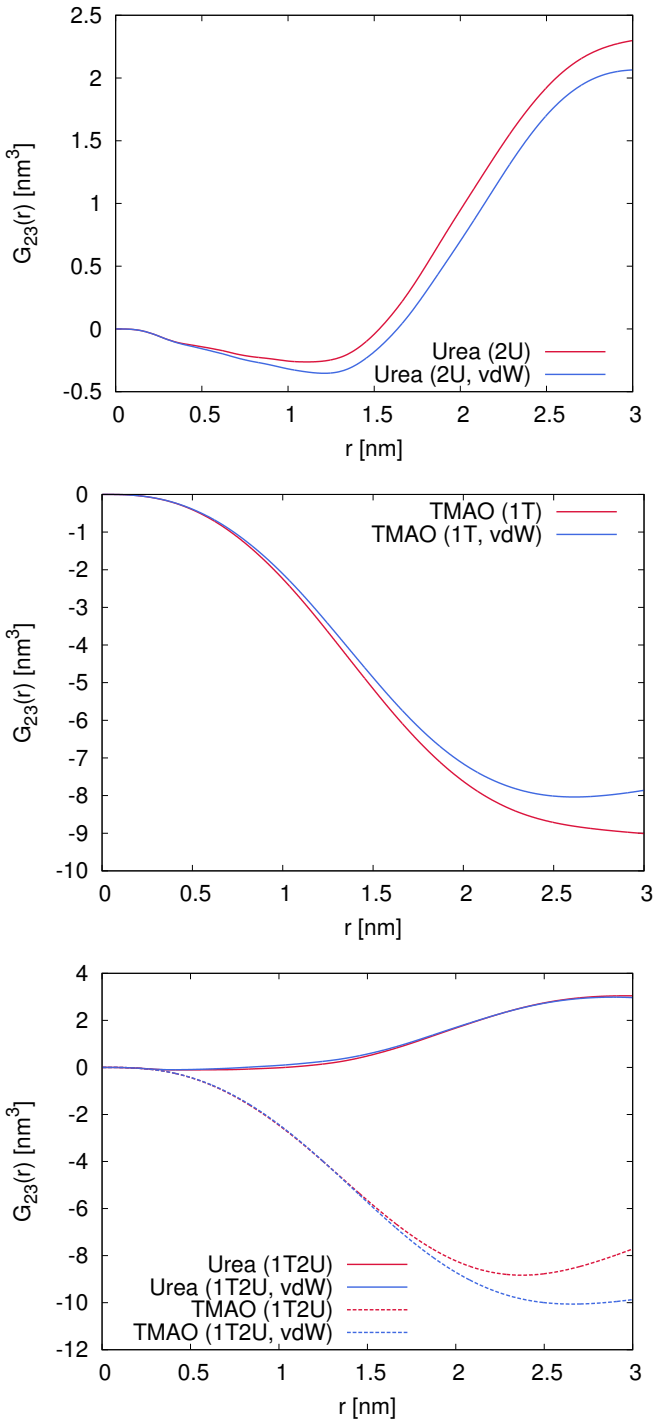


Figure 42: Kirkwood-Buff integrals (KBIs) for urea and TMAO around the DNA G-quadruplex. Top: Results for urea from 2U solutions. Middle: Results for TMAO from 1T solutions. Bottom: Results for TMAO and urea from 1T2U solutions. Red lines denote the KBI values for the ParmBSC1 force field and blue lines the corresponding results for the modified ParmBSC1_vdW force field. The dotted lines in the bottom panel denote the corresponding results for TMAO whereas solid lines represent the outcomes for urea. Reprinted with permission from [1].

One can observe a strong accumulation of urea molecules around the DNA within distances $r \geq 3$ nm, which is in agreement with the local/bulk partition coefficients. For all distances $r \geq 1.5$ nm the KBIs take negative values, which can be associated with the steric repulsion effects between the DNA and the urea molecules. Taking into consideration the highly positive local/bulk partition coefficients (Fig. 41), it can be concluded that the exclusion of water molecules is more pronounced in comparison to urea. This can be also considered as a co-solute accumulation effect. Contrary to urea, the KBI values for TMAO are negative at all distances, which points out the preferential exclusion. It is worth noting that the net values of the KBIs for individual solutions (2U and 1T) are in good agreement with the corresponding KBI values for the mixture (1T2U), which allows to postulate a pairwise additivity of co-solute contributions from the single component solutions such that cross-correlation effects in the mixture can be disregarded.

Analogous conclusions can be drawn also with regard to the sum of the individual KBIs. The corresponding results are shown in Fig. 43.

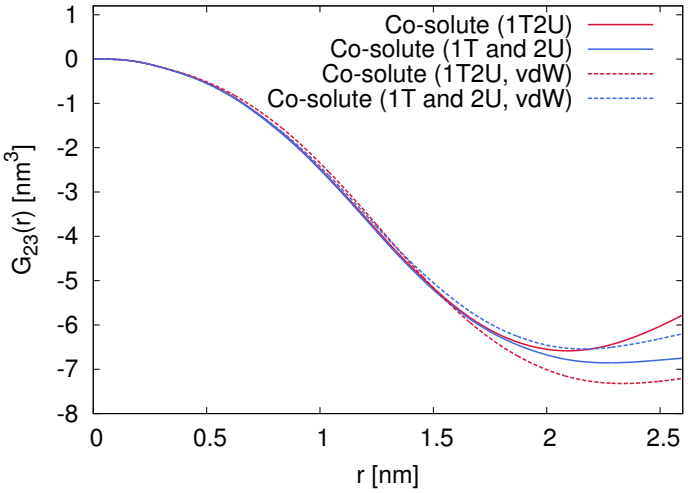


Figure 43: Sum of the combined KBIs from urea and TMAO in single co-solute solutions (1T and 2U) as marked by the red solid line in comparison to the combined co-solute approach in the two co-solute solution (1T2U) shown as the blue solid line. The dotted lines show the corresponding results for the modified ParmBSC1 force field approach with the same color code. Reprinted with permission from [1].

It can be noted that the combination of the KBIs from the single co-solute solutions (1T and 2U) according to $G_{23}^{\mathrm{comb}}(r) = G_{23}^{\mathrm{urea,2U}}(r) + G_{23}^{\mathrm{TMAO,1T}}(r)$ yields the results, which show a high agreement with the combined KBIs $G_{23}^{2,\mathrm{comb}}(r) = G_{23}^{\mathrm{urea,1T2U}}(r) + G_{23}^{\mathrm{TMAO,1T2U}}(r)$, for TMAO and urea in the mixture (1T2U), such that $G_{23}^{\mathrm{comb}}(r) \approx G_{23}^{2,\mathrm{comb}}(r)$. This indicates that the local composition of urea and TMAO in 1T2U is approximately identical to the co-solutes in the individual 1T and 2U mixtures. This rationalizes the additivity of the resulting influences according to linear superposition. Consequently, the higher order cross-correlation effects can be neglected. Although slight differences can be observed for the modified ParmBSC1 force field, the general trends is maintained.

The consequence of the aforesaid observation is that the destabilization impact in TMAO:urea mixtures is less pronounced in comparison to individual urea solutions due to the already mentioned compensation effect attributed to TMAO. Hence, TMAO can be considered to counteract the destabilizing effect of urea, although the exact compensation is not achieved. This can be related to the high net charge of the DNA, which modifies significantly the binding behavior of certain co-solutes when compared to uncharged molecules [4]. Taking all these findings into consideration it can be concluded that the resulting destabilization tendency is 2U > 1T2U > 1T and that the solvent-induced effects in terms of kosmotropic and chaotropic influences are negligible.

6.2.3 Interaction energies and intermolecular hydrogen bonds

Co-solutes can exert the effect on biomolecule via indirect mechanisms by influencing the solvation shells near the biomolecule and the structure of bulk water [49], or by forming directly hydrogen bonds with the biomolecule [3]. To get a deeper insight into this problem, I calculated the number of hydrogen bonds between DNA G-quadruplex and TMAO and urea as well as with water. The results are presented in Table 4 and shown in Fig. 44.

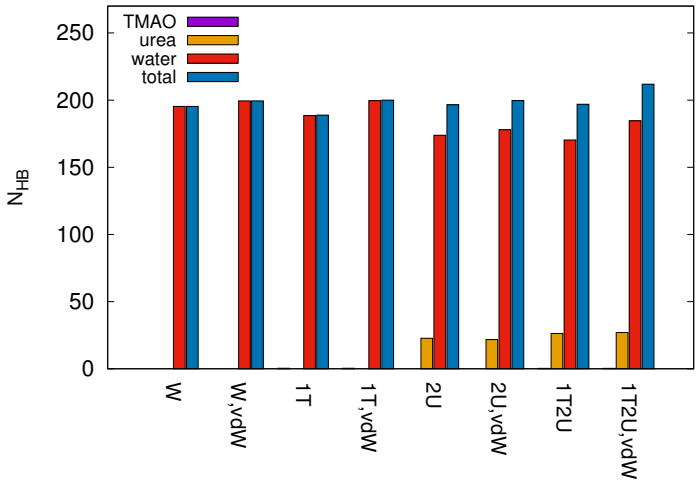


Figure 44: Number of hydrogen bonds between the DNA G-quadruplex and both co-solute or water molecules, respectively. The total number of hydrogen bonds refers to the sum of all hydrogen bonds. Definitions of short-hand notations are given in Table 3. Reprinted with permission from [1].

Table 4: Number of hydrogen bonds N_{HB} between the DNA G-quadruplex and both co-solute or water molecules, respectively. The total number of hydrogen bonds refers to the sum of all hydrogen bonds. Definitions of short-hand notations are given in Table 3. Reprinted with permission from [1].

System	DNA-TMAO	DNA-urea	DNA-water	total
W	0	0	195.32	195.32
W,vdW	0	0	199.47	199.47
1T	0.35	0	188.46	188.81
1T,vdW	0.38	0	199.60	199.99
2U	0	22.74	173.79	196.53
2U,vdW	0	21.74	177.95	199.70
1T2U	0.27	26.30	170.34	196.91
1T2U,vdW	0.28	27.01	184.56	211.85

As expected from the previous results concerning the negative preferential binding coefficients, which predicted TMAO exclusion to the bulk, TMAO does not form any signifcant amount of hydrogen bonds with DNA. At the same time, the number of hydrogen bonds between DNA and urea is nearly six times higher than the number of DNA–TMAO H-bonds, but approximately five times lower than the number of H-bonds between DNA and water molecules. These findings are valid both for single co-solute solutions and for the TMAO–urea mixture. It can be thus speculated that the influence of TMAO molecules in the 1:2 TMAO:urea mixture at large distances around the DNA (Fig. 38) on the binding properties between urea and DNA is only marginal. Furthermore, the addition of urea (2U) does not cause the loss of total hydrogen bonds in comparison to the pure water reference solution (W). Also the individual numbers of hydrogen bonds between DNA and TMAO or urea in the mixtures remain unchanged when compared to the single co-solute solutions (1T and 2U), such that no significant changes between mixture and the single co-solute solutions

are detectable. However, when comparing pure water (W) to the individual urea solution (2U) and to the mixture (1T2U), one can observe a change in the number of hydrogen bonds between water molecules and the DNA. The decrease in number of water-related hydrogen bonds can be rationalized by the replacement of water molecules around DNA by an equivalent amount of urea molecules. In contrast, the numbers of hydrogen bonds between water and DNA are roughly comparable for the pure water (W) and the single TMAO solution (1T), which can be explained by large distance of TMAO around the DNA.

To answer the question whether the protecting effect of TMAO against urea is fulfilled via direct interactions between both co-solutes, I analyzed the number of hydrogen bonds between TMAO and urea as well as between each of them and water. The results are shown in Table 5 and in Fig. 45.

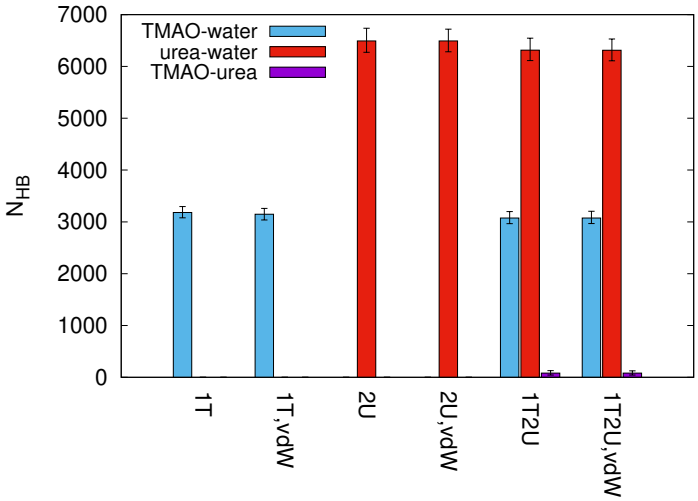


Figure 45: Number of hydrogen bonds between co-solutes and water and co-solutes themselves for the simulations in unmodified and vdW-modified force field. Definitions of short-hand notations are given in Table 3. Reprinted with permission from [1].

Table 5: Average number of hydrogen bonds between particular co-solutes and water for the systems simulated in unmodified and vdW-modified (index "vdW") ParmBSC1 force field. Definitions of short-hand notations are given in Table 3. Reprinted with permission from [1].

System	N_{HB} **TMAO-water**	N_{HB} **urea-water**	N_{HB} **TMAO-urea**
1T	3181.655	0	0
1T,vdW	3147.033	0	0
2U	0	6492.919	0
2U,vdW	0	6492.117	0
1T2U	3075.079	6315.182	81.150
1T2U,vdW	3074.557	6312.543	81.158

It can be noticed that there are nearly no hydrogen bonds formed between TMAO and urea, which is in agreement with other studies [36, 6, 172, 82]. Urea forms double as much H-bonds with water as TMAO, which is associated with its molecular structure. TMAO has only one H-bond acceptor group, whereas urea has both acceptor and donor groups. Moreover, the number of H-bonds formed between each of the co-solutes and water remains approximately the same, regardless if the co-solutes are present in individual solutions or they coexist in 1:2 molar mixtures. This lack of apparent differences in hydrogen bonding pattern between co-solutes themselves, together with potential stabilization of DNA structure by TMAO in the presence of urea, indicates that the stabilizing effect of TMAO is exerted via some other (indirect) mechanisms than by direct hydrogen

bonding with DNA residues. The mechanism can possibly involve the cooperative strengthening of the hydrogen bond network of water molecules and consequent reduction of water dynamics in TMAO and urea solution in agreement with the previous study published in Ref. [6]. Although there are slight differences between the number of hydrogen bond calculated for ParmBSC1 and ParmBSC1_vdW force fields, it can be seen that all values are roughly comparable. This shows that the individual properties of the co-solutes remain unchanged in the mixture, which indicates the linear additivity of the individual contributions from urea and TMAO to the compensation mechanism.

To get a deeper insight into the energetic aspects of the interaction between DNA G-quadruplex molecule and other key components of the solution, I calculated the energies of electrostatic and dispersion interactions in terms of Coulomb and short-range Lennard-Jones (LJ) energies. The results are shown in Fig. 46 and in Tables 6 and 7.

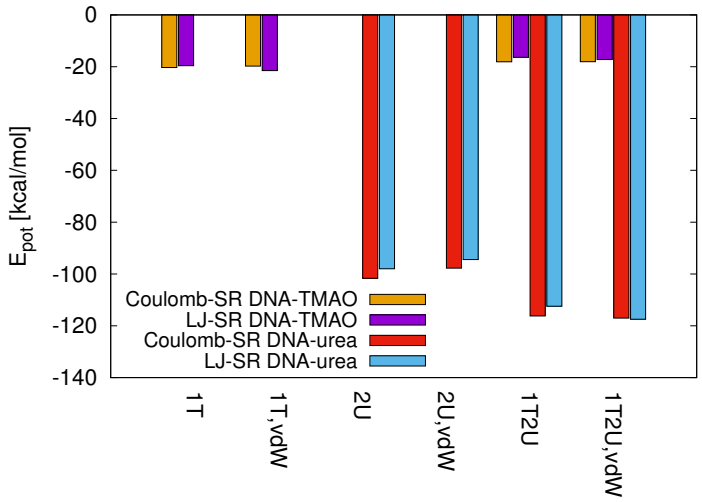


Figure 46: Short-range LJ and Coulomb interaction energies between the DNA G-quadruplex and TMAO and urea, respectively. Definitions of short-hand notations are given in Table 3. Reprinted with permission from [1].

Table 6: Electrostatic and Lennard-Jones interaction energies between DNA and other components of the solution. Reprinted with permission from [1].

System	Coulomb-SR DNA-TMAO [kcal/mol]	Coulomb-SR DNA-urea [kcal/mol]	Coulomb-SR DNA-water [kcal/mol]
pure water	-	-	-1756.31
1M TMAO	-20.32	-	-1696.96
2M urea	-	-101.66	-1580.46
1M TMAO + 2M urea	-18.09	-116.20	-1535.17
System	LJ-SR DNA-TMAO [kcal/mol]	LJ-SR DNA-urea [kcal/mol]	LJ-SR DNA-water [kcal/mol]
pure water	-	-	-78.20
1M TMAO	-19.61	-	-58.15
2M urea	-	-97.96	-29.49
1M TMAO + 2M urea	-16.34	-112.43	-28.37

Table 7: Electrostatic and Lennard-Jones interaction energies between DNA and other components of the solution in the systems simulated with ParmBSC1 (vdW) force field. Reprinted with permission from [1].

System	Coulomb-SR DNA-TMAO [kcal/mol]	Coulomb-SR DNA-urea [kcal/mol]	Coulomb-SR DNA-water [kcal/mol]
pure water (vdW)	-	-	-1760.44
1M TMAO (vdW)	-19.77	-	-1750.91
2M urea (vdW)	-	-97.72	-1594.26
1M TMAO + 2M urea (vdW)	-18.06	-117.03	-1621.63
System	LJ-SR DNA-TMAO [kcal/mol]	LJ-SR DNA-urea [kcal/mol]	LJ-SR DNA-water [kcal/mol]
pure water (vdW)	-	-	-65.95
1M TMAO (vdW)	-21.50	-	-62.78
2M urea (vdW)	-	-94.41	-32.98
1M TMAO + 2M urea (vdW)	-17.23	-117.44	-32.45

It can be noticed that both electrostatic and Lennard-Jones energies reveal negative values, which indicates the attraction of both co-solutes to DNA G-quadruplex. Although it is an expected outcome with regard to my previous findings concerning urea, it is indeed surprising for TMAO with regard to its preferential exclusion behavior. Notably, TMAO exhibits only weak LJ and Coulomb attraction to DNA, resulting possibly from its chemical structure and the relatively few molecules, which are accumulated in close vicinity around the DNA ($r < 1$ nm). Under the simulation conditions, TMAO takes a zwitterionic form with pronounced negative charge at oxygen and positive charge at nitrogen atoms, which is responsible for the attraction to negatively charged G-quadruplex backbone phosphorus atoms. On the other hand, both electrostatic and dispersion interactions between DNA and urea are nearly one order of magnitude stronger than corresponding DNA–TMAO interaction energies. This indicates a strong attraction of urea to DNA in agreement with Ref. [3] and with aforementioned preferential binding and exclusion behavior. The most pronounced electrostatic attraction is observed between DNA and water molecules, and is an order of magnitude stronger than corresponding DNA–urea interaction energies and two orders of magnitude stronger than corresponding DNA–TMAO interaction energies. One can notice that electrostatic attraction between DNA and water is the strongest in pure water solution and decreases upon the addition of TMAO, to reach the lowest value in the mixtures where urea is present, either alone or together with TMAO. This shows that although DNA remains in principle well hydrated, the presence of co-solutes weakens the hydration shell of biomolecule, where the strongest dehydrating impact is exerted by urea. This is also associated with amount effects, where the number of water molecules is reduced in the co-solute systems and replaced by corresponding co-solute molecules. Although the energetic differences are not so pronounced between pure TMAO and water solutions and amount to ca. 10-50 kcal/mol, for the solutions with urea (both alone and in combination with TMAO) these differences add up to 200 kcal/mol. This corresponds to my previous findings indicating that urea not only directly interacts with DNA, but also influences biomolecular hydration pattern, thus contributing to DNA structure destabilization [3]. Although there exist small differences between electrostatic energies of DNA–water interaction in pure urea and in urea/TMAO solutions, they are not very pronounced, thus suggesting a minor role of TMAO in modulating the energies of DNA interactions with other components of the simulated systems.

Analogous effects in certain simulated systems are observed when LJ energies are concerned. Looking at TMAO and urea binding to DNA, one can notice that the energetic contribution from dispersion interaction is nearly the same as the energetic contribution from electrostatic interaction. This suggests that co-solute binding is enthalpy driven. On the contrary, LJ energies of water molecules binding to DNA are nearly two orders of magnitude lower than corresponding coulombic energies. Since LJ energies are much more short-ranged than electrostatic ones, they are responsible for DNA solvation only at shortest distances, within the first solvation shell. After

the first solvation shell is filled, water molecules interacting with DNA via Coulomb forces provide good hydration of DNA G-quadruplex also at larger distances from molecular surface.

It can be however observed that short-ranged LJ energies of urea binding are approximately 3-4 times higher than corresponding LJ energies of water–DNA attraction. Thus one can conclude that urea binding dominates binding characteristics at short distances, possibly competing with water molecules for the binding sites within the first solvation shell in agreement with Ref. [3]. In case of TMAO, both LJ and Coulomb energies have relatively low values, contributing to the previous findings indicating preferential repulsion of TMAO from biomolecular surface.

What is remarkable, the corresponding interaction energies for single components and in the mixture are roughly comparable. There are slight differences observed only for the mixtures (1T2U), where the electrostatic energies take more negative values in comparison to single urea solution (2U). Notwithstanding, the presence of both co-solutes in the mixture does not lead to a mutual influence on the accumulation behavior, which agrees with my previous results. Taking into consideration also the results for the hydrogen bonds, it can be assumed that the individual contributions from the co-solutes add pairwise linearly according to the compensation mechanism. Furthermore, the dispersion and electrostatic interactions have similar magnitude for each co-solute, which suggests an enthalpy-driven accumulation behavior around the DNA in terms of favorable interaction energies.

The Tables 8 and 9 show the electrostatic and Lennard-Jones interaction energies between TMAO, urea and water in the systems simulated with ParmBSC1 (Table 8) and ParmBSC1 with vdW correction (Table 9) force fields without the presence of DNA molecule.

Table 8: Electrostatic and Lennard-Jones interaction energies between the components of the solution in the systems simulated in ParmBSC1 force field. Reprinted with permission from [1].

System	Coulomb-SR TMAO-urea [kcal/mol]	Coulomb-SR TMAO-water [kcal/mol]	Coulomb-SR urea-water [kcal/mol]	Coulomb-SR water-water [kcal/mol]
pure water	-	-	-	-5.12e+05
1M TMAO	-	-29776.82	-	-4.47e+05
2M urea	-	-	-32740.02	-4.30e+05
1M TMAO + 2M urea	-715.65	-28662.36	-31555.79	-3.67e+05
System	LJ-SR TMAO-urea [kcal/mol]	LJ-SR TMAO-water [kcal/mol]	LJ-SR urea-water [kcal/mol]	LJ-SR water-water [kcal/mol]
pure water	-	-	-	71540.74
1M TMAO	-	-905.53	-	64493.19
2M urea	-	-	-4271.56	62384.95
1M TMAO + 2M urea	-819.75	-509.86	-3781.60	55243.49

The energies of cross-interactions between TMAO, urea and water are much higher than the interaction energies of either of them with DNA, and count up to four orders of magnitude. The highest attractive electrostatic energies are observed between water molecules themselves, and they are an order of magnitude higher than coulombic attraction energies between TMAO or urea and water. Consequently, electrostatic energies of TMAO–urea interaction are roughly two orders of magnitude lower than the energies between either TMAO or urea and water. This corresponds to the earlier observations from hydrogen bond analysis, indicating that both co-solutes do not interact with themselves (see: Ref. [6]). Interestingly, dispersion energies of TMAO or urea interaction with water are significantly lower than corresponding electrostatic energies. The dominance of Coulombic energies points out the pronounced role of partial charges in co-solute attraction. This is especially remarkable for zwitterionic TMAO, where the difference between short-ranged LJ and Coulomb energies add up to two orders of magnitude.

Table 9: Electrostatic and Lennard-Jones interaction energies between the components of the solution in the systems simulated in vdW-modified ParmBSC1 force field. Reprinted with permission from [1].

System	Coulomb-SR TMAO-urea [kcal/mol]	Coulomb-SR TMAO-water [kcal/mol]	Coulomb-SR urea-water [kcal/mol]	Coulomb-SR water-water [kcal/mol]
pure water (vdW)	-	-	-	-5.12e+05
1M TMAO (vdW)	-	-29402.63	-	-4.47e+05
2M urea (vdW)	-	-	-32722.34	-4.30e+05
1M TMAO + 2M urea (vdW)	-715.87	-28654.24	-31521.86	-3.67e+05
System	LJ-SR TMAO-urea [kcal/mol]	LJ-SR TMAO-water [kcal/mol]	LJ-SR urea-water [kcal/mol]	LJ-SR water-water [kcal/mol]
pure water (vdW)	-	-	-	71536.44
1M TMAO (vdW)	-	-920.10	-	64529.51
2M urea (vdW)	-	-	-4276.56	62391.64
1M TMAO + 2M urea (vdW)	-821.04	-511.62	-3786.33	55218.16

Furthermore, the energies of interaction between both co-solutes and water are remarkably higher than the interaction energies of either of them with DNA, in terms of both LJ and Coulomb contributions. This points out that TMAO and urea influence DNA structural stability rather by indirect mechanism than by direct interactions with the biomolecule.

6.2.4 Insights into the compensation mechanism

The results presented in this work reveal the influence of structural arrangement of the co-solutes on DNA stability. In case of urea and TMAO, the individual results in one-component co-solute solutions 2U and 1T, exhibit the positive and negative values for the KBIs, respectively (Fig. 42). With rerefence to the Kirkwood-Buff theory, this indicates a strong destabilizing effect on DNA induced by urea, whereas TMAO exerts a stabilizing impact on the corresponding reference structure. These effects are reversible and do not induce any chemical modification of the DNA structure. Since the variable electrostatic or dispersion energies as well as short- and long-range interactions for hydrogen bonds between the co-solutes in particular mixtures can be disregarded, these effects are not strong enough to provide a rationale for the compensation mechanism. Nonetheless, the clear ordering of TMAO and urea can be rationalized with respect to the the differences in the interaction energies between the DNA and both co-solutes.

Taking into consideration the additivity of the local arrangement, the compensation effect might be promoted by the co-solute accumulation and exclusion behavior. This conclusion is grouned by the lower values of the preferential binding coefficients for combined co-solutes in comparison to single urea species (see: Fig. 40). This points out a slight weakening of the destabilizing influence of urea in the presence of TMAO. The outcome for the combined KBIs (see: Fig. 43) also highlights the additivity of the influences of both co-solutes. Thus, the local arrangement of TMAO and urea around DNA does not vary significantly in the mixtures of both when compared to the individual solutions, such that the cross-correlation influences are of minor importance. These findings may be applicable for highly charged macromolecules which exhibit a strong attraction of dipolar co-solutes. In consequence, the binding of urea would be more favorable energetically in comparison to TMAO (Fig. 46). This explains the pronounced separation between urea and TMAO shells around DNA molecule. Since the nucleotides are buried inside the G-quadruplex compact structure, the presence of further favorable attractive dispersion interactions with TMAO is disfavored. All these outcomes provide a rationale for the postulated compensation mechanisms for nucleic acid structures.

7 Interaction profile of TMAO and urea with DNA G-quadruplex structure

In this chapter, I study the antiparallel basket type human telomeric G-quadruplex (PDB: 2KF7 [307, 306]) of outstanding biological interest [313] in the presence of TMAO and urea in near-physiological concentrations of 1 M and 2 M accordingly, as well as in the 1:2 M mixture of both co-solutes and in the pure water [2]. Under physiological pH 7.0, TMAO adopts the zwitterionic form, which is also employed for the purpose of these computer simulations. The structure of the simulated DNA conformation is shown in Fig. 37 in Chapter 6.

I apply the atomistic MD simulations in the microsecond scale, which is long enough to reflect on the phenomena happening in nature. With the means of computational methods, the detailed insights into the structural and conformational characteristics of a G-quadruplex type DNA in pure water as well as in the presence of stabilizing and destabilizing agents like TMAO and urea can be provided. The impact of thermal stress associated with the simulation temperatures higher than the melting temperature will also be investigated.

The mechanism of G-quadruplex and co-solutes interaction from thermodynamic perspective has been already described in details in Chapter 6. This study is a significant extension of my previous findings and gets into details of the interaction mechanisms between the DNA and co-solutes. In this chapter, the interaction phenomena are presented from a visual and more physiological perspective. The study described in the previous section focuses on the stabilization and destabilization aspects in terms of the molecular theory of solutions. Here, I focus on the corresponding molecular implications and a further support for the previous findings. Since the stabilization and destabilization effects are often manifested by the structural effects in the main solute, the monitoring of DNA conformational fluctuations as a means to get insight into the nature of DNA and co-solute interactions is the main focus of this research.

7.1 Methodology

The structural characteristics of the initial input structure of the antiparallel basket G-quadruplex (PDB entry code: 2KF7 [307, 306], structure obtained via NMR) has been described in details in the previous chapter. The original 2KF7 DNA G-quadruplex has 8-bromoguanine at the position 7, which was replaced by guanine with the use of VMD (version 1.9.3) software [308, 309]. Prior to MD simulations, several short MD runs of 100 ns were performed in the presence of 1, 2 and 3 K^+ cations in order to establish the configuration of stabilizing ions optimal for this study. For a two tetrad molecule, all three configurations are theoretically possible. The analyses of the molecular stability, trajectory and the energetic properties of the simulated solution show that the application of two K^+ ions provides the most stable environment for further study. Hence, the structure was complexed with 2 K^+ ions placed in the central channel between the tetrads. Then the structure was energy minimized with the Steepest Descend algorithm including harmonic restraints on the ions with $k = 1000$ kJ·mol^{-1}·nm^{-1} to achieve a stable position of K^+ ions for a native DNA conformation. The initial input structure is shown in Fig. 47.

All-atomistic molecular dynamics (MD) simulations were performed with the Gromacs 2018 software package [314, 275]. Four molecular systems have been prepared: DNA G-quadruplex in pure water solution, in the presence of 1 M TMAO, 2 M urea and in 1 M TMAO and 2 M urea mixture. The negative net charge of each system was neutralized with the addition of 19 K^+ counterions. Additionally, each simulation box contains 0.1 M KCl in order to mimic the experimental conditions. The selected ionic strength of the buffer has been shown to influence positively the stability of the position of two K^+ channel ions in G-quadruplex structure [315].

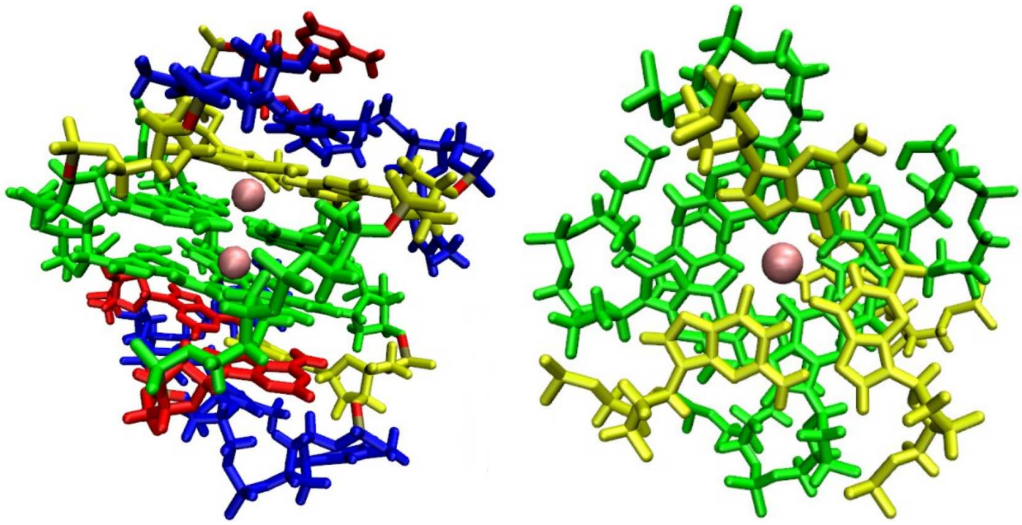


Figure 47: Antiparallel basket type DNA G-quadruplex (PDB code: 2KF7) in horizontal (left) and vertical (right) projection. Color code: G-tetrads are marked in green, dG not forming the tetrads in yellow, dA in red and dT in blue. The K^+ ions are depicted as light red spheres. Reprinted with permission from [2].

I applied ParmBSC1 force field [228] for DNA and Joung-Cheatham ion parameters for ions [316], Kirkwood-Buff force field (KBFF) for urea [277] and Garcia force field for TMAO [310] in combination with TIP3P water model [287]. Recent studies on the validation of force fields for co-solutes proved that this combination of force fields provides the most biologically accurate outcome of the interaction pattern between co-solutes and biomolecules [279]. Particle Mesh Ewald (PME) method [280, 281] was employed to treat long-range electrostatic interactions. Short-range electrostatic as well as Lennarg-Jones interactions were truncated with a cutoff of 1.2 nm. Prior to MD run, the systems were pre-equilibrated for 50 ns (time step Δt = 2 fs) in the NVT ensemble with DNA heavy atoms being position restrained, followed by 50 ns (time step Δt = 2 fs) pre-equilibration in NpT ensemble and the same position restraints. LINCS algorithm [282] was applied to constrain all bonds. The temperature of 300 K was maintained with the use of Nose-Hoover thermostat [244, 245], and Parrinello-Rahman barostat [247] with reference pressure of 1 bar was applied to control the pressure. Consequently, all systems were were equilibrated for 100 ns (time step Δt = 2 fs) in the NVT ensemble (T = 300 K) followed by 100 ns equilibration in NpT ensemble, without position restraints on DNA atoms. I performed 1 μs (time step Δt = 2 fs) production run of the equilibrated systems at T = 300 K in the NpT ensemble without position restraints on DNA heavy atoms. Additionally to the simulation in physiological temperature of 300 K, I performed high temperature simulations of the equilibrated systems in T = 360 K, which is slightly higher than the melting temperature for this DNA structure. The melting temperature T_m has been calculated according to the general formula related to the nucleobase content:

$$T_m = 4 \times (G + C) + 2 \times (A + T)$$

This formula has been shown to be universal for most of DNA sequences. In case of 2KF7 DNA quadruplex, the corresponding melting temperature yields $T_m \approx 68°\text{C} = 341$ K. Therefore the temperature of 360 K has been chosen as the temperature sufficiently higher than T_m in order to make it possible for thermal phenomena to occur, but at the same time to prevent thermal denaturation of the DNA.

The high temperature simulations aim to analyze the distribution of stabilizing (TMAO) and destabilizing (urea) co-solutes in mildly denaturing conditions, but without inducing thermal unfolding of DNA structure. Under these condition one can expect the slight destabilization of DNA conformation associated with thermal fluctuations, which allows to observe whether the co-solutes of interest show an increased affinity to certain DNA parts, which are otherwise buried inside rigidly stacked tetrads and not available for interactions. Furthermore, mild thermal destabilization of DNA structure enables the observation whether stabilizing co-solutes are able to counteract the impact of temperature and help to maintain the DNA structural stability. Thus, with reference to

the outcomes described in the previous chapter, it can be observed whether TMAO has the ability to counteract also the 'chemical' destabilization by urea, additionally to the temperature-induced destabilization effects.

7.2 Results and discussion

7.2.1 DNA flexibility descriptors

The B-factor, known also as temperature factor or Debye-Waller factor [317, 318, 319], is a commonly applied measure to give insight into the fexibility of atoms, chains or even the whole regions of biomolecules. It is often applied to indicate qualitatively the relative vibrational motion of the parts of a biomolecule. Rigid and well-ordered regions of a biomolecule are characterized by low B-factors, whereas high B-factors indicate the residues which are very flexible. In biophysics, B-factors are described by the Eqn. 72 [320]:

$$B = \frac{8\pi^2}{3}\langle RMSF^2 \rangle \tag{72}$$

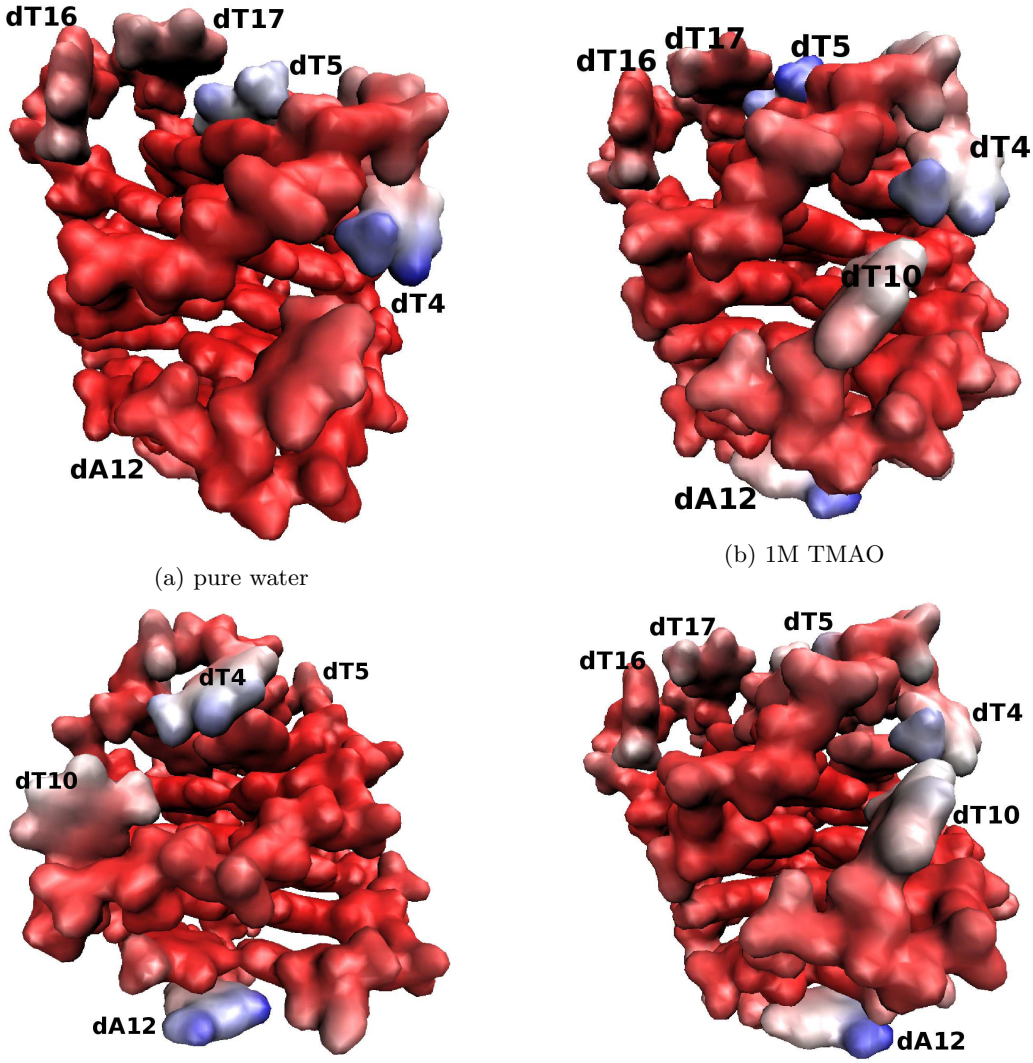
where RMSF denotes Root Mean Square Fluctuation, defined as the deviation of atomic position with respect to the position of the corresponding reference atoms over time. The detailed analysis of RMSF for all simulated systems will be shown in the further part of this section.

For all the simulated systems, I calculated B-factors relative to a reference structure, which is the energy minimized structure in pure water, with the use of Gromacs 2018 software package [314, 275] and visualized using VMD [308, 309]. The coordinates of particular residues were averaged over the whole trajectory. B-factors for 2KF7 G-quadruplex structure simulated in 300 K in pure water and in the solutions of urea and TMAO are illustrated in Fig. 48. The most flexible regions are colored in blue, whereas the most rigid regions are marked in red, with white colored parts corresponding to intermediate flexibility.

It can be noticed that the residues dT4, dT5 and dA12 maintain an enhanced flexibility in all simulated systems, although individual differences between the systems can be observed. For DNA G-quadruplex simulated in pure water (Fig. 48a), residues dT4 and dT5 constitute the most flexible parts. Although residues dA12, dT16 and dT17 also exhibit an enhanced flexibility, it is considerably reduced in comparison to the first thymidines. Analogously, dT4 and dT5 are also the most flexible parts of G-quadruplex simulated in the presence of TMAO (Fig. 48b), but in these simulation conditions the flexibility of dA12 is comparable to that of 5'-terminal thymidines. Slightly enhanced flexibility is observed also for dT10 residue. In the presence of 2 M urea (Fig. 48c), the highest flexibility is observed also for dT4 and dA12 residues, and slightly enhanced but lower than for those both nucleotides also for dT10. However, in contrast to simulations in 1 M TMAO, dT5 residue remains relatively rigid, with only some atoms showing an increased flexibility. Although residues dT4 and dA12 remain the most flexible in the mixture of 1 M TMAO and 2 M urea (Fig. 48d), in this system dT10 demonstrates comparable flexibility to those two residues. Like in 2 M urea simulations, for dT5 only some atoms feature an increased flexibility. It is worth noting that dT16 and dT17 residues demonstrate an increased flexibility only in the simulations in pure water. For all the other systems, only some atoms of dT16 and dT17 nucleotides are more flexible. Hence, one can notice that in the presence of 1 M TMAO the overall flexibility of G-quadruplex molecule is decreased in comparison to the simulation in pure water, which is demonstrated by increased area of the highest rigidity. However, same observation can be done for G-quadruplex simulated in 2 M urea, where the global flexibility is decreased not only in comparison to pure water, but also to the simulation in 1 M TMAO.

In the 1:2 mixture of both co-solutes, the global DNA flexibility is enhanced in comparison to that in pure TMAO or pure urea, but still notably reduced in comparison to the simulation in pure water. This allows to conclude that the presence of co-solutes generally stabilizes DNA structure, in agreement with aforementioned experimental observations for piezophilic organisms. This is also reflected in reduced RMSF for the flexible parts of G-quadruplex located outside the rigid stem (see: Fig. 50) in the systems containing co-solutes in comparison to simulations in pure water. This concerns especially the first 5'-terminal TTA triplet, which is stabilized in TMAO and urea tertiary and quaternary solutions. Although some differences in RMSF of particular TTA

triplet regions exist between systems simulated in pure TMAO, pure urea and TMAO:urea mix, they are not very pronounced.



(a) pure water

(b) 1M TMAO

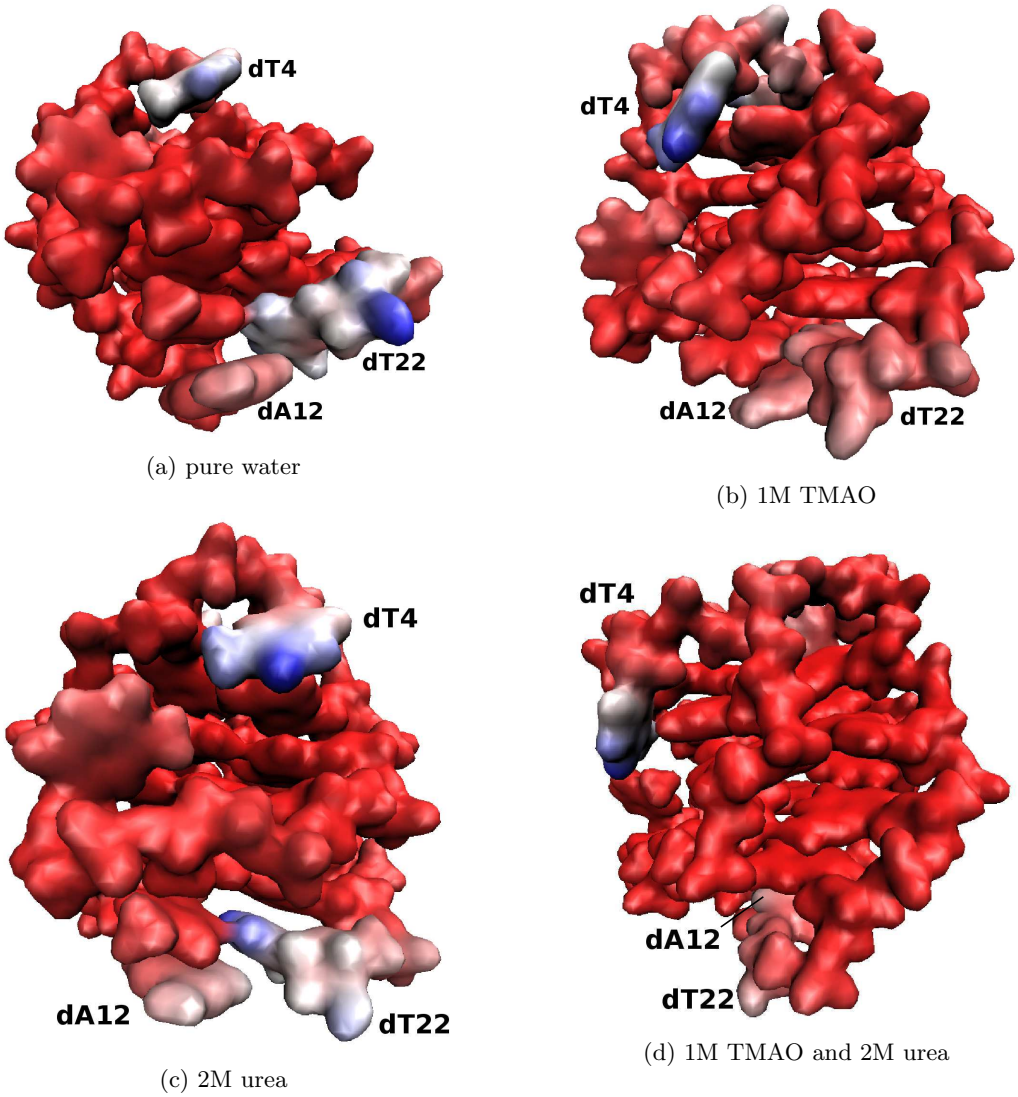
(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 48: Visualizations of B-factors for DNA G-quadruplex structures simulated at $T = 300$ K in pure water (top left), 1 M TMAO (top right), 2 M urea (bottom left) and 1 M TMAO and 2 M urea solution (bottom right). The most flexible regions of the molecule are marked in blue and the most rigid in red. Reprinted with permission from [2].

In general, for all simulated systems guanine triplets constitute the most rigid regions of G-quadruplex molecule, which corresponds to the formation of stable G-tetrads, as well as stable G-rich non-tetrad regions.

I performed analogous B-factor calculations also for high temperature simulations, and visualized them in VMD. The corresponding G-quadruplex structures with the most flexible and the most rigid regions marked with the same color code as for 300 K simulations are shown in Fig. 49.



(a) pure water

(b) 1M TMAO

(c) 2M urea

(d) 1M TMAO and 2M urea

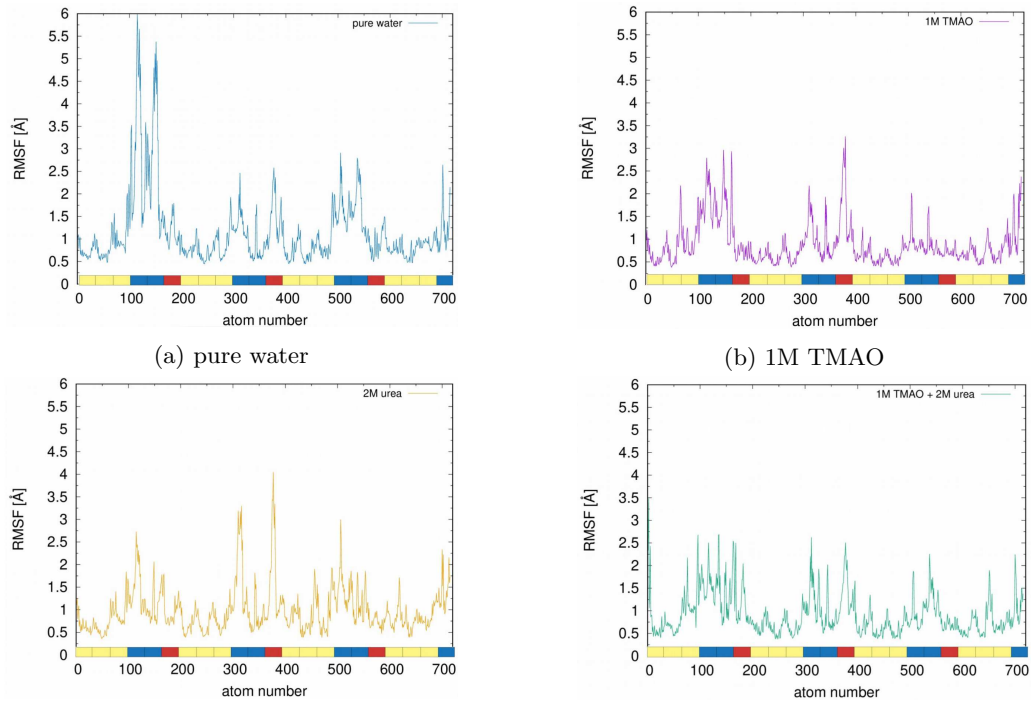
Figure 49: Visualizations of B-factors for DNA G-quadruplex structures simulated at $T = 360$ K in pure water (top left), 1 M TMAO (top right), 2 M urea (bottom left) and 1 M TMAO and 2 M urea solution (bottom right). The most flexible regions of the molecule are marked in blue and the most rigid in red. Reprinted with permission from [2].

Although relatively rigid in 300 K simulations, the dT22 residue exhibits significant flexibility in 360 K simulations in all analyzed systems. The flexibility of dT22 is especially pronounced for DNA structures simulated in pure water and in 2 M urea. In the systems containing 1 M TMAO, both alone and in the mixture with urea, dT22 is slightly less flexible, which indicates a stabilizing impact of TMAO on DNA structure. It has to be noted that the decrease in flexibility of the terminal residues only can not always be directly translated into the increase in stability of the whole structure. However, as a "rule of thumb", the decreased mobility of the residues corresponds to the more stable overall structure [321]. On the contrary, more fluctuations of the residues increase the risk of the eventual unwinding of the entire structure [321]. Thus, any observed shift in the fluctuation pattern allows to speculate on the potential stability of the whole DNA conformation.

Analogously to 300 K simulations, dT4 residue remains very flexible in all simulated solutions. Flexibility of dA12 is slightly reduced in comparison to that of dT4, but it still remains one of the most flexible residues. The only exception is the system containing 1:2 TMAO:urea mixture, where dA12 remains relatively rigid. As in 300 K simulations, dG nucletide residues are the most rigid parts of DNA structures, which points out the high stability of G-quadruplex tetrads. The rigidity of G-stem region is also reflected by relatively low RMSF values in comparison to neighboring DNA parts (see: Fig. 51). However, in contradiction to 300 K simulations, in higher temperatures

G-quadruplexes simulated in pure water appear to be more stable than the structures simulated in the presence of co-solutes, with RMSF values for flexible TTA tertads being approximately 2 Å lower than in pure TMAO or urea solutions up till approximately 4 Å lower than in TMAO:urea mixture. This points out only minor influence of co-solutes on DNA stability when thermal fluctuations due to an increased temperature are involved.

For both 300 K and 360 K simulations I calculated also the RMSF in relation to the energy minimized reference structure in pure water. Color code on the X-axis of each figure refers to particular nucleic residue: dG - yellow, dT - blue, dA - red. The results are shown in Fig. 50 and 51, accordingly.



(a) pure water

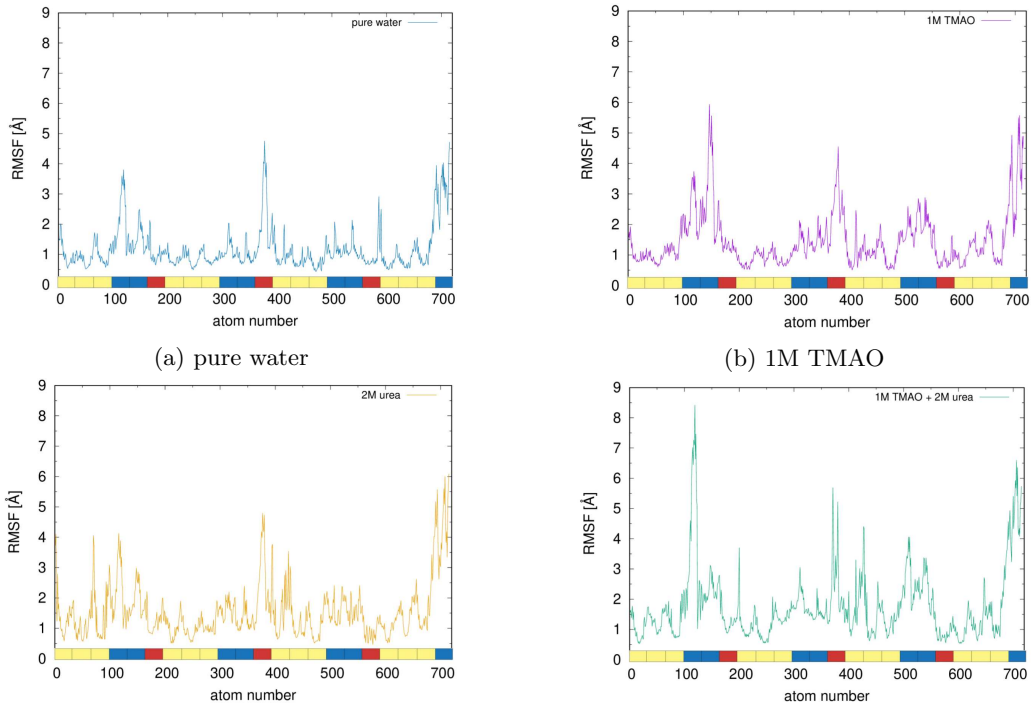
(b) 1M TMAO

(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 50: RMSF of DNA G-quadruplex structure simulated in 300 K in different solutions. Reprinted with permission from [2].

It can be observed that in 300 K the flexibility of 5-terminal nucleotide triplet TTA is reduced in the solution containing 1 M TMAO as compared to the simulation in pure water. In high temperature simulations in 360 K the fluctuations of all TTAs are increased in comparison to the simulations in 300 K. The highest fluctuations are observed in the mixed solution of both 1 M TMAO and 2 M urea in 360 K. GGG triplets exhibit considerably reduced flexibility, which indicates the formation of stable tetrads.



(a) pure water

(b) 1M TMAO

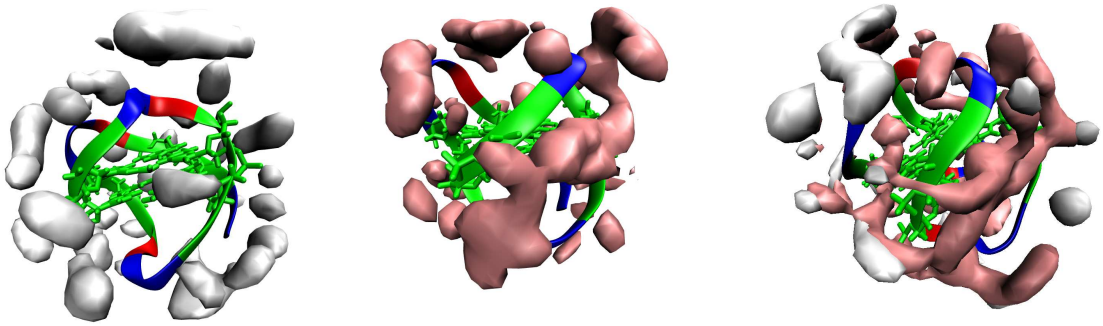
(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 51: RMSF of DNA G-quadruplex structure simulated in 360 K in different solutions. Reprinted with permission from [2].

7.2.2 Density distribution of co-solutes and water around DNA

The density distribution of TMAO, urea and water around DNA G-quadruplex structure has been visualized in the form of volumetric maps. DNA backbone is represented by cartoon model, whereas the two G-tetrads are represented by stick model. Particular nucleic residues are marked with the following color code: dG - green, dA - red and dT - blue. For the co-solutes, the following color code has been applied: TMAO - white, urea - pink, water - cyan. Volumetric maps for G-quadruplex DNA simulated in 300 K are shown in Fig. 53. For the purpose of transparency, I visualized the density distribution of TMAO and urea around DNA also with exclusion of water molecules (Fig. 52).



(a) 1M TMAO

(b) 2M urea

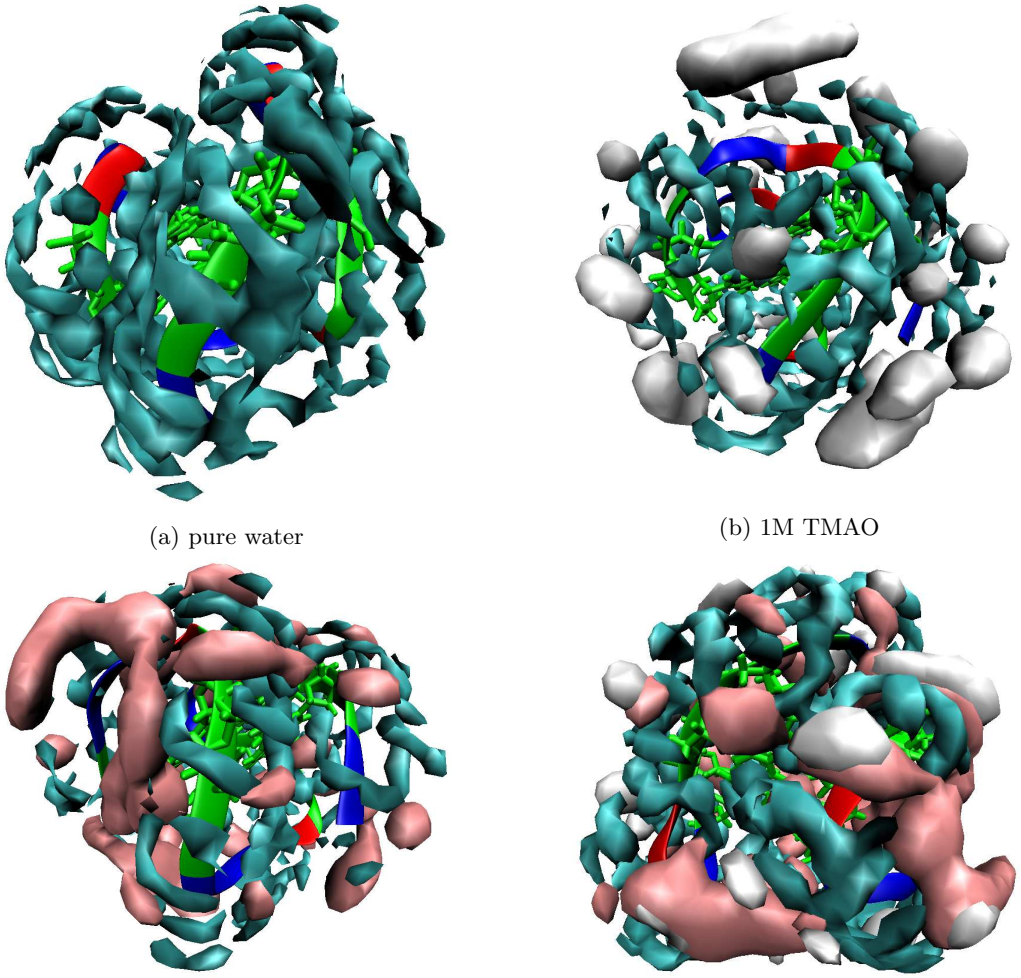
(c) 1M TMAO and 2M urea

Figure 52: Density distribution of TMAO and urea around DNA G-quadruplex structures simulated at 300 K in 1 M TMAO, 2 M urea and 1:2- molar TMAO:urea solutions. TMAO is colored in white and urea in pink. Reprinted with permission from [2].

It can be observed that TMAO locates around G-quadruplex superficially and does not enter DNA grooves (Fig. 52a). TMAO forms a dense layer around DNA, but without showing preference towards specific nucleic residues. Vast amount of TMAO, instead of being attracted towards

biomolecular surface, is dissipated in the bulk. The absence of *intercalation* effects for TMAO molecules can be a result of its molecular structure. In the simulation condition, corresponding to pH = 7 in nature, TMAO adopts a zwitterionic form, which results in repulsion between its negatively charged oxygen atoms and highly negative DNA backbone. Additionaly, the molecular size of TMAO hinders it from entering DNA grooves. Although under these simulation conditions the proper intercalation process, which describes mainly the phenomena in the microscale, can be excluded when the large co-solute molecules are involved, in this work I apply this expression to point out the analogy between the molecular intercalation event and the allocation of TMAO and urea in the vicinity of particular DNA parts in the macroscale.

In contrast, urea molecules locate much nearer to the DNA surface than TMAO (Fig. 52b), inserting DNA grooves. Urea molecules accumulate mostly in G-quadruplex major groove, formed by the residues dG15-dG19 and dG14-dG20. To smaller extent, they occupy also the second groove, composed of dG19-dG7 and dG20-dG8. The ability of urea to enter nucleic structures is favored by its smaller molecular size in comparison to TMAO and the lack of atomic charge. Only reduced number of urea molecules are dissipated in the bulk, which emphasizes its preference to localize near DNA surface. In the 1:2 mixture of TMAO and urea (Fig. 52c), certain amount of urea still enters DNA grooves, whereas some TMAO remains more distant to DNA surface. However, most of urea molecules mix with TMAO, thus forming a fused co-solute layer of co-solutes surrounding DNA G-quadruplex.



(a) pure water

(b) 1M TMAO

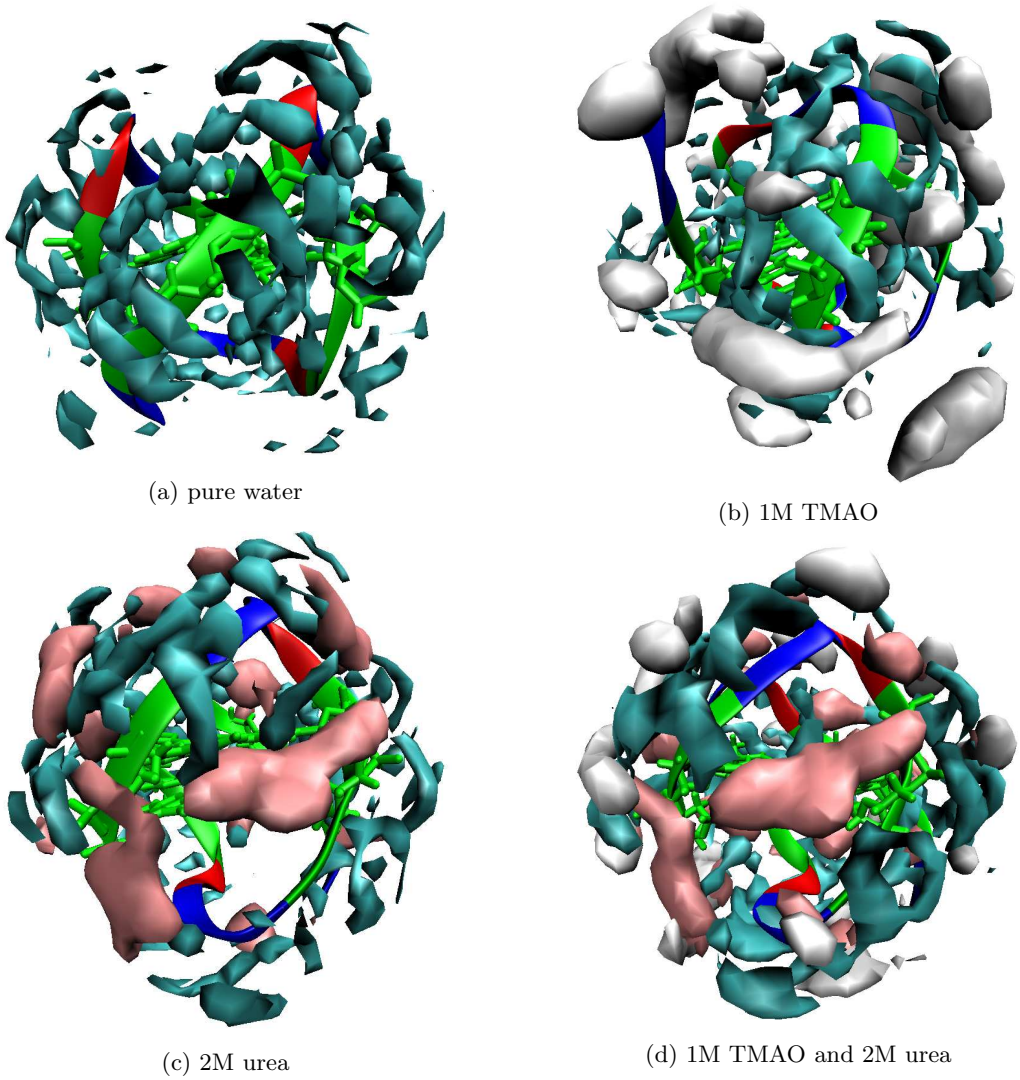
(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 53: Density distribution of TMAO, urea and water around DNA G-quadruplex structures simulated at 300 K in pure water, 1 M TMAO, 2 M urea and 1:2-molar TMAO:urea solutions. TMAO is colored in white, urea in pink and water in cyan. Reprinted with permission from [2].

The density distribution of water around DNA G-quadruplex is illustrated in Fig. 53. Without co-solutes, water occupies tightly the free space between nucleotide residues, providing good solvation of DNA molecule (Fig. 53a). Analogous phenomenon is observed in 1 M TMAO solution (Fig. 53b). In the presence of 2 M urea, water competes with urea molecules for the space inside DNA grooves (Fig. 53c). However, in 1:2 mixture of TMAO and urea one can observe a hybridized layer of water and co-solutes around DNA rather than formation of separate layers of molecules, approaching DNA residues to different extent (Fig. 53d). Although in separate TMAO nad urea solutions one can notice clearly that urea tends to enter DNA grooves whereas TMAO remains mainly outside, in their 1:2 mixture both co-solutes appear to form a combined layer, intruding DNA structure only to a very small degree. What is remarkable, the density of water molecules between DNA residues appears to be reduced in co-solutes mixture in comparison to pure water, pointing out a decreased hydration tendency. This suggests the possible dewetting properties of TMAO, which in turn promotes intermolecular interactions within DNA quadruplex in agreement with Ref. [158].

Analogous volumetric maps were constructed for G-quadruplex mixtures simulated in T = 360 K, with (Fig. 54) and without (Fig. 55) volumetric density of water shown.



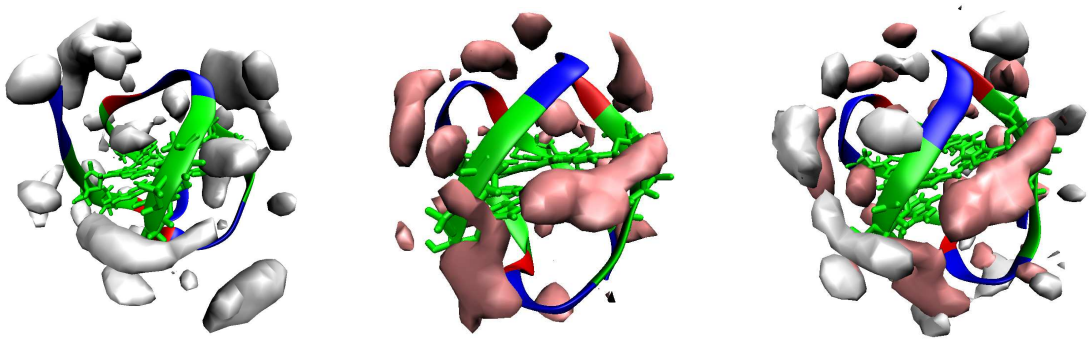
(a) pure water

(b) 1M TMAO

(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 54: Density distribution of TMAO, urea and water around DNA G-quadruplex structures simulated at 360 K in pure water, 1 M TMAO, 2 M urea and 1:2-molar TMAO:urea solutions. TMAO is colored in white, urea in pink and water in cyan. Reprinted with permission from [2].



(a) 1M TMAO (b) 2M urea (c) 1M TMAO and 2M urea

Figure 55: Density distribution of TMAO and urea around DNA G-quadruplex structures simulated at 360 K in 1 M TMAO, 2 M urea and 1:2-molar TMAO:urea solutions. TMAO is colored in white and urea in pink. Reprinted with permission from [2].

Similarly to systems simulated in 300 K, in pure TMAO solution TMAO tends to form outermost layer surrounding DNA surface (Fig. 55a), whereas urea shows some degree of insertion into the DNA main grooves (Fig. 55b).

In TMAO:urea mixture, both co-solutes form a mixed layer around DNA (Fig. 55c), with certain level of grooves insertion by urea. A good solvation of G-quadruplex by water is observed in pure water and in 1 M TMAO solution (Fig. 54a and 54b), whereas in 2 M urea water forms mainly an external solvation layer in favour of urea, which enters both DNA grooves effectively. In TMAO:urea mixture, similarly to systems simulated in 300 K, both co-solutes form a combined co-solute layer, which interferes with solvation water shell (Fig. 54d). However, the density of water molecules located between DNA residues appears to be greater than in case of mixed TMAO:urea system simulated in 300 K and system containing 2 M urea only. This suggests that TMAO protects DNA structure from detrimental impact of urea by securing an efficient hydration of the biomolecule.

7.2.3 DNA groove parameters: twist angle and groove size

With the application of DSSR software of 3DNA server [322], I analyzed G-tetrad twist angle (step parameter) and the size of quadruplex DNA grooves, expressed as the distance between consecutive phosphorus atoms (pink) of the residues forming G-tetrads: PP1, PP2, PP3 and PP4 as shown in Fig. 56. G-tetrad step parameters: twist and rise, allow to quantify the molecular *intercalation*. For *intercalating* co-solutes, small twist and large rise are observed. Additionally, the twist angle between two adjacent tetrads allows the analysis of the polymorphism of the G-quadruplex structure and its classification [102, 323] as well as the tension in the structure of quadruplex DNA based on its correlation with the twist angle [102].

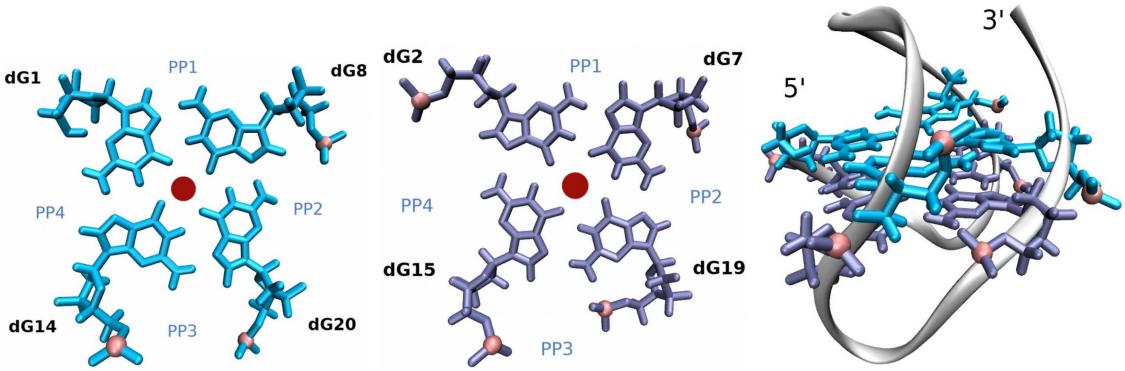
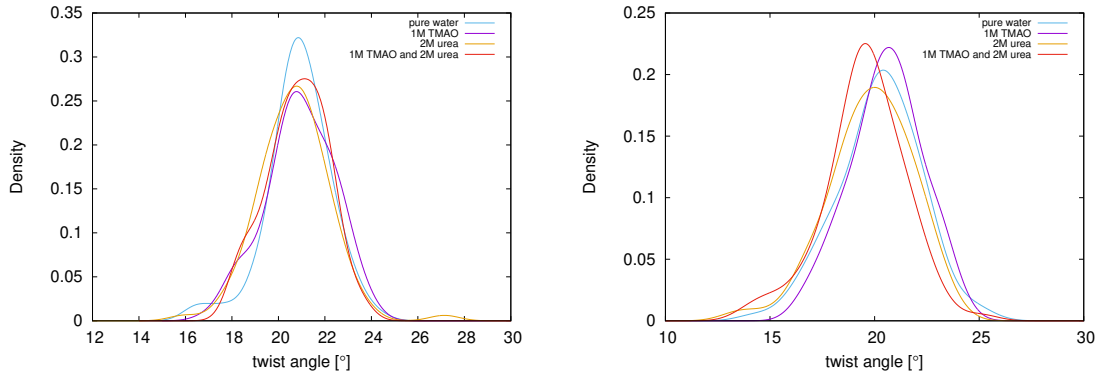


Figure 56: DNA G-quadruplex tetrad 1 (left) and tetrad 2 (middle) in vertical and in horizontal view (right). Reprinted with permission from [2].

The kernel distribution of twist angles between both tetrads of 2KF7 G-quadruplex simulated in 300 K and 360 K in different solutions are shown in Fig. 57.



(a) 300 K (b) 360 K

Figure 57: Distribution of twist angles between two adjacent tetrads of DNA G-quadruplex simulated in $T = 300$ K (left) and $T = 360$ K (right). Reprinted with permission from [2].

In native conditions, represented by 300 K simulation in pure water, most of twist angles between adjacent G-tetrads are confined in the range 20° - 22° (Fig. 57, left). It can be noticed that the presence of co-solutes in the simulation box generally extends the range of available twist angles. In pure TMAO solution, most of twist angles overlap with those observed in water. This supports the hypothesis that TMAO stabilizes the native conformation of biomolecules. In 2 M urea, the range of available twist angles is shifted towards slightly lower values of 19° to 21°. Smaller twist is the sign of possible insertion of urea molecules into DNA structure. Although the distribution of twist angles observed in 1:2 TMAO:urea mixture is slightly broader than in pure water, the general shape of kernel distribution curve is the same like in the nearly-native conditions. This indicates a potential stabilizing role of TMAO on DNA structure in the presence of denaturant urea.

The kernel distribution of twist angles for DNA systems simulated in 360 K is generally broader than for the simulation in near-physiological temperature 300 K (Fig. 57, right). Due to a thermal destabilization of DNA structure there are more twist angle conformations available. It can be noticed that kernel distribution of twist angles for simulations in 1 M TMAO is very similar to the twist distribution in 300 K for G-quadruplexes in pure water and in 1 M TMAO, with the maximum located around 20°. This indicates that TMAO stabilizes DNA native twist angle at ca. 20°. For the simulation in 360 K in pure water, most available twist angles are still confined in the range 20°-22°, but the distribution curve is generally broader than in 300 K, indicating increased flexibility of G-quadruplex structure. In general, the presence of co-solutes in the simulation box changes the range of available twist conformations by shifting the maximum of kernel density distribution. In 2 M urea, most twist angles are in the range 17°-23°, which is significantly broader than in 300 K and in 360 K simulation in pure water. This points out additional instability of G-quadruplex structure caused by denaturant urea, which adds up to thermal instability induced by increased temperature. In the simulation box with urea, where also TMAO is present, the maximum of twist angle distribution is still shifted towards smaller angles than in 300 K simulations or high temperature simulations in pure water or pure TMAO, but the distribution curve is not so broad like in pure urea solution. This suggests that even though urea causes the shift of favored twist angles towards lower values, the presence of TMAO still stabilizes DNA structure - although with new preferred twist angle, lower than in the native structure.

Fig. 58 shows kernel density distribution of groove sizes for G-quadruplex simulated in 300 K in different solutions. Size of the DNA grooves is represented by the distances between corresponding phosphorus atoms, where PP2 and PP3 correspond to the size of DNA major grooves, and PP1 and PP4 belong to the minor groove region.

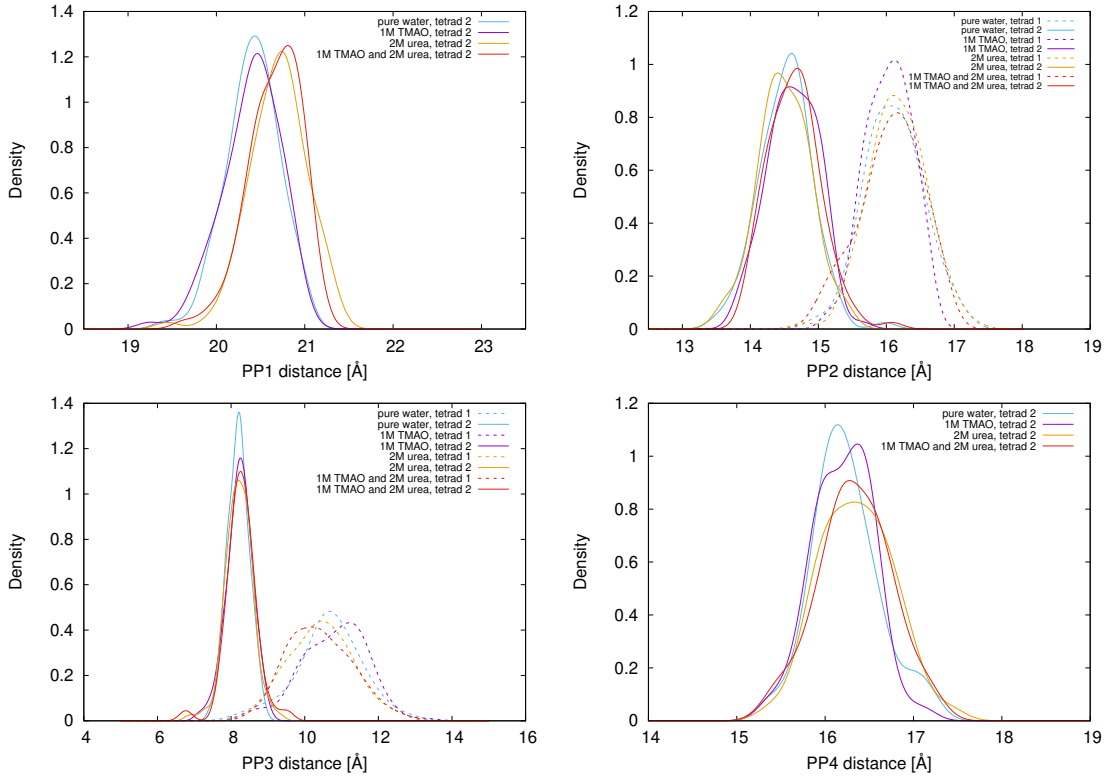


Figure 58: Size of DNA G-quadruplex grooves PP1 (top left), PP2 (top right), PP3 (bottom left) and PP4 (bottom right) for simulations at 300 K. Dashed lines represent the tetrad 1 and solid lines the tetrad 2. Reprinted with permission from [2].

The dimension of G-quadruplex major grooves, defined by the distances PP2 and PP3 between consecutive phosphorus atoms, seem not to be influenced by the presence of TMAO and urea in the simulation box. There is only minor influence of co-solutes on PP2 distance for both quadruplex tetrads (Fig. 58, top right), whereas the distance PP3 for tetrad 2 is very stable for all simulated systems and not influenced by the co-solutes (Fig. 58, bottom left). In tetrad 1, co-solutes cause slight shift in favored value of PP3 distance, which is directly translated into the changes in tetrad twist angle.

Unlike DNA major grooves, on which the impact of co-solutes is nearly negligible, the dimension of minor groove region defined by the distances PP1 and PP4 seem to be influenced by TMAO and urea to some degree. The favored PP1 distance is the same in 1 M TMAO solution as in pure water, with the maximum of its kernel distribution around 20.5 Å (Fig. 58, top left). The presence of urea, both alone and in the mixture with TMAO, shifts the preferred PP1 distance towards higher values of ca. 21 Å. It hints that urea causes slight swelling of G-quadruplex molecules, probably inserting into DNA minor groove regions, as suggested already by the analysis of urea density distribution in the previous chapter. The kernel distribution curve of allowed PP4 distances has similar shape and width in pure water and pure TMAO solution, with slight shift of the maximum towards bigger distances for TMAO (Fig. 58, bottom right). In the solutions containing urea, both alone and with TMAO, the distribution curve is broader, comprising higher range of available PP4 distances. This points out increased instability of DNA structure in the presence of urea. Moreover, in urea solutions the maximum of PP4 distances distribution is shifted towards slightly higher values, indicating swelling of DNA molecules possibly associated with urea insertion between its residues.

The analysis of groove size in the course of MD simulation was performed also for the systems simulated in high temperature conditions. The results obtained for 360 K are shown in Fig. 59.

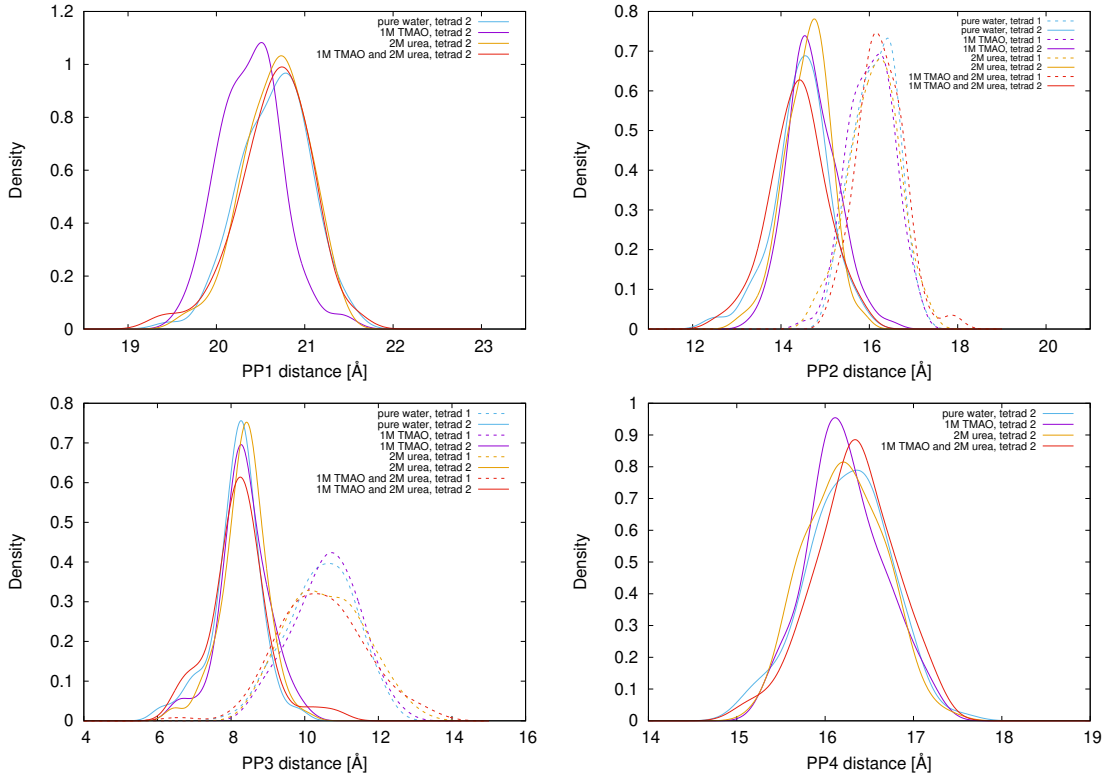


Figure 59: Size of DNA G-quadruplex grooves PP1 (top left), PP2 (top right), PP3 (bottom left) and PP4 (bottom right) for the high temperature simulations at 360 K. Dashed lines represent the tetrad 1 and solid lines the tetrad 2. Reprinted with permission from [2].

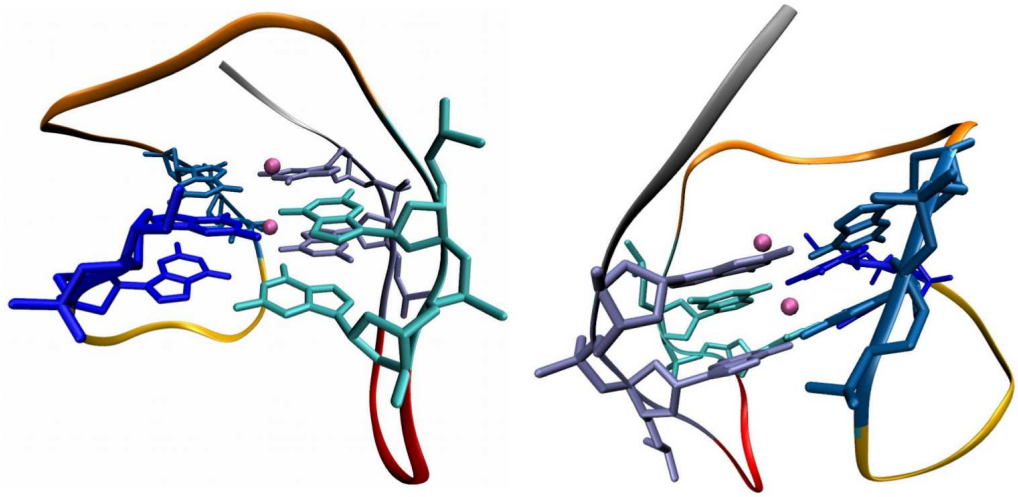
Similarly to the simulations in 300 K, also in high temperature simulated systems the influence of co-solutes on the size of the major grooves (defined by distances PP2 and PP3) is only minor, although considerably more pronounced than at room temperature. Like in 300 K, also in 360 K there is a remarkable impact of urea and TMAO on the size of minor groove region. In 300 K, the distance PP1 for most DNA conformations was confined in the range 20-21 Å for the systems containing urea, whereas in 360 K the favorable PP1 distance of 20-21 Å is observed not only in the solutions of 2 M urea and 1:2 TMAO:urea mix, but also in pure water (Fig. 59, top left). Shift towards higher values of PP1 distance for G-quadruplex simulated in near-native solution of pure water in high temperature suggests swelling of DNA molecule in comparison to the system simulated in 300 K. Moreover, the kernel distribution of possible PP1 conformations in water, urea and TMAO:urea mix is remarkably broader than in 300 K. The availability of more conformations indicates higher structural instability of DNA in high temperature conditions than in room temperature. What is interesting, the distribution of PP1 distances for G-quadruplex in 1 M TMAO is significantly more narrow than in pure water or in the presence of urea, and the shape of kernel distribution curve with maximum at ca. 20.5 Å resembles the curves recorded in 300 K for both 1 M TMAO and pure water. This indicates that TMAO stabilizes G-quadruplex structure, both against potentially denaturing agents like urea and against thermal destabilization in high temperature conditions. Contrarily to PP1 distance, for PP4 distance there are not much differences observed between the simulations in 300 K and 360 K, both in the shape of the kernel distribution curve and in the location of the distribution maximum (Fig. 59, bottom right). Also co-solutes seem to have only minor influence on the shape of this minor groove region.

When G-quadruplex grooves are concerned, the favored PP2 distances are generally similar to those observed for 300 K (Fig. 59, top right). In tetrad 1, co-solutes appear to have nearly no influence on PP2 distance, whereas in tetrad 2 the maximum of kernel distribution of PP2 distance is the same for water and 1 M TMAO, but shifted towards higher values in 2 M urea and towards lower values in TMAO:urea mixture. In terms of PP3 distance, groove size in 360 K is in principle similar like in 300 K, but the distributions of possible PP3 distances for both tetrads are broader (Fig. 59, bottom left). This shows that due to an increased DNA flexibility associated

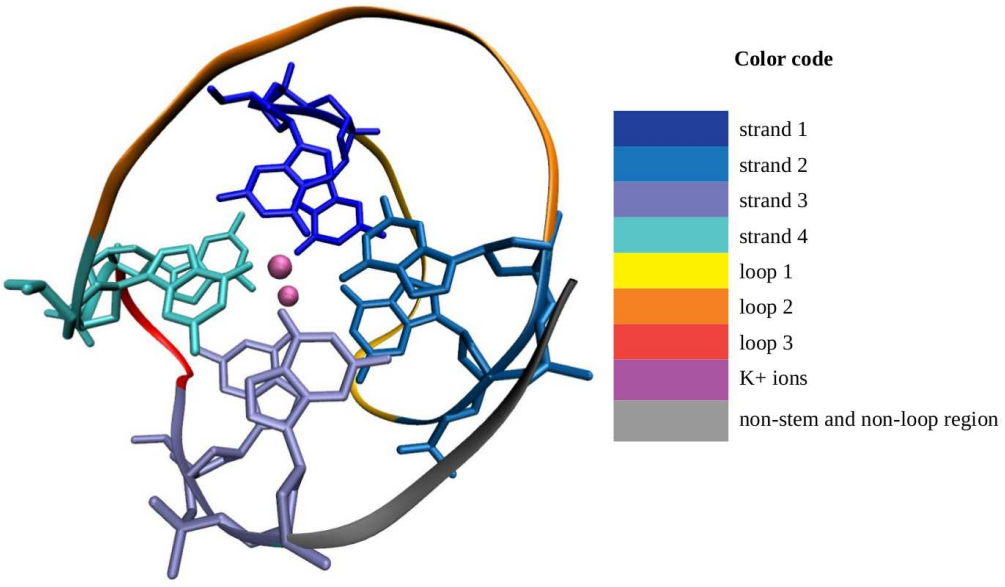
with its structural destabilization by high temperature, there are more conformations available. For tetrad 2, the presence of co-solutes does not change significantly the favored PP3 distance. However, for tetrad 1 the addition of urea alone or in combination with TMAO shifts the preferred PP3 towards lower values.

7.2.4 Inter- and intramolecular hydrogen bonds

I analyzed the number of hydrogen bonds formed between the molecules of water, TMAO and urea and particular parts of 2KF7 DNA G-quadruplex: four strands forming the stem and three loops, as well as A, T, G bases and phosphates. Together with intermolecular hydrogen bonds, I analyzed also the number of H-bonds formed within DNA structure in the course of MD simulation. The detailed analysis of G-quadruplex structure with separation into stem and loops is shown in Table 10 and in Fig. 60.



(a) side view



(b) top view

Figure 60: Structure of 2KF7 DNA G-quadruplex. Particular quadruplex stem and loops are marked with appropriate color code. Reprinted with permission from [2].

Table 10: Residues forming stem, loops and non-stem or non-loop regions of 2KF7 DNA G-quadruplex

G-quadruplex structure

Stem	residues	Loops	residues
strand 1	DG1, DG2	**1**	DG3, DT4, DT5, DA6
strand 2	DG8, DG7	**2**	DG9, DT10, DT11, DA12, DG13
strand 3	DG20, DG19	**3**	DT16, DT17, DA18
strand 4	DG14, DG15		

Residues forming neither stem nor loops
DG21, DT22 and stabilizing ions K23 and K24

The histograms of intramolecular hydrogen bonds are shown in Fig. 61.

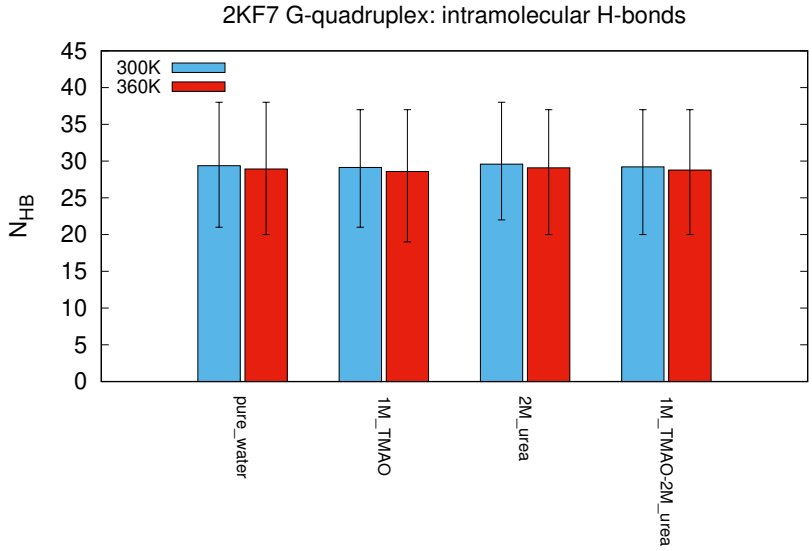


Figure 61: Number of intramolecular hydrogen bonds for DNA G-quadruplex simulated in 300 K and 360 K in water, 1 M TMAO, 2 M urea and 1:2 M TMAO:urea mixture. Reprinted with permission from [2].

The number of intramolecular hydrogen bonds within G-quadruplex structure appears to be affected by thermal phenomena to some degree. Raising the temperature from 300 K to 360 K results in slight reduction of the number of intramolecular H-bonds. Although urea is considered as biomolecular structure destabilizer, its presence in the simulation box does not influence the number of intramolecular H-bonds remarkably, regardless if it is present in standalone solution or in a mixture with TMAO. The presence of TMAO in the solution apparently also does not affect the average number of intramolecular H-bonds within G-quadruplex structure. This shows that any potential stabilization effects of TMAO and destabilization by urea are not associated with direct impact on DNA structure stabilizing hydrogen bonds, but rather a result of other kind of interactions, possibly affecting water solvation shell around DNA as suggested in the previous chapter discussing volumetric maps.

The intermolecular hydrogen bond numbers for all analyzed systems are shown in Fig. 62, 63, 64 and 65.

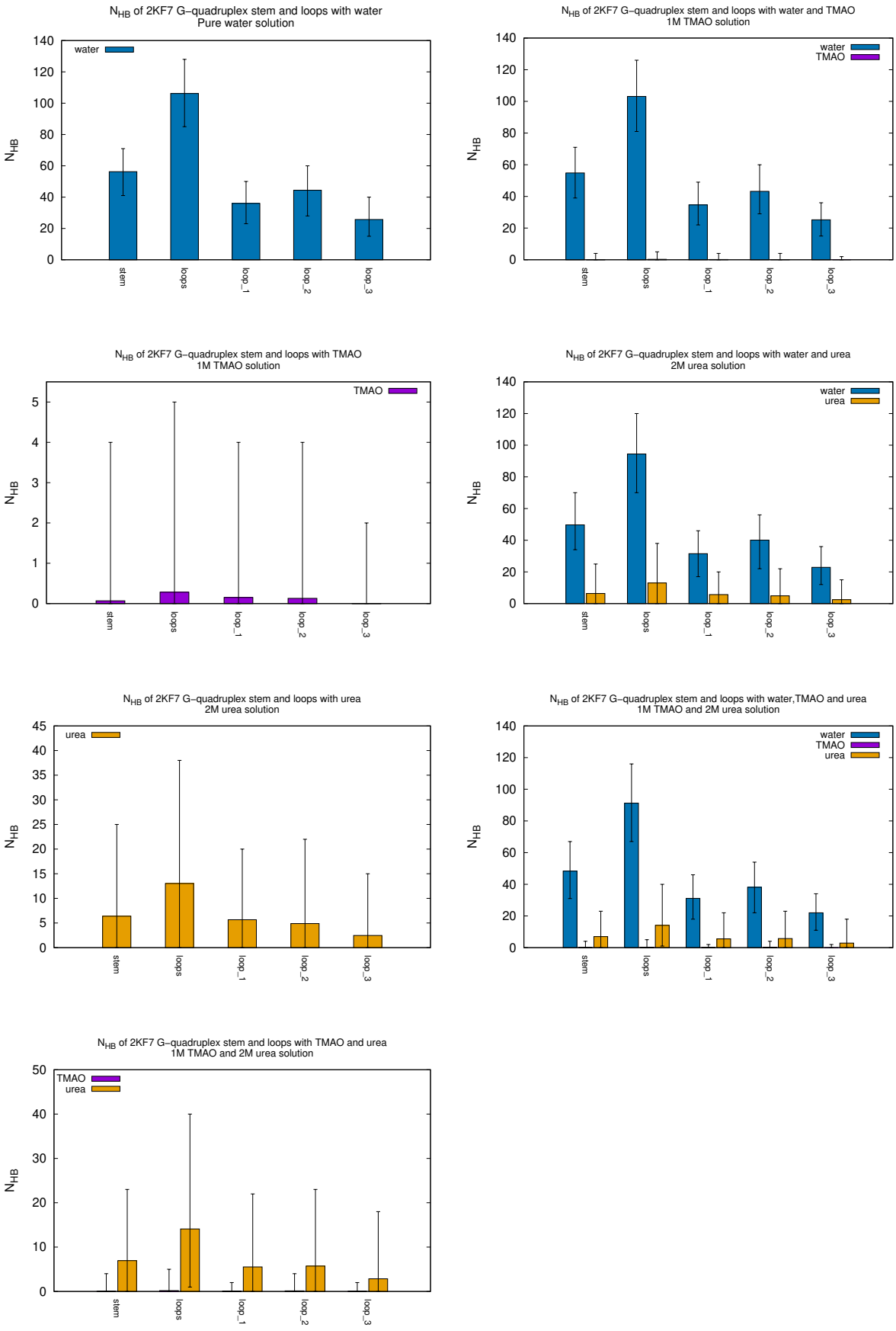


Figure 62: Number of hydrogen bonds between DNA stem and loops and co-solutes at 300 K. Reprinted with permission from [2].

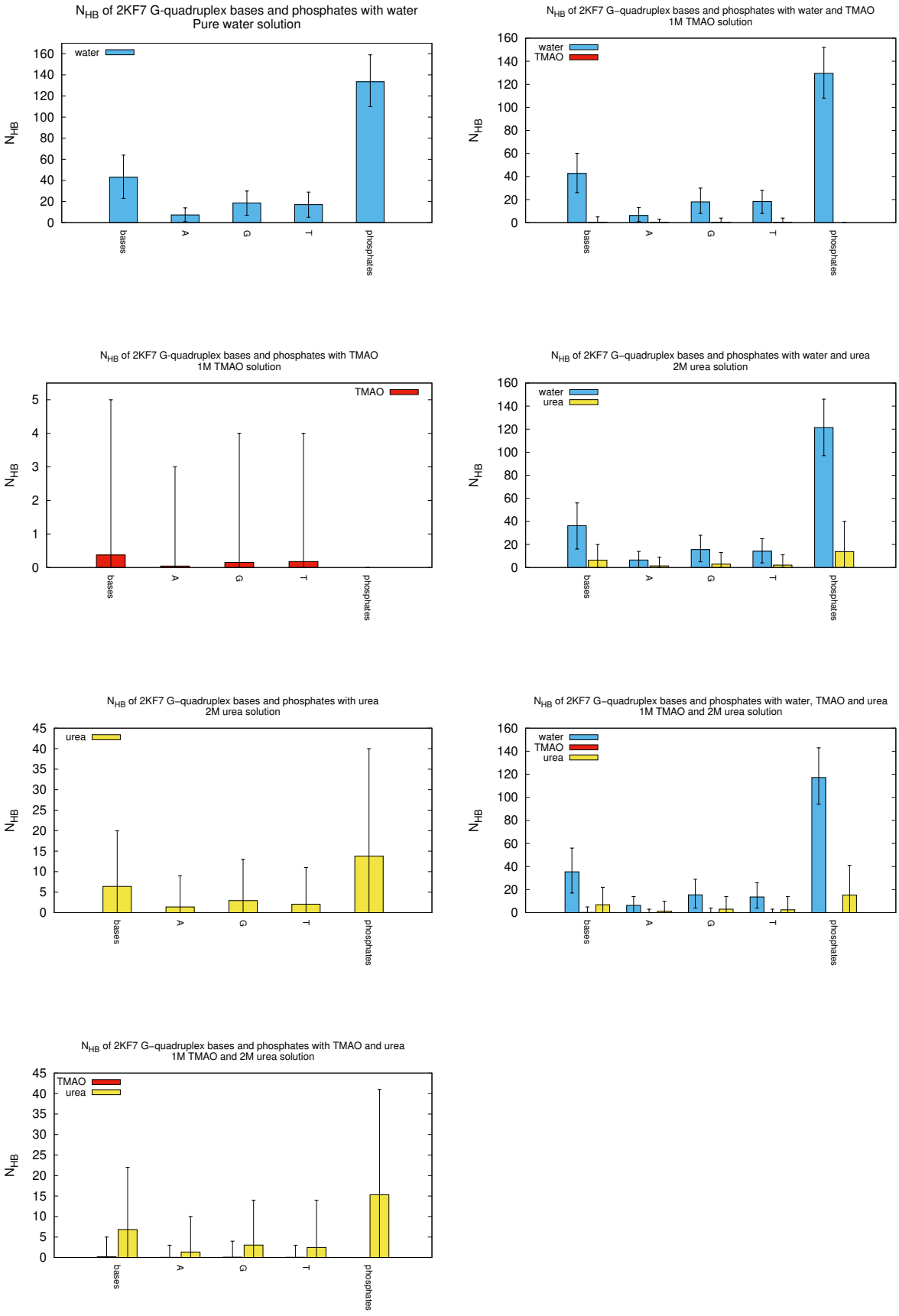


Figure 63: Number of hydrogen bonds between DNA bases and phosphates and co-solutes at 300 K. Reprinted with permission from [2].

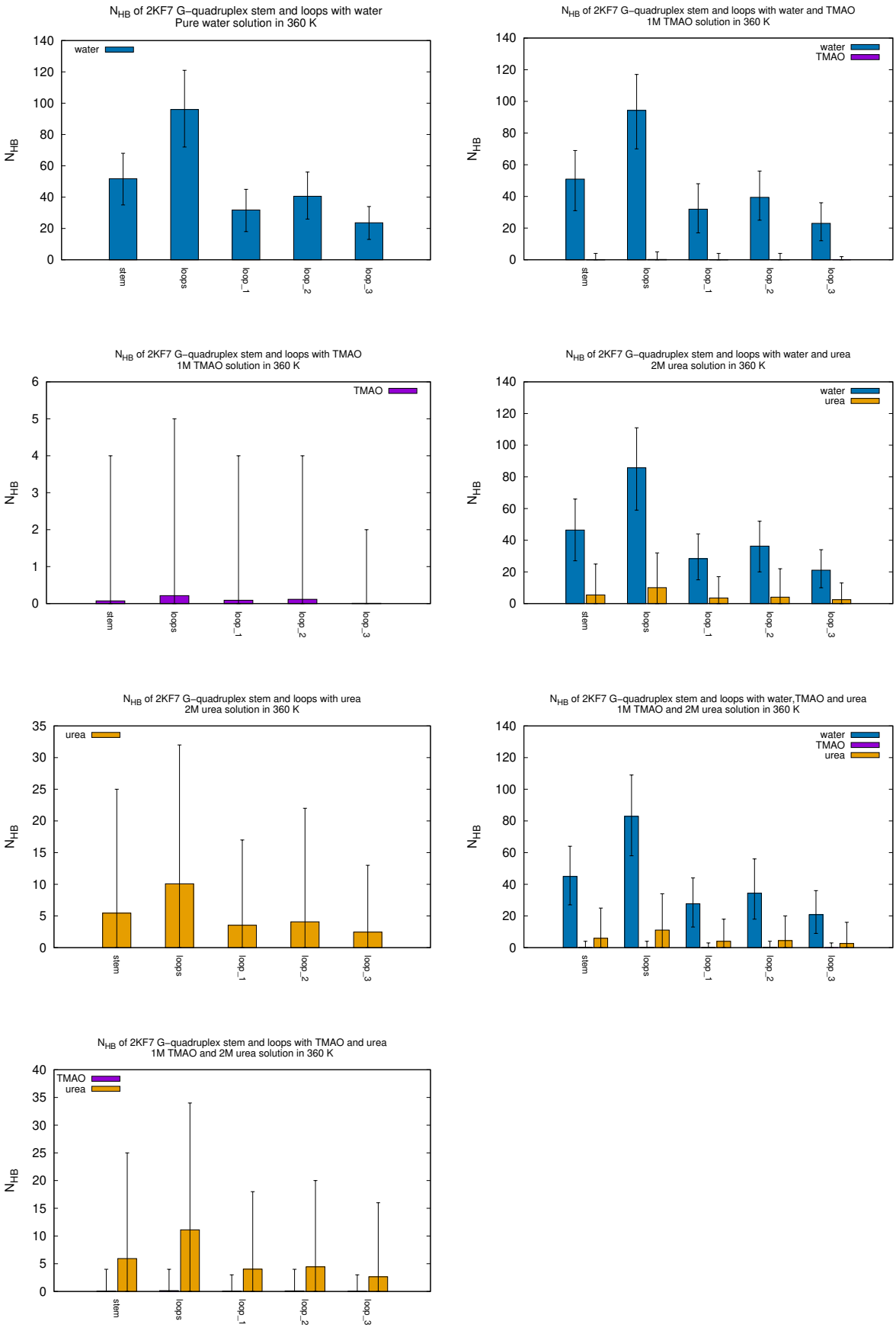


Figure 64: Number of hydrogen bonds between DNA stem and loops and co-solutes at 360 K. Reprinted with permission from [2].

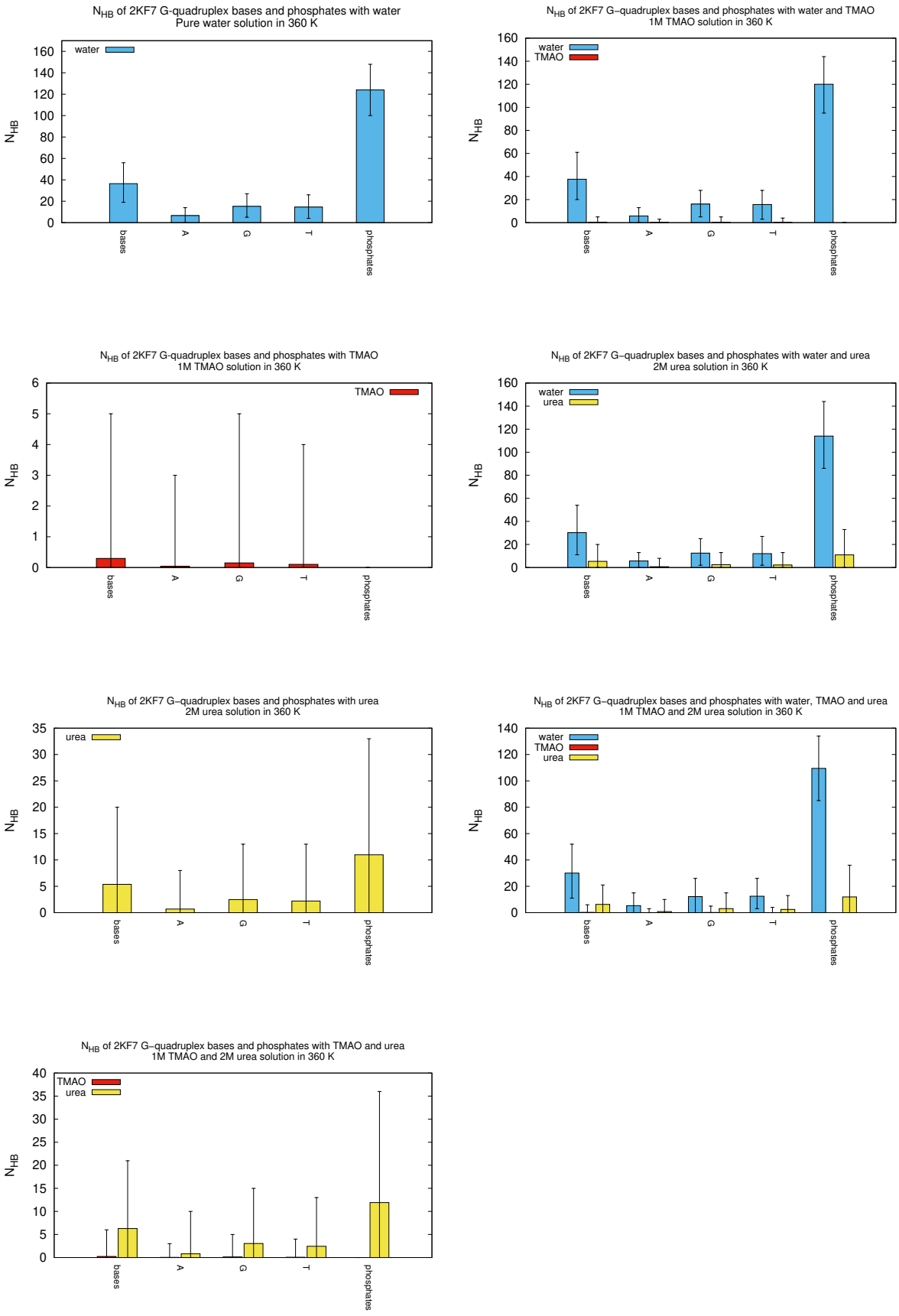


Figure 65: Number of hydrogen bonds between DNA bases and phosphates and co-solutes at 360 K. Reprinted with permission from [2].

Water molecules are found to form hydrogen bonds with all G-quadruplex parts preferentially over urea or TMAO, providing good hydration of both stem and loop regions as well as the regions in the vicinity of bases and phosphates. Phosphates are preferentially solvated over bases since they are more exposed to the environment than nucleic bases. The number of hydrogen bonds formed by water with particular bases is the highest for G and the lowest for A, which can be easily clarified by the sequence of 2KF7 DNA G-quadruplex, where the number of guanine residues (12 G) dominates over thymines (7 T) or adenines (3 A). Nonetheless, the observed formation of H-bonds with guanines is not self evident due to the fact that most G residues are involved in tetrad formation, which limits their exposition to the external environment and thus the capability to form H-bonds to some extent. The same binding tendency towards bases is observed also for TMAO and urea. As far as G-quadruplex stem and loops are concerned, water forms hydrogen bonds preferentially with loops than with stem region, with slight preference towards loop 2. In contrary to water, the number of hydrogen bonds formed with DNA by co-solutes is remarkably lower. TMAO forms hydrogen bonds with DNA only sporadically and their stability is very low. Some bonds are formed with G-quadruplex loops, preferentially over stem region. This can be associated with steric hindrance, since due to its size and pronounced negative charge the access of TMAO to the DNA stem is obstructed and the possibilities to hydrogen bond with G-tetrads is very limited. Since TMAO is missing hydrogen bond donor groups, it cannot bind with phosphates and H-bonds are being formed only with nucleic bases. In general, the number of H-bonds formed between urea and particular DNA parts is higher than for TMAO, but still much lower than for water. It can be observed that water preferentially hydrates DNA (see: Fig. 53), thus blocking possible binding sites from interaction with co-solutes. Urea generally prefers to form hydrogen bonds with phosphates than bases, and with loops rather than with stem. Indeed, the difference in the number of formed H-bonds is not so much pronounced like for water molecules. The formation of hydrogen bonds with more buried DNA parts, like the stem region or some more hidden regions of the loops, is facilitated by urea missing charge and smaller size in comparison to TMAO. For that reason, urea can enter the stem region and bind with the bases comparatively well as with phosphates or nucleic bases forming the loops. These observations are valid regardless if TMAO or urea are present in the DNA solution togeter or separately.

Binding preferences towards stem and loops or bases and phosphates are for both co-solutes and water the same independent on the temperature of the simulation. Although the number of H-bonds formed between water, TMAO and urea and particular DNA part is slightly lower in 360 K in comparison to the simulations in 300 K, the difference is not much pronounced and the same general binding preferences like in 300 K are observed. This suggests that all G-quadruplex parts remain same well hydrated also in high temperature conditions. Furthermore, the core DNA structure remains most possibly intact in the temperature 360 K due to the fact that the stem region and the bases forming it remain buried in the structure interior.

7.2.5 Number of contacts between DNA G-quadruplex and co-solutes

I calculated the average number of contacts between the atoms of G-quadruplex stem and loop regions and the centers of mass of TMAO and urea using FastNS (Fast Neighbor Search) utility of MdAnalysis [324]. The results are shown in Table 11.

These findings show that for both TMAO and urea there are slightly more contacts formed in the temperature 300 K in comparison to 360 K. This corresponds to the results of H-bonds analysis, where the number of formed hydrogen bonds has been found to be slightly decreased in high temperature. It can be observed that urea forms more contacts than TMAO with both stem and loops. However, the difference is not very pronounced when considering that the molarity of urea in the solution (2 M) is double as high as the molarity of TMAO (1 M). In general, for both co-solutes the number of contacts formed with DNA loops is higher than with stem region, mainly due to the better solvent accessibility of the loops in comparison to densely packed stacked G-tetrads forming the stem. Notably, there are no major differences in the number of contacts formed between TMAO or urea and DNA regardless if they are present in the simulation box separately or together as a mixture. Only the number of contacts between urea and both stem and loops appears to be slightly higher when urea is present in 1:2 mixture with TMAO than alone in water. However, the discrepancy between both types of solution is not far-reaching.

Table 11: Number of contacts between DNA G-quadruplex stem and loops and the co-solutes TMAO and urea in 300 K and 360 K molecular dynamics simulations. Reprinted with permission from [2].

System	300 K			
	stem		loops	
	TMAO	urea	TMAO	urea
1M TMAO	35.05 ± 0.01	-	44.90 ± 0.02	-
2M urea	-	82.03 ± 0.03	-	102.42± 0.03
1M TMAO and 2M urea	34.81 ± 0.01	85.67 ± 0.03	45.89 ± 0.02	107.12 ± 0.03
	360 K			
	stem		loops	
	TMAO	urea	TMAO	urea
1M TMAO	33.46 ± 0.01	-	43.08 ± 0.02	-
2M urea	-	75.49 ± 0.02	-	94.16 ± 0.03
1M TMAO and 2M urea	33.73 ± 0.01	79.28 ± 0.03	43.85 ± 0.02	98.87 ± 0.03

The number of contacts between DNA and other components of the solution within a sphere of arbitrarily chosen radius R can be also approached by the analysis of cumulative number radial distribution function (CN-RDF). Thus, I calculated the averaged number of TMAO, urea and water molecules at the distance R around near-spherical G-quadruplex molecule for the simulations in 300 K and 360 K. As the cutoff distance R, I selected the distance where the radial distribution of all co-solute and water molecules converges to the bulk (RDF = 1). Since for all the simulated systems and all the co-solute molecules the distance R ranged between 2.3 and 2.8 nm, for the reason of simplification for all the systems the R = 2.5 nm has been selected. The resulting CN-RDF values for the whole DNA molecule, approximated to the nearest integer, are presented in Table 12.

Table 12: Cumulative number radial distribution function of water, TMAO and urea around the whole DNA G-quadruplex molecule. Reprinted with permission from [2].

System	300 K		
	CN-RDF water	CN-RDF TMAO	CN-RDF urea
pure water	1972	-	-
1M TMAO	1840	38	-
2M urea	1739	-	86
1M TMAO and 2M urea	1611	38	90
System	360 K		
	CN-RDF water	CN-RDF TMAO	CN-RDF urea
pure water	1831	-	-
1M TMAO	1725	36	-
2M urea	1641	-	80
1M TMAO and 2M urea	1523	36	83

In high temperature simulations at T = 360 K, the results show slightly lower number of TMAO and urea molecules within the distance of 2.5 nm from DNA molecule in comparison to 300 K. Though, the differences are not large. However, the number of water molecules in the vicinity of DNA is remarkably lower in 360 K than in 300 K. This suggests that elevated temperature leads to slight dehydration of G-quadruplex, which as a consequence can lead to destabilization of its biomolecular structure. Notwithstanding, the solvation shell of the radius R = 2.5 nm around DNA molecule is filled mostly with the molecules of water, mainly due to the prevalence of water over co-solutes in the simulated systems. Similarly, nearly twice as much urea than TMAO molecules enter that solvation shell, which can be rationalized by the double molar density of urea in comparison to TMAO in the simulated systems (2 M urea vs. 1 M TMAO).

In addition to the calculations for the whole DNA structure, I calculated the averaged number of TMAO, urea and water molecules at the distance R around stem and loops of G-quadruplex for the simulations in 300 K and in 360 K. As the cutoff distance R, I chose arbitrarily R = 1.5 nm as the standard distance representing the close proximity of most spherical molecules. The resulting CN-RDF values approximated to the nearest integer are presented in Tables 13 and 14.

Table 13: Cumulative number radial distribution function of water, TMAO and urea around stem and loops of DNA G-quadruplex molecule simulated at the temperature 300 K. Reprinted with permission from [2].

System in 300 K	**stem**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	322	-	-
1M TMAO	302	5	-
2M urea	277	-	17
1M TMAO and 2M urea	258	5	18
System in 300 K	**loops**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	348	-	-
1M TMAO	325	6	-
2M urea	301	-	18
1M TMAO and 2M urea	208	6	19

Table 14: Cumulative number radial distribution function of water, TMAO and urea around stem and loops of DNA G-quadruplex molecule simulated at the temperature 360 K. Reprinted with permission from [2].

System in 360 K	**stem**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	299	-	-
1M TMAO	283	5	-
2M urea	262	-	15
1M TMAO and 2M urea	244	5	16
System in 360 K	**loops**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	322	-	-
1M TMAO	305	6	-
2M urea	283	-	16
1M TMAO and 2M urea	264	6	17

It can be noticed that the average number of molecules accumulated within the distance R = 1.5 nm from DNA surface is slightly higher for loops than for stem, possibly due to the higher exposition to the environment and consequently better solvent accessibility of quadruplex loops in comparison to relatively buried stem region. For both stem and loops, the number of accumulated water molecules is slightly lower in high temperature simulations in comparison to simulations at 300 K. At the same time, no major differences between simulations in 300 K and 360 K are observed for TMAO and urea. Nonetheless, approximately 3 times more urea than TMAO molecules are accumulated at the distance R = 1.5 nm from both stem and loops. This confirms the hypothesis of TMAO preferential exclusion from the DNA surface, whereas urea is being preferentially attracted.

8 Impact of the force field on the interaction profile of co-solutes with DNA

With reference to the molecular modelling in the biophysical context, the choice of the relevant force field is of utmost importance for the reliable estimation of inter- and intramolecular forces between atoms and the potential energy of a system. The interatomic potentials, constituting a basis for every force field, can be derived experimentally or in combination with quantum mechanical calculations [243]. Beside the function of the potential within the studied systems, force field defines also the parameters like dihedral angles, bonds and non-bond interactions or out-of-plane interactions, to name only few [243]. The wide variety of parameters described by a force field underpins the importance of its proper selection in order to secure the reliability and transferability of the results obtained via computer simulations.

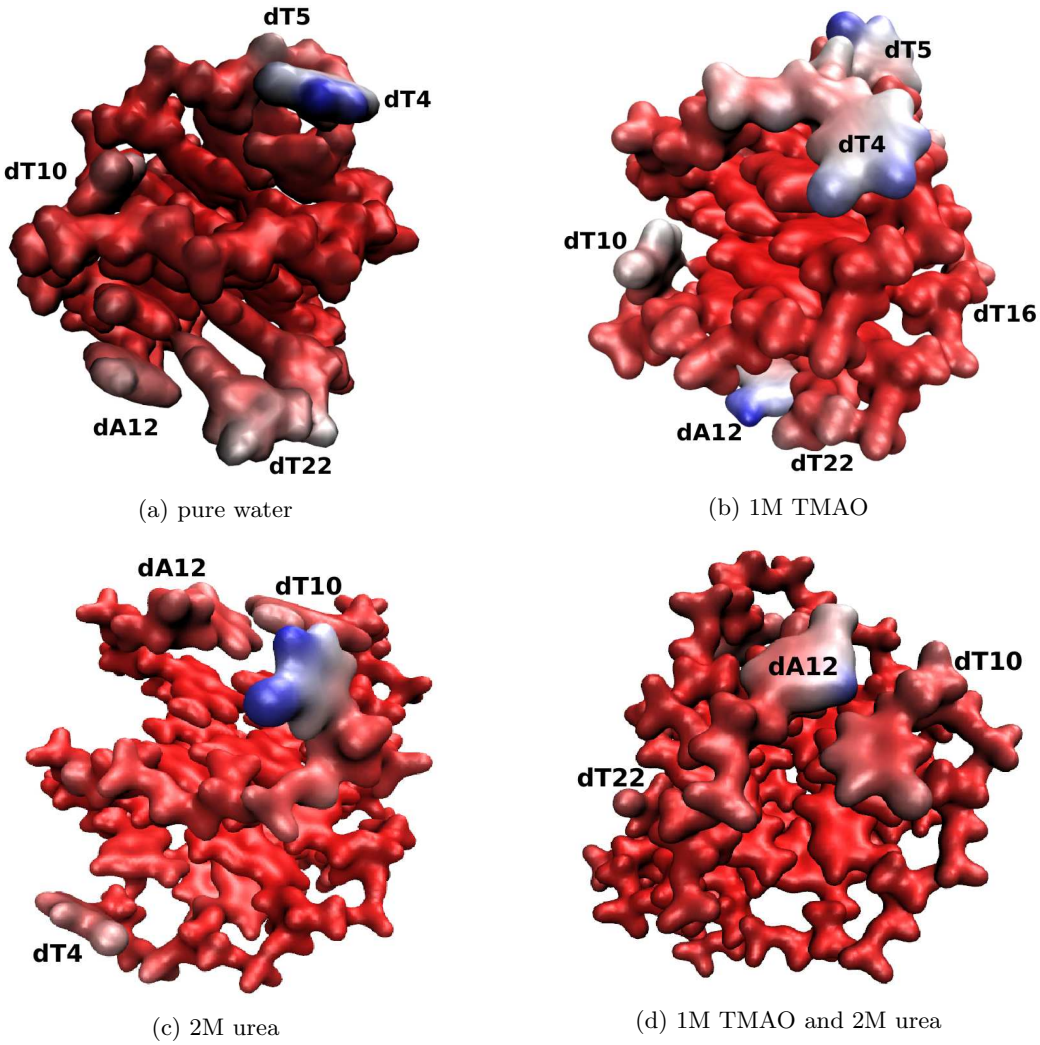
Despite extensive fitting procedures and their empirical validations, the definition of each force field encompasses certain limitations, which can exert an impact on the obtained outcomes [243, 325]. Hence, due to the restraints of quantum calculations and experimental procedures, the force field parameters of large macromolecules, such as proteins or DNA structures, are often derived from the studies of more accessible small organic molecules. Although such approach provides solid and reproducible results in the majority of computational studies, some more sensitive research, like calculations of free energy landscapes, may require additional force field refinement according to empirical data. An example of such study has been described in Ref. [312]. In this work, the classical ParmBSC1 force field for the simulation of nucleic acid structures [228] has been modified to reproduce better the experimental thermodynamic stability of the studied non-canonical DNA G-quadruplex structure [312]. The refined force field involved scaling down the van der Waals parameters for the DNA atom types O, NA, NB, NC and N2 by the factor of 0.975. It is worth noting that scaling down affects only the intramolecular interactions, leaving the original BSC1 values for the intermolecular interactions between DNA and water or ions [312]. The modified van der Waals radii and the atom types, for which the modification has been applied, are shown in Table 1 and in Fig. 1 in Ref. [312].

In Chapter 6, I studied the thermodynamic aspects of DNA G-quadruplex interactions with selected co-solutes basing on the molecular dynamics simulations in classical ParmBSC1 force field [228] as well as in vdW-refined force field described in Ref. [312]. In this chapter, I study the interaction of the same 2KF7 DNA G-quadruplex with TMAO and urea from biomolecular and more visual perspective basing on the MD simulations in vdW-modified ParmBSC1 force field [312]. The computer simulation procedure as well as the performed analyses have been described in details in Chapter 7. In contrast to the simulation in Chapter 7, for the purpose of this study I applied the refined ParmBSC1_vdW force field. The aim of this work is not to research on the possible limitations or benefits of certain force field, but to have insight into the impact of the force field selection on the MD simulation outcome. With this study, I intend to get an overview of the possible influence of the force field parametrization nuances on the validity and reproducibility of the computer simulation results as well as on the universality of selected force field for certain biomolecules.

8.1 Results and discussion

8.1.1 DNA flexibility descriptors in vdW-modified force field

The visualizations of B-factors (Eqn. 72) for DNA G-quadruplex structures simulated at $T = 300$ K in vdW-modified ParmBSC1 force field are shown in Fig. 66. The results of analogous computer simulations performed for all corresponding systems in unmodified ParmBSC1 force field are shown in Fig. 48.



(a) pure water

(b) 1M TMAO

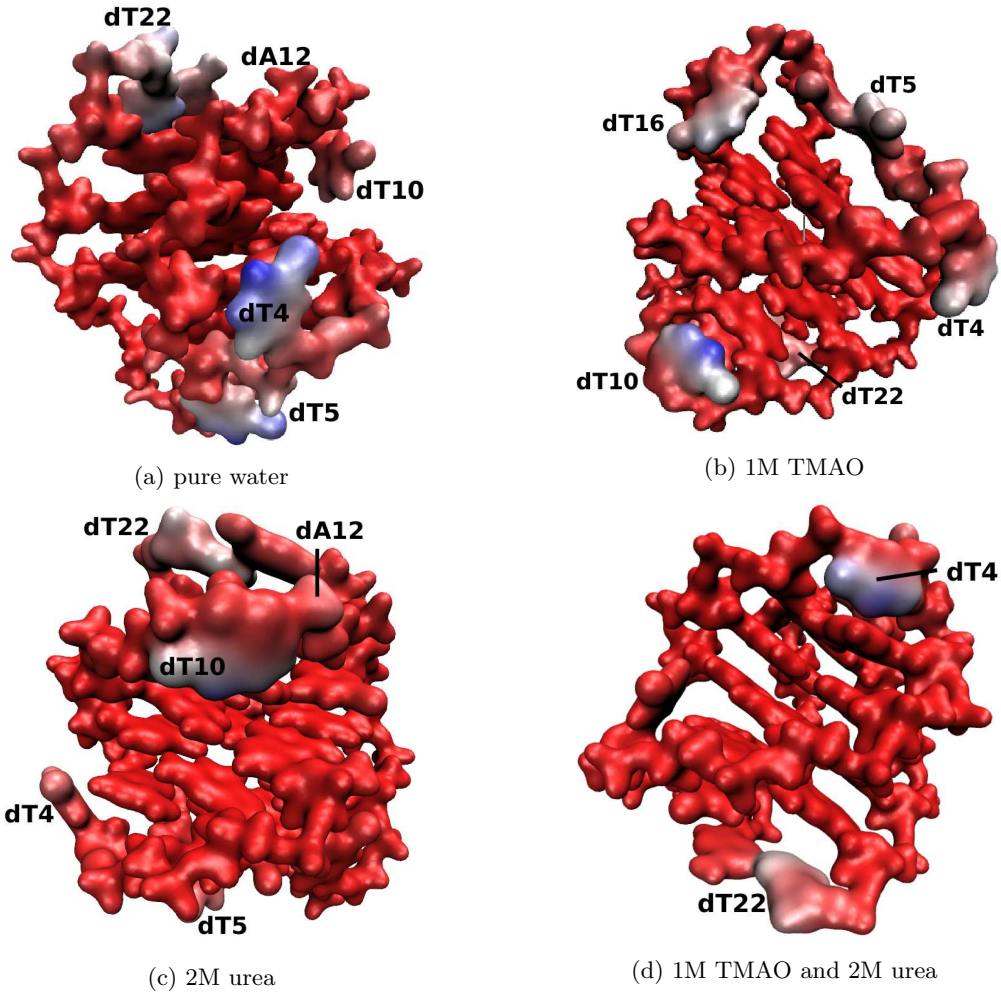
(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 66: Visualizations of B-factors for DNA G-quadruplex structures simulated in vdW-modified ParmBSC1 force field at $T = 300$ K in pure water (top left), 1 M TMAO (top right), 2 M urea (bottom left) and 1 M TMAO and 2 M urea solution (bottom right). The most flexible regions of the molecule are marked in blue and the most rigid in red.

Although some minor differences between individual systems simulated in ParmBSC1 and ParmBSC1_vdW force field exist, it can be noted that the general flexibility pattern is maintained regardless of the applied force field. For the systems simulated in the unmodified ParmBSC1 force field, the residues dT4, dT5 and dA12 showed an enhanced flexibility in pure water and in the presence of co-solutes, and analogous phenomemon is observed also in ParmBSC1_vdW force field. In pure water (Fig. 66a), similarly to ParmBSC1 simulations, the residues dT4 and dT5 are the most flexible, with reduced - although still remarkable - flexibiity registered for dA12, dT16, dT17 and some atoms of dT10. It is worth noting that also the flexibility of dT22 is enhanced in ParmBSC1_vdW simulations, although this residue remains relatively rigid in original force field. Analogously, the residues dT4, dT5 and dA12 are the most flexible in the ternary systems containing 1 M TMAO simulated in both force fields, with dT10 showing significantly reduced flexibility (Fig. 66b). In the mixtures of 1 M TMAO and 2 M urea, dA12 remains the most flexible residue regardless of the force field (Fig. 66d). However, dT4, whose flexibility was very pronounced in ParmBSC1 force field, appears to be less flexible in ParmBSC1_vdW simulations. For both force fields, residues dT5 and dT10 are characterized by enhanced flexibility, although lower than dA12. The remarkable differences in the residues flexibility between both force fields are observed only in the ternary systems with 2 M urea (Fig. 66c). In ParmBSC1_vdW simulations, dT10 appears to be the most flexible residue, whereas the flexibility of dT4 and dA12 is noticeably reduced. This is exactly the opposite phenomenon than observed in ParmBSC1 simulations, where dT4 and dA12 are the most flexible, and dT10 in comparison more rigid.

Analogously to the flexibility analysis in high temperature simulations in ParmBSC1 force field (Fig. 49), I analyzed the flexibility of the corresponding systems simulated in T = 360 K in vdW-modified force field. The results are shown in Fig. 67.



(a) pure water

(b) 1M TMAO

(c) 2M urea

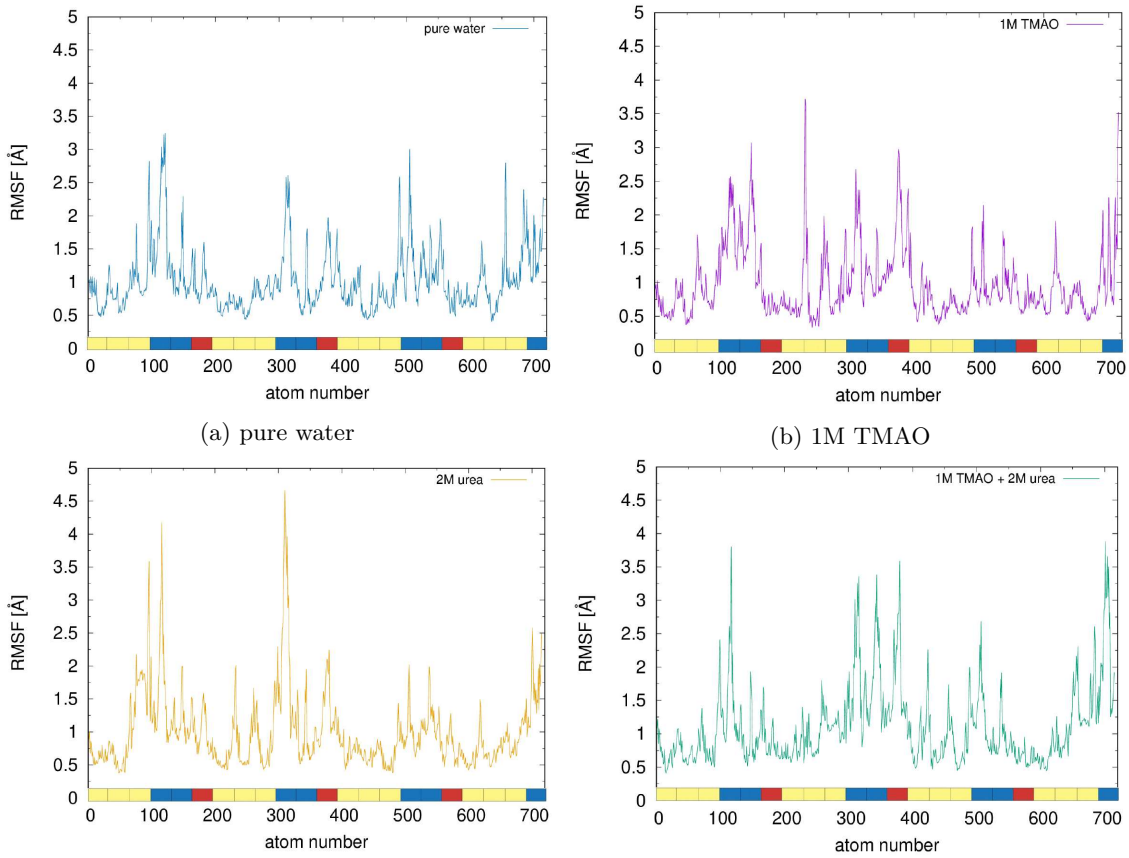
(d) 1M TMAO and 2M urea

Figure 67: Visualizations of B-factors for DNA G-quadruplex structures simulated in vdW-modified ParmBSC1 force field at $T = 360$ K in pure water (top left), 1 M TMAO (top right), 2 M urea (bottom left) and 1 M TMAO and 2 M urea solution (bottom right). The most flexible regions of the molecule are marked in blue and the most rigid in red.

Similarly to ParmBSC1 simulations, also in vdW-modified force field the residue dT22 exhibits enhanced flexibility in all studied systems, in particular in pure water and single 2 M urea solution. However, the differences between residue flexibility in both force fields for individual systems are more pronounced than in 300 K simulations. Although in pure water dT4 and dT22 are generally the most flexible residues, in ParmBSC1_vdW simulations also the enhanced flexibility of dT5 is revealed (Fig. 67a). This residue was relatively rigid in ParmBSC1 force field. Similarly to the simulations in unmodified force field, some atoms of dT10 and dA12 exhibit an enhanced flexibility. In 1:2 TMAO:urea mixture, for both force fields the highest flexibility is observed for dT4 (Fig. 67d). In vdW-modified force field, slightly enhanced flexibility is observed also for dT22. However, contrary to ParmBSC1 simulations, dA12 remains relatively rigid. Although relatively rigid in ParmBSC1 force field, the residue dT10 is the most flexible one in 1 M TMAO solution simulated in ParmBSC1_vdW force field (Fig. 67b). The residues dT4 and dA12, being the most flexible in unmodified force field, in vdW-modified one show reduced flexibility, together with dT5 and dT16. For ternary system with 2 M urea, dT22 is one of the most flexible residues regardless of the force field (Fig. 67c). However, dT10, which was relatively rigid in unmodified force field, shows here the flexibility comparable to dT22. The residues dT4, dT5 and dA12 are in contrary

characterized by only minor flexibility.

Consequently, for the simulations in T = 300 K in vdW-modified force field I calculated the RMSF profiles analogous to those presented in Fig. 50. The new results are shown in Fig. 68. Color code on the X-axis of each figure refers to particular nucleic residues: dG - yellow, dT - blue, dA - red.



(a) pure water

(b) 1M TMAO

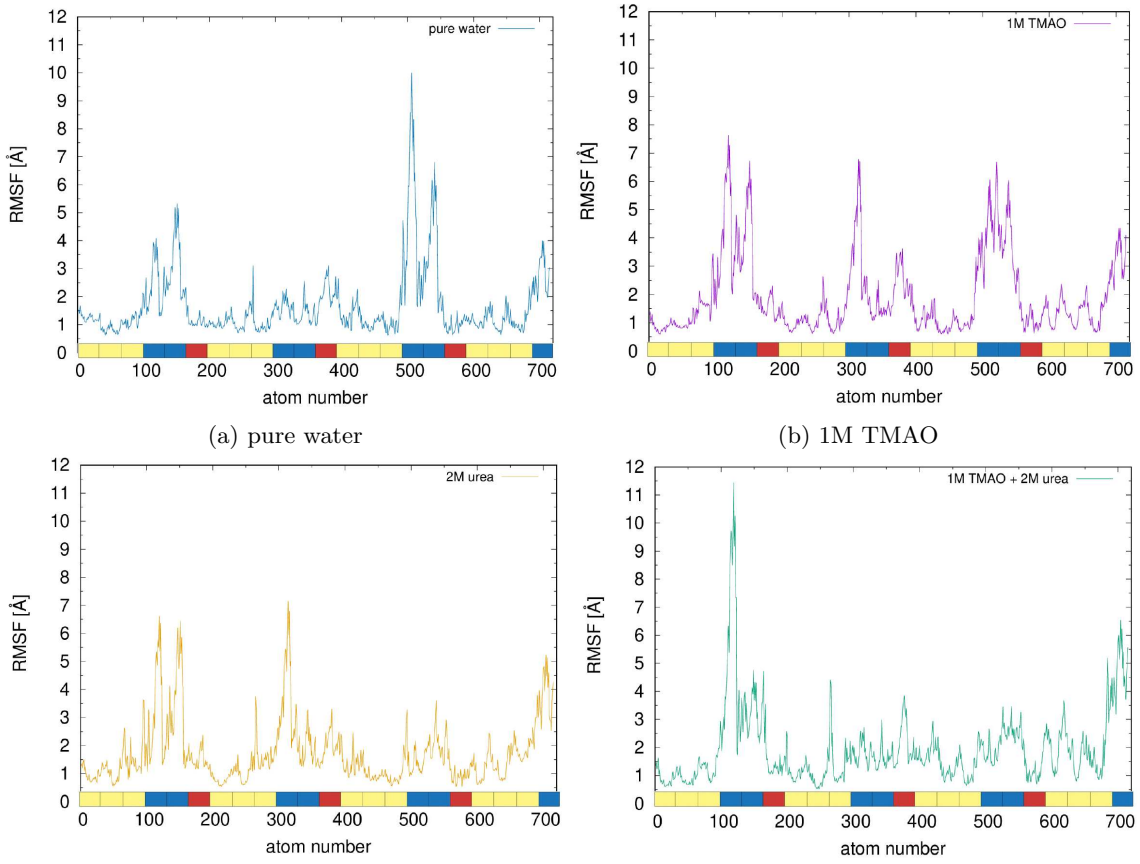
(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 68: RMSF of DNA G-quadruplex structure simulated at 300 K in different solutions in vdW-modified ParmBSC1 force field.

It can be noticed that the general RMSF profile for corresponding systems is preserved regardless of the applied force field, although the differences between the heights of the individual RMSF peaks exist. Simulations in both ParmBSC1 and ParmBSC1_vdW force fields reveal the enhanced flexibility of TTA triplets, whereas the flexibility of guanines is significantly reduced. For G-quadruplex simulated in pure water, the highest flexibility is observed for the first 5'-terminal TTA triplet, although the magnitude of the fluctuations is nearly double as high in ParmBSC1 force field as in vdW-modified one (6 Å compared to 3.7 Å). The magnitude of the fluctuations of remaining TTAs is similar in both force fields and counts up to approximately 3 Å. For ternary systems with 1 M TMAO, the magnitude of the corresponding TTA triplets fluctuations are the same regardless of the applied force field. However, in ParmBSC1_vdW simulations one can observe the enhanced flexibility of dG8 residue, which is not present in ParmBSC1 simulations. Analogously to the systems in pure water, also for the simulations in single 2 M urea solution the biggest discrepancy is observed in the fluctuation magnitude of the first 5'-terminal TTA triplet (2.7 Å in ParmBSC1 vs. 4 Å in its vdW-modified version), whereas the fluctuations of the second and third TTAs remain relatively similar and count up to 4 Å vs. 4.5 Å and 3 Å vs. 2 Å, correspondingly, starting from 5'-terminus. For the simulations in 1:2 TMAO:urea mixture, the magnitude of the fluctuations recorded in ParmBSC1 force field is generally lower than the fluctuations registered in ParmBSC1_vdW for the corresponding residues, although the overall fluctuation profile remains the same.

In the next step, I calculated the RMSF profile for high temperature simulations, analogous to those shown in Fig. 51. The results for the simulations in vdW-modified ParmBSC1 force field in $T = 360$ K are shown in Fig. 69.



(a) pure water

(b) 1M TMAO

(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 69: RMSF of DNA G-quadruplex structure simulated at 360 K in different solutions in vdW-modified ParmBSC1 force field.

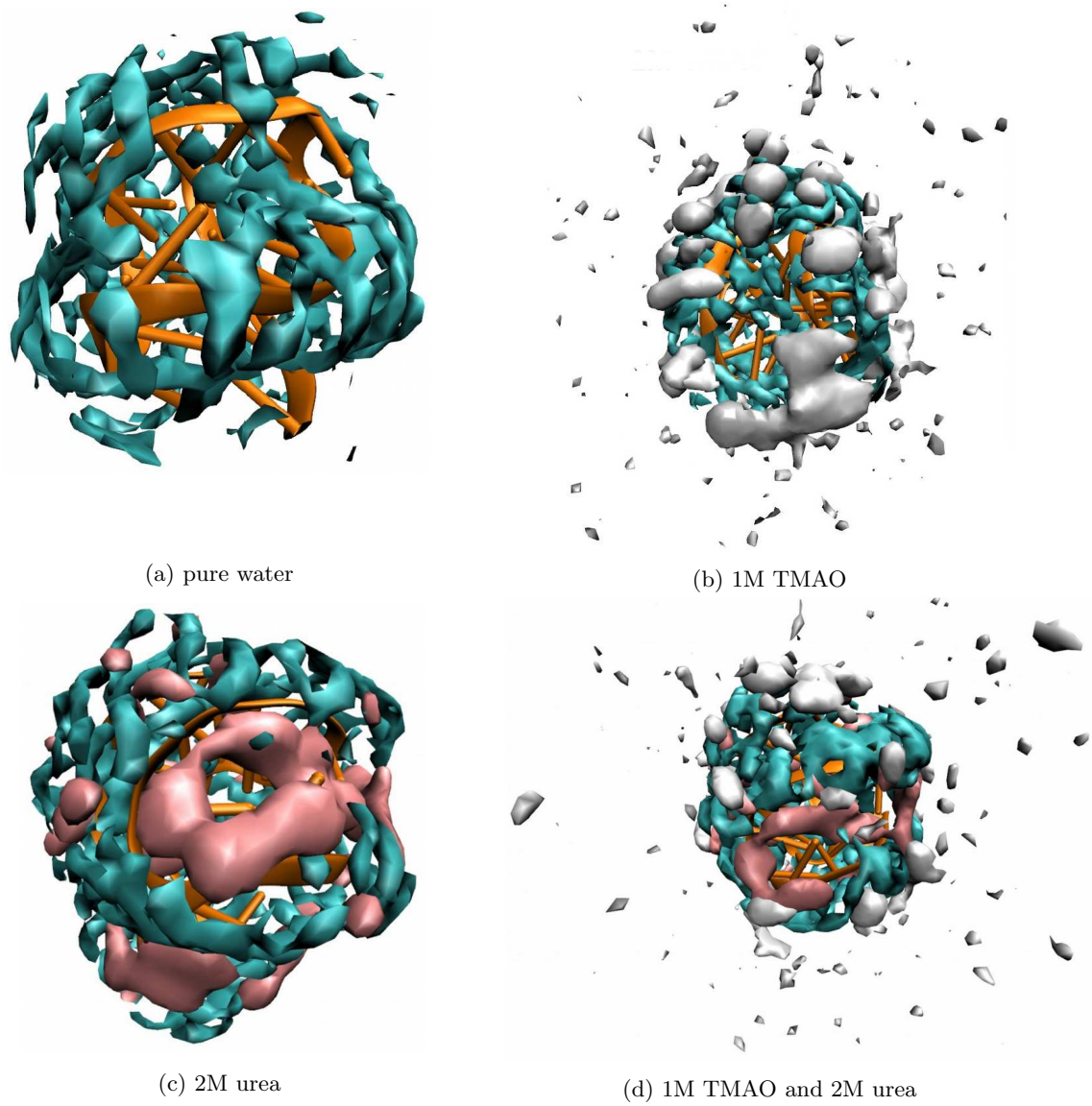
Similarly to the B-factor analysis, also the corresponding RMSF profiles exhibit significantly greater discrepancies between ParmBSC1 and ParmBSC1_vdW outcomes for the simulations in 360 K than in 300 K. Although TTA triplets appear to be the most flexible for all simulated systems in both force fields, the individual differences are not negligible. For G-quadruplex in pure water, in ParmBSC1 force field the fluctuation magnitude is comparable for all TTAs, whereas in ParmBSC1_vdW the fluctuation of 3'-terminal TTA is double as high as for 5'-terminal TTA and three times as high as for the central TTA. On the contrary, for G-quadruplex in 1 M TMAO solution the fluctuation altitudes of TTAs are comparable in ParmBSC1_vdW force field and count up to approximately 7 Å, which is more than the highest fluctuation reported in ParmBSC1 force field. In unmodified force field, 5'-terminal TTA exhibits the highest fluctuation with the altitude double as high as the 3'-terminal one. In single 2 M urea solution, the RMSF profile of the corresponding TTAs is similar for both force fields, although the altitudes observed in vdW-modified force field are on average 2 Å higher than in unmodified one. Similar relation can be observed for the TMAO:urea mixture, where the general fluctuation profile is analogous for both force fields. However, the fluctuations of the 5'-terminal TTA are higher in ParmBSC1_vdW force field, whereas two other TTAs fluctuate more in ParmBSC1.

These results show that the temperature of the simulation has the greatest impact on the results obtained for different force fields when the biomolecular flexibility is concerned. In room temperature, the results obtained for both force fields do not differ significantly and the minor discrepancies between force field parameters do not influence the general conclusion. However,

for high temperature studies the choice of appropriate force field reaches on importance. This is especially significant for the computer simulation methods involving temperature alterations within the systems, such as Well-Tempered Metadynamics (WT-MetaD) [267] or replica exchange molecular dynamics [326].

8.1.2 Density distribution profiles in vdW-modified force field

For the systems simulated in vdW-modified ParmBSC1 force field, the density distribution of TMAO, urea and water around DNA G-quadruplex structure has been visualized in the form of volumetric maps, analogous to those shown in Fig. 53 and Fig. 52. DNA backbone is represented by cartoon model, whereas the two G-tetrads are represented by stick model. Particular nucleic residues are marked with the following color code: dG - green, dA - red and dT - blue. For the co-solutes, the following color code has been applied: TMAO - white, urea - pink, water - cyan. The results for ParmBSC1_vdW force field are shown in Fig. 70 and in Fig. 71, where water molecules are excluded for the purpose of transparency.



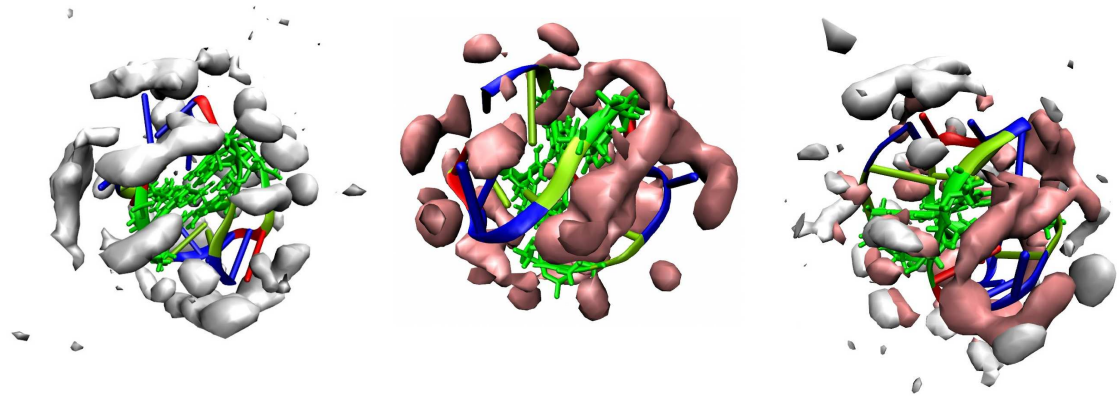
(a) pure water

(b) 1M TMAO

(c) 2M urea

(d) 1M TMAO and 2M urea

Figure 70: Density distribution of TMAO, urea and water around DNA G-quadruplex structures simulated in vdW-modified ParmBSC1 force field at 300 K in pure water, 1 M TMAO, 2 M urea and 1:2-molar TMAO:urea solutions. TMAO is colored in white, urea in pink and water in cyan.

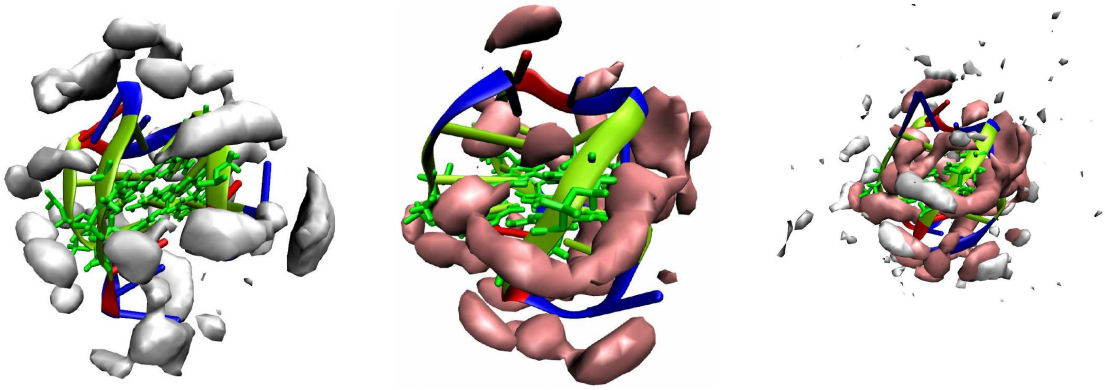


(a) 1M TMAO (b) 2M urea (c) 1M TMAO and 2M urea

Figure 71: Density distribution of TMAO and urea around DNA G-quadruplex structures simulated in vdW-modified ParmBSC1 force field at 300 K in 1 M TMAO, 2 M urea and 1:2-molar TMAO:urea solutions. TMAO is colored in white and urea in pink.

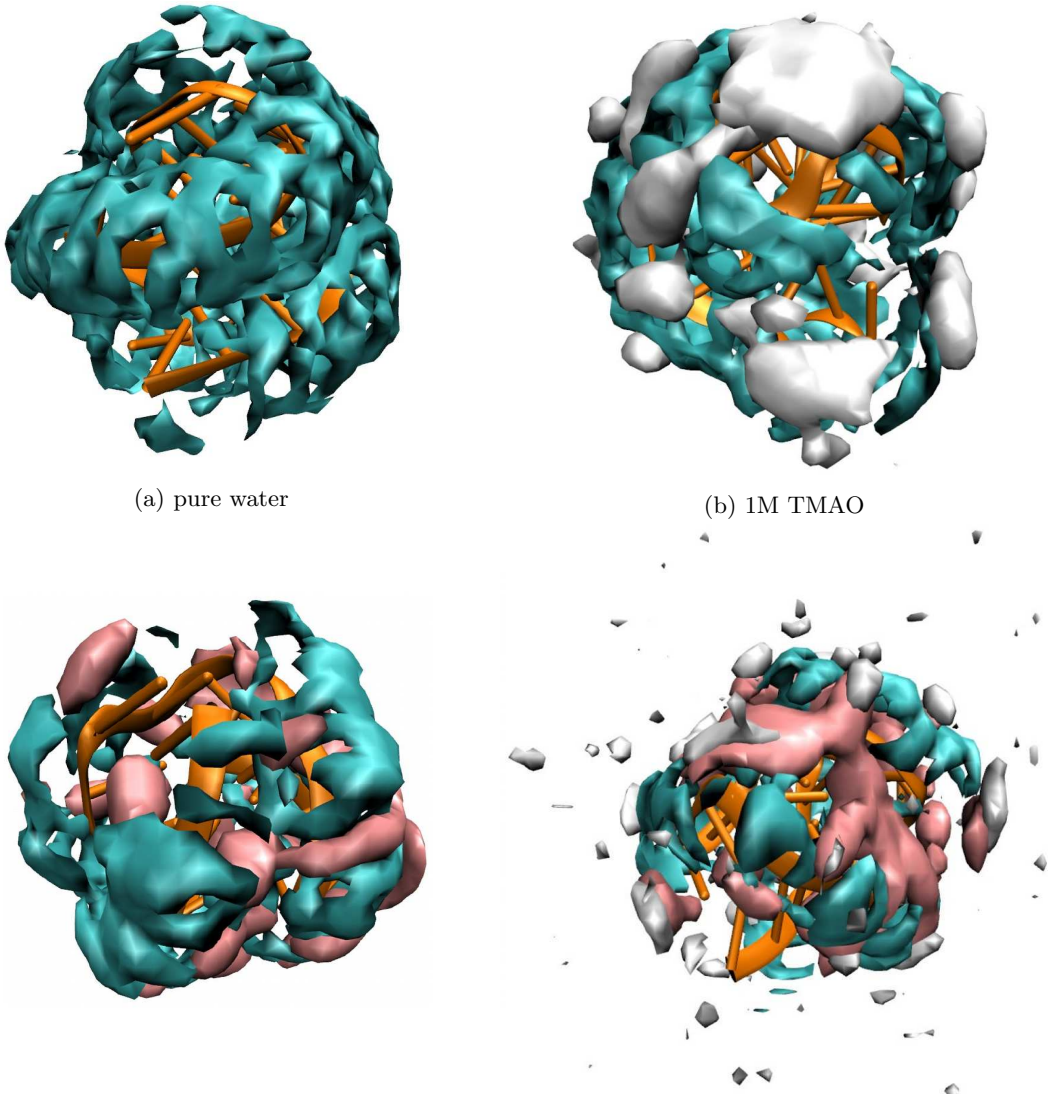
It can be noticed that the force field choice does not influence remarkably the observed distribution of co-solutes around DNA structure. Both volumetric maps created without (Fig. 71) and with the presence of water molecules (Fig. 71) represent the distribution of water and co-solutes around DNA G-quadruplex analogous to that reported in ParmBSC1 force field for corresponding systems. This shows that the conclusion drawn from the analysis of the volumetric maps for a given solution is independent on the applied force field relevant for the studied biomolecule.

I created also the volumetric maps for the systems simulated in $T = 360$ K in ParmBSC1_vdW force field. The corresponding results, analogous to those shown in Fig. 55 and Fig. 54 are shown in Fig. 72 and Fig. 73.



(a) 1M TMAO (b) 2M urea (c) 1M TMAO and 2M urea

Figure 72: Density distribution of TMAO and urea around DNA G-quadruplex structures simulated in vdW-modified ParmBSC1 force field at 360 K in 1 M TMAO, 2 M urea and 1:2-molar TMAO:urea solutions. TMAO is colored in white and urea in pink.



(a) pure water

(b) 1M TMAO

(c) 2M urea

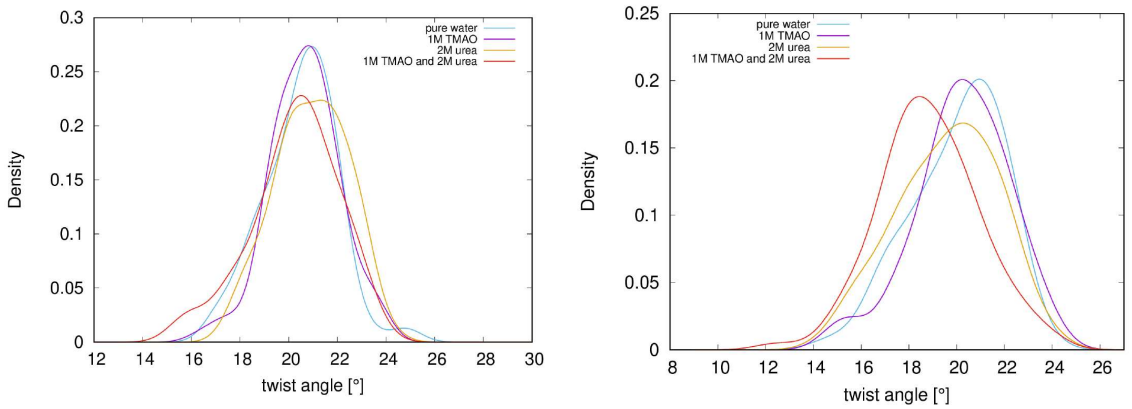
(d) 1M TMAO and 2M urea

Figure 73: Density distribution of TMAO, urea and water around DNA G-quadruplex structures simulated in vdW-modified ParmBSC1 force field at 360 K in pure water, 1 M TMAO, 2 M urea and 1:2-molar TMAO:urea solutions. TMAO is colored in white, urea in pink and water in cyan.

Similarly to the simulations in 300 K, also in high temperatures the force field choice does not influence the observed distribution of co-solutes around the central solute. However, it has to be noted that the differences between ParmBSC1 and ParmBSC1_vdW force fields are only minor and affect the van der Waals radii of selected atoms, leaving other parameters unchanged. Thus, the impact of slight vdW-modification on the obtained simulation results may not be so much pronounced as when some other DNA-relevant force field, like eg. CHARMM or BMS, are applied [327, 328]. This highlights the importance of the proper force field choice depending on the investigated parameters as well as the individual properties of the studied biomolecules [327, 328].

8.1.3 DNA groove parameters in vdW-modified force field

With the use of DSSR software of 3DNA server [322], I analyzed G-tetrad twist angle and the size of the quadruplex DNA grooves (see: Fig. 56) for G-quadruplex solutions simulated in vdW-modified ParmBSC1 force field. I applied the same analysis procedure as for corresponding systems simulated in unmodified ParmBSC1 force field (Figs. 57 and 58). The kernel distribution of twist angles between both tetrads of 2KF7 G-quadruplex simulated in 300 K and 360 K in different solutions in ParmBSC1_vdW force field are shown in Fig. 74.



(a) 300 K

(b) 360 K

Figure 74: Distribution of twist angles between two adjacent tetrads of DNA G-quadruplex simulated in $T = 300$ K (left) and $T = 360$ K (right) in vdW-modified ParmBSC1 force field.

Similar to the twist angle distribution in ParmBSC1 force field, also for the simulation in its vdW-modified version in 300 K most twist angles are confined within the range 20°-22°. In both force fields, the presence of co-solutes in the simulation box extends the range of available twist angles. It can be noticed that in the ParmBSC1_vdW force field the profile of twist angle distribution for single 1 M TMAO solution overlaps with the distribution profile in pure water more ideally than in the simulations in ParmBSC1 force field, thus providing even stronger confirmation of the stabilizing impact of TMAO on the native conformation at ca. 21° twist angle. However, in ParmBSC1 force field the profiles of twist angle distribution in ternary 2 M urea solution and in TMAO:urea mixture overlap very closely, whereas in ParmBSC1_vdW force field the overlap of corresponding profiles is not so exact. One can notice that in 2 M urea solution the maximum of the distribution is shifted towards larger angles than in ParmBSC1 force field. For 1:2 TMAO:urea mixture, the simulations in ParmBSC1_vdW force field reports the DNA conformations with smaller twist angles, which are not available in ParmBSC1 simulations.

The differences between the results for ParmBSC1 force field and its vdW-modified version observed in 300 K simulations become remarkably smaller when high temperature simulations in 360 K are considered (Fig. 74b and 57b). In both force fields, the distribution of twist angles is broader in 360 K than in 300 K, with more conformations being available. It can be observed that in ternary solutions with 1 M TMAO, the co-solute stabilizes the native angle at ca. 21° in both force fields. In 2 M urea and 1:2 TMAO:urea mixtures, the maximum of twist angle distribution is shifted towards smaller angles in comparison to the simulations in pure water. This observation is valid both for ParmBSC1 force field and its vdW-modified version. Slight differences are observed only for single 1 M TMAO mixture. In comparison to the simulations in pure water, in ParmBSC1_vdW force field the maximum of twist angle distribution is shifted towards smaller angles, whereas in ParmBSC1 force field the maximum is located at the same angles as in pure water.

The kernel density distribution of groove sizes for G-quadruplex simulated in 300 K in ParmBSC1_vdW force field is shown in Fig. 75. Analogously to the results for ParmBSC1 force field (Fig. 58), the size of DNA grooves is represented by the distances between corresponding phosphorus atoms, where PP2 and PP3 correspond to the size of DNA major grooves, and PP1 and PP4 belong to the minor groove region.

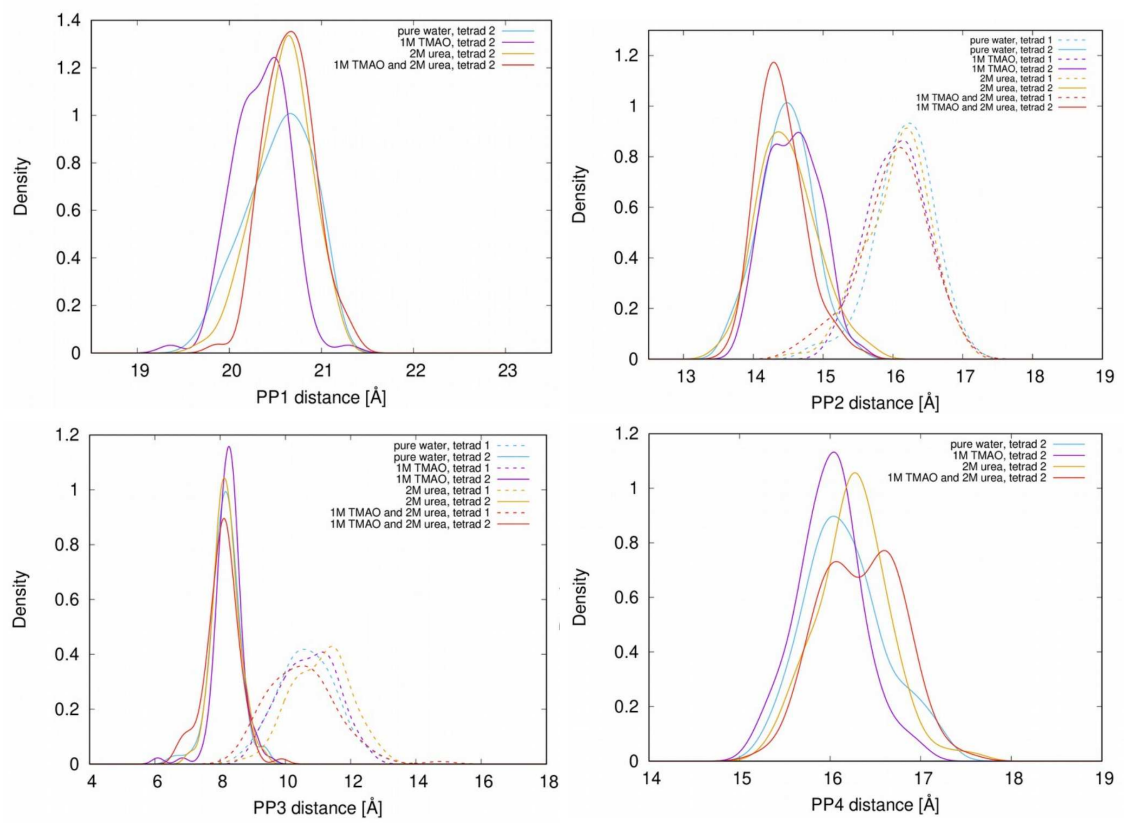


Figure 75: Size of DNA G-quadruplex grooves PP1 (top left), PP2 (top right), PP3 (bottom left) and PP4 (bottom right) for simulations in 300 K in vdW-modified ParmBSC1 force field.

In contrary to the twist angle distribution, there are remarkable differences between the distribution of groove sizes for corresponding systems simulated in both force fields. Although the PP1 groove size for native conformation as expressed by the simulation in pure water is analogous in both force fields and counts up to 20-21 Å, for the mixtures with co-solutes the differences in PP1 distributions between force fields are very pronounced. Although the distributions of PP1 distances in 2 M urea and TMAO:urea mixture overlap in both force fields, the distributions for pure water and 1 M TMAO obtained in the simulations in ParmBSC1_vdW force field do not show the overlap observed in ParmBSC1 force field. Analogously, also the distribution profiles of PP4 distances, which constitute the minor groove region together with PP1, show remarkable divergences between both force fields. However, the differences between distribution profiles become smaller when the major grooves are considered. The distribution profiles of both PP2 and PP3 exhibit the same characteristics and the location of the distribution maxima regardless of the applied force field. These observations point out the importance of the proper force field parametrization depending on the studied biomolecular sructure as well as on the analyzed parameters.

The kernel density distribution of groove sizes for G-quadruplex simulated in 360 K in ParmBSC1_vdW force field, analogous to those obtained for ParmBSC1 force field (Fig. 59), is shown in Fig. 76.

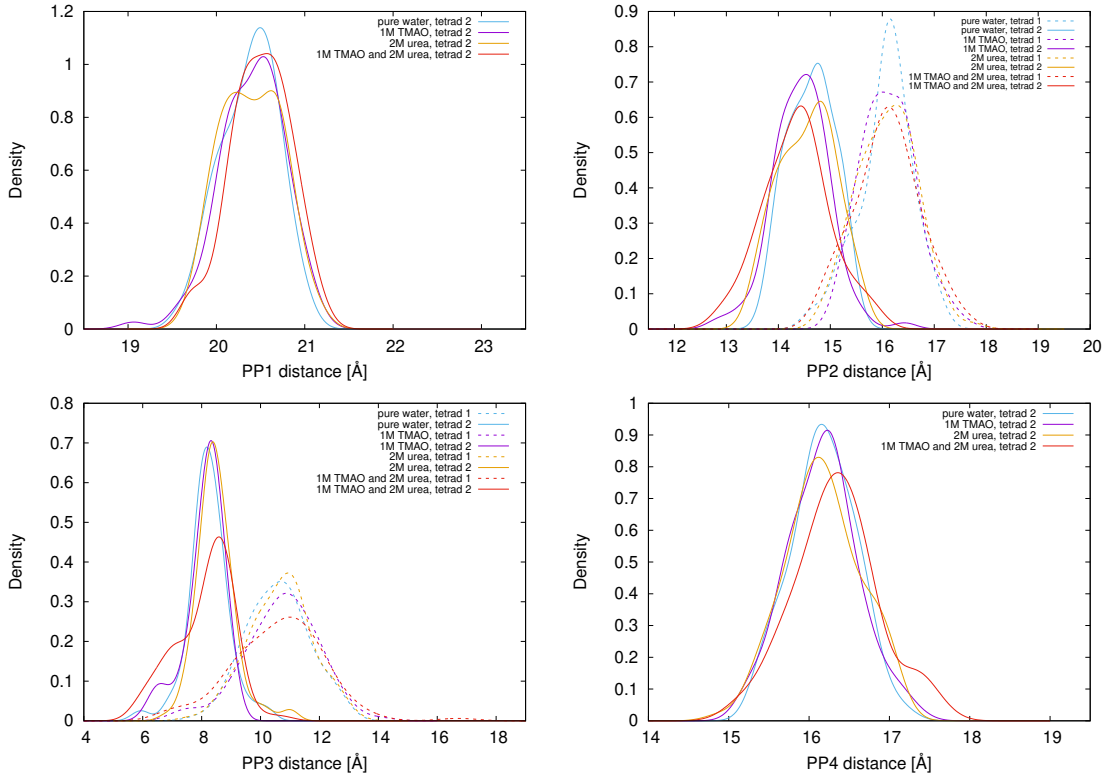


Figure 76: Size of DNA G-quadruplex grooves PP1 (top left), PP2 (top right), PP3 (bottom left) and PP4 (bottom right) for simulations in 360 K in vdW-modified ParmBSC1 force field.

Similarly to the simulations in 300 K, also in high temperatures the differences between the distribution of major groove sizes (distances PP2 and PP3) for both force fields are not so much pronounced as the differences in the distribution of size of the non-groove region (distances PP1 and PP4). Both ParmBSC1 and ParmBSC_vdW force fields reveal analogous distribution profiles of PP2 and PP3 distances for corresponding mixtures. When PP1 and PP4 distances are concerned, the simulations in ParmBSC1 force field show a shift towards lower values for G-quadruplex in 1 M TMAO solution in comparison to the simulations in pure water. This shift is not observed in vdW-modified force field. This confirms the previous conclusion about the importance of proper force field parametrization depending on the intended application.

8.1.4 Inter- and intramolecular hydrogen bonds in vdW-modified force field

The histograms of intramolecular hydrogen bonds for the systems simulated in ParmBSC1_vdW force field are shown in Fig. 77. Corresponding results obtained for the simulations in ParmBSC1 force field are shown in Fig. 61. The analysis of G-quadruplex structure with separation into stem and loops is shown in Table 10 and in Fig. 60.

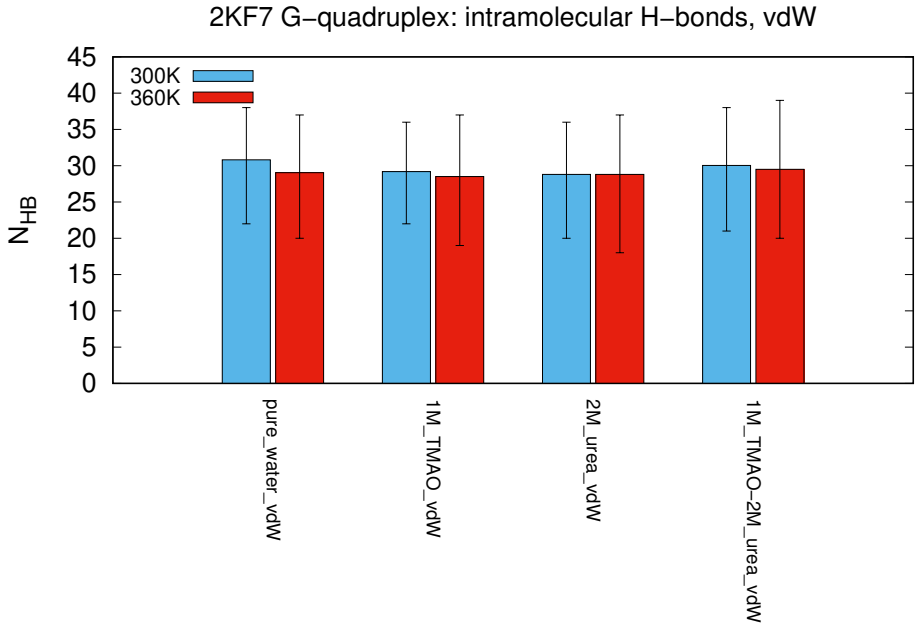


Figure 77: Number of intramolecular hydrogen bonds for DNA G-quadruplex simulated in 300 K and 360 K in water, 1 M TMAO, 2 M urea and 1:2M TMAO:urea mixture in vdW-modified ParmBSC1 force field.

The general trend of the intramolecular hydrogen bonds number reported for the simulations in ParmBSC1_vdW force field does not show any remarkabke differences to that reported for the simulations in ParmBSC1 force field for corresponding solutions. Hence, the same conclusion can be drawn from the simulations in both force fields.

As the next step, I calculated the numbers of intermolecular hydrogen bonds for all analyzed systems simulated in vdW-modified ParmBSC1 force field. The results for particular parts of the DNA structure are shown in Figs. 78, 79, 80 and 81. The corresponding results obtained for ParmBSC1 force field are shown in Figs. 62, 63, 64 and 65.

For the simulations in $T = 300$ K, there are no apparent differences in the number of hydrogen bonds formed between particular components of the solution and DNA stem and loops or bases and phosphates depending on the applied force field. Both ParmBSC1 and its vdW-modified version reflect the same histogram profiles for corresponding mixtures, indicating identical proportions between the number of hydrogen bonds formed with certain loop, stem, bases or phosphates for a given solution.

The same lack of dependency of the results on the applied force field can be observed in high temperature simulations in $T = 360$ K (Figs. 80 and 81). This shows that both ParmBSC1 force field and its vdW-modified version reflect properly the number of hydrogen bonds formed in the studied systems and have no major influence on the obtained results.

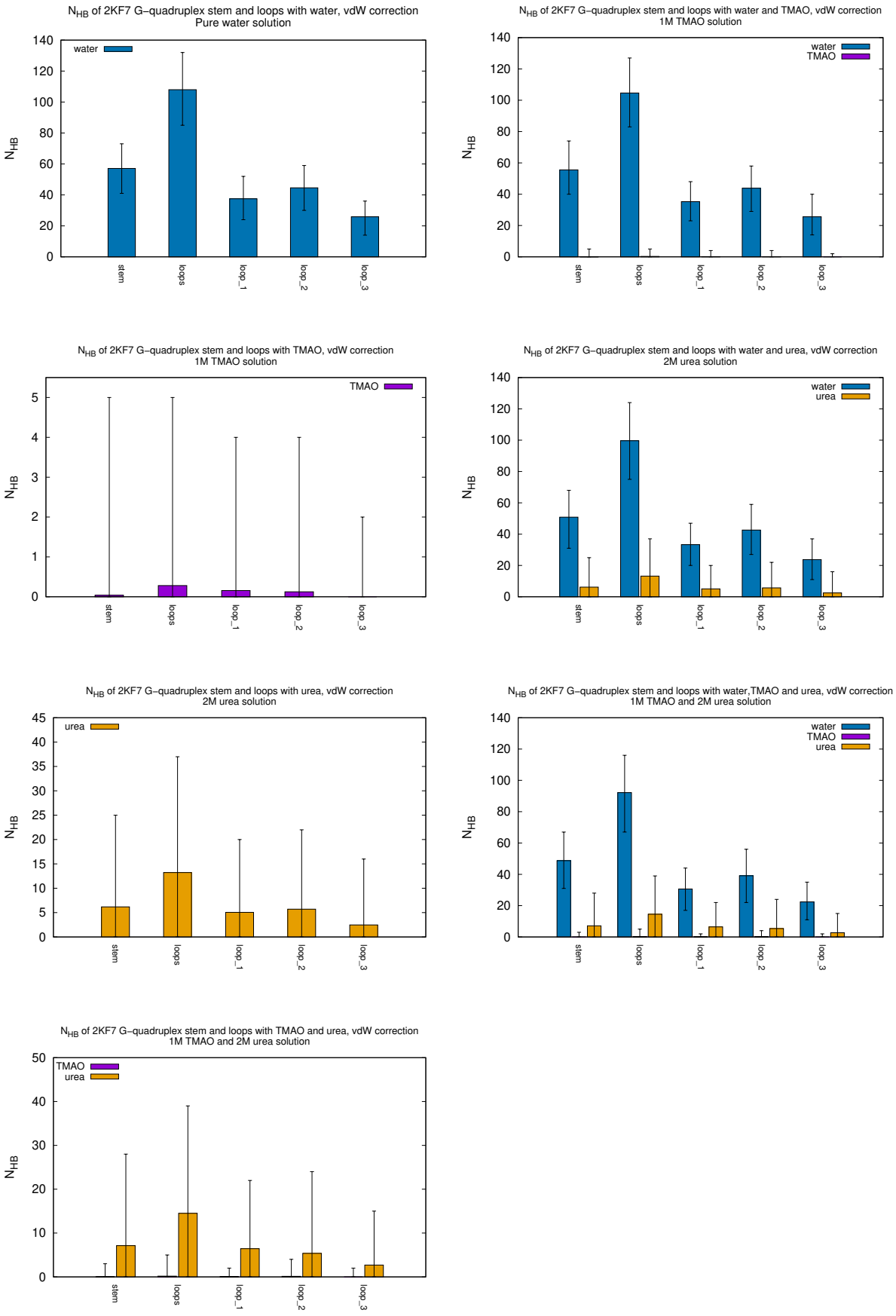


Figure 78: Number of hydrogen bonds between DNA stem and loops and co-solutes at 300 K in vdW-modified ParmBSC1 force field.

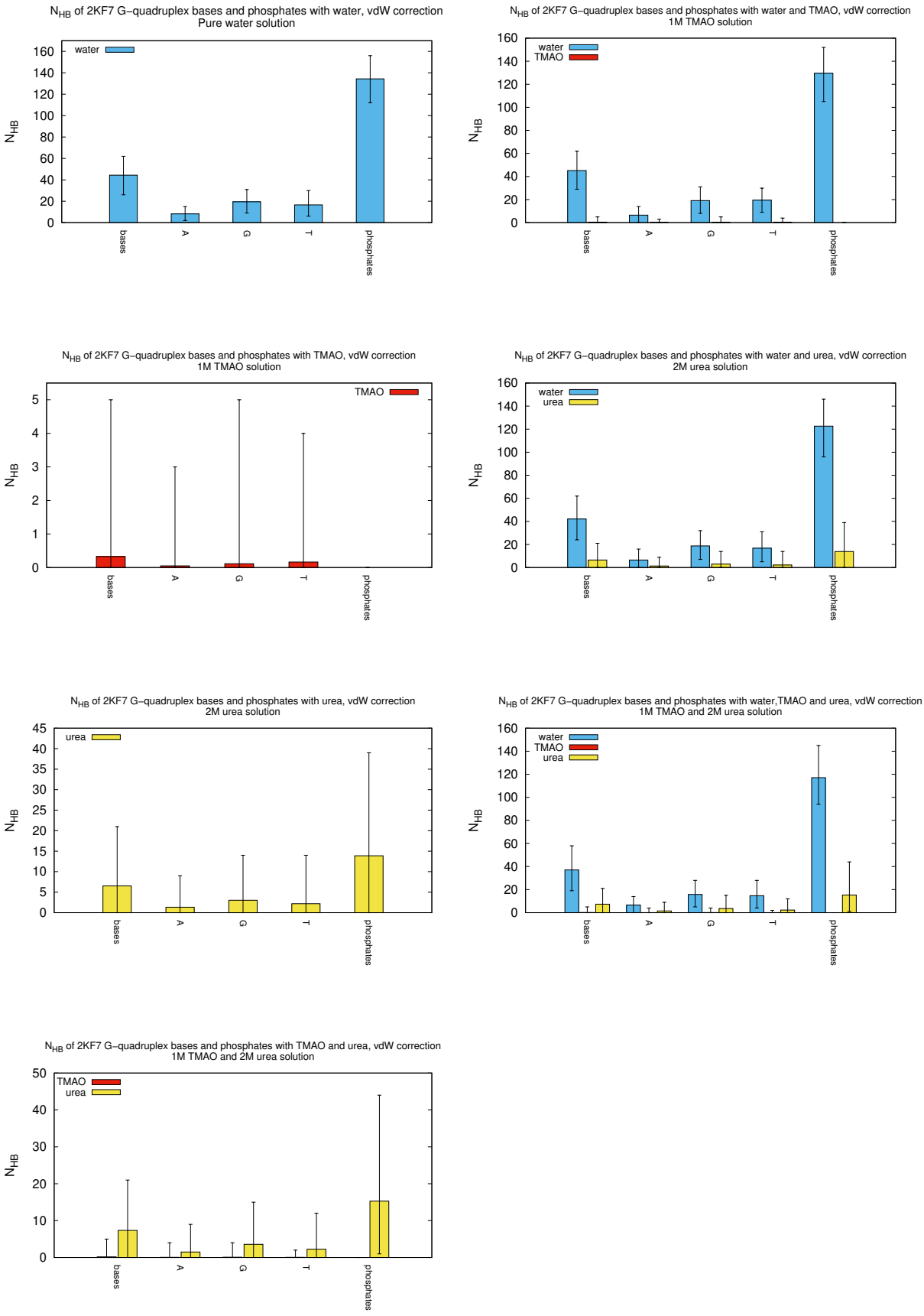


Figure 79: Number of hydrogen bonds between DNA bases and phosphates and co-solutes at 300 K in vdW-modified ParmBSC1 force field.

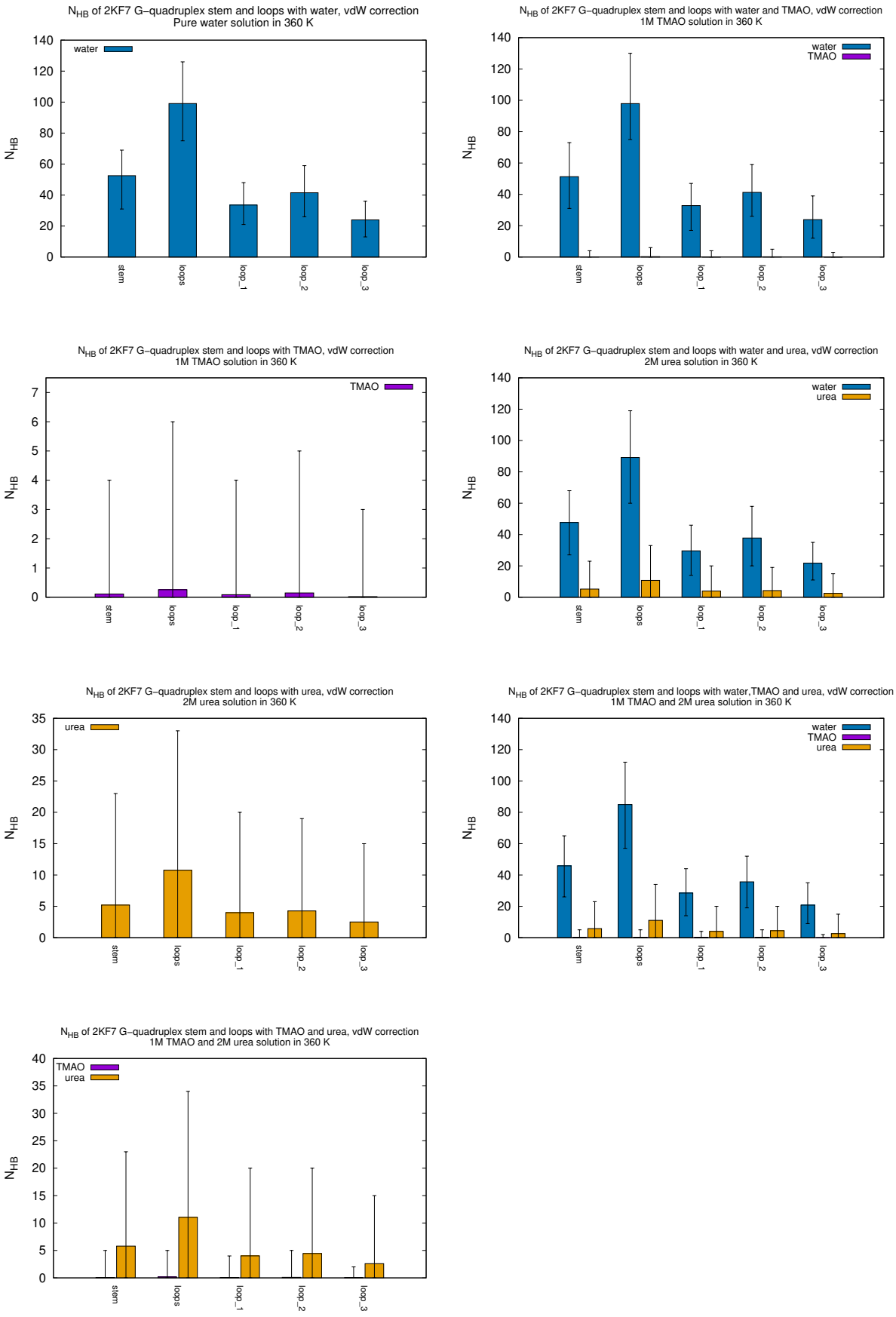


Figure 80: Number of hydrogen bonds between DNA stem and loops and co-solutes at 360 K in vdW-modified ParmBSC1 force field.

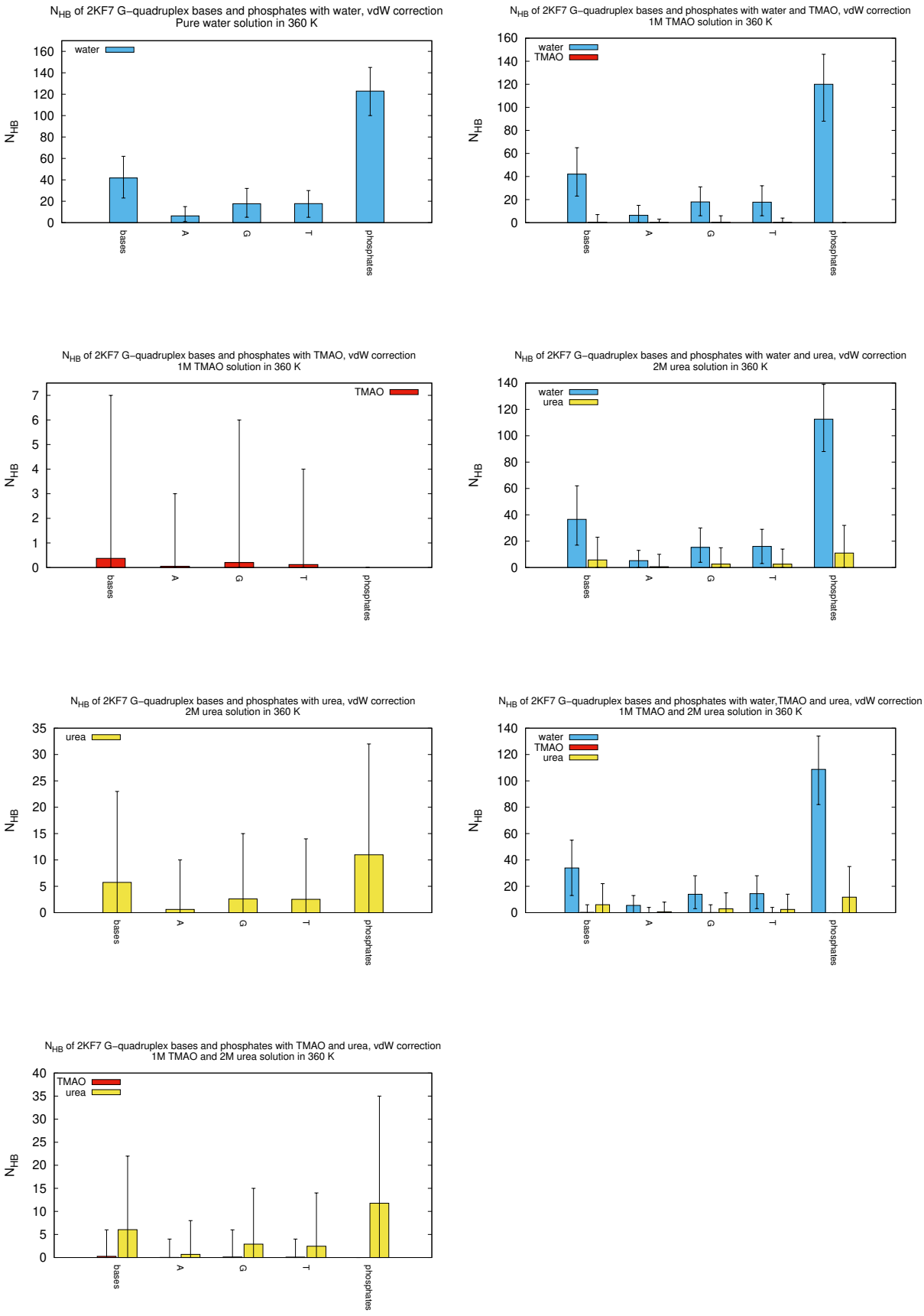


Figure 81: Number of hydrogen bonds between DNA bases and phosphates and co-solutes at 360 K in vdW-modified ParmBSC1 force field.

8.1.5 Number of contacts in vdW-modified force field

For the simulations in ParmBSC1_vdW force field, I calculated the average number of contacts between the atoms of G-quadruplex stem and loop regions and the centers of mass of TMAO and urea using FastNS (Fast Neighbor Search) utility of MdAnalysis [324]. The results are shown in Table 15. The corresponding results obtained from the simulations in unmodified ParmBSC1 force field are shown in Table 11.

Table 15: Number of contacts between DNA G-quadruplex stem and loops and the co-solutes TMAO and urea in 300 K and 360 K molecular dynamics simulations in vdW-modified ParmBSC1 force field.

System, vdW	**300 K**			
	stem		**loops**	
	TMAO	**urea**	**TMAO**	**urea**
1M TMAO	34.67 ± 0.01	-	44.62 ± 0.02	-
2M urea	-	81.75 ± 0.03	-	103.81± 0.03
1M TMAO and 2M urea	34.69 ± 0.01	85.74 ± 0.03	45.48 ± 0.02	106.92 ± 0.03
	360 K			
	stem		**loops**	
	TMAO	**urea**	**TMAO**	**urea**
1M TMAO	33.33 ± 0.01	-	43.54 ± 0.02	-
2M urea	-	75.82 ± 0.02	-	95.12 ± 0.03
1M TMAO and 2M urea	33.61 ± 0.01	79.02 ± 0.02	43.36 ± 0.02	98.11 ± 0.03

The differences between the number of contacts reported for ParmBSC1 and ParmBSC1_vdW force fields are within the statistical error associated with the simulation procedure. This observation is valid both for the simulations in 300 K and 360 K. One can thus conclude that the modification of certain van der Waals parameters in the force field parametrization does not influence the observed number of contacts formed in the studied systems, analogously to the number of formed hydrogen bonds.

In accordance with the simulations in unmodified ParmBSC1 force field, also for the simulations in its vdW-modified version I calculated the cumulative number radial distribution function (CN-RDF) as the measure of the number of contacts between DNA and other components of the solution within a sphere of arbitrarily chosen radius R. The cutoff distance R has been set to R = 2.5 nm to conform with the calculations for ParmBSC1 force field. The resulting CN-RDF values for the whole DNA molecule for 300 K and 360 K, approximated to the nearest integer, are shown in Table 16. The corresponding results for unmodified ParmBSC1 force field are shown in Table 12.

Table 16: Cumulative number radial distribution function of water, TMAO and urea around the whole DNA G-quadruplex molecule simulated in the vdW-modified ParmBSC1 force field.

System, vdW	**300 K**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	1953	-	-
1M TMAO	1840	37	-
2M urea	1740	-	86
1M TMAO and 2M urea	1614	37	90
System, vdW	**360 K**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	1849	-	-
1M TMAO	1741	36	-
2M urea	1642	-	80
1M TMAO and 2M urea	1513	36	82

One can notice that the differences between CN-RDF values calculated in both force fields are only minor and do not influence the general conclusion. Therefore I calculated also the averaged number of TMAO, urea and water molecules at the distance R around stem and loops of G-quadruplex for the simulations in 300 K and 360 K in ParmBSC1_vdW force field, in addition to the calculations for the whole DNA structure. Similarly to the simulations in unmodified ParmBSC1 force field, the cutoff distance has been set to R = 1.5 nm, which is the standard distance representing the close proximity of most spherical molecules. The resulting CN-RDF values approximated to the nearest integer are presented in Tables 17 and 18. The corresponding results obtained for ParmBSC1 force field are shown in Tables 13 and 14.

Table 17: Cumulative number radial distribution function of water, TMAO and urea around stem and loops of DNA G-quadruplex simulated at 300 K in vdW-modified ParmBSC1 force field.

System in 300 K, vdW	**stem**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	321	-	-
1M TMAO	303	5	-
2M urea	280	-	17
1M TMAO and 2M urea	259	5	18
System in 300 K, vdW	**loops**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	347	-	-
1M TMAO	326	6	-
2M urea	305	-	18
1M TMAO and 2M urea	281	6	19

Table 18: Cumulative number radial distribution function of water, TMAO and urea around stem and loops of DNA G-quadruplex simulated at 360 K in vdW-modified ParmBSC1 force field.

System in 360 K, vdW	**stem**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	303	-	-
1M TMAO	288	5	-
2M urea	264	-	15
1M TMAO and 2M urea	244	5	16
System in 360 K, vdW	**loops**		
	CN-RDF water	**CN-RDF TMAO**	**CN-RDF urea**
pure water	327	-	-
1M TMAO	311	6	-
2M urea	286	-	16
1M TMAO and 2M urea	263	6	17

For the simulations in T = 300 K, there are nearly no differences between CN-RDF values obtained in ParmBSC1 and ParmBSC1_vdW force field. The discrepancies between the results obtained for both force fields are slightly more pronounced in high temperature simulations, but they do not affect the general conclusion. This observation, together with the outcomes from H-bond and number of contacts calculations allow to conclude that the modification of selected van der Waals parameters in the applied force field does not affect the reported interaction pattern between simulated DNA structure and the other components of the solution. However, one has to have in mind that the differences between ParmBSC1 and ParmBSC1_vdW force field are only minor and involve only selected parameters. Thus, it can be expected that the choice of the force field class different to Amber, such as CHARMM or GROMOS, could affect the simulation results dramatically.

9 Summary and conclusion

The interactions between biomolecules such as proteins or nucleic acids with various osmolytes have been a subject of intensive research in the recent years. Among them, the agents stabilizing the native structure of the biomolecule as well as those destabilizing it gain in particular importance. Although the effects of those co-solutes on proteins have been widely explored and reviewed to date, their impact on nucleic acids remained often vague. Since many of those co-solutes are not only omnipresent in the cellular environment of the nucleic acids but also widely applied in laboratories for biomolecular studies, there is an essential need to establish a clear picture of co-solvent effects in nucleic acids systems. Beside canonical helical DNA structures, which are considered as the most standard carriers of genetic information, the genome of numerous organisms contains many non-canonical structures, whose biological role reaches far beyond the typical expression of genes. Among them, DNA G-quadruplexes reach particular importance due to their presence in transcriptional regulatory regions of multiple genes and oncogenes as well as in chromosomal telomeres, and their resulting regulatory role in fundamental biochemical processes, such as gene expression and functional genomics [15, 18, 19, 20].

In this work, one approaches the topic of the co-solutes interaction with various non-canonical DNA structures, such as short DNA hairpins or DNA G-quadruplex, as well as with canonical Watson-Crick DNA helix. To get a deeper insight into the problem of DNA–co-solute interactions, I applied the computational method of all-atomistic Molecular Dynamics (MD) simulations in combination with Molecular Theory of Solutions. The advantage of computer simulations lays in their ability to provide the details of co-solvent effects on atomic level, which is usually not reachable in experiment [8]. The experimental approaches, although considered as more direct way of biomolecular research, often cannot provide the molecular details of the DNA–co-solvent interactions or the resulting unfolding phenomena. However, despite numerous advantages, the computational methods carry several limitations. One of them is the available simulation time scale, which is typically shorter than the time scale of several spontaneously occurring biochemical processes, such as nucleic acids folding or base pair opening. Therefore certain molecular events are typically not observed even in the longest MD runs.

As the first approach to study the interaction of the co-solutes with nucleic acid structures, I investigated the mutual interplay between various components of the pure solution, such as water and co-solutes, themselves. In Chapter 3, I studied the properties of TMAO and urea interactions in aqueous binary and ternary solutions. The obtained results confirm a rather negligible association behavior between the two co-solutes even at higher concentrations, which is evidenced by the low stability and number of TMAO–urea hydrogen bonds. It has been shown that TMAO forms preferentially the hydrogen bonds with water, and they are allegedly the most stable of all hydrogen bonds formed in aqueous TMAO–urea mixtures. TMAO has been shown also to stabilize the hydrogen bonds of water significantly, whereas the impact of urea on the dynamics of water hydrogen bond is only minor. Both co-solutes are slowing down water dynamics, which is demonstrated by longer water dipolar relaxation times and consequent changes of hydrogen bond forward lifetimes. In general, TMAO has been found to exert much stronger impact on the water characteristics than urea.

Despite the classical division into chaotropes ("structure breakers") and kosmotropes ("structure makers"), the presented results reveal that both co-solutes exhibit kosmotropic properties in aqueous solution. Although the effects slightly differ, both urea and TMAO strengthen the hydrogen bond network of water molecules, which is elucidated by the increase in the relative number of total hydrogen bonds and a decrease in water fluctuations. This results in an increased dipolar relaxation times and longer hydrogen bond forward lifetimes. Furthermore, no compensation of the influence of urea and TMAO on the water structure and dynamics in the presence of both co-solutes has been observed. Due to the TMAO zwitterionic structure, the hydrogen bonds between water and TMAO molecules are substantially stronger in comparison to water and urea, which leads to the stable hydration shells around TMAO. Hence, disperse and relatively instable TMAO–urea complexes can be considered as solvent–shared co-solute pairs with regard to their individual molecular properties. Taking all these findings into consideration, this study shows the absence of the clear chaotropic properties of urea, which sheds a new light on the importance of

water–mediated effects for the stabilization or destabilization of macromolecules. Since both co-solutes behave predominantly like kosmotropes, one can conclude that the global influence on the water structure and dynamics as well as the water–mediated interactions are of minor importance for stabilization or destabilization phenomena. This indicates that organic co-solutes can exert the structural modification effects on macromolecules via the differences in their accumulation pattern around the central solute, in agreement with hitherto findings within the framework of local/bulk partitioning approaches [61, 62, 63, 64, 69, 4]. With regard to these results, the distinction of organic co-solutes to kosmotropes or chaotropes should be carefully reconsidered to explain the effects on macromolecular structure. Furthermore, it can be expected that also other organic co-solutes reveal analogous behavior to that reported for TMAO and urea.

In Chapter 4, I studied the interaction between a short 7-bp DNA oligonucleotide with the sequence d(GCGAAGC) in its native and unfolded form in aqueous solution with different concentrations of urea. The presented results show the preferential binding of urea to DNA regardless of its conformation, which stands in agreement with recent experimental findings concerning urea binding to DNA and RNA [329, 330, 331]. In addition, the results clearly indicate that urea favors the unfolded DNA state, which is consistent with hitherto findings on the influence of urea on proteins [47, 48, 46], RNA tetraloops [332] and DNA hairpins [119]. These results pinpoint the unspecific binding mechanism between urea and nucleic acids conformations, which can be qualitatively transferable to different DNA forms, as suggested in Ref. [162], even though the DNA–urea interaction strength may potentially differ between non-canonical DNA hairpins and DNA double helix [160]. Thus, this study gives the insights into the nature of the preferential interactions in ternary systems of DNA, urea and water and provides a consistent view on the energetic contributions and the local interactions between DNA and urea.

In Chapter 5, I investigated the interactions between two DNA structures: a short 7-bp DNA oligonucleotide with the sequence d(GCGAAGC) and a canonical 24-bp B-DNA helix, with ectoine. Due to its cell-protecting properties, this co-solute is commonly used in the pharmaceutical industry as an ingredient in a broad range of health products. This research focuses on the ectoine–DNA binding behavior as an approach to study the co-solute influence on DNA structure. Although ectoine has been found to stabilize generally the structure of proteins [72, 73, 193, 148, 55, 192], the obtained results indicate the strong and unspecific binding of ectoine to DNA, prompted by enthalpic mechanisms involving Lennard-Jones and Coulomb short-range interactions. Thus, ectoine can be considered as a DNA denaturing agent in agreement with recent experimental results [197]. The denaturing impact of ectoine can be attributed mainly to the highly negative charge of the DNA phosphodiester backbone, which encourages the preferential binding of zwitterionic ectoine molecules via strong electrostatic interactions combined with pronounced dispersion energies. At the same time, water structure around DNA molecule seems not to be affected even by high ectoine concentrations, which counts against the entropic contribution to the binding mechanism. In terms of statistical mechanical theories of solutions, I determined the related change in the melting temperature of DNA and proved its qualitative agreement with experiment. The observed strong binding of ectoine to DNA rationalizes also the recently reported protection mechanism preventing irradiation damages of nucleic acids [195, 196]. In this aspect, the preferential binding behavior of ectoine promoting the formation a stable protective shell around DNA overweights its potential denaturing impact on DNA. It can be speculated that this stable ectoine shell can be interpreted as an efficient barrier for irradiation damages, providing the protection of the nucleic acids. Hence, this study provides a coherent view on the molecular aspects of DNA and ectoine interactions in aqueous systems and allows to shed more light onto its nature.

In Chapter 6, I investigated a basket-type DNA G-quadruplex of polymorphic nature. This type of G-quadruplex is composed of two G-tetrads, which are being responsible for its unique stability. I performed MD simulations in the biologically relevant concentrations of co-solutes like TMAO and urea, referring to the findings for piezophilic organisms, who acquire TMAO concentrations of 0.6–1 M in order to adapt to osmotic stress. With reference to the studies suggesting that in 1:2 TMAO and urea mixtures TMAO can fully compensate the denaturing effect of urea, I simulated the systems also in 2 M urea and 1:2 molar mixed TMAO and urea solutions. In particular, I focued on DNA binding behavior with both co-solutes and with water, which gives an insight into their influence on G-quadruplex structural stability. With the application of Kirkwood-Buff

theory, the molecular nature of DNA–co-solutes interaction mechanism has been approached. The presented results evidence for the preferential binding of urea and the preferential exclusion for TMAO molecules around the DNA. The corresponding findings are comparable between the single component solutions and TMAO:urea mixture. Analogously, there is no significant change in the accumulation behavior of the co-solutes, number of hydrogen bonds and interaction energies in the mixture in comparison to the single co-solute solutions. These results indicate a pairwise linear additivity of the individual contributions from the co-solutes. Despite the data showing that 1 M TMAO can entirely compensate the denaturing effect of 2 M urea on protein structure [155, 156, 46, 154], my studies on DNA G-quadruplex show that the compensation effect is not complete when nucleic acids are considered. Although the binding behavior of urea to DNA structure is much less pronounced in the presence of TMAO than in single urea solution, it is not entirely nullified, which can be translated to the destabilizing impact on G-quadruplex structure. In more details, my results reveal that urea can be considered as a strong DNA destabilizer whereas TMAO stabilizes the DNA structure in terms of the Kirkwood-Buff theory of solutions, where urea tends to accumulate closely around DNA, and TMAO is being repelled to the bulk phase. In fact, both TMAO and urea layers exhibit only weak mutual influences or interactions. However, the combined consideration of TMAO and urea as an artificial co-solute shows that the preferential binding coefficient is slightly lowered in comparison to ternary urea solutions. As a consequence, the destabilization effect by urea is not so strongly pronounced in presence of TMAO due to the structure–stabilizing compensation mechanism involving TMAO. The clear separation of individual co-solute layers and the absence of substantial mixing effects between urea and TMAO around the DNA, being responsible for the pairwise additivity of the individual preferential binding coefficients, ensures the presence of the compensation effects. As it was already shown in my previous study, urea exerts its denaturing effect on nucleic acids via indirect mechanism [3]. The current findings confirm this conclusion, showing that also DNA structural stabilization by TMAO occurs rather via indirect mechanism involving interference with water structure than by the formation of direct bonds with nucleic acids. The mechanism of TMAO stabilization of nucleic acids is thus similar to that for proteins [155, 36], although there are also studies that evidence on the direct TMAO–protein interaction [163, 137].

To extend and significantly complement the research presented in Chapter 6, in Chapter 7 I investigate the 2KF7 DNA G-quadruplex composed of only two tetrads from more visual perspective, involving MD simulations together with molecular modelling. The study presented in Chapter 6 tackles the topic of DNA–co-solutes interactions from a purely thermodynamical point of view. The work described in Chapter 7 gives a detailed insight into biomolecular aspects of the described phenomena to complete the picture achieved numerically. Whereas the aforementioned work discusses mainly the stabilization and destabilization aspects in terms of the Kirkwood-Buff theory of solutions, in this chapter I present the corresponding molecular implications and a further rationale for the previous findings. By means of Molecular Dynamics (MD) simulations, I attempt to reconstruct the near-native environment of biomolecules, which includes not only water, but also essential co-solutes like TMAO and urea. With reference to the recent findings on the nature of piezophilic microorganisms, who adopt certain concentrations and proportions of urea and TMAO in order to modulate their cellular response to osmotic stress, I studied the DNA G-quadruplex in the presence of near-physiological concentrations of both co-solutes alone as well as in 1:2 molar TMAO:urea mixture. This combination of both co-solutes has been proved to provide a full compensation of urea denaturing impact on proteins, at the same time giving the protection against osmotic stress in deep sea microorganisms. This work provides a detailed qualitative analysis of the interaction between TMAO and urea and DNA G-quadruplex, with regard to its stem and loop regions together with nucleic bases and phosphate backbone. Through the analysis of density distribution functions one can get insight into the changes in DNA solvation pattern upon the addition of co-solutes in conjunction with increased temperature higher than the melting temperature T_m for this type of DNA. One can can observe the insertion of urea molecules into DNA grooves in most of the simulation conditions, whereas TMAO predominantly stays away from the DNA surface. In the mixture of urea and TMAO, TMAO is pulling urea away from DNA structural interior and thus provides the solvation shell sufficient to stabilize the biomolecule. Notwithstanding, my results demonstrate that contrary to proteins DNA G-quadruplex structure is relatively resistant to the influence of co-solutes in terms of both stabilization and destabilization. While most proteins undergo certain degree of denaturation after exposition to biologically

relevant concentrations of urea, G-quadruplex appears to be rather unaffected by the presence of urea. Analogously, whereas TMAO has been widely proved to stabilize protein structure against detrimental impact of both urea or temperature, the stability of G-quadruplex structure remains vastly unaltered under physiological concentrations of TMAO. This points out the extraordinary stability of DNA G-quadruplexes, which is expressed not only as thermal resistance, but also as invulnerability to the influence of co-solutes. Additionally, the presented findings suggest that possible impact of co-solutes on G-quadruplex structural stability is rather not associated with the formation of hydrogen bonds nor the direct contacts between co-solutes and DNA. Neither the DNA structure-stabilizing intramolecular hydrogen bonds seem to be affected by the temperature or the presence of co-solutes. Thus, the changes in flexibility pattern of particular parts of the G-quadruplex structure can be attributed to the hydrating role of water, together with thermal instability in high temperature simulations. In this aspect this study presents a novel insight into numerous aspects of the interplay between TMAO and urea and non-canonical DNA structures, which - contrary to most proteins - bear a pronounced negative charge.

The interaction between fundamental biomolecules such as proteins or nucleic acids with molecular agents stabilizing and destabilizing their native structure has been a subject of intensive research in the recent years. Since certain co-solutes, such as urea, ectoine or TMAO, are not only omnipresent in the native environment of most nucleic acids, but also commonly applied in medicine, industry or biotechnology, the necessity to establish a clear picture of their behavior in nucleic acids systems gains on crucial importance. The work presented in this thesis enables to get the insight into the molecular details of the interactions between selected nucleic acid structures of biological importance with relevant co-solutes, and therefore to hint on their potential application for DNA structural modulations in modern biophysics, chemistry and biotechnology. Moreover, this work demonstrates a novel implementation of the Kirkwood–Buff theory for studying the biomolecular systems. Besides the mathematical extension and refinement of the theory, this study introduces an approach to solving the essential topics of biomolecular physics, such as the molecular details of particle–particle interactions in the systems of biological interest. Particular emphasis is placed on the employment of the Kirkwood–Buff theory to extract the thermodynamic characteristics of the simulated systems, which are not available in experimental setup. Furthermore, this research focuses on the practical application and the implementation of the theory to study ternary and quaternary systems, which approximate well the real environment of the living cells. In this aspect, this work constitutes an important contribution to the area of applied physics, with particular regard to the biophysics and physical biochemistry.

Primary references

[1] E. A. Oprzeska-Zingrebe and J. Smiatek, "Interactions of a DNA G-quadruplex with TMAO and urea: a molecular dynamics study on co-solute compensation mechanisms," *Phys. Chem. Chem. Phys.*, vol. 23, no. 2, pp. 1254–1264, 2021.

[2] E. A. Oprzeska-Zingrebe and J. Smiatek, "Basket-type G-quadruplex with two tetrads in the presence of TMAO and urea: A molecular dynamics study," *J. Mol. Struct.*, vol. 1274, no. 2, p. 134375, 2023.

[3] E. A. Oprzeska-Zingrebe and J. Smiatek, "Preferential binding of urea to single-stranded DNA structures: a molecular dynamics study," *Biophys. J.*, vol. 114, no. 7, pp. 1551–1562, 2018.

[4] E. A. Oprzeska-Zingrebe, S. Meyer, A. Roloff, H.-J. Kunte, and J. Smiatek, "Influence of compatible solute ectoine on distinct DNA structures: thermodynamic insights into molecular binding mechanisms and destabilization effects," *Phys. Chem. Chem. Phys.*, vol. 20, no. 40, pp. 25861–25874, 2018.

[5] E. A. Oprzeska-Zingrebe and J. Smiatek, "Aqueous ionic liquids in comparison with standard co-solutes," *Biophys. Rev.*, vol. 10, no. 3, pp. 809–824, 2018.

[6] E. A. Oprzeska-Zingrebe and J. Smiatek, "Aqueous mixtures of urea and trimethylamine-N-oxide: Evidence for kosmotropic or chaotropic behavior?," *J. Phys. Chem. B*, vol. 123, no. 20, pp. 4415–4424, 2019.

[7] E. A. Oprzeska-Zingrebe, M. Kohagen, J. Kästner, and J. Smiatek, "Unfolding of DNA by co-solutes: insights from Kirkwood–Buff integrals and transfer free energies," *Europ. Phys. J. Spec. Top.*, vol. 227, pp. 1665–1679, 2019.

[8] E. A. Oprzeska-Zingrebe and J. Smiatek, "Some notes on the thermodynamic accuracy of coarse-grained models," *Front. Mol. Biosci.*, vol. 6, no. 87, 2019.

Secondary references

[9] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid," *Nature*, vol. 171, pp. 737–738, 1953.

[10] S. Benabou, A. Aviñó, R. Eritja, C. González, and R. Gargallo, "Fundamental aspects of the nucleic acid i-motif structures," *RSC Advances*, vol. 4, no. 51, pp. 26956–26980, 2014.

[11] E. P. Wright, J. L. Huppert, and Z. A. E. Waller, "Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH," *Nucleic Acids Res.*, vol. 45, no. 6, pp. 2951–2959, 2017.

[12] M. Zeerati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger, and D. Christ, "I-motif DNA structures are formed in the nuclei of human cells," *Nat. Chem.*, vol. 10, pp. 631–637, 2018.

[13] A. T. Phan and J.-L. Mergny, "Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix," *Nucleic Acids Res.*, vol. 30, no. 21, pp. 4618–4625, 2002.

[14] M. Guéron and J.-L. Leroy, "The i-motif in nucleic acids," *Curr. Opin. Struct. Biol.*, vol. 10, no. 3, pp. 326–331, 2000.

[15] H. Han and L. H. Hurley, "G-quadruplex DNA: a potential target for anti-cancer drug design," *Trends Pharmacol. Sci.*, vol. 21, no. 4, pp. 136–142, 2000.

[16] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, and S. Neidle, "Quadruplex DNA: sequence, topology and structure," *Nucleic Acids Res.*, vol. 34, no. 19, pp. 5402–5415, 2006.

[17] H. J. Lipps and D. Rhodes, "G-quadruplex structures: *in vivo* evidence and function," *Trends Cell Biol.*, vol. 19, no. 8, pp. 414–422, 2009.

[18] M. Kaushik, S. Kaushik, A. Bansal, S. Saxena, and S. Kukreti, "Structural diversity and specific recognition of four stranded G-quadruplex DNA," *Curr. Mol. Med.*, vol. 11, pp. 744–769, 2011.

[19] M. L. Bochman, K. Paeschke, and V. A. Zakian, "DNA secondary structures: stability and function of G-quadruplex structures," *Nat. Rev. Genet.*, vol. 13, no. 11, pp. 770–780, 2012.

[20] D. Rhodes and H. J. Lipps, "G-quadruplexes and their regulatory roles in biology," *Nucleic Acids Res.*, vol. 43, no. 18, pp. 8627–8637, 2015.

[21] A. W. Oliver, I. Bogdarina, E. Schroeder, I. A. Taylor, and G. G. Kneale, "Preferential binding of fd gene 5 protein to tetraplex nucleic acid structures," *J.Mol.Biol.*, vol. 301, no. 3, pp. 575–584, 2000.

[22] H. Arthanari and P. H. Bolton, "Functional and dysfunctional roles of quadruplex DNA in cells," *Chem. Biol.*, vol. 8, no. 3, pp. 221–230, 2001.

[23] D. Japrung, M. Henricus, Q. Li, G. Maglia, and H. Bayley, "Urea facilitates the translocation of single-stranded DNA and RNA through the alpha-hemolysin nanopore," *Biophys. J.*, vol. 98, no. 9, pp. 1856–1863, 2010.

[24] A. N. Lane, J. B. Chaires, R. D. Gray, and J. O. Trent, "Stability and kinetics of G-quadruplex structures," *Nucleic Acids Res.*, vol. 36, no. 17, pp. 5482–5515, 2008.

[25] D. J. Patel, A. T. Phan, and V. Kuryavyi, "Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics," *Nucleic Acids Res.*, vol. 35, no. 22, pp. 7429–7455, 2007.

[26] H. Liu and D. Liu, "DNA nanomachines and their functional evolution," *Chem. Commun.*, vol. 19, pp. 2625–2636, 2009.

[27] A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan, "Challenges and opportunities for structural DNA nanotechnology," *Nat. Nanotechnol.*, vol. 6, no. 12, pp. 763–772, 2011.

[28] J. R.Williamson, M.K.Raghuraman, and T. R.Cech, "Monovalent cation-induced structure of telomeric DNA: The G-quartet model," *Cell*, vol. 59, no. 5, pp. 871–880, 1989.

[29] A. L. Fink, "Chaperone-mediated protein folding," *Physiol. Rev.*, vol. 79, no. 2, pp. 425–449, 1999.

[30] C. M. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, pp. 884–890, 2003.

[31] K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes, and H. J. Lipps, "Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*," *Nat. Struct. Mol. Biol.*, vol. 12, pp. 847–854, 2005.

[32] A. D. Brown, "Compatible solutes and extreme water stress in eukaryotic micro-organisms," *Adv. Microb. Physiol.*, vol. 17, pp. 181–242, 1978.

[33] E. A. Galinski, "Compatible solutes of halophilic eubacteria: molecular principles, water-solute interactions, stress protection," *Experientia*, vol. 49, pp. 487–496, 1993.

[34] G. Lentzen and T. Schwarz, "Extremolytes: Natural compounds from extremophiles for versatile applications," *Appl. Microbiol. Biotechnol.*, vol. 72, no. 4, pp. 623–634, 2006.

[35] A. Oren, "Microbial life at high salt concentrations: phylogenetic and metabolic diversity," *Saline Syst.*, vol. 4, no. 2, p. 2, 2008.

[36] W. J. Xie, S. Cha, T. Ohto, M. Bonn, J. Hunger, and Y. N. et al., "Large hydrogen-bond mismatch between TMAO and urea promotes their hydrophobic association," *Chem*, vol. 4, pp. 2615–2627, 2018.

[37] P. H. Yancey, "Water stress, osmolytes and proteins," *Am. Zool.*, vol. 41, no. 4, pp. 699–709, 2001.

[38] P. H. Yancey, "Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses," *J. Exp. Biol.*, vol. 208, pp. 2819–2830, 2005.

[39] M. F. Roberts, "Organic compatible solutes of halotolerant and halophilic microorganisms," *Saline Syst.*, vol. 1, no. 5, pp. 1–30, 2005.

[40] H.-J. Kunte, "Osmoregulation in bacteria: Compatible solute accumulation and osmosensing," *Environ. Chem.*, vol. 3, no. 2, pp. 94–99, 2006.

[41] J. C. Lee and S. N. Timasheff, "The stabilization of proteins by sucrose," *J. Biol. Chem.*, vol. 256, no. 14, pp. 7193–7201, 1981.

[42] T. Arakawa and S. N. Timasheff, "The stabilization of proteins by osmolytes," *Biophys. J.*, vol. 47, no. 3, pp. 411–414, 1985.

[43] S. N. Timasheff, "Protein hydration, thermodynamic binding, and preferential hydration," *Biochemistry*, vol. 41, no. 46, pp. 13473–13482, 2002.

[44] P. Lamosa, D. L. Turner, R. Ventura, C. Maycock, and H. Santos, "Protein stabilization by compatible solutes: Effect of diglycerol phosphate on the dynamics of *Desulfovibrio gigas* rubredoxin studied by NMR," *Eur. J. Biochem.*, vol. 270, no. 23, pp. 4606–4614, 2003.

[45] T. O. Street, D. W. Bolen, and G. D. Rose, "A molecular mechanism for osmolyte-induced protein stability," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 38, pp. 13997–14002, 2006.

[46] D. R. Canchi and A. E. García, "Cosolvent effects on protein stability," *Annu. Rev. Phys. Chem.*, vol. 64, pp. 273–293, 2013.

[47] B. J. Bennion and V. Daggett, "The molecular basis for the chemical denaturation of proteins by urea," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 5142–5147, 2003.

[48] R. Zangl, R. Zhou, and B. J. Berne, "Urea's action on hydrophobic interactions," *J. Am. Chem. Soc.*, vol. 131, pp. 1535–1541, 2009.

[49] Y. Zhang and P. S. Cremer, "Chemistry of hofmeister anions and osmolytes," *Annu. Rev. Phys. Chem.*, vol. 61, pp. 63–83, 2010.

[50] Z. Zhang, N. I. Dmitrieva, J.-H. Park, R. L. Levine, and M. B. Burg, "High urea and NaCl carbonylate proteins in renal cells in culture and in vivo, and high urea causes 8-oxoguanine lesions in their DNA," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 25, pp. 9491–9496, 2004.

[51] I. Lidbury, J. C. Murrell, and Y. Chen, "Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 7, pp. 2710–2715, 2014.

[52] J. Smiatek, R. K. Harishchandra, O. Rubner, H.-J. Galla, and A. Heuer, "Properties of compatible solutes in aqueous solution," *Biophys. Chem.*, vol. 160, no. 1, pp. 62–68, 2012.

[53] J. Smiatek, R. K. Harishchandra, H.-J. Galla, and A. Heuer, "Low concentrated hydroxyectoine solutions in presence of dppc lipid bilayers: a computer simulation study," *Biophys. Chem.*, vol. 180, pp. 102–109, 2013.

[54] J. Smiatek, "Osmolyte effects: impact on the aqueous solution around charged and neutral spheres," *J. Phys. Chem. B*, vol. 118, no. 3, pp. 771–782, 2014.

[55] M. B. Hahn, T. Solomun, R. Wellhausen, S. Hermann, H. Seitz, S. Meyer, H.-J. Kunte, J. Zeman, F. Uhlig, J. Smiatek, and H. Sturm, "Influence of the compatible solute ectoine on the local water structure: Implications for the binding of the protein G5P to DNA," *J. Phys. Chem. B*, vol. 119, no. 49, pp. 15212–15220, 2015.

[56] M. B. Hahn, F. Uhlig, T. Solomun, J. Smiatek, and H. Sturm, "Combined influence of ectoine and salt: spectroscopic and numerical evidence for compensating effects on aqueous solutions," *Phys. Chem. Chem. Phys.*, vol. 18, no. 41, pp. 28398–28402, 2016.

[57] M. A. Schroer, J. Michalowsky, B. Fischer, J. Smiatek, and G. Grübel, "Stabilizing effect of TMAO on globular PNIPAM states: preferential attraction induces preferential hydration," *Phys. Chem. Chem. Phys.*, vol. 18, no. 46, pp. 31459–31470, 2016.

[58] K. D. Collins, "Ions from the hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process," *Methods*, vol. 34, no. 3, pp. 300–311, 2004.

[59] Y. Marcus, "Effect of ions on the structure of water: structure making and breaking," *Chem. Rev.*, vol. 109, no. 3, pp. 1346–1370, 2009.

[60] Y. Marcus, *Ions in Solution and their Solvation*. John Wiley & Sons, 2015.

[61] E. S. Courtenay, M. W. Capp, C. F. Anderson, and M. T. Record, "Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro.," *Biochemistry*, vol. 39, no. 15, pp. 4455–4471, 2000.

[62] S. Shimizu, "Estimating hydration changes upon biomolecular reactions from osmotic stress, high pressure, and preferential hydration experiments," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 5, pp. 1195–1199, 2004.

[63] S. Shimizu and D. J. Smith, "Preferential hydration and the exclusion of cosolvents from protein surfaces," *J. Chem. Phys.*, vol. 121, no. 2, pp. 1148–1154, 2004.

[64] S. Shimizu and C. L. Boon, "The Kirkwood–Buff theory and the effect of cosolvents on biochemical reactions," *J. Chem. Phys.*, vol. 121, no. 18, pp. 9147–9155, 2004.

[65] M. Aburi and P. E. Smith, "A combined simulation and Kirkwood–Buff approach to quantify cosolvent effects on the conformational preferences of peptides in solution," *J. Phys. Chem. B*, vol. 108, no. 22, pp. 7382–7388, 2004.

[66] P. E. Smith, "Chemical potential derivatives and preferential interaction parameters in biological systems from Kirkwood–Buff theory," *Biophys. J.*, vol. 91, no. 3, pp. 849–856, 2006.

[67] J. Rösgen, B. M. Pettitt, and D. W. Bolen, "Protein folding, stability, and solvation structure in osmolyte solutions," *Biophys. J.*, vol. 89, no. 5, pp. 2988–2997, 2005.

[68] P. E. Smith, "Cosolvent interactions with biomolecules:relating computer simulation data to experimental thermodynamic data," *J. Phys. Chem. B*, vol. 108, pp. 18716–18724, 2004.

[69] J. Smiatek, "Aqueous ionic liquids and their effects on protein structures: an overview on recent theoretical and experimental results," *J. Phys. Condens. Matt.*, vol. 29, no. 23, p. 233001, 2017.

[70] T. Arakawa and S. N. Timasheff, "Preferential interactions of proteins with solvent components in aqueous amino acid solutions," *Arch. Biochem. Biophys.*, vol. 224, no. 1, pp. 169–177, 1983.

[71] J. A. Schellman, "Protein stability in mixed solvents: a balance of contact interaction and excluded volume," *Biophys. J.*, vol. 85, no. 1, pp. 108–125, 2003.

[72] I. Yu and M. Nagaoka, "Slowdown of water diffusion around protein in aqueous solution with ectoine," *Chem. Phys. Lett.*, vol. 388, no. 4-6, pp. 316–321, 2004.

[73] I. Yu, Y. Jindo, and M. Nagaoka, "Microscopic understanding of preferential exclusion of compatible solute ectoine: direct interaction and hydration alteration," *J. Phys. Chem. B*, vol. 111, no. 34, pp. 10231–10238, 2007.

[74] S. Shimizu and N. Matubayasi, "Preferential hydration of proteins: A Kirkwood-Buff approach," *Chem. Phys. Lett.*, vol. 420, pp. 518–522, 2006.

[75] S. Pieraccinia, S. Conti, S. Chaurasia, and M. Sironi, "Modelling the effect of osmolytes on peptide mechanical unfolding," *Chem. Phys. Lett.*, vol. 578, pp. 138–143, 2013.

[76] S. N. Timasheff, "Water as ligand: preferential binding and exclusion of denaturants in protein unfolding," *Biochemistry*, vol. 31, no. 41, pp. 9857–9864, 1992.

[77] S. N. Timasheff, "Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 15, pp. 9721–9726, 2002.

[78] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.

[79] P. Ball and J. E. Hallsworth, "Water structure and chaotropicity: their uses, abuses and biological implications," *Phys. Chem. Chem. Phys.*, vol. 17, no. 13, pp. 8297–8305, 2015.

[80] K. D. Collins, "Charge density-dependent strength of hydration and biological structure.," *Biophys. J.*, vol. 72, no. 1, p. 65, 1997.

[81] S. Paul and G. N. Patey, "Structure and interaction in aqueous urea–trimethylamine-N-oxide solutions," *J. Am. Chem. Soc.*, vol. 129, no. 14, pp. 4476–4482, 2007.

[82] C. J. Sahle, M. A. Schroer, I. Juurinen, and J. Niskanen, "Influence of TMAO and urea on the structure of water studied by inelastic X-ray scattering," *Phys. Chem. Chem. Phys.*, vol. 18, no. 24, pp. 16518–16526, 2016.

[83] S. G. Zetterholm, G. A. Verville, L. Boutwell, C. Boland, J. C. Prather, J. Bethea, J. Cauley, K. E. Warren, S. A. Smith, D. H. Magers, and N. I. Hammer, "Noncovalent interactions between trimethylamine N-oxide (TMAO), urea, and water," *J. Phys. Chem. B*, vol. 122, no. 38, pp. 8805–8811, 2018.

[84] J. G. Kirkwood and F. P. Buff, "The statistical mechanical theory of solutions. I," *J. Chem. Phys.*, vol. 19, no. 6, pp. 774–777, 1951.

[85] A. Ben-Naim, *Statistical thermodynamics for chemists and biochemists.* Berlin, Germany: Springer Science & Business Media, 2013.

[86] P. E. Smith, E. Matteoli, and J. P. O'Connell, *Fluctuation theory of solutions: applications in chemistry, chemical engineering, and biophysics*. CRC Press, 2013.

[87] V. Pierce, M. Kang, M. Aburi, S. Weerasinghe, and P. E. Smith, "Recent applications of Kirkwood-Buff theory to biological systems," *Cell Biochem. Biophys.*, vol. 50, no. 1, pp. 1–22, 2008.

[88] A. Ben-Naim, "Inversion of the kirkwood–buff theory of solutions: application to the water–ethanol system," *J. Chem. Phys.*, vol. 67, no. 11, pp. 4884–4890, 1977.

[89] T. Kobayashi, J. E. Reid, S. Shimizu, M. Fyta, and J. Smiatek, "The properties of residual water molecules in ionic liquids: a comparison between direct and inverse Kirkwood–Buff approaches," *Phys. Chem. Chem. Phys.*, vol. 19, no. 29, pp. 18924–18937, 2017.

[90] P. E. Smith, "On the Kirkwood–Buff inversion procedure," *J. Chen. Phys.*, vol. 129, no. 12, p. 124509, 2008.

[91] S. Shimizu and N. Matubayasi, "A unified perspective on preferential solvation and adsorption based on inhomogeneous solvation theory," *Physica A*, vol. 492, pp. 1988–1996, 2018.

[92] A. N. Krishnamoorthy, C. Holm, and J. Smiatek, "The influence of co-solutes on the chemical equilibrium - a Kirkwood-Buff theory for ion pair association-dissociation processes in ternary electrolyte solutions," *J. Phys. Chem. C*, vol. 122, pp. 10293–10302, 2018.

[93] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Chromosomal DNA and its packaging in the chromatin fiber," in *Molecular Biology of the Cell, 4th edition*, Garland Science, 2002. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26834/.

[94] D. Svozil, J. Kalina, M. Omelka, and B. Schneider, "DNA conformations and their sequence preferences," *Nucleic Acids Res.*, vol. 36, pp. 3690–3706, 2008.

[95] N. Maizels and L. T. Gray, "The G4 genome," *PLoS Genet.*, vol. 9, no. 4, p. e1003468, 2013.

[96] G. Biffi, D. Tannahill, J. McCafferty, and S. Balasubramanian, "Quantitative visualization of DNA G-quadruplex structures in human cells," *Nat. Chem.*, vol. 5, pp. 182–186, 2013.

[97] N. Maizels, "G4-associated human diseases," *EMBO Rep.*, vol. 16, pp. 910–922, 2015.

[98] R. F. Hoffmann, Y. M. Moshkin, S. Mouton, N. A. Grzeschik, R. D. Kalicharan, J. Kuipers, A. H. Wolters, K. Nishida, A. V. Romashchenko, J. Postberg, H. Lipps, E. Berezikov, O. C. Sibon, B. N. Giepmans, and P. M. Lansdorp, "Guanine quadruplex structures localize to heterochromatin," *Nucleic Acids Res.*, vol. 44, no. 1, pp. 152–163, 2016.

[99] T. Simonsson, "G-quadruplex DNA structures variations on a theme," *Biol. Chem.*, vol. 382, no. 4, pp. 621–628, 2001.

[100] J. T. Davis, "G-quartets 40 years later: From 5'-GMP to molecular biology and supramolecular chemistry," *Angew. Chem. Int. Ed.*, vol. 43, no. 6, pp. 668–698, 2004.

[101] S. Neidle and G. N. Parkinson, "The structure of telomeric DNA," *Curr. Opin. Struct. Biol.*, vol. 13, no. 3, pp. 275–283, 2003.

[102] R. Reshetnikov, A. Kopylov, and A. Golovin, "Classification of G-Quadruplex DNA on the basis of the quadruplex twist angle and planarity of G-quartets," *Acta Naturae*, vol. 2, no. 4, p. 7281, 2010.

[103] S. A. Dvorkin, A. I. Karsisiotis, and M. W. da Silva, "Encoding canonical DNA quadruplex structure," *Sci. Adv.*, vol. 4, p. eaat3007, 2018.

[104] G. N. Parkinson, M. P. H. Lee, and S. Neidle, "Crystal structure of parallel quadruplexes from human telomeric DNA," *Nature*, vol. 417, pp. 876–880, 2002.

[105] A. I. Karsisiotis, C. O'Kane, and M. W. da Silva, "DNA quadruplex folding formalism a tutorial on quadruplex topologies," *Methods*, vol. 64, no. 1, pp. 28–35, 2013.

[106] W. E. Wright, V. M. Tesmer, K. E. Huffman, S. D. Levene, and J. W. Shay, "Normal human chromosomes have long G-rich telomeric overhangs at one end," *Genes and Dev.*, vol. 11, pp. 2801–2809, 1997.

[107] A. L. Moye, K. C. Porter, S. B. Cohen, T. Phan, K. G. Zyner, N. Sasaki, G. O. Lovrecz, J. L. Beck, and T. M. Bryan, "Telomeric G-quadruplexes are a substrate and site of localization for human telomerase," *Nat. Commun.*, vol. 6, p. 7643, 2015.

[108] R. M. Dirks, M. Lin, E. Winfree, and N. A. Pierce, "Paradigms for computational nucleic acid design," *Nucleic Acids Res.*, vol. 32, no. 4, pp. 1392–1403, 2004.

[109] S. Yoshizawa, G. Kawai, K. Watanabe, K.-I. Miura, and I. Hirao, "GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins," *Biochemistry*, vol. 36, no. 16, pp. 4761–4767, 1997.

[110] D. W. Cowing, J. C. A. Bardwell, E. A. Craig, and C. Woolford, "Consensus sequence for Escherichia coli heat shock gene promoters," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 82, no. 9, pp. 2679–2683, 1985.

[111] I. Hirao, G. Kawai, S. Yoshizawa, Y. Nishimura, Y. Ishido, K. Watanabe, and K.-I. Miura, "Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat," *Nucleic Acids Res.*, vol. 22, no. 4, pp. 576–582, 1994.

[112] P. Elias and I. R. Lehman, "Interaction of origin binding protein with an origin of replication of herpes simplex virus 1," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 85, no. 9, pp. 2959–2963, 1988.

[113] J. Smiatek, C. Chen, D. Liu, and A. Heuer, "Stable conformations of a single stranded deprotonated DNA i-motif," *J.Phys.Chem. B*, vol. 115, pp. 13788–13795, 2011.

[114] J. Smiatek, D. Janssen-Müller, R. Friedrich, and A. Heuer, "Systematic detection of hidden complexities in the unfolding mechanism of a cytosine-rich DNA strand," *Physica A*, vol. 394, pp. 136–144, 2014.

[115] J. Smiatek, D. Liu, and A. Heuer, "High temperature unfolding simulations of a single-stranded DNA i-motif," *Curr. Phys. Chem.*, vol. 2, pp. 115–123, 2012.

[116] I. Hirao, Y. Nishimura, Y.-I. Tagawa, K. Watanabe, and K.-I. Miura, "Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing," *Nucl. Acids Res.*, vol. 20, no. 15, pp. 3891–3896, 1992.

[117] G. Portella and M. Orozco, "Multiple routes to characterize the folding of a small DNA hairpin," *Angew. Chem. Int. Edt.*, vol. 49, no. 42, pp. 7673–7676, 2010.

[118] P. Padrta, R. Štefl, L. Králík, L. Žídek, and V. Sklenář, "Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings," *J. Biomol. NMR*, vol. 24, pp. 1–14, 2002.

[119] S. Patra, C. Anders, N. Erwin, and R. Winter, "Osmolyte effects on the conformational dynamics of a DNA hairpin at ambient and extreme environmental conditions," *Angew. Chem. Int. Edt.*, vol. 56, no. 18, pp. 5045–5049, 2017.

[120] J. Smiatek and A. Heuer, "Deprotonation mechanism of a single-stranded DNA i-motif," *RSC Adv.*, vol. 4, pp. 17110–17113, 2014.

[121] H. Drew, R. Wing, T. Takano, C. Broka, K. Tanaka, S. Itakura, and R. Dickerson, "RCSB protein data bank ID: 1BNA." `https://www.rcsb.org/structure/1bna`, 1981. [Online; accessed 24-January-2022].

[122] P. Padrta, R. Štefl, L. Žídek, and V. Sklenář, "RCSB protein data bank ID: 1KR8." `https://www.rcsb.org/structure/1KR8`, 2002. [Online; accessed 24-January-2022].

[123] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, and A. T. Phan, "RCSB protein data bank ID: 2KF8." `https://www.rcsb.org/structure/2KF8`, 2009. [Online; accessed 24-January-2022].

[124] L. Sapier and D. Harries, "Wisdom of the crowd," *Bunsen-Magazin*, vol. 19, pp. 152–162, 2017.

[125] D. A. C. Beck, B. J. Bennion, D. O. V. Alonso, and V. Daggett, "Simulations of macromolecules in protective and denaturing osmolytes: properties of mixed solvent systems and their effects on water and protein structure and dynamics," *Methods Enzymol.*, vol. 428, pp. 373–396, 2007.

[126] R. Zangi, R. Zhou, and B. Berne, "Urea's action on hydrophobic interactions," *J. Am. Chem. Soc.*, vol. 131, no. 4, pp. 1535–1541, 2009.

[127] D. Horinek and R. R. Netz, "Can simulations quantitatively predict peptide transfer free energies to urea solutions? Thermodynamic concepts and force field limitations," *J. Phys. Chem. A*, vol. 115, no. 23, pp. 6125–6136, 2011.

[128] E. J. Guinn, L. M. Pegram, M. W. Capp, M. N. Pollock, and T. M. Record Jr, "Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer," *Proc. Natl. Acad. Sci. USA*, vol. 108, no. 41, pp. 16932–16937, 2011.

[129] S. Micciulla, J. Michalowsky, M. A. Schroer, C. Holm, R. von Klitzing, and J. Smiatek, "Concentration dependent effects of urea binding to poly (N-isopropylacrylamide) brushes: a combined experimental and numerical study," *Phys. Chem. Chem. Phys.*, vol. 18, no. 7, pp. 5324–5335, 2016.

[130] M. Haberler, C. Schröder, and O. Steinhauser, "Hydrated ionic liquids with and without solute: The influence of water content and protein solutes," *J. Chem. Theory Comput.*, vol. 8, no. 10, pp. 3911–3928, 2012.

[131] V. Lesch, A. Heuer, V. A. Tatsis, C. Holm, and J. Smiatek, "Peptides in the presence of aqueous ionic liquids: tunable co-solutes as denaturants or protectants?," *Phys. Chem. Chem. Phys.*, vol. 17, no. 39, pp. 26049–26053, 2015.

[132] D. Diddens, V. Lesch, A. Heuer, and J. Smiatek, "Aqueous ionic liquids and their influence on peptide conformations: Denaturation and dehydration mechanisms," *Phys. Chem. Chem. Phys.*, vol. 31, 2017.

[133] S. Ghosh, S. Dey, M. Patel, and R. Chakrabarti, "Can an ammonium-based room temperature ionic liquid counteract the urea-induced denaturation of a small peptide?," *Phys. Chem. Chem. Phys.*, vol. 19, no. 11, pp. 7772–7787, 2017.

[134] S. Sarkar, S. Ghosh, and R. Chakrabarti, "Ammonium based stabilizers effectively counteract urea-induced denaturation in a small protein: insights from molecular dynamics simulations," *RSC Advances*, vol. 7, no. 83, pp. 52888–52906, 2017.

[135] J. Mondal, G. Stirnemann, and B. Berne, "When does trimethylamine N-oxide fold a polymer chain and urea unfold it?," *J. Phys. Chem. B*, vol. 117, no. 29, pp. 8723–8732, 2013.

[136] J. Mondal, D. Halverson, I. T. Li, G. Stirnemann, G. C. Walker, and B. J. Berne, "How osmolytes influence hydrophobic polymer conformations: A unified view from experiment and theory," *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 30, pp. 9270–9275, 2015.

[137] Y.-T. Liao, A. C. Manson, M. R. DeLyser, W. G. Noid, and P. S. Cremer, "Trimethylamine N-oxide stabilizes proteins via a distinct mechanism compared with betaine and glycine," *PNAS*, vol. 114, no. 10, pp. 2479–2484, 2017.

[138] A. M. Bhattacharyya and P. Horowitz, "Alteration around the active site of rhodanese during urea-induced denaturation and its implications for folding," *J. Biol. Chem.*, vol. 275, no. 20, pp. 14860–14864, 2000.

[139] A. K. Bhuyan, "Protein stabilization by urea and guanidine hydrochloride," *Biochemistry*, vol. 41, no. 45, pp. 13386–13394, 2002.

[140] J. Smiatek, "Osmolyte effects: Impact on the aqueous solution around charged and neutral spheres," *J. Phys. Chem. B*, vol. 118, no. 3, pp. 771–782, 2014.

[141] F. Rodríguez-Ropero and N. F. A. van der Vegt, "On the urea induced hydrophobic collapse of a water soluble polymer," *Phys.Chem.Chem.Phys.*, vol. 17, pp. 8491–8498, 2015.

[142] S. Goyal, A. Chattopadhyay, K. Kasavajhala, and U. D. Priyakumar, "Role of urea–aromatic stacking interactions in stabilizing the aromatic residues of the protein in urea-induced denatured state," *J. Am. Chem. Soc.*, vol. 139, no. 42, pp. 14931–14946, 2017.

[143] D. Nayar, A. Folberth, and N. van der Vegt, "Molecular origin of urea driven hydrophobic polymer collapse and unfolding depending on side chain chemistry," *Phys. Chem. Chem. Phys.*, vol. 19, pp. 18156–18161, 2017.

[144] J. Rösgen, B. M. Pettitt, and D. W. Bolen, "An analysis of the molecular origin of osmolyte-dependent protein stability," *Protein Sci.*, vol. 16, no. 4, pp. 733–743, 2007.

[145] E. Schneck, D. Horinek, and R. R. Netz, "Insight into the molecular mechanisms of protein stabilizing osmolytes from global force-field variations," *J. Phys. Chem. B*, vol. 117, no. 28, pp. 8310–8321, 2013.

[146] J. M. Schurr, D. P. Rangel, and S. R. Aragon, "A contribution to the theory of preferential interaction coefficients," *Biophys. J.*, vol. 89, no. 4, pp. 2258–2276, 2005.

[147] J. Rösgen, B. M. Pettitt, and D. W. Bolen, "Uncovering the basis for nonideal behavior of biological molecules," *Biochemistry*, vol. 43, no. 45, pp. 14472–14484, 2004.

[148] A. Eiberweiser, A. Nazet, S. E. Kruchinin, M. V. Fedotova, and R. Buchner, "Hydration and ion binding of the osmolyte ectoine," *J. Phys. Chem. B*, vol. 119, no. 49, pp. 15203–15211, 2015.

[149] T.-Y. Lin and S. N. Timasheff, "Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein," *Biochemistry*, vol. 33, pp. 12695–12701, 1994.

[150] P. H. Yancey and J. F. Siebenaller, "Trimethylamine oxide stabilizes teleost and mammalian lactate dehydrogenases against inactivation by hydrostatic pressure and trypsinolysis," *J. Exp. Biol.*, vol. 202, no. 24, pp. 3597–3603, 1999.

[151] P. H. Yancey, W. R. Blake, and J. Conley, "Unusual organic osmolytes in deep-sea animals: adaptations to hydrostatic pressure and other perturbants," *Comp. Biochem. Physiol. A Mol. Integr. Physiol*, vol. 133, no. 3, pp. 667–676, 2002.

[152] M. T. Velasquez, A. Ramezani, A. Manal, and D. S. Raj, "Trimethylamine N-oxide: The good, the bad and the unknown," *Toxins (Basel)*, vol. 8, no. 11, p. 326, 2016.

[153] B. J. Bennion and V. Daggett, "Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: A chemical chaperone at atomic resolution," *Proc. Natl. Acad. Sci. USA*, vol. 101, no. 17, pp. 6433–6438, 2004.

[154] Z. A. Levine, L. Larini, N. E. LaPointe, S. C. Feinstein, and J.-E. Shea, "Regulation and aggregation of intrinsically disordered peptides," *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 9, pp. 2758–2763, 2015.

[155] Q. Zou, B. J. Bennion, V. Daggett, , and K. P. Murphy, "The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea," *J. Am. Chem. Soc.*, vol. 124, no. 7, pp. 1192–1202, 2002.

[156] P. Ganguly, T. Hajari, J.-E. Shea, and N. F. A. van der Vegt, "Mutual exclusion of urea and trimethylamine N-oxide from amino acids in mixed solvent environment," *J. Phys. Chem. Lett.*, vol. 6, no. 4, pp. 581–585, 2015.

[157] G. Graziano, "How does trimethylamine N-oxide counteract the denaturing activity of urea?," *Phys. Chem. Chem. Phys.*, vol. 13, no. 39, pp. 17689–17695, 2011.

[158] S. Paul and G. N. Patey, "Structure and interaction in aqueous ureatrimethylamine-N-oxide solutions," *J. Am. Chem. Soc.*, vol. 129, no. 14, pp. 4476–4482, 2007.

[159] U. D. Priyakumar, C. Hyeon, D. Thirumalai, and A. D. MacKerell Jr, "Urea destabilizes RNA by forming stacking interactions and multiple hydrogen bonds with nucleic acid bases," *J. Am. Chem. Soc.*, vol. 131, no. 49, pp. 17759–17761, 2009.

[160] J. Hong, M. W. Capp, C. F. Anderson, R. M. Saecker, D. J. Felitsky, M. W. Anderson, and M. T. Record, "Preferential interactions of glycine betaine and of urea with DNA: Implications for DNA hydration and for effects of these solutes on DNA stability," *Biochemistry*, vol. 43, no. 46, pp. 14744–14758, 2004.

[161] H. Klump and W. Burkart, "Calorimetric measurements of the transition enthalpy of DNA in aqueous urea solutions," *Biochim. Biophys. Acta*, vol. 475, no. 4, pp. 601–604, 1977.

[162] E. J. Guinn, J. J. Schwinefus, H. K. Cha, J. L. McDevitt, W. E. Merker, R. Ritzer, G. W. Muth, S. W. Engelsgjerd, K. E. Mangold, P. J. Thompson, M. J. Kerins, and T. M. Record Jr, "Quantifying functional group interactions that determine urea effects on nucleic acid helix formation," *J. Am. Chem. Soc.*, vol. 135, no. 15, pp. 5828–5838, 2013.

[163] S. S. Cho, G. Reddy, J. E. Straub, and D. Thirumalai, "Entropic stabilization of proteins by TMAO," *J. Phys. Chem. B*, vol. 115, no. 45, pp. 13401–13407, 2011.

[164] J. Ma, I. M. Pazos, and F. Gai, "Microscopic insights into the protein-stabilizing effect of trimethylamine N-oxide (TMAO)," *PNAS*, vol. 111, no. 23, pp. 8476–8481, 2014.

[165] Y. L. A. Rezus and H. J. Bakker, "Observation of immobilized water molecules around hydrophobic groups," *Phys. Rev. Lett.*, vol. 99, 2007.

[166] A. Panuszko, P. Bruzdziak, J. Zielkiewicz, D. Wyrzykowski, and J. Stangret, "Effects of urea and trimethylamine-N-oxide on the properties of water and the secondary structure of hen egg white lysozyme," *J. Phys. Chem. B*, vol. 113, no. 44, pp. 14797–14809, 2009.

[167] J. Hunger, K.-J. Tielrooij, R. Buchner, M. Bonn, and H. J. Bakker, "Complex formation in aqueous trimethylamine-N-oxide (TMAO) solutions," *J. Phys. Chem. B.*, vol. 116, no. 16, pp. 4783–4795, 2012.

[168] L. Larini and J.-E. Shea, "Double resolution model for studying TMAO/water effective interactions," *J. Phys. Chem. B*, vol. 117, no. 42, pp. 13268–13277, 2013.

[169] L. Knake, G. Schwaab, K. Kartaschew, and M. Havenith, "Solvation dynamics of trimethylamine N-oxide in aqueous solution probed by terahertz spectroscopy," *J. Phys. Chem. B*, vol. 119, no. 43, pp. 13842–13851, 2015.

[170] F. Meersman, D. Bowron, A. K. Soper, and M. H. Koch, "Counteraction of urea by trimethylamine N-oxide is due to direct interaction," *Biophys. J.*, vol. 97, no. 9, pp. 2559–2566, 2009.

[171] F. Meersman, D. Bowron, A. K. Soper, and M. H. Koch, "An X-ray and neutron scattering study of the equilibrium between trimethylamine N-oxide and urea in aqueous solution," *Phys. Chem. Chem. Phys.*, vol. 13, no. 30, pp. 13765–13771, 2011.

[172] J. Hunger, N. Ottosson, K. Mazur, M. Bonna, and H. J. Bakker, "Water-mediated interactions between trimethylamine-N-oxide and urea.," *Phys. Chem. Chem. Phys.*, vol. 17, no. 1, pp. 298–306, 2015.

[173] M. A. Schroer, M. Paulus, C. Jeworrek, C. Krywka, S. Schmacke, Y. Zhai, D. C. F. Wieland, C. J. Sahle, M. Chimenti, C. A. Royer, B. Garcia-Moreno, M. Tolan, and R. Winter, "High-pressure SAXS study of folded and unfolded ensembles of proteins," *Biophys. J.*, vol. 99, no. 10, pp. 3430–3437, 2010.

[174] M. A. Schroer, Y. Zhai, D. C. F. Wieland, C. J. Sahle, J. Nase, M. Paulus, M. Tolan, and R. Winter, "Exploring the piezophilic behavior of natural cosolvent mixtures," *Angew. Chem. Int. Ed.*, vol. 50, no. 48, pp. 11413–11416, 2011.

[175] C. Hölzl, D. Horinek, P. Kibies, R. Frach, S. M. Kast, S. Imoto, D. Marx, S. Suladze, and R. Winter, "Design principles for high-pressure force fields: Aqueous TMAO solutions from ambient to kilobar pressures," *J. Chem. Phys.*, vol. 144, no. 14, p. 144104, 2016.

[176] S. S. Cho, G. Reddy, J. E. Straub, and D. Thirumalai, "Entropic stabilization of proteins by TMAO," *J. Phys. Chem. B*, vol. 115, no. 45, pp. 13401–13407, 2011.

[177] Y.-T. Liao, A. C. Manson, M. R. DeLyser, W. G. Noid, and P. S. Cremer, "Trimethylamine N-oxide stabilizes proteins via a distinct mechanism compared with betaine and glycine," *Proc. Natl. Acad. Sci. USA*, vol. 114, no. 10, pp. 2479–2484, 2017.

[178] N. Smolin, V. P. Voloshin, A. V. Anikeenko, A. Geiger, R. Winter, and N. N. Medvedev, "TMAO and urea in the hydration shell of the protein SNase," *Phys. Chem. Chem. Phys.*, vol. 19, no. 9, pp. 6345–6357, 2017.

[179] P. M. Reddy, M. Taha, P. Venkatesu, A. Kumar, and M. J. Lee, "Destruction of hydrogen bonds of poly(N-isopropylacrylamide) aqueous solution by trimethylamine N-oxide," *J. Chem. Phys.*, vol. 136, 2012.

[180] F. Rodríguez-Ropero and N. F. van der Vegt, "Direct osmolyte–macromolecule interactions confer entropic stability to folded states," *J. Phys. Chem. B*, vol. 118, no. 26, pp. 7327–7334, 2014.

[181] I. Tah and J. Mondal, "How does a hydrophobic macromolecule respond to a mixed osmolyte environment?," *J. Phys. Chem. B*, vol. 120, no. 42, pp. 10969–10978, 2016.

[182] P. Ganguly, P. Boserman, N. F. van der Vegt, and J.-E. Shea, "Trimethylamine N-oxide counteracts urea denaturation by inhibiting protein–urea preferential interaction," *J. Am. Chem. Soc.*, vol. 140, no. 1, pp. 483–492, 2017.

[183] D. Nayar and N. F. van der Vegt, "Cosolvent effects on polymer hydration drive hydrophobic collapse," *J. Phys. Chem. B*, vol. 122, no. 13, pp. 3587–3595, 2018.

[184] I. Baskakov, A. Wang, and D. Bolen, "Trimethylamine-N-oxide counteracts urea effects on rabbit muscle lactate dehydrogenase function: a test of the counteraction hypothesis," *Biophys. J.*, vol. 74, no. 5, pp. 2666–2673, 1998.

[185] A. J. Wang and D. W. Bolen, "A naturally occurring protective system in urea-rich cells: Mechanism of osmolyte protection of proteins against urea denaturation," *Biochemistry*, vol. 36, pp. 9101–9108, 1997.

[186] P. Ganguly, N. F. van der Vegt, and J.-E. Shea, "Hydrophobic association in mixed urea–TMAO solutions," *J. Phys. Chem. Lett.*, vol. 7, no. 15, pp. 3052–3059, 2016.

[187] P. Venkatesu, M.-J. Lee, and H.-m. Lin, "Osmolyte counteracts urea-induced denaturation of α-chymotrypsin," *J. Phys. Chem. B*, vol. 113, no. 15, pp. 5327–5338, 2009.

[188] R. Sarma and S. Paul, "Exploring the molecular mechanism of trimethylamine-N-oxides ability to counteract the protein denaturing effects of urea," *J. Phys. Chem. B*, vol. 117, no. 18, pp. 5691–5704, 2013.

[189] J. M. Pastor, M. Salvador, M. Argandoña, V. Bernal, M. Reina-Bueno, L. N. Csonka, J. L. Iborra, C. Vargas, J. J. Nieto, and M. Cánovas, "Ectoines in cell stress protection: uses and biotechnological production," *Biotechnol. Adv.*, vol. 28, no. 6, pp. 782–801, 2010.

[190] H. J. Kunte, G. Lentzen, and E. A. Galinski, "Industrial production of the cell protectant ectoine: protection mechanisms, processes, and products," *Curr. Biotechnol.*, vol. 3, pp. 10–25, 2014.

[191] L. Czech, L. Hermann, N. Stöveken, A. A. Richter, A. Höppner, S. H. Smits, J. Heider, and E. Bremer, "Role of the extremolytes ectoine and hydroxyectoine as stress protectants and nutrients: Genetics, phylogenomics, biochemistry, and structural analysis," *Genes*, vol. 9, no. 4, p. 177, 2018.

[192] G. Zaccai, I. Bagyan, J. Combet, G. J. Cuello, B. Demé, Y. Fichou, F.-X. Gallat, V. M. G. Josa, S. von Gronau, M. Haertlein, A. Martel, M. Moulin, M. Neumann, M. Weik, and D. Oesterhelt, "Neutrons describe ectoine effects on water H-bonding and hydration around a soluble protein and a cell membrane," *Sci. Rep.*, vol. 6, p. 31434 EP, 2016.

[193] A. Roychoudhury, D. Haussinger, and F. Oesterhelt, "Effect of the compatible solute ectoine on the stability of the membrane proteins," *Protein and Peptide Letters*, vol. 19, no. 8, pp. 791–794, 2012.

[194] J. Buenger and H. Driller, "Ectoin: an effective natural substance to prevent UVA-induced premature photoaging," *Skin Pharmacol. Physiol.*, vol. 17, pp. 232–237, 2004.

[195] M. B. Hahn, S. Meyer, M.-A. Schröter, H.-J. Kunte, T. Solomun, and H. Sturm, "DNA protection by ectoine from ionizing radiation: molecular mechanisms," *Phys. Chem. Chem. Phys.*, vol. 19, no. 37, pp. 25717–25722, 2017.

[196] M.-A. Schröter, S. Meyer, M. B. Hahn, T. Solomun, H. Sturm, and H.-J. Kunte, "Ectoine protects DNA from damage by ionizing radiation," *Sci. Rep.*, vol. 7, no. 1, p. 15272, 2017.

[197] S. Meyer, M.-A. Schröter, M. B. Hahn, T. Solomun, H. Sturm, and H. J. Kunte, "Ectoine can enhance structural changes in DNA *in vitro*," *Sci. Rep.*, vol. 7, p. 7170, 2017.

[198] M. Schnoor, P. Voß, P. Cullen, T. Böking, H.-J. Galla, E. A. Galinski, and S. Lorkowski, "Characterization of the synthetic compatible solute homoectoine as a potent PCR enhancer," *Biochem. Biophys. Res. Commun.*, vol. 322, pp. 867–872, 2004.

[199] S. Ghosh, H. Dixit, and R. Chakrabarti, "Ion assisted structural collapse of a single stranded DNA: A molecular dynamics approach," *Chem. Phys.*, vol. 459, pp. 137–147, 2015.

[200] P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus, and G. N. Somero, "Living with water stress: evolution of osmolyte systems," *Science*, vol. 217, no. 4566, pp. 1214–1222, 1982.

[201] J. Rösgen and R. Jackson-Atogi, "Volume exclusion and H-bonding dominate the thermodynamics and solvation of trimethylamine-N-oxide in aqueous urea," *J. Am. Chem. Soc.*, vol. 134, no. 7, pp. 3590–3597, 2012.

[202] A. Ben-Naim, *Molecular Theory of Solutions*. New York, United States: Oxford University Press, 2006.

[203] A. Ben-Naim, *Molecular Theory of Water and Aqueous Solutions. Part I: Understanding Water*. Singapore: World Scientific Publishing, 2009.

[204] E. A. Ploetz and P. E. Smith, "Local fluctuations in solution: Theory and applications," *Adv. Chem. Phys.*, vol. 153, pp. 311–372, 2013.

[205] F. P. Buff and R. Brout, "Molecular formulation of thermodynamic functions encountered in solution theory," *J. Chem. Phys.*, vol. 23, no. 3, pp. 458–465, 1955.

[206] D. Hall, "Kirkwood-Buff theory of solutions. An alternative derivation of part of it and some applications," *Transact. Farad. Soc.*, vol. 67, pp. 2516–2524, 1971.

[207] K. E. Newman, "Kirkwood–Buff solution theory: derivation and applications," *Chem. Soc. Rev.*, vol. 23, no. 1, pp. 31–40, 1994.

[208] I. L. Shulgin and E. Ruckenstein, "The Kirkwood-Buff theory of solutions and the local composition of liquid mixtures," *J. Phys. Chem. B*, vol. 110, no. 25, pp. 12707–12713, 2006.

[209] P. E. Smith, "Computer simulation of cosolvent effects on hydrophobic hydration," *J. Phys. Chem. B*, vol. 103, no. 3, pp. 525–534, 1999.

[210] R. Chitra and P. E. Smith, "Preferential interactions of cosolvents with hydrophobic solutes," *J. Phys. Chem. B*, vol. 105, no. 46, pp. 11513–11522, 2001.

[211] I. L. Shulgin and E. Ruckenstein, "Various contributions to the osmotic second virial coefficient in protein- water- cosolvent solutions," *J. Phys. Chem. B*, vol. 112, no. 46, pp. 14665–14671, 2008.

[212] P. W. Atkins and J. de Paula, *Physical Chemistry*. Oxford (UK): Oxford Univ. Press, 2010.

[213] A. N. Krishnamoorthy, J. Zeman, C. Holm, and J. Smiatek, "Preferential solvation and ion association properties in aqueous dimethyl sulfoxide solutions," *Phys. Chem. Chem. Phys.*, vol. 18, pp. 31312–31322, 2016.

[214] B. M. Baynes and B. L. Trout, "Proteins in mixed solvents: a molecular-level perspective," *J. Phys. Chem. B*, vol. 107, no. 50, pp. 14058–14067, 2003.

[215] D. K. Hore, D. S. Walker, and G. L. Richmond, "Water at hydrophobic surfaces: when weaker is better," *J. Am. Chem. Soc.*, vol. 130, no. 6, pp. 1800–1801, 2008.

[216] D. J. Felitsky and M. T. Record, "Application of the local-bulk partitioning and competitive binding models to interpret preferential interactions of glycine betaine and urea with protein surface," *Biochemistry*, vol. 43, no. 28, pp. 9276–9288, 2004.

[217] A. Luzar and D. Chandler, "Hydrogen-bond kinetics in liquid water," *Nature*, vol. 379, no. 6560, p. 55, 1996.

[218] A. Luzar, "Resolving the hydrogen bond dynamics conundrum," *J. Chem. Phys.*, vol. 113, no. 23, pp. 10663–10675, 2000.

[219] D. van der Spoel, P. J. van Maaren, P. Larsson, and N. Tîmneanu, "Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media," *J. Phys. Chem. B*, vol. 110, no. 9, pp. 4393–4398, 2006.

[220] P. Kumar, S. V. Buldyrev, F. W. Starr, N. Giovambattista, and H. E. Stanley, "Thermodynamics, structure, and dynamics of water confined between hydrophobic plates," *Phys. Rev. E*, vol. 72, no. 5, p. 051503, 2005.

[221] D. Laage, G. Stirnemann, F. Sterpone, R. Rey, and J. T. Hynes, "Reorientation and allied dynamics in water and aqueous solutions," *Annu. Rev. Phys. Chem.*, vol. 62, pp. 395–416, 2011.

[222] S. W. de Leeuw, J. W. Perram, and E. R. Smith, "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants," *Proc. Royal Soc. A*, vol. 373, pp. 27–56, Oct. 1980.

[223] M. Neumann, "Dipole moment fluctuation formulas in computer simulations of polar systems," *Mol. Phys.*, vol. 50, no. 4, pp. 841–858, 1983.

[224] J. Caillol, D. Levesque, and J. Weis, "Theoretical calculation of ionic solution properties," *J. Chem. Phys.*, vol. 85, no. 11, pp. 6645–6657, 1986.

[225] A. R. Leach, *Molecular modeling: principles and applications*. Great Britain: Pearson Education Limited, 2001.

[226] J. W. Ponder and D. A. Case, "Force fields for protein simulations," *Adv. Prot. Chem.*, vol. 66, pp. 27–85, 2003.

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

[228] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, *et al.*, "Parmbsc1: a refined force-field for DNA simulations," *Nature Meth.*, vol. 13, no. 1, p. 55, 2016.

[229] W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*. Groningen, The Netherlands: Biomos, 1987.

[230] W. R. P. Scott, P. H. Huenenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krueger, and W. F. van Gunsteren, "The GROMOS biomolecular simulation program package," *J. Phys. Chem. A*, vol. 103, pp. 3596–3607, 1999.

[231] A. D. MacKerell Jr, B. Brooks, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus, *CHARMM: The Energy Function and Its Parametrization with an Overview of the Program*, pp. 271–277. Chichester, United Kingdom: Schleyer, P.v.R., John Wiley and Sons, 1998.

[232] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, 1998.

[233] A. D. Mackerell Jr, M. Feig, and C. L. Brooks III, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations," *J. Comput. Chem.*, vol. 25, no. 11, pp. 1400–1415, 2004.

[234] A. D. MacKerell Jr, N. Banavali, and N. Foloppe, "Development and current status of the CHARMM force field for nucleic acids," *Biopolymers*, vol. 56, no. 4, pp. 257–265, 2001.

[235] W. L. Jorgensen, D. S. Maxwell, , and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc.*, vol. 118, no. 45, pp. 11225–11236, 1996.

[236] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides," *J. Phys. Chem. B*, vol. 105, no. 28, pp. 6474–6487, 2001.

[237] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, 1995.

[238] A. R. Leach, *Energy Minimisation and Related Methods for Exploring the Energy Surface*, pp. 253 – 264. Great Britain: Pearson Education Limited, 2001.

[239] L. Verlet, "Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," *Phys. Rev.*, vol. 159, no. 1, pp. 98–103, 1967.

[240] K. Binder, J. Horbach, W. Kob, W. Paul, and F. Varnik, "Molecular dynamics simulations," *J. Phys.: Condens. Matter*, vol. 16, p. S429, 2004.

[241] T. Hansson, C. Oostenbrink, and W. van Gunsteren, "Molecular dynamics simulations," *Curr. Opin. Struct. Biol.*, vol. 12, no. 2, pp. 190–196, 2002.

[242] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nat. Struct. Biol.*, vol. 9, no. 9, pp. 646–652, 2002.

[243] D. Frenkel and B. Smit, *Understanding molecular simulation : from algorithms to applications*. San Diego, California, United States: Academic Press, 2007.

[244] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *J. Chem. Phys.*, vol. 81, no. 1, pp. 511–519, 1984.

[245] W. G. Hoover, "Canonical dynamics: equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, no. 3, p. 1695, 1985.

[246] H. C. Andersen, "Molecular dynamics simulations at constant pressure and \or temperature," *J. Chem. Phys.*, vol. 72, no. 4, pp. 2384–2393, 1980.

[247] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, 1981.

[248] S. Nosé and M. L. Klein, "Constant pressure molecular dynamics for molecular systems," *Mol. Phys.*, vol. 50, no. 5, pp. 1055–1076, 1983.

[249] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Mol. Phys.*, vol. 87, no. 5, pp. 1117–1157, 1996.

[250] A. Laio and M. Parrinello, "Escaping free-energy minima," *PNAS*, vol. 99, no. 20, pp. 12562–12566, 2002.

[251] A. Laio and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Rep.Progr.Phys.*, vol. 71, pp. 126601–126622, 2008.

[252] D. Branduardi, G. Bussi, and M. Parrinello, "Metadynamics with adaptive gaussians," *J.Chem.Theory Comput.*, vol. 8, pp. 2247–2254, 2012.

[253] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *WIREs Comput. Mol. Sci.*, vol. 1, pp. 826–843, 2011.

[254] G. Bussi and A. Laio, "Using metadynamics to explore complex free-energy landscapes," *Nat. Rev. Phys.*, vol. 2, pp. 200–212, 2020.

[255] L. Sutto, S. Marsili, and F. L. Gervasio, "New advances in metadynamics," *WIREs Comput. Mol. Sci.*, vol. 2, pp. 771–779, 2012.

[256] B. Peters, G. T. Beckham, and B. L. Trout, "Extensions to the likelihood maximization approach for finding reaction coordinates," *J. Chem. Phys.*, vol. 127, no. 3, p. 034109, 2007.

[257] P. Tiwary and B. J. Berne, "Spectral gap optimization of order parameters for sampling complex molecular systems," *Proc. Natl Acad. Sci. USA*, vol. 113, no. 11, pp. 2839–2844, 2016.

[258] D. Branduardi, F. L. Gervasio, and M. Parrinello, "From A to B in free energy space," *J. Chem. Phys.*, vol. 126, no. 5, p. 054103, 2007.

[259] G. D. Leines and B. Ensing, "Path finding on high-dimensional free energy landscapes," *Phys. Rev. Lett.*, vol. 109, p. 020601, 2012.

[260] A. P. de Alba Ortíz, J. Vreede, and B. Ensing, "The adaptive path collective variable: A versatile biasing approach to compute the average transition path and free energy of molecular transitions," *Methods Mol Biol.*, vol. 2022, pp. 255–290, 2019.

[261] M. M. Sultan and V. S. Pande, "Automated design of collective variables using supervised machine learning," *J. Chem. Phys.*, vol. 149, no. 9, p. 094106, 2018.

[262] V. Rizzi, D. Mendels, E. Sicilia, and M. Parrinello, "Blind search for complex chemical pathways using harmonic linear discriminant analysis," *J. Chem. Theory Comput.*, vol. 15, no. 8, pp. 4507–4515, 2019.

[263] G. M. Piccini and M. Parrinello, "Accurate quantum chemical free energies at affordable cost," *J. Phys. Chem. Lett.*, vol. 10, no. 13, pp. 3727–3731, 2019.

[264] J. McCarty and M. Parrinello, "A variational conformational dynamics approach to the selection of collective variables in metadynamics," *J. Chem. Phys.*, vol. 147, no. 20, p. 204109, 2017.

[265] M. M. Sultan and V. S. Pande, "tICA-metadynamics: Accelerating metadynamics by using kinetically selected collective variables," *J. Chem. Theory Comput.*, vol. 13, no. 6, pp. 2440–2447, 2017.

[266] W. Chen and A. L. Ferguson, "Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration," *J. Comput. Chem.*, vol. 39, no. 25, pp. 2079–2102, 2018.

[267] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," *Phys. Rev. Lett.*, vol. 100, p. 020603, 2008.

[268] R. Galvelis and Y. Sugita, "Replica state exchange metadynamics for improving the convergence of free energy estimates," *J. Comput. Chem.*, vol. 36, no. 19, pp. 1446–1455, 2015.

[269] S. Piana and A. Laio, "A bias-exchange approach to protein folding," *J. Phys. Chem. B*, vol. 111, no. 17, pp. 4553–4559, 2007.

[270] A. Gil-Ley and G. Bussi, "Enhanced conformational sampling using replica exchange with collective-variable tempering," *J. Chem. Theory Comput.*, vol. 11, no. 3, pp. 1077–1085, 2015.

[271] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, "Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics," *J. Phys. Chem. B*, vol. 110, no. 8, pp. 3533–3539, 2006.

[272] G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, "Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics," *J. Am. Chem. Soc.*, vol. 128, no. 41, pp. 13435–13441, 2006.

[273] M. Bonomi and M. Parrinello, "Enhanced sampling in the well-tempered ensemble," *Phys. Rev. Lett.*, vol. 104, no. 19, p. 190601, 2010.

[274] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "Gromacs: Fast, flexible, and free," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, 2005.

[275] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *J.Chem. Theory Comput.*, vol. 4, no. 3, pp. 435–447, 2008.

[276] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.

[277] S. Weerasinghe and P. E. Smith, "A Kirkwood-Buff derived force field for mixtures of urea and water," *J. Phys. Chem. B*, vol. 107, no. 16, pp. 3891–3898, 2003.

[278] H. Berendsen, J. Grigera, and T. Straatsma, "The missing term in effective pair potentials," *J. Phys. Chem.*, vol. 91, no. 24, pp. 6269–6271, 1987.

[279] D. Markthaler, J. Zeman, J. Baz, J. Smiatek, and N. Hansen, "Validation of trimethylamine-N-oxide (TMAO) force fields based on thermophysical properties of aqueous TMAO solutions," *J. Phys. Chem. B*, vol. 121, no. 47, pp. 10674–10688, 2017.

[280] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N log (N) method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993.

[281] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, 1995.

[282] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, 1997.

[283] "Molinformation cheminformatics." https://www.molinspiration.com/cgi-bin/properties. last accessed on February 1st, 2019.

[284] Y. L. A. Rezus and H. J. Bakker, "Effect of urea on the structural dynamics of water," *Proc. Natl. Acad. Sci. USA*, vol. 103, no. 49, pp. 18417–18420, 2006.

[285] V. Kocherbitov, V. Veryazov, and O. Soederman, "Hydration of trimethylamine-N-oxide and of dimethyldodecylamine-N-oxide: An ab initio study," *J. Mol. Struct. Theochem.*, vol. 808, no. 1-3, pp. 111–118, 2007.

[286] D. J. Evans and B. L. Holian, "The Nose-Hoover thermostat," *J. Chem. Phys.*, vol. 83, no. 8, pp. 4069–4074, 1985.

[287] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.

[288] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Ann. Phys.*, vol. 369, no. 3, pp. 253–287, 1921.

[289] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, pp. 37–52, 1987.

[290] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comput. Stat.*, vol. 2, pp. 433–459, 2010.

[291] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: new feathers for an old bird," *Comput. Phys. Commun.*, vol. 185, no. 2, pp. 604–613, 2014.

[292] L. J. Nordstrom, C. A. Clark, B. Anderson, S. M. Champlin, and J. J. Schwinefus, "Effect of ethylene glycol, urea, and N-methylated glycines on DNA thermal stability: The role of DNA base pair composition and hydration," *Biochemistry*, vol. 45, no. 31, pp. 9604–9614, 2006.

[293] J. Rösgen, B. M. Pettitt, and D. W. Bolen, "Uncovering the basis for nonideal behavior of biological molecules," *Biochemistry*, vol. 43, no. 45, pp. 14472–14484, 2004.

[294] C.-H. Lee, H. Mizusawa, and T. Kakefuda, "Unwinding of double-stranded DNA helix by dehydration," *Proc. Natl. Acad. Sci. USA*, vol. 78, no. 5, pp. 2838–2842, 1981.

[295] W. Gilkerson and K. Srivastava, "The dipole moment of urea," *J. Phys. Chem.*, vol. 64, no. 10, pp. 1485–1487, 1960.

[296] T. A. Larsen, M. L. Kopka, and R. E. Dickerson, "Crystal structure analysis of the B-DNA dodecamer CGTGAATTCACG," *Biochemistry*, vol. 30, no. 18, pp. 4443–4449, 1991.

[297] T. A. Larsen, M. L. Kopka, and R. E. Dickerson, "RCSB protein data bank ID: 1D29." `https://www.rcsb.org/structure/1d29`, 1991. [Online; accessed 02-March-2018].

[298] "Sigma-aldrich product information: ectoine." `https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/e2271pis.pdf`. [Online; accessed 02-March-2018].

[299] "PubChem compound summary for CID 6993223: ectoine zwitterion." `https://pubchem.ncbi.nlm.nih.gov/compound/6993223#section=Top`, 2006. [Online; accessed 07-March-2018].

[300] "Sigma aldrich: Ectoine product information." `https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/e2271pis.pdf`, 2007. [Online; accessed 27-August-2018].

[301] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–1174, 2004.

[302] J. Wang, W. W. P. A., Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *J. Mol. Graph. Model.*, vol. 25, no. 2, pp. 247–260, 2006.

[303] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The Amber biomolecular simulation programs," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1668–1688, 2005.

[304] A. W. S. da Silva and W. F. Vranken, "ACPYPE-Antechamber python parser interface," *BMC Res. Notes*, vol. 5, no. 1, p. 367, 2012.

[305] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, and T. E. Cheatham III, "Assessing the current state of AMBER force field modifications for DNA," *J. Chem. Theo. Comput.*, vol. 12, no. 8, pp. 4114–4127, 2016.

[306] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, and A. T. Phan, "Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers," *J. Am. Chem. Soc.*, vol. 131, pp. 4301–4309, 2009.

[307] "RCSB protein data bank ID: 2KF7." `http://www.rcsb.org`. [last accessed on February 2^{nd}, 2019].

[308] W. Humphrey, A. Dalke, and K. Schulten, "VMD - visual molecular dynamics," *J. Molec. Graphics*, vol. 14, pp. 33–38, 1996.

[309] "VMD - visual molecular dynamics." `http://www.ks.uiuc.edu/Research/vmd/`. [last accessed on February 2^{nd}, 2019].

[310] D. R. Canchi, P. Jayasimha, D. C. Rau, G. I. Makhatadze, and A. E. Garcia, "Molecular mechanism for the preferential exclusion of TMAO from protein surfaces," *J. Phys. Chem. B*, vol. 116, no. 40, pp. 12095–12104, 2012.

[311] J. Zeman, C. Holm, and J. Smiatek, "The effect of small organic cosolutes on water structure and dynamics," *J. Chem. Eng. Data*, vol. 65, no. 3, pp. 1197–1210, 2019.

[312] C. Yang, M. Kulkarni, M. Lim, and Y. Pak, "*In silico* direct folding of thrombin-binding aptamer G-quadruplex at all-atom level," *Nucl. Acids Res.*, vol. 45, no. 22, pp. 12648–12656, 2017.

[313] J. Knop, S. Patra, B. Harish, C. A. Royer, and R. Winter, "The deep sea osmolyte trimethylamine N-oxide and macromolecular crowders rescue the antiparallel conformation of the human telomeric G-quadruplex from urea and pressure stress," *Chem. Eur. J.*, vol. 24, no. 54, pp. 14346–14351, 2018.

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

[315] M. Rebič, A. Laaksonen, J. Šponer, J. Uličný, and F. Mocci, "Molecular dynamics simulation study of parallel telomeric DNA quadruplexes at different ionic strengths: Evaluation of water and ion models," *J. Phys. Chem. B*, vol. 120, no. 30, pp. 7380–7391, 2016.

[316] I. S. Joung and T. E. Cheatham III, "Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations," *J. Phys. Chem. B*, vol. 112, no. 30, pp. 9020–9041, 2008.

[317] P. Debye, "Interferenz von Röntgenstrahlen und Wärmebewegung," *Ann. Phys.*, vol. 348, no. 1, pp. 49–92, 1913.

[318] I. Waller, "Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen," *Z. Phys. A*, vol. 17, no. 1, pp. 398–408, 1923.

[319] Z. Sun, Q. Liu, G. Qu, Y. Feng, , and M. T. Reetz, "Utility of B-factors in protein science: Interpreting rigidity, flexibility, and internal motion and engineering thermostability," *Chem. Rev.*, vol. 119, no. 3, pp. 1626–1665, 2019.

[320] I. Bahar, A. R. Atilgan, and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential," *Fold. Des.*, vol. 2, no. 3, pp. 173–181, 1997.

[321] D. S. Tomar, G. Licari, J. Bauer, S. K. Singh, L. Li, and S. Kumar, "Stress-dependent flexibility of a full-length human monoclonal antibody: Insights from molecular dynamics to support biopharmaceutical development," *J. Pharm. Sci*, pp. 1–10, 2021.

[322] "3DNA: a suite of software programs for the analysis, rebuilding and visualization of 3-dimensional nucleic acid structures." `http://x3dna.org`. [last accessed on May 28^{th}, 2020].

[323] V. Tsvetkov, G. Pozmogova, and A. Varizhuk, "The systematic approach to describing conformational rearrangements in G-quadruplexes," *J. Biomol. Struct. Dyn.*, vol. 34, no. 4, pp. 705–715, 2016.

[324] "MDAnalysis: A python package for the analysis of molecular dynamics simulations.." `https://www.mdanalysis.org/`. [last accessed on May 30^{th}, 2020].

[325] C. Kramer, A. Spinn, and K. R. Liedl, "Charge anisotropy: Where atomic multipoles matter most," *J. Chem. Theory Comput.*, vol. 10, no. 10, pp. 4488–4496, 2014.

[326] R. Qi, G. Wei, B. Ma, and R. Nussinov, "Replica exchange molecular dynamics: A practical application protocol with solutions to common problems and a peptide aggregation and self-assembly example," *Methods Mol. Biol.*, vol. 1777, pp. 101–119, 2018.

[327] S. Y. Reddy, F. Leclerc, and M. Karplus, "DNA polymorphism: A comparison of force fields for nucleic acids," *Biophys. J.*, vol. 84, no. 3, pp. 1421–1449, 2003.

[328] V. Minhas, T. Sun, A. Mirzoev, N. Korolev, A. P. Lyubartsev, and L. Nordenskiöld, "Modeling DNA flexibility: Comparison of force fields from atomistic to multiscale levels," *J. Phys. Chem. B*, vol. 124, no. 1, pp. 38–49, 2020.

[329] E. D. Holmstrom, N. F. Dupuis, and D. J. Nesbitt, "Kinetic and thermodynamic origins of osmolyte-influenced nucleic acid folding," *J. Phys. Chem. B*, vol. 119, no. 9, pp. 3687–3696, 2015.

[330] D. Lambert and D. E. Draper, "Denaturation of RNA secondary and tertiary structure by urea: simple unfolded state models and free energy parameters account for measured m-values," *Biochemistry*, vol. 51, no. 44, pp. 9014–9026, 2012.

[331] J. Yoon, D. Thirumalai, and C. Hyeon, "Urea-induced denaturation of preQ1-riboswitch," *J. Am. Chem. Soc.*, vol. 135, no. 32, pp. 12112–12121, 2013.

[332] J. C. Miner and A. E. García, "Equilibrium denaturation and preferential interactions of an RNA tetraloop with urea," *J. Phys. Chem. B*, vol. 121, pp. 3734–3746, 2017.

List of Figures

List of Tables