UNIVERSITY OF
BIRMINGHAM

DNA as a Vehicle for Chemistry and Assembly

A Thesis Submitted to
the University of Birmingham
for the degree of

Doctor of Philosophy

Student Name: Benjamin Allott
Supervisors: Prof. Rachel O'Reilly
Dr. Dwaipayan Chakrabarti

Date: 3rd December, 2024
Academic Year: 2024/2025
Word Count: 33819

School of Chemistry
College of Engineering and Physical Sciences
University of Birmingham

Contents

List of Figures

List of Tables

Acronyms

DMF N,N-Dimethylformamide

DMSO Dimethyl Sulfoxide

DNA Deoxyribonucleic Acid

DTS DNA Templated Synthesis

FRET Fluorescence Resonance Energy Transfer

LCMS Liquid Chromatography-Mass Spectrometry

MC Monte Carlo

MD Molecular Dynamics

NCL Native Chemical Ligation

NMR Nuclear Magnetic Resonance

PAGE Polyacrylamide Gel Electrophoresis

PD Palladium

RAFT Reversible Addition-Fragmentation Chain-Transfer

RDRP Reversible Deactivation Radical Polymerization

RNA Ribonucleic Acid

RP-HPLC Reversed Phase High Performance Liquid Chromatography

SN2 Bimolecular Nucleophilic Substitution

SPDP 3-(2-Pyridyldithio)propionic acid *N*-hydroxysuccinimide ester

ssDNA Single-Stranded DNA

TAMRA 6-Carboxytetramethylrhodamine

TCEP Tris(2-carboxyethyl)phosphine

tRNA Transfer RNA

Abstract

This thesis delves into the intricacies of DNA Templated Synthesis (DTS) and investigates the self-assembly of octahedral patchy colloids in computer simulations. A common theme for both strands of research is DNA, which can be used to functionalise the patches to encode a hierarchy of interactions that could as well be specific for programming distinct self-assembly pathways. Our experimental study reveals and measures how a unique protective effect present in DTS can shield a labile thioester-TAMRA from hydrolysis. We show that protection is enhanced when the abasic site is located across the minor groove, specifically the -3 position. It is proven that this protective effect occurs *via* the interaction of the TAMRA moiety with an abasic site, with variations in the degree of protection according to the relative position of the abasic site to that of the TAMRA linker. This protective effect was discovered and measured in the hopes that it could enhance the synthesis of long-chain polymers *via* DTS. This experimental chapter is complemented by the following molecular dynamics study concerned with modelling this unique effect.

The study employs models of the short-TAMRA linker at abasic sites +3 and -3, with TAMRA represented either as the lactone or zwitterion resonance form at neutral or physiological salt concentrations. Neutral salt concentration conditions involve sufficient sodium ions to neutralise the negative charge of the DNA phosphate backbone. In contrast, physiological salt concentration not only neutralises this charge but also includes additional sodium and chloride ions in equal proportions to mimic physiological conditions.

We hypothesise, based on both experimental and computational studies, that the protective effect necessitates physiological salt concentration and initially requires the lactone resonance form for insertion.

This dual approach of experimental and computational research provided a detailed understanding of how the protection mechanism is manifested. The study of self-assembly of octahedral patchy colloids explores how different self-assembly pathways can be programmed to yield simple cubic crystals and whether the quality of the crystals self-assembled differs in those cases. To this end, the four patches on a plane are distinguished from the remaining two across the plane, the former referred to as equatorial patches (labelled E) and the latter axial patches (labelled A), in terms of the strength of interactions between the patches within a given type. Different scenarios are considered in computer simulations: equatorial-bias (E-E interaction strengths 5 times stronger than A-A interaction strengths) and axial-bias (A-A interaction strengths 5 times stronger than E-E interaction strengths, with and without interactions between different types of patches, along with the control scenario (no distinction between patch-patch interactions). We show marginally better-quality crystals are produced *via* equatorial biasing with no (E-A interaction), with the control scenario of no biasing being a close second and characterisations of the crystallisation pathways in the different scenarios. Our findings suggest that uniform growth of the largest crystalline cluster in all three dimensions favours the formation of good-quality simple cubic crystals. This study elucidates key parameters influencing the self-assembly of simple cubic crystals from octahedral patchy colloids. This thesis not only advances knowledge of DTS and colloidal self-assembly but sets the stage for future explorations, developing these phenomena for material and drug development.

Acknowledgements

I would like to express my deepest gratitude to Rachel O'Reilly for her consistent and invaluable support and guidance. Despite her busy schedule, Rachel always arranged scientific meetings where she would probe my trajectory of scientific inquiry to ensure high-quality and air-tight research.

I am also immensely thankful to Dwaipayan for his warmth, patience, wisdom, and the time he dedicated to mentoring me in computational simulations. Creating a nurturing, supportive ecosystem within the Chakrabarti group.

Special thanks to Andrew Turbefield and John Bath for their intellectual guidance in the Ribosome project meetings between the University of Birmingham and Oxford. Among the other mentors, I am super grateful to Thomas Wilks for the myriad of stimulating scientific discussions we shared in the lab and his unwavering personal support throughout my PhD journey.

My time with the O'Reilly group has been filled with fun, friendship, and memorable experiences, for which I am deeply thankful. The Chakrabarti group has provided not only humorous moments both in and out of the office but also deep, tangential conversations spanning science, life and the Welsh.

A special acknowledgement goes to Martin Burgh, who, despite having a family and a full-time job, generously offered his time to discuss and guide me through molecular dynamics simulations and Gaussian scans. His support has been invaluable to my research progress.

CHAPTER 1

Introduction

1.1 | The Structure and Composition of DNA

Deoxyribonucleic acid (DNA) is a fundamental biomolecule present in all living organisms, serving as the genetic blueprint for life.[1] Its structure allows for replication, modification, and storage of genetic information.[2] Ribose nucleic acid (RNA) allows for the translation of the genetic code into proteins. The nucleic acid's ability (DNA and RNA) to instruct the ribosomes in protein synthesis stems from its sequence of nucleobases. Both sets of nucleic acids host adenine, cytosine, and guanine; however, DNA has thymine, whereas RNA hosts uracil. These nucleobases, two pyrimidines and two purines, pair specifically (adenine with thymine/uracil, and cytosine with guanine) following Watson-Crick base pairing rules (Figure 1.1).[3] The sequence of these bases dictates the sequence of amino acids in proteins, with every three nucleobases encoding a specific amino acid.[4]

Both nucleic acids are comprised of a phosphate backbone, ribose sugar, and nucleobase, the difference stemming from the identity of one nucleobase and the identity of the ribose sugar. RNA is comprised of ribose, whereas DNA is comprised of deoxyribose. This difference in ribose structure transpires to a differences in structure and reactivity: the presence of the 2'-hydroxy group in ribose makes RNA more chemically reactive and less stable than DNA. This hydroxyl group makes RNA more prone to hydrolysis, especially in alkaline conditions, as it facilitates the cleavage of phosphodiester bonds.[5] the presence of the 2'-hydroxyl group allows RNA to adopt a variety of secondary and tertiary structures. These structures, including hairpin loops and pseudoknots, are essential for RNA's diverse functions, such as acting in catalysis (e.g., ribozymes) and facilitating gene expression and regulation.[6]

The central dogma of molecular biology explains the flow of genetic information within cells: DNA is first transcribed into RNA, and then the RNA is translated into proteins. This process is fundamental for gene expression, ensuring that the genetic code stored in DNA is accurately conveyed to produce functional proteins, which carry out vital roles in the cell. .[7]

The geometric aspects of DNA and RNA are also crucial for their function. DNA typically forms a double helix with two antiparallel strands, meaning the strands run in opposite directions. This antiparallel arrangement is essential for the replication and stability of the DNA molecule.[8] In contrast, RNA can form various secondary structures, including hairpins and loops, due to intramolecular base pairing.[9]

The duplex structure of DNA is stabilised by hydrogen bonds between the nucleobases and by base-stacking interactions. The antiparallel nature of the DNA strands allows for the complementary base pairing necessary for the double helix structure. RNA duplexes, although less common, also exhibit antiparallel strands and are stabilised similarly, though they are typically shorter and less stable than DNA duplexes.[10]

Understanding these structural and functional nuances is essential for discussing the results appropriately, as they highlight the importance of nucleotide sequence and structure in genetic information processing and stability.

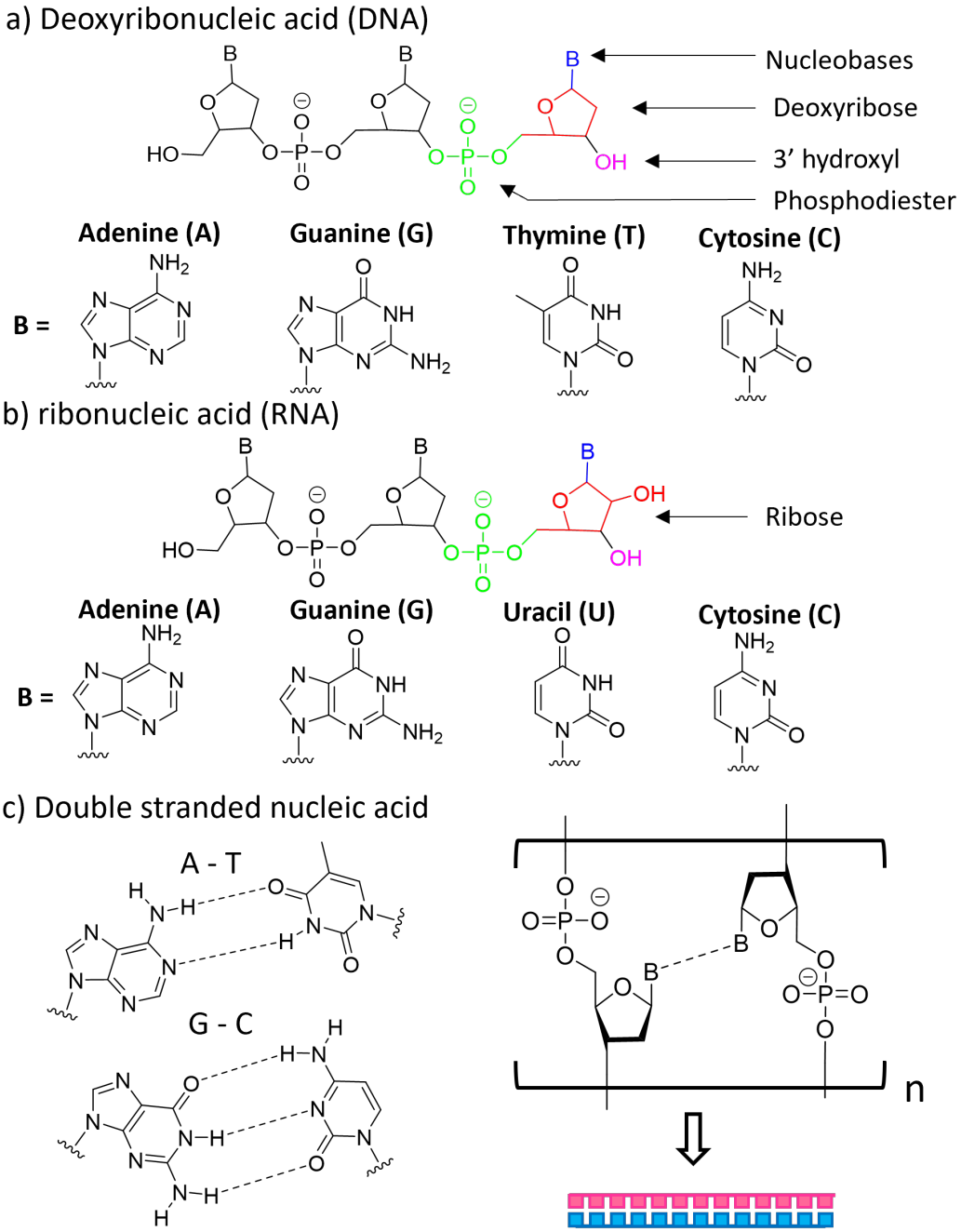


Fig. 1.1. Skeletal diagram of DNA and RNA displaying each respective nucleobase. a) is the schematic representation of DNA with each constituent component: sugar, phosphate and base, with each various base represented underneath. b) exhibits the schematic representation of RNA with each base displayed underneath. c) displays Watson-Crick hydrogen bonding between the bases of DNA and shows the cartoon representation of the duplex used throughout the thesis.

1.2 | Nature's Utilisation of DNA for Synthesis (Ribosome)

Nature's synthesis of proteins is a remarkable methodology that is programmable, autonomous, and sequence-specific, and it harbours the capability for molecular evolution.[11,12] To construct a long-chain molecule utilising this methodology necessitates the ability to store, read, and implement synthetic instructions sequentially.[13] This implies that reacting monomers must bear a DNA sequence that is coherent with the synthesis instructions.

The synthesis program unfolds through intricate steps of transcription and translation.[13–16] Transcription is the intricate process of converting genetic information into synthetic instructions, whereas translation refers to the formation of the corresponding protein, abiding by these instructions.

The genetic information is meticulously stored in a hierarchical assembly of DNA housed within the chromosomes located in the cell nucleus.[17] In eukaryotic organisms, such as plants, animals, and fungi, chromosomes are located in the cell nucleus, whereas in prokaryotic organisms like bacteria and archaea, the genetic material is typically found in a single, circular chromosome located in the nucleoid region, which is not membrane-bound[18,19]. The specific region of genetic information intended for transcription is demarcated by proteins—transcription factors—that bind to specific loci on the DNA.[20] These transcription factors attract RNA polymerase to the transcription site.[21]

RNA polymerase navigates the upstream 5 _ 3 DNA strand, transcribing the nucleotide sequence using free nucleotides as the complementary base pairs, and subsequently, re-anneals the strand as it traverses the DNA. On the 5 end of the transcribed RNA, a 7-methylguanosine cap (5 G cap) is added, which protects the RNA from degradation and is involved in ribosome binding during translation. The 3 end of the RNA is extended with a polyadenylate (poly A) tail, which also protects the RNA and aids in the termination of transcription (Figure 1.2).[21] Introns are excised, and exons are amalgamated through RNA splicing, mediated by SNURPS, which discern the commencement and termination of the sequence slated for excision.[22] Subsequently, the spliceosome degrades the introns for recycling and ligates the exons.[23]

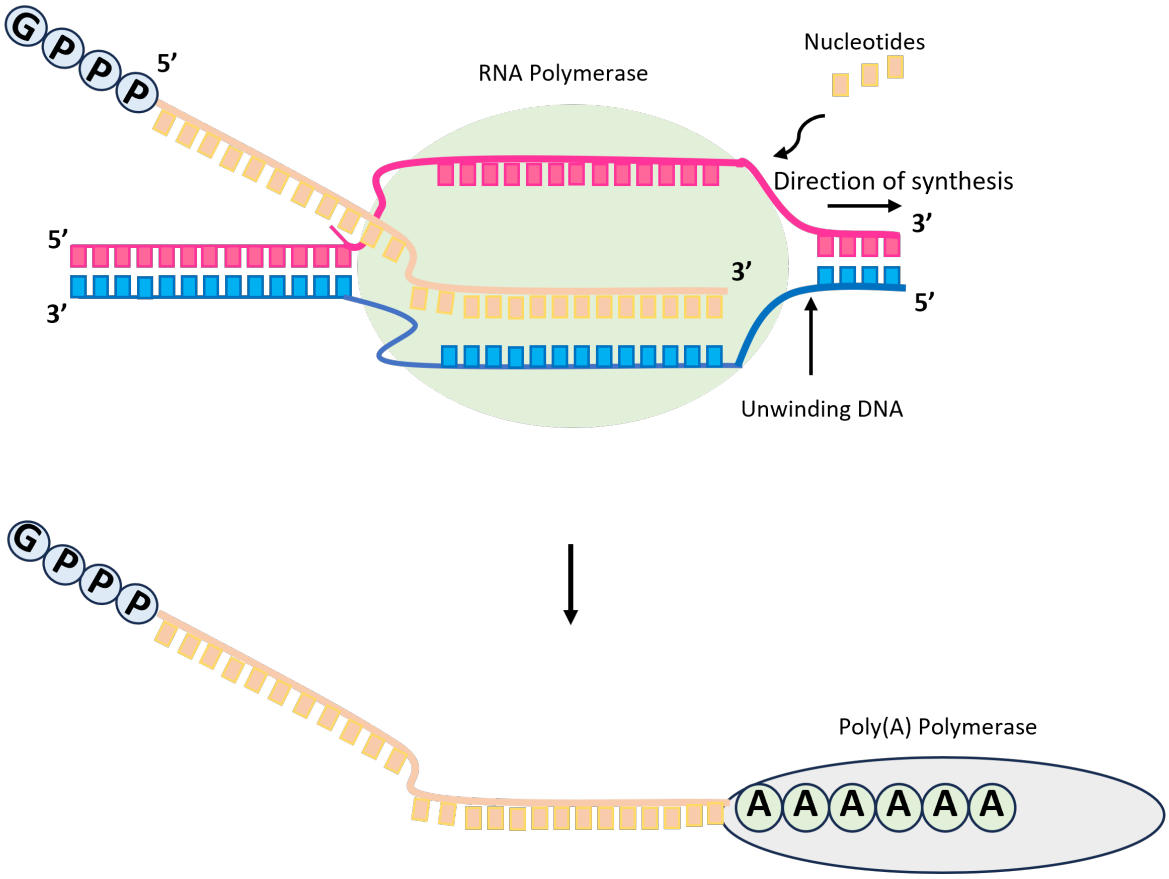


Fig. 1.2. Transcription and RNA processing: RNA polymerase synthesises a complementary RNA strand from the DNA template, moving in the 5 to 3 direction. A 5 G cap is added to the RNA for protection and ribosome binding. Following transcription, a polyadenylate (poly A) tail is added to the 3 end by poly(A) polymerase, enhancing RNA stability and assisting in termination.

1.2.1 The Ribosome: Architect of Proteins

The ribosome is a substantial ribonucleoprotein, composed of two heterogeneous subunits: one large and one small.[16] Both subunits present two binding sites, each capable of accommodating t-RNA molecules. The ribosome, functioning as a catalyst, orchestrates itself around the mRNA strand, meticulously reading the sequence of nucleic acids embedded within the mRNA.[24] The sequence of nucleic acids on the mRNA is read in sets of three nucleotides, known as codons. Each codon specifies a particular amino acid, or a stop signal, in the genetic code. For instance, the codon AUG codes for the amino acid methionine and also serves as the start codon, signaling the beginning of translation. Other examples include UUU coding for phenylalanine, and UGA, UAA, and UAG serving as stop codons, which signal the termination of translation. For a comprehensive list of codons and their corresponding amino acids, refer to the standard codon table (Figure 1.3) Every set of three nucleic acids on the mRNA template strand forms a codon, each linked to a corresponding t-RNA molecule.[25]

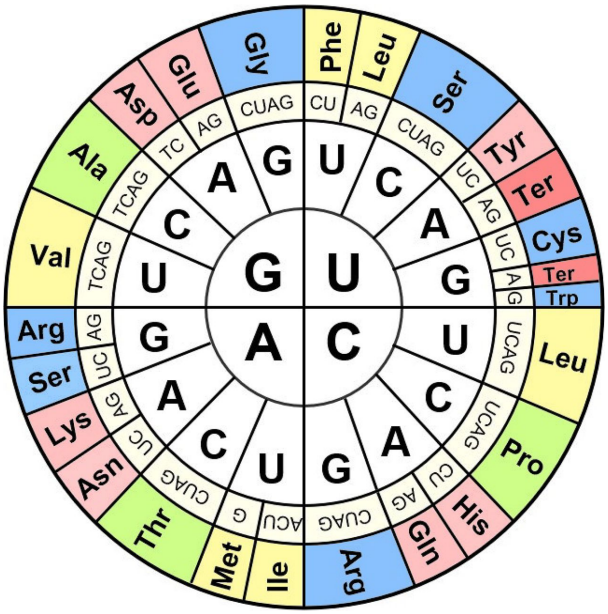


Fig. 1.3. The codon chart illustrates how sequences of three RNA nucleotide bases determine specific amino acids. The chart's three outer rings represent the different possibilities for each base in the codon. The innermost ring shows the first base (A, U, G, or C), the middle ring represents the second base, and the outermost ring shows the third base. By following the bases outward from the center, you can determine the corresponding amino acid or stop signal. For instance, the codon "AUG" codes for methionine (Met), while "UAA" codes for a termination signal (Ter). Figure adapted from Ref.[26]

The chemistry behind this precise matching begins before translation, when aminoacyl-tRNA synthetases catalyse the attachment of each amino acid to its respective t-RNA, a process

called aminoacylation. This reaction is powered by ATP and results in a high-energy bond between the amino acid and the tRNA, preparing the amino acid for its role in translation. Once charged, the tRNA delivers the amino acid to the ribosome, specifically binding at the ribosome's A site during elongation.[27]

Peptide bond formation occurs at the ribosome's peptidyl transferase center, an RNA-based enzymatic site within the large ribosomal subunit. During this reaction, the amino group of the incoming amino acid attacks the carboxyl group of the growing peptide chain, forming a new peptide bond. This process is facilitated by the ribosomal RNA (rRNA) and involves a dehydration synthesis, releasing water and linking amino acids in the growing polypeptide chain.[28]

The ribosome's smaller subunit aligns the mRNA to facilitate the efficient reading of codon sets,[29] while the larger subunit extracts the amino acid from the t-RNA and incorporates it into the elongating protein chain (Figure 1.4).

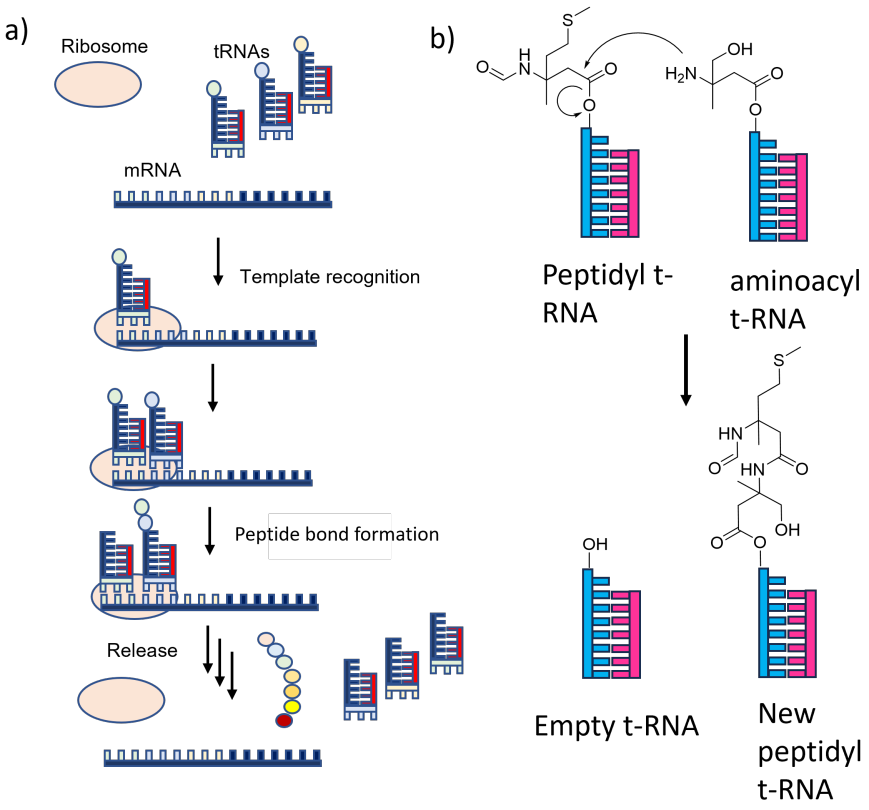


Fig. 1.4. Schematic representation of the translation process involved in ribosomal synthesis. The mRNA strand encodes the chemical composition of the desired protein in a sequence of codons which correspond to a unique a t-RNA molecule. The corresponding t-RNA molecule hybridises to the templating m-RNA strand *via* the ribosome. This process repeats, placing the t-RNA units in close proximity, consequently encouraging the acyl transfer of the proteins to take place.

1.2.2 Functionality of Ribosomal Sites

The ribosome accommodates three pivotal sites: the A site, the P site, and the E site.[15] The journey of the amino acid commences at the A site, where the integrity of the codon – anti-codon match is meticulously scrutinised. Upon a successful match, the amino acid migrates to the P site, where it merges into the burgeoning protein chain through a meticulous insertion mechanism, enabling the assimilation of the incoming monomer into the distal end of the chain. Subsequently, the depleted t-RNA trans-locates to the E site, where it is expelled from the ribosome, poised for recycling(Figure 1.5).

This sophisticated ribosomal method of synthesis enables the selective templating of large, multicomponent proteins by enhancing the proximity of t-RNA molecules through strategic binding to the mRNA strand.[30,31] For the ribosome to selectively template chemistries through this mechanism, the concentrations of the free t-RNAs within the cell must be sufficiently attenuated to render off-template reactions highly improbable.[32]

This biomolecular methodology exemplifies an exceedingly efficient mechanism of sequence-specific, programmable synthesis, tailor-made for in-vitro protein synthesis. The adaptation of this synthesis methodology for in-vitro systems, utilising monomers divergent from amino acids, could empower scientists to proficiently fabricate sequence-defined polymers.

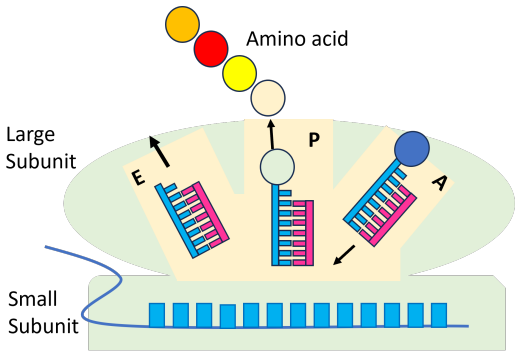


Fig. 1.5. The ribosome consists of three key sites: the A site, P site, and E site. The process begins at the A site, where the codon-anticodon pairing is verified. The amino acid then moves to the P site, where it is incorporated into the growing protein chain. Finally, the tRNA, now empty, moves to the E site and exits the ribosome, ready for recycling.

DNA-Recorded Synthesis and Directed Evolution

One of the first mechanisms applied to replicate to some degree the synthetic prowess of the ribosome was DNA-recorded synthesis.[33] Using a unique DNA sequence associated with a different molecule of interest combinatorial libraries are produced with the resulting polymer ligated to a unique DNA duplex describing its composition.[34] This technique has been utilised in industry to describe and screen libraries of more than 10^9 molecules. This technique can occur in solution[35] or *via* the use of split pool approaches (Figure 1.6).[36] The advantage of this method is the ability to handle huge quantities of compounds simultaneously due to the tracking and identification enabled *via* the DNA tag. However, this synthesis and screening does not fully encapsulate the synthetic capabilities of the ribosome as it's not a programmable nor automated method of synthesis.

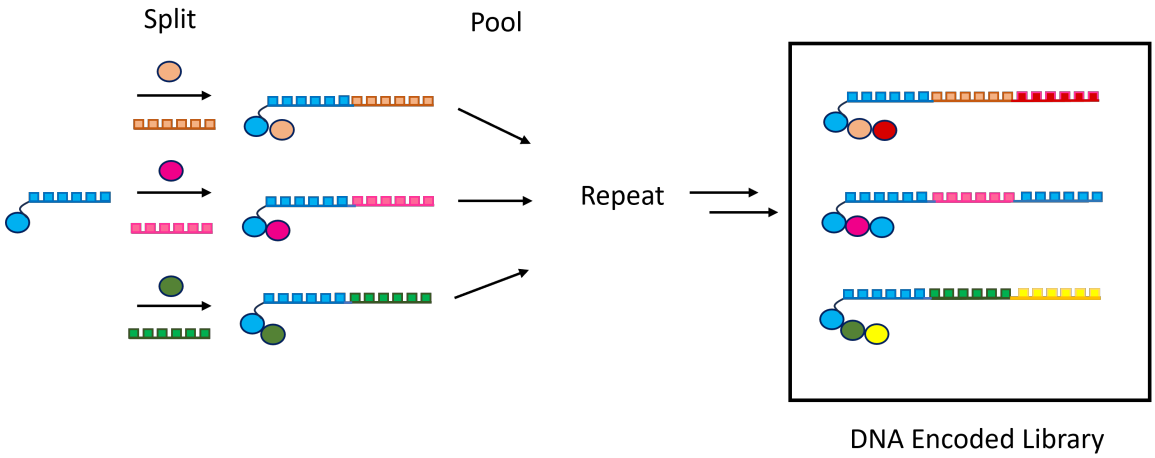


Fig. 1.6. Demonstrating the split-pool approach to the formation of a DNA-encoded library. In the split step, a diverse set of chemical building blocks is attached to DNA fragments, each uniquely tagged with a specific DNA sequence. These DNA-conjugated molecules are then pooled together. In the subsequent repeat cycles, the pool is split again, and new building blocks are added, each step appending unique DNA tags to the growing library. This process is iterated multiple times, resulting in a combinatorial library of compounds, each uniquely identifiable by its DNA sequence tag, forming the DNA-encoded library.

1.3 | DNA Routing

DNA can also be utilised as a director for specific chemical synthesis. In DNA routing, desired functioning monomers are ligated to a unique single-strand DNA strand. This single-strand DNA acts as a gene for this particular molecule with a bespoke sequence of nucleotides describing characteristics, for example in the scenario of utilising amino acids, the nucleotide

sequence can describe things such as (side chain identity, chirality, backbone length etc).[37] These DNA tags are then subjected to a range of reaction stations, with each reaction station (resin, column or bead) ligated to a particular anticodon designed to hybridise to a particular subset of the library (Figure 1.7).[38] Each of these various reaction points is associated with a particular reaction. After each subsequent reaction, the DNA strands are washed allowing for the addition and hybridisation to sequential reacting stations. Thereby through a process of sequential steps of splitting, synthesis and pooling, DNA-directed synthesis is achieved. This process enables both precise identification and control over the final product enabled by the sequence of the DNA tag. DNA-directed synthesis therefore not only describes the product but also dictates its reaction route. This method has been successful in routing the synthesis of short peptides.[39] However, again this process doesn't fully capture the capability of the ribosome, as the reactions do not occur in one reaction vessel and the polymerisation is not completely autonomous.

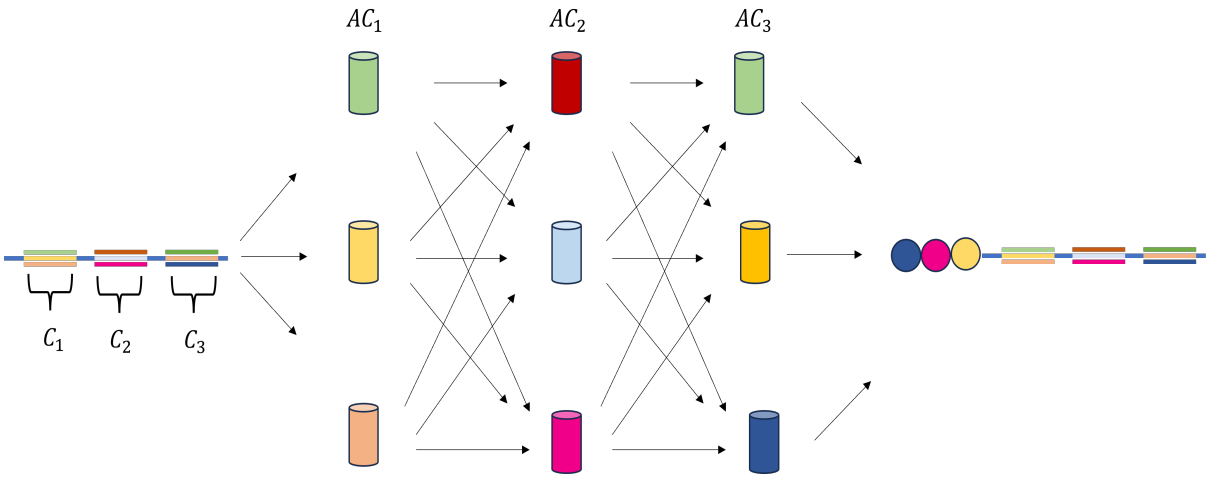


Fig. 1.7. Illustration of DNA routed synthesis showing hybridisation of ssDNA genes to complementary anticodon sequences attached to solid support.

DNA Architectures for Reaction Templating

To emulate the ribosome's autonomous, sequence-specific, and programmable synthetic capabilities, DNA can be utilised as a fundamental component. In ribosomes, t-RNA molecules host the reacting monomers, with their sequence determined by m-RNA. To apply this mechanism for external synthetic applications and non-amino acid chemistries, certain modifications are necessary. One critical alteration is the substitution of RNA with DNA, thereby preventing side reactions caused by RNA's hydroxyl groups acting as nucleophiles.

As previously mentioned, regarding the stability of RNA, RNA is inherently less stable than DNA due to the presence of the 2'-hydroxyl group on the ribose sugar, which makes RNA more prone to hydrolysis. This instability, combined with the nucleophilic nature of the hydroxyl group, can lead to side reactions and degradation, complicating its use in synthetic applications. Substituting RNA with DNA improves stability, as DNA lacks the 2'-hydroxyl group, making it less reactive and more suitable for long-term and controlled synthetic processes.

RNA's ability to act as a nucleophile is largely attributed to the presence of the 2'-hydroxyl (2'-OH) group in its ribose sugar. This 2'-OH group can participate in nucleophilic attacks, which are crucial for several biological processes. For instance, in self-splicing introns—specifically group I and group II introns—the 2'-OH group initiates a nucleophilic attack on the phosphodiester bond within the RNA backbone, leading to the excision of the intron and ligation of exons.[40] Similarly, ribozymes such as the hammerhead ribozyme utilize the 2'-OH group for nucleophilic attack during the cleavage of RNA molecules, resulting in the formation of a 2',3'-cyclic phosphate and a 5'-hydroxyl terminus.[41] Moreover, during RNA processing events like tRNA splicing and rRNA maturation, the 2'-OH group acts as a nucleophile, facilitating various cleavage and ligation reactions.[42] Additionally, the nucleophilic nature of RNA's 2'-OH group is leveraged in laboratory settings under specific conditions to induce RNA cleavage. The presence of divalent metal ions, such as Mg^{2+}, can enhance this nucleophilicity by stabilizing the transition state of the reaction.[43]

Furthermore, a thorough investigation into the range of chemical reactions that can be conducted on DNA is imperative for complete control over the design of such mechanisms. This includes exploring the conformations and geometries of DNA modified to template reactions, as well as identifying the extent of chemical substitutions and reactions achievable with modified DNA. A diverse array of DNA architectures that successfully template chemical reactions has been identified, broadening the scope of potential synthetic applications (Figure 1.8).

End-of-Helix Architecture One such architecture is the end-of-helix. In this design, the reactive groups are localised at the complementary 5' and 3' termini of the DNA strands.

T Architecture The T architecture hosts a modification attached *via* a nucleobase, as opposed to an hydroxyl strand terminus. In this design, an advantantagous non-hybridized overhang is exhibited, this is useful for carrying out reaction *via* DNA as the strand can be

utilised to trigger more complex nucleic acid displacement reactions *via* toe-hold mediated strand displacement.

Ω-Architecture Similar to the End-of-Helix architecture, the Ω-Architecture attaches reactive groups to complementary 3' and 5' termini; however, in this architecture, a non-hybridised loop region is included.

Minor-Groove Coupling A distinctive arrangement of reactants *via* templating DNA is the minor-groove architecture. The minor-groove coupling allows both reactive groups to be attached to either 3' or 5' termini, allowing for the same directionality whilst enabling a close spatial co-localization of the reactive moieties.

Yoctoreactor Design The yoctoreactor design employs a three-way junction to template chemical reactions through internal modification of nucleic acid strands. This junction acts as a scaffold, organising reactants into a precise spatial arrangement necessary for reactions.

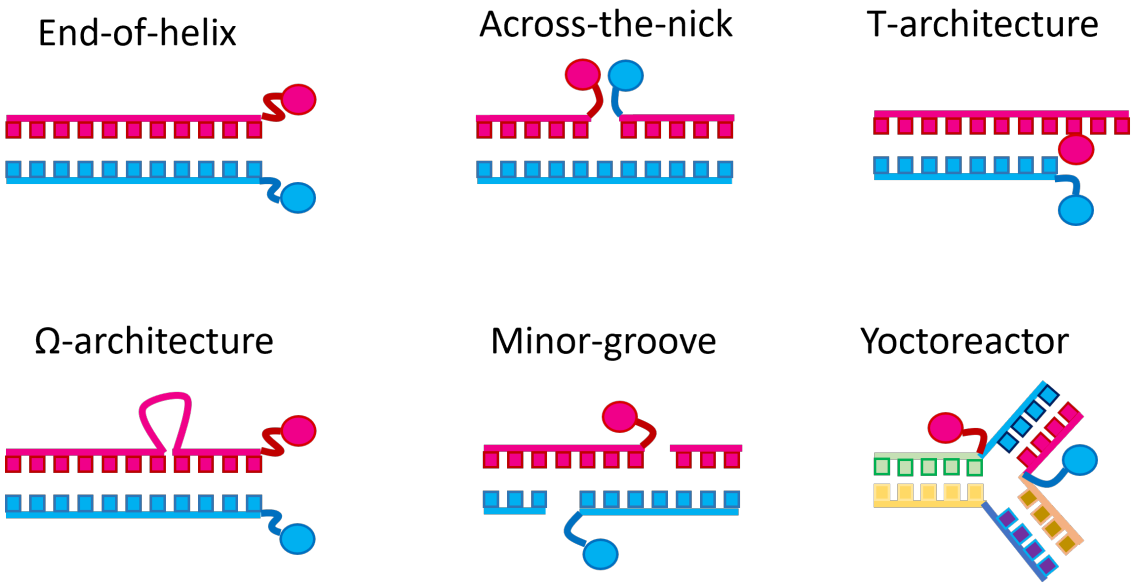


Fig. 1.8. The variety of DNA architectures successfully implemented to template chemical reactions.

Using these designs, a myriad of chemistry reactions have been successfully facilitated using DNA-templated chemistry. To name a few, these reactions include S_N2, conjugate addition, reductive amination, oxazolidine formation, nitro-aldol, thiol-Michael addition, native chemical ligation, and Pd-catalyzed cross-coupling (Figure 1.9).

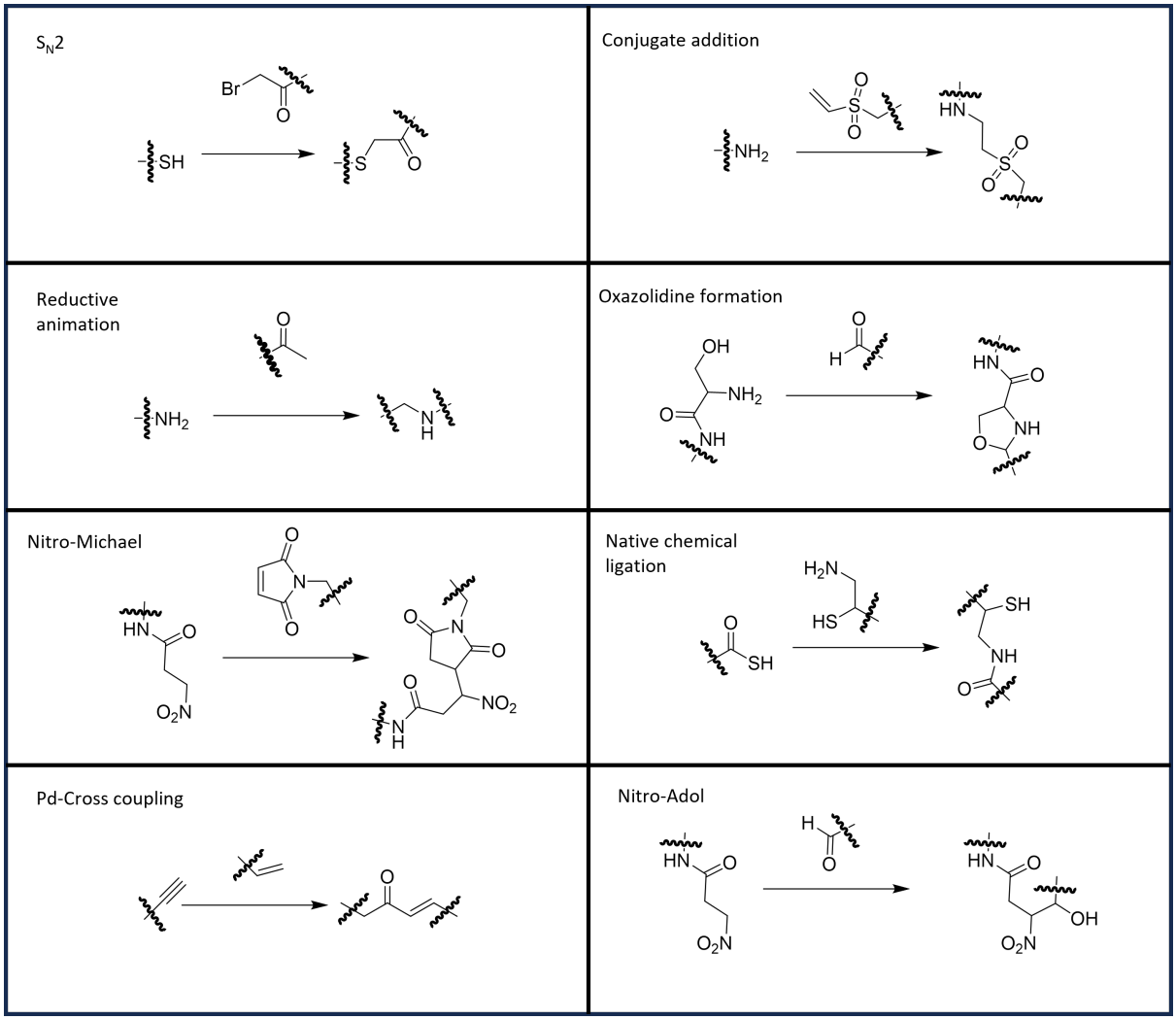


Fig. 1.9. Mechanisms that have proven to function in the context of dynamic combinatorial chemistry (DTS): The figure illustrates various chemical reaction mechanisms, including Sn2 substitution, conjugate addition, reductive amination, oxazolidine formation, nitro-Michael addition, native chemical ligation, Pd-cross coupling, and nitro-Aldol reaction. These reactions are integral in facilitating the dynamic transformations and assembly processes within the framework of DTS.

DNA Templated Synthesis: Emulating Ribosomal Capabilities

DNA Templated Synthesis (DTS), first pioneered by Liu and Gartner,[44] is a method that closely emulates the ribosome's synthetic capabilities. DTS facilitates programmable, autonomous, and sequence-specific synthesis. Utilising various chemistries and architectures found in chemically modified DNA, DTS has been used to autonomously create long-chain polymers in a one-pot synthesis without external interference.

In a manner akin to DNA routing, specific nucleotide sequence subsections encode for certain monomers, enabling both the programmability and identification of the resulting product. The pre-defined programmed modified oligomers allow DNA strands to hybridise, forming a planned architecture that brings modified substituents into proximity, thereby enhancing reaction likelihood.

The specificity of Watson-Crick base pairing underpins the autonomous nature of DTS, as modified oligomers are used in concentrations low enough to prevent non-templated reactions. This specificity is closely related to the concept of Effective Molarity, which refers to the effective concentration of reactive sites within a localised region. In the context of DTS, the Effective Molarity is increased due to the proximity of complementary strands, which enhances the likelihood of templated reactions occurring over non-templated ones. By keeping the concentration of modified oligomers low, the system ensures that only those reactions guided by Watson-Crick base pairing are favored, thus maintaining the specificity and efficiency of the process (Figure 1.10).[45]

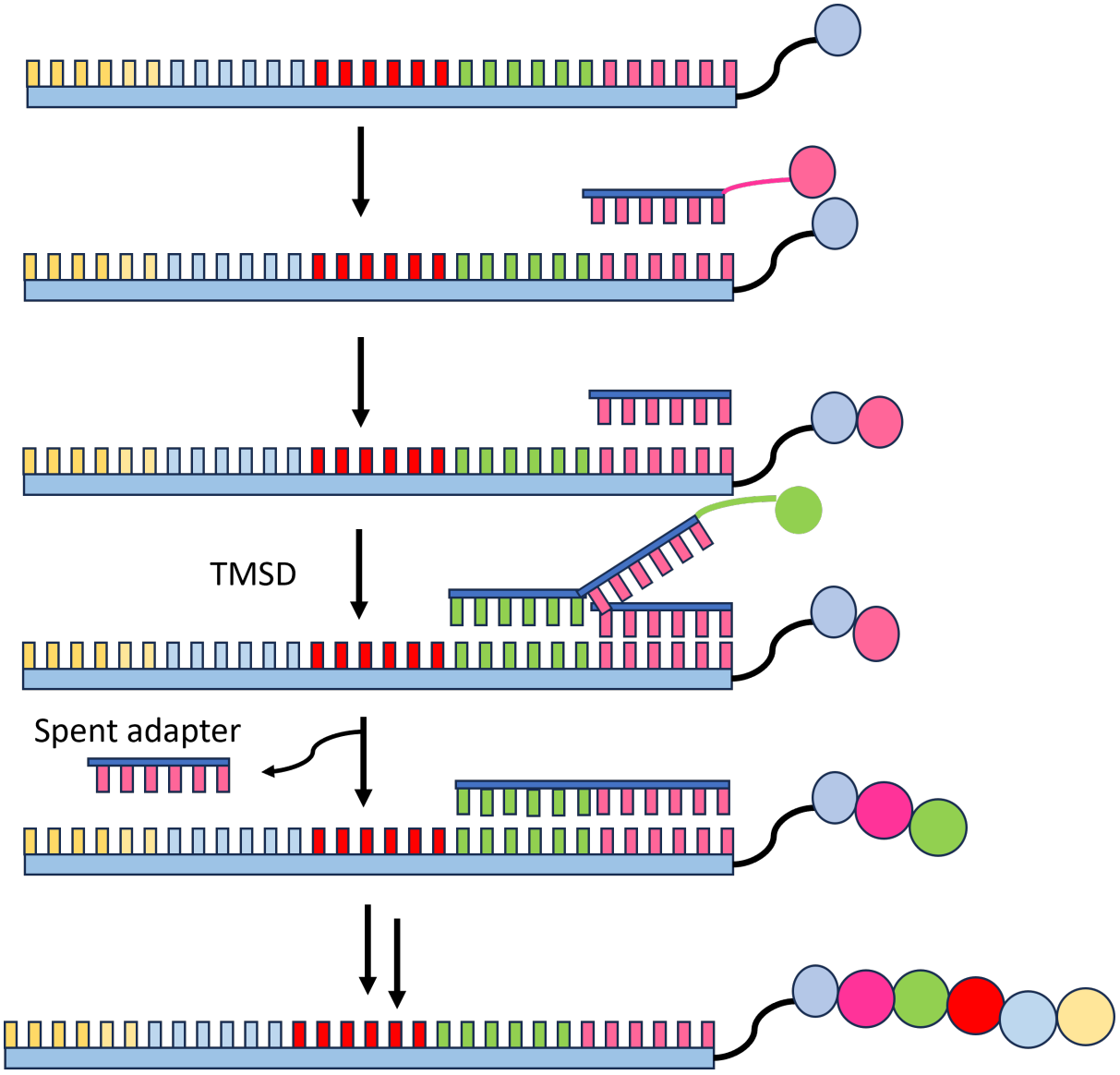


Fig. 1.10. An example of DNA templated synthesis (DTS) polymerisation, demonstrates how toehold-mediated strand displacement (TMSD) shifts the equilibrium to drive the reaction forward. In the process, an initial adapter binds to the template, facilitating polymer assembly. As the reaction progresses, a new adapter displaces the initial one *via* TMSD, enabling further chain growth and continuation of the polymerisation reaction.

For the mechanism to progress autonomously, an energetic driving force is required, enabling the growing polymer chain to advance and hybridise to subsequent modified oligomers. One such way to progress the mechanism is *via* toe-hold-mediated stand displacement (TMSD) (Figure 1.11). TMSD drives the exchange of adapters and thus the progression of the mechanism in a thermodynamic fashion by forming a greater number of Watson and Crick base pairs to the target oligomer.

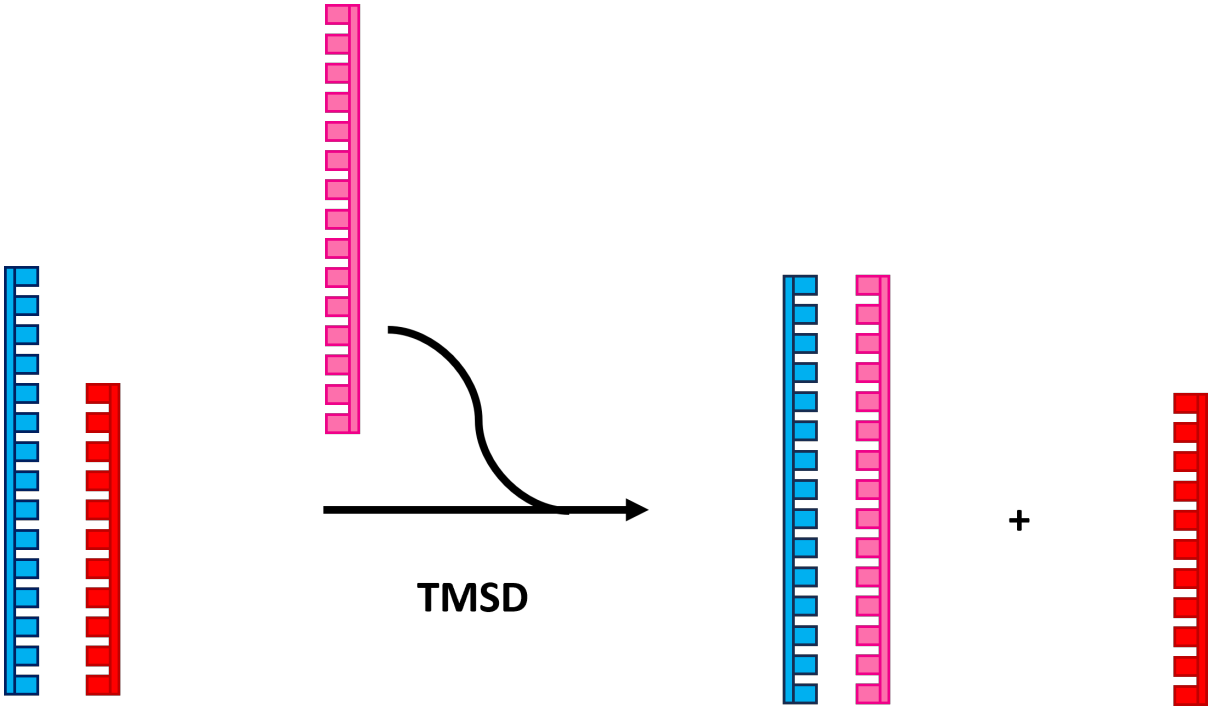


Fig. 1.11. Scheme demonstrating Toe-hold-mediated-strand-displacment where a new duplex is formed for resulting in increased base-pair interactions.

1.4 | DNA Origami

The structural versatility of DNA has enabled its applications to extend beyond biological functions. It's frequently employed in the realm of nanotechnology functioning as a structural scaffold. One example of such an application is DNA origami, a process that allows the creation of diverse nanoscale shapes and structures (Figure 1.13).[46–48] DNA origami involves folding a single-stranded DNA scaffold into specific shapes using short single-stranded 'staple' oligonucleotides(Figure 1.12).[49] This technique enables the precise arrangement of thousands of nucleotides to form objects with nanometer-scale dimensions.

The core element of DNA origami is the Watson-Crick base pair, which provides sequence recognition and specificity. However, the formation of double-helical DNA domains of defined length is not solely dependent on base pairing. The stability and structure of the DNA duplex also result from several other factors, including the chirality of the ribose sugar, which determines the right-handed helical structure, pi-stacking interactions between adjacent bases, which contribute to the stability and compactness of the helix, and electrostatic repulsion between the negatively charged phosphate backbones, which is mitigated by cations in the environment. Together, these factors ensure the correct folding and stability of the DNA origami structures.[50,51]

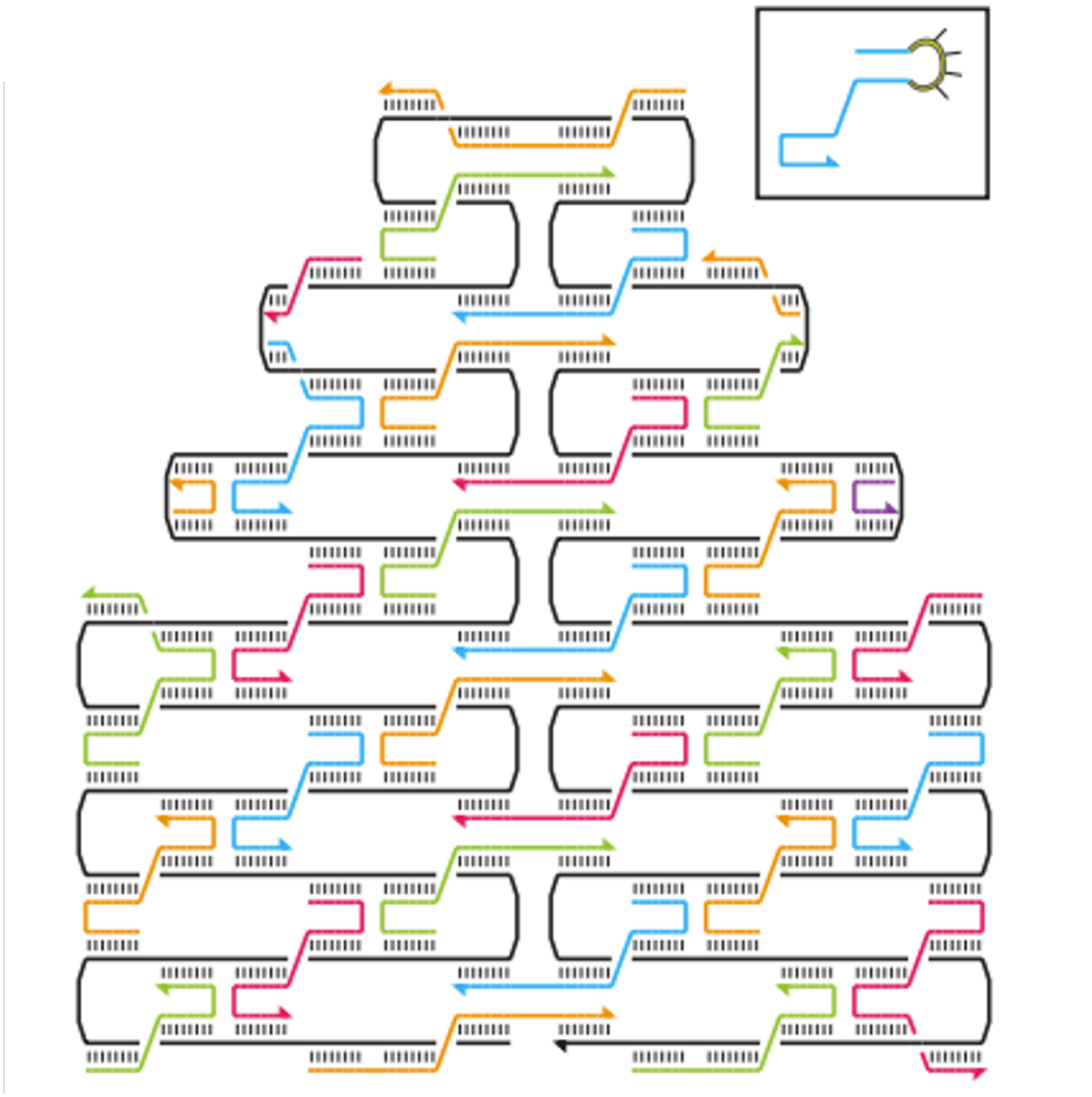


Fig. 1.12. This figure shows the final DNA structure formed by folding a long, single-stranded DNA scaffold, stabilized by staple strands. Staple strands bind via Watson-Crick base pairing, creating periodic crossovers between helices, which define the shape. The scaffold strand is folded through these staple interactions, creating the desired nanoscale structure. Modifications such as a 4-T loop (a loop consisting of four thymine bases that increases flexibility at specific points in the structure) strengthen and stabilize the origami. The staples bridge multiple helices, holding the scaffold in place and creating high-precision nanoscale shapes. Figure adapted from Ref.[52]

These double-helical domains can be connected to adjacent domains by immobilised Holliday junctions formed by antiparallel cross-overs of either the staple or scaffold strand.[49]

The design of DNA-origami objects is akin to creating a blueprint for a building, where the location and shape of each 'brick' (double-helical DNA domain) need to be specified.[53,54] The

DNA origami design process can be optimised through the use of computational software like caDNAno and computer-aided engineering for DNA origami (CanDo), which predicts the 3D structure of DNA origami designs.[55]

One of the most striking features of DNA origami is its sub-nanometer precision.[56] Such precision is indispensable in fields like nanoelectronics[57] and nanorobotics,[58] where exact spatial configurations are crucial. Owing to its inherent modular precision, DNA origami has enabled the construction of larger, more complex assemblies from DNA (Figure 1.13).[59,60]

The versatility of DNA scaffolds is further enhanced by their ability to be functionalised with various molecules, including proteins[61] and nanoparticles[62]. This functionalisation paves the way for creating nano-devices tailored for specific applications, such as targeted drug delivery and biosensing.[63,64] Additionally, DNA origami structures can exhibit dynamic behaviours in response to external stimuli, making them suitable for creating responsive nanomachines.[65,66]

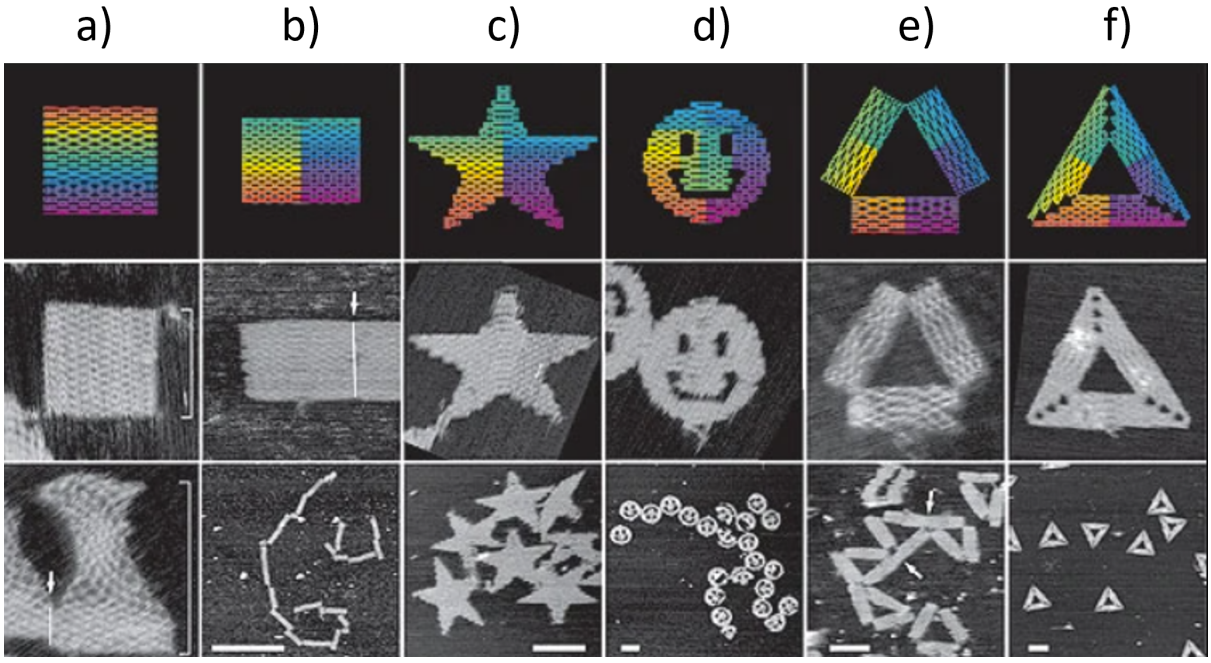


Fig. 1.13. Examples of DNA origami achieved from various folding pathways:(a) square, (b) rectangle, (c) star, (d) disk with three holes, (e) triangle with rectangular domains, and (f) sharp triangle with trapezoidal domains. Insets show unfolding sequences with colour indicating base-pair index (red for the 1st base, purple for the 7,000th). Middle row: Diagrams depicting helical bending at crossovers. Bottom two rows: AFM images, with key features highlighted (white lines for blunt-end stacking, brackets in (a) showing stretched square dimensions, and white hairpins in (f)). All images without scale bars are 165nm × 165nm. Scale bars: (b) 1μm; (c–f) 100nm. Figure adapted from Ref.[52]

1.5 | Models in Science

Computation has become indispensable to science. It's been able to expand theoretical understanding through simulations and data analysis (Figure 1.14). Simulations have enabled scientists to probe the interactions of complex systems and phenomena. By implementing simulations, scientists can gain succinct molecular insight that otherwise would be inaccessible. By enabling scientists to create models which test hypotheses and allow visualisation of out-of-reach phenomena our understanding of the great and small is sufficiently enhanced. From molecular interactions in biology[67,68] to galactic dynamics in astrophysics[69,70] simulations and computational techniques have expanded our understanding of the universe.

Through the embedding of computation in science, large sets of empirical data are now ubiquitous in all areas of scientific enquiry. The implementation of models and code is therefore paramount in providing insights into these large data sets. Software has enabled meaningful patterns to be extracted from gargantuan datasets, providing a deeper and more accurate understanding of the natural world.

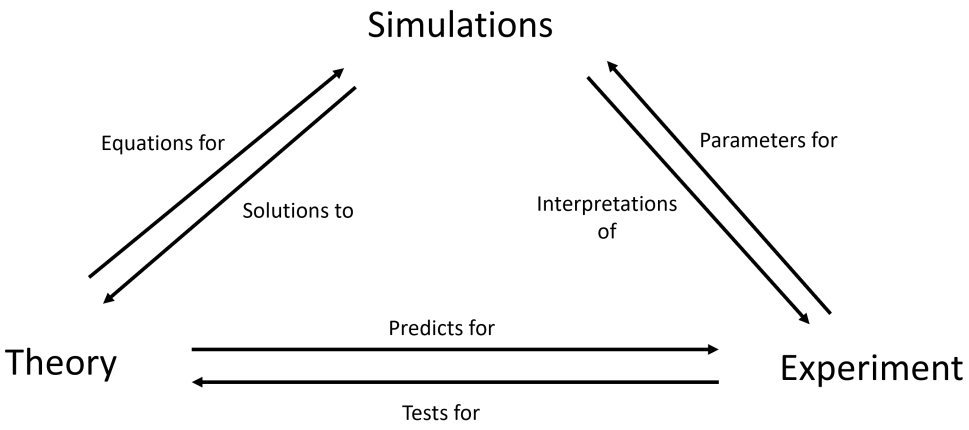


Fig. 1.14. Simulations bridge theory and experiment: theoretical equations are parameterized by experimental data, enabling simulations to provide solutions for theoretical predictions and interpretations of experimental results. In turn, experiments validate theoretical models, which then inform further experimentation.

Molecular Dynamics (MD) simulations have great relevance in molecular biology and drug discovery.[71] They can capture the behaviour of biomolecules in full atomic detail with fine temporal resolution.[72] The implementation of MD has deeply enhanced our understanding of bio-molecular events, such as the functional mechanisms in proteins,[73] the structural basis for diseases[74], and the design and optimisation of small molecules,[75] peptides,[76] and proteins.[77]

The pioneering utilisation of MD dates back to the 1950s and modelled the dynamics of simple gasses.[78] later the use of MD expanded in the 1970s to model protein dynamics.[79] These pioneering works laid the groundwork for the development and implementation of molecular dynamics. MD simulations are now frequently used to interpret and enhance the understanding derived from experimental results. The use of molecular dynamics has even expanded to the field of neuroscience, where it plays a crucial role in studying proteins critical to neuronal signalling.[80]

MD simulations function by calculating the force exerted on each atom within a system. MD uses Newton's laws of motion to predict the trajectory and consequently positions of each atom for a particular time-step (usually on the scale of a few femtoseconds).[81] This allows for a detailed three-dimensional representation of the system's atomic configuration. The distinctive advantage of MD is its ability to capture every atom's position and motion at each time point, and under precisely controlled conditions, this is unmatched by experimental techniques.

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2} \tag{1.1}$$

Here, $F(t)$ is the force on an atom at time t, m is its mass, and $x(t)$ is its position. The force is used to determine the velocity and subsequent position of each component within a timestep, thus revealing the system's dynamic evolution.

Force Fields and Accuracy

MD simulations employ molecular force fields to calculate the inter-atomic forces present in the system.[82] The force-field ultimately describes the inter-atomic potential energy between the atoms as obtained *via* fitted to quantum mechanical calculations and experimental measurements.[83] While these force fields have improved substantially, they remain approximations, and their limitations must be considered when analysing simulation results.[84] The precision of MD simulations therefore is contingent upon the force fields used. Force fields can be further refined *via* appropriate and bespoke quantum mechanical calculations.[85,86]

$$E_{\text{eff}} = \sum E_{\text{bonds}} + \sum E_{\text{angles}} + \sum E_{\text{dihedrals}} + \sum E_{\text{elec}} + \sum E_{\text{vdw}}$$

(1.2)

Here, U is the potential energy, r and θ represent bond lengths and angles, and K_r and K_θ are force constants. Additionally, classical MD simulations do not allow for the formation or breaking of covalent bonds. Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) simulations are often used to study reactions involving changes to covalent bonds or light-driven processes.

1.6 | MD ensmebles

The operation of MD can occur *via* different regimes. These are microcanonical (NVE), canonical (NVT), or isothermal-isobaric (NPT). The choice of which regime to use is crucial as it defines the conditions under which the system is simulated and thus the thermodynamic properties which are conserved.[87]

Regarding the microcanonical NVE the number of molecules (N), the volume (V), and the total energy (E) are constant, this sort of ensemble represents an isolated system. This ensemble is particularly useful for exploring energy-conserved systems but is less ideal for studying equilibration due to its inability to control temperature.[88]

The canonical ensemble (NVT) Ensemble is defined by a constant number of particles (N), volume (V), and temperature (T). This consequently simulates a system in thermal equilibrium with a heat bath. Energy in the NVT ensemble fluctuates according to the Boltzmann distribution, and the partition function encompasses all possible energy levels. The thermodynamic potential in the NVT ensemble is related to the Helmholtz free energy, this function is minimised in equilibrium. The Helmholtz free energy accounts for both the most likely microstate (global minimum of energy) and the increasing density of states, illustrating the balance between energy minimisation and entropy maximisation.[89] This is the appropriate ensemble for when temperature control is essential.[90]

The Isothermal-Isobaric ensemble (NPT) conserves the number of particles (N), pressure (P), and temperature (T). This simulation regime is therefore used with constant temperature and pressure conditions. Differing from the NVT ensemble the partition function for the NPT ensemble takes into account the fluctuating volume of the system. The associated thermodynamic potential in NPT is proportional to the Gibbs free energy, with the Gibbs free energy being defined to be at a minimum for equilibrium. The Gibbs free energy combines

internal energy, pressure-volume work, and entropy, making the *NPT* ensemble particularly suitable for simulating chemical reactions and processes that occur at constant pressure in laboratories.[91]

Each ensemble serves a unique purpose in molecular dynamics simulations, allowing researchers to explore and understand the behaviour of molecular systems under different thermodynamic conditions. The choice of ensemble depends on the specific requirements of the simulation and the physical properties of interest.

1.7 | Molecular Dynamics application

As molecular dynamics describes the inter and intra-molecular conformation trajectory, one of the most intuitive applications of MD simulations is in the bio-molecular field.[72,92–94] In this context MD has provided insights beyond the average structures yielded by methods like X-ray crystallography and cryo-EM, revealing the dynamic behaviour of biomolecule as well as water molecules and ions which are often crucial in understanding protein function and ligand binding.[95]

MD simulations are particularly valuable in determining how a bio-molecular system responds to perturbations. This includes the effects of removing or replacing ligands,[96] mutating amino acid residues,[97] and applying external forces or environmental changes.[98] These simulations offer a deeper understanding of the structural basis for various molecular events. For larger more complex molecular systems, Coarse-grained simulations have been implemented which abstract away finer atomic details.[99] By approximating what is deemed to be unnecessary or predictable atomistic and molecular behaviour, this computational technique has enabled the study of larger bio-molecular assemblies over extended timescales.

The understanding of molecular events such as the interactions between drugs and DNA has been greatly enhanced by computational modelling and molecular dynamics.[100,101] Simulations have gleaned insight into complex molecular interactions which are paramount for drug development and the understanding of molecular biology. Through the implementation of molecular dynamics and computation, scientists can elucidate the mode and site of binding of a drug-DNA interaction.[102] Understanding of this event can be extended to the relative energetics of the binding process[103,104] and even the hydration sphere around drug-DNA complexes.[105] The insights enable scientists to perfect a specific drug-DNA interaction which

ultimately ensures the correct physiological response.[106] This is crucial to design and ensure the desired binding for particular drugs.

1.8 | Use of DNA in Programming Self-Assembly

DNA has demonstrated stability in a diverse range of architectures and has been successfully used in DNA origami to create and template complex patterns. Because of its structural flexibility and bonding specificity *via* Watson-Crick base pairing.[107,108] Specifically, the implementation of DNA as a structural scaffold was first demonstrated by Mirkin and coworkers, with the assembly of gold nanocrystals using DNA-coated colloids.[109] This aggregation demonstrated temperature dependence, adding another dimension to the control of DNA-mediated assembly processes. Alivisato demonstrated the potential for assembling gold clusters into specific structures such as dimers and trimers, further highlighting DNA's role in crystal engineering.[107] Challenges in creating ordered structures from micron-sized DNA colloidal clusters (DNACCs) were initially significant, as early attempts often resulted in amorphous aggregates.[110] However, a major breakthrough occurred in 2005 when Crocker *et al.* successfully formed crystal structures from DNACCs, marking a significant advancement in the field.[110]

The use of DNA offers great versatility in designing binding configurations. For instance, DNA duplexes can be linked by single-stranded DNA linkers that connect the overhangs of two duplexes.[110,111] The strength of these interactions, driven by DNA hybridisation, can be finely tuned by adjusting the grafting density and the specific DNA sequences used. This ability to control the binding interactions allows for precise assembly pathways, leading to the formation of well-defined structures.[112]

DNA has also found success as a structural scaffold in colloidal molecules. Colloidal molecules are hierarchical colloids, found by the assembly of two or more colloids, with these colloids varying in size and composition.[113] The use of DNA in this pursuit offers specific, precise and tuneable interaction between the colloids forming the colloidal molecule.[107,114]

Colloidal self-assembly offers an efficient bottom-up route for creating diverse nano-structured materials.[115] Advances in synthetic techniques have enabled the creation of colloids with anisotropic shapes and interactions, particularly through DNA-mediated binding. This tunability in inter-particle interactions makes colloidal particles ideal for building materials with controlled properties.[116] Colloidal self-assembly is crucial for applications in areas like optical engineering, where it's used for creating photonic band gaps in materials[117,118], and in producing colloidal

clusters or "molecules" that mimic molecular structures on a larger scale[119] with applications in optics.[120,121] The ability to engineer these properties through colloidal self-assembly highlights its significance in material science and technology.

The concept of bottom-up assembly using colloids is a significant area of research in material science, particularly for creating smart, nanostructured materials. This approach involves assembling nanometer- or micrometre-sized particles into complex secondary structures, whose properties are highly responsive to environmental stimuli.[122] This method offers several advantages, such as the ability to control the size, shape, and arrangement of particles, leading to materials with tunable physical and chemical properties. The colloids can be manipulated through various interactions, including electrostatic forces, van-der-Waals forces, and DNA-based interactions, allowing for precise control over the assembly process.[116] This technique opens up possibilities in creating advanced materials for various applications, ranging from colour displays[123] and filters[124] to bio-sensors.[125]

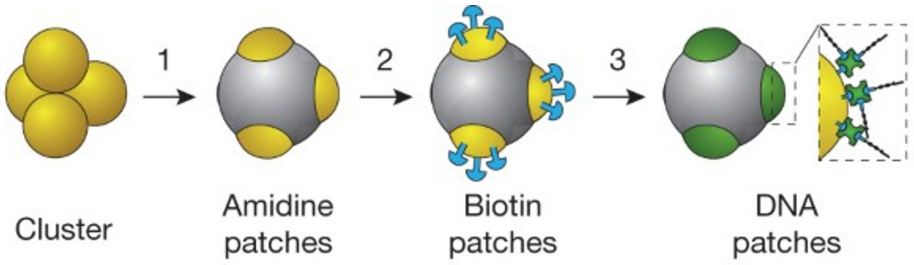


Fig. 1.15. Demonstrates the synthesis of tetrahedral, DNA-functionalised colloidal particles. Step 1 involves swelling a polystyrene microsphere cluster in styrene, followed by polymerization to create patches.Step 2: biotin is attached to the patches. Step 3: DNA oligomers attach to the biotin functinalised patches *via* biotin-streptavidin linkage. Figure adapted from ref.[119].

1.9 | Hierarchical. Self Assembley

Hierarchical self-assembly, a critical concept in molecular and materials science, plays a pivotal role in the formation of complex biological structures.[126] This process is evident in a range of biological systems such as collagen formation[127] to the assembly of the tobacco mosaic virus[128]. Hierarchical assembly is required to organise these molecular sub-units into complex structures at the macroscopic scales.[129] These natural examples inspire the laboratory construction of mesoscopic structures with unique properties and functions.[130]

Using colloids, scientists have been able to implement hierarchical self-assembly to create desirable crystals with minimal to no defects.[131] Such well-structured crystals can potentially exhibit unique physical, chemical, and optical properties, making them suitable for advanced applications in photonics,[132] drug delivery,[133] and sensor technology.[134] In photonics, for example, colloidally assembled crystals can be used to create photonic bandgap materials, which can manipulate light in novel ways, enabling the development of highly efficient optical fibers, waveguides, and even invisibility cloaks.[117] The ability to precisely control the periodicity of the crystal lattice is key to achieving the desired photonic properties.[135]

In drug delivery, hierarchical self-assembly allows for the creation of nanoparticle-based carriers that can encapsulate drugs and release them in a controlled manner. These carriers can be engineered to respond to specific stimuli, such as pH or temperature changes, ensuring that the drug is released at the target site, thereby increasing the therapeutic efficacy while minimizing side effects.[136,137]

In sensor technology, the use of colloidally assembled crystals can lead to the development of highly sensitive and selective sensors. For instance, crystals with specific pore sizes can be designed to detect the presence of particular molecules or ions, making them useful for environmental monitoring or medical diagnostics. The hierarchical assembly process allows for the fine-tuning of these materials to achieve the desired sensitivity and selectivity.[138,139]

Hierarchical assembly is crucial for the controlled formation of complex bespoke material, which opens new avenues in the design and fabrication of innovative materials.

Manipulation of the assembly of such macro-molecules has been achieved *via* a range of methods. One such method involves the implementation of isotropic long-range electrostatic repulsion and isotropic short-range attraction to control the assembly of colloidal diamond.[140] Another such hierarchical assembly is implemented *via* anisotropic particles.[141,142]

One such study implements hierarchical self-assembly *via* biasing patch strengths and widening the patch angle and patch interaction distance (Figure 1.9).[143] This approach works by exploiting a hierarchy of interaction strengths between the patches on the colloidal particles. Specifically, the study used triblock patchy particles that have two distinct attractive patches, A and B, positioned at the poles of a charged band. By assigning different interaction strengths to these patches (e.g., $\epsilon_{AA} > \epsilon_{AB} > \epsilon_{BB}$), the assembly process can be staged: stronger interactions (e.g., ϵ_{AA}) drive the formation of initial clusters (such as tetrahedra or octahedra), while weaker interactions (e.g., ϵ_{BB}) facilitate the assembly of these clusters into larger, ordered structures. Widening the patch angle increases the surface area available for interaction, promoting the formation of larger and more stable clusters, while increasing the patch interaction distance (controlled by the parameter s) allows for more flexible and diverse assembly pathways. This careful control over patch properties enables the creation of complex, defect-free crystal structures with high precision.

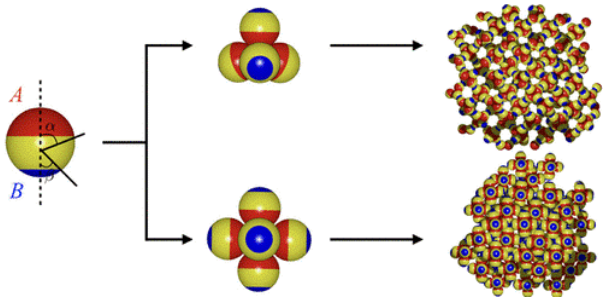


Fig. 1.16. Hierarchical assembly is directed variation in patch angle angles. Two regimes in this study both implemented patch strengths of: $\epsilon_{AA} = 5, \quad \epsilon_{BB} = 1$. Here, ϵ_{AA} represents the interaction strength between two A patches when they face each other, while ϵ_{BB} represents the interaction strength between two B patches. These values indicate the depth of the potential well for the respective interactions, with higher values corresponding to stronger attractive forces.
For tetrahedral clustering, patches with a shallower angle ($\alpha = 80°$, $\beta = 40°$ and interaction range $s = 5$) lead to a ring-like structure with tetrahedral subunits, progressing towards a cubic diamond formation. Larger half-angle for patches and smaller interaction range ($\alpha = 85°$, $\beta = 40°$, $s = 1$) promotes the formation of octahedral subclusters, which assemble into a body-centered cubic colloidal crystal. Taken with permission from Ref.[143]

1.10 | Monte Carlo

Aside from molecular dynamics, Monte Carlo (MC) simulations have also proven an invaluable technique in probing molecular interactions; Monte Carlo simulations have been utilised in modelling molecular self-assembly and probing the evolution of complex molecular structures.[144,145] Monte Carlo simulations differ from molecular dynamics in that they employ stochastic sam-

pling methods to explore the configurational space of a system. This is achieved through random sampling of possible states, which allows the algorithm to navigate the potential energy landscape effectively, identifying stable and metastable states that might be missed by deterministic methods like molecular dynamics.[146]

In the context of molecular bonding, Monte Carlo operates *via* a statistical algorithm that navigates through the potential energy landscape *via* a controlled set of biased random moves. The algorithm is biased towards moves that lower the system's energy, thereby favoring transitions that lead to more stable configurations.[147,148]

The likelihood of each move in a Monte Carlo simulation is determined by the Boltzmann factor, which weighs the energy change resulting from the move against the thermal energy of the system. The Boltzmann factor is given by $\exp(-\Delta E/k_B T)$, where ΔE is the energy difference between the current and proposed states, k_B is the Boltzmann constant, and T is the temperature. This factor ensures that moves which decrease the system's energy are always accepted, while moves that increase the energy are accepted with a probability that decreases exponentially with ΔE. This probabilistic approach allows the system to escape local energy minima and explore a broader range of configurations, ultimately converging towards a global or near-global minimum.[149]

For example, in a study on molecular self-assemblies by Perlstein (1992), Monte Carlo simulations were used to predict the structure of one-dimensional translation aggregates. The system started in a high-energy conformation, and random moves were proposed to adjust the molecular orientations and positions. Each move was evaluated using the Boltzmann factor. If the move resulted in a lower energy, it was accepted unconditionally. However, if the energy increased, the move was accepted with a probability determined by the Boltzmann factor. This allowed the simulation to avoid getting trapped in local minima and to explore various conformations that could potentially lead to the global energy minimum. The outcome was a prediction of aggregate structures that closely matched experimental observations, demonstrating the effectiveness of the Boltzmann factor in guiding the system towards stable, low-energy configurations.[150]

Monte Carlo simulations of molecular self-assembly typically start with a random distribution of molecules in a defined system size, enclosed by periodic boundary conditions. These simulations can be performed in different statistical ensembles, such as the *NVT* (constant number of particles, volume, and temperature) or *NPT* (constant number of particles, pressure, and

temperature) ensembles, which can be selected depending on the specific physical conditions to be simulated.[151,152]

An example of an *NVT* Monte Carlo simulation is found in the study by Martsinovich and Troisi (2010), where they modeled the self-assembly of benzenedicarboxylic acids on surfaces. In this case, the *NVT* ensemble was chosen to maintain a constant number of molecules, volume, and temperature, which is crucial for accurately simulating the self-assembly process at a fixed temperature (Figure 1.17). The study demonstrated how different arrangements of the benzenedicarboxylic acids on a surface could be predicted based on the competition between intermolecular interactions and interactions with the substrate. The outcomes of these simulations provided insights into the design of organic frameworks and surface functionalisation techniques, showing that self-assembly can be directed by carefully tuning the interaction parameters.[151]

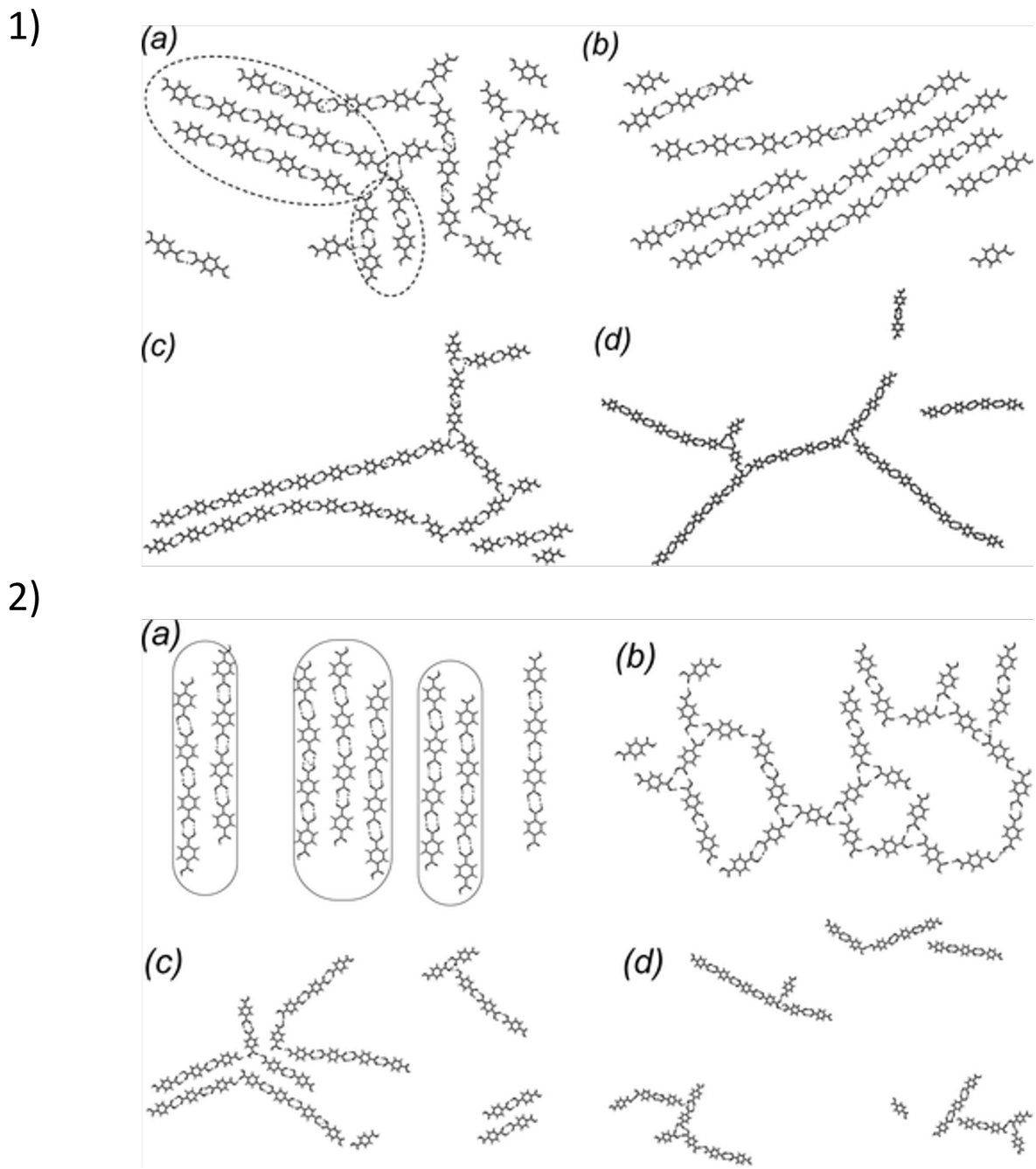


Fig. 1.17. (1) shows the final structures from molecular dynamics (MD) simulations, and (2) shows results from Monte Carlo (MC) simulations, both in the NVT ensemble under identical conditions. In each set, (a) and (b) depict dense cells (64%) with ordered and disordered starting positions of TPA molecules, respectively. (c) and (d) show intermediate (28%) and large cells (8%) with ordered starting positions. Figure adapted from Ref.[151]

In contrast, *NPT* Monte Carlo simulations are used when it is important to allow the system to change its volume in response to external pressure. For instance, in the study by Ghobadi et al. (2011), *NPT* Monte Carlo simulations were employed to investigate the solubility of SO_2 and CO_2 in imidazolium-based ionic liquids (Figure 1.18). The use of the *NPT* ensemble allowed the simulation to account for the volumetric changes that occur when gases dissolve

in the ionic liquids under constant pressure, which is critical for accurately predicting the solubility behavior. The results showed that the solubility of SO_2 was highest in ionic liquids with smaller anions, providing valuable insights for designing more efficient gas capture and separation processes.[153]

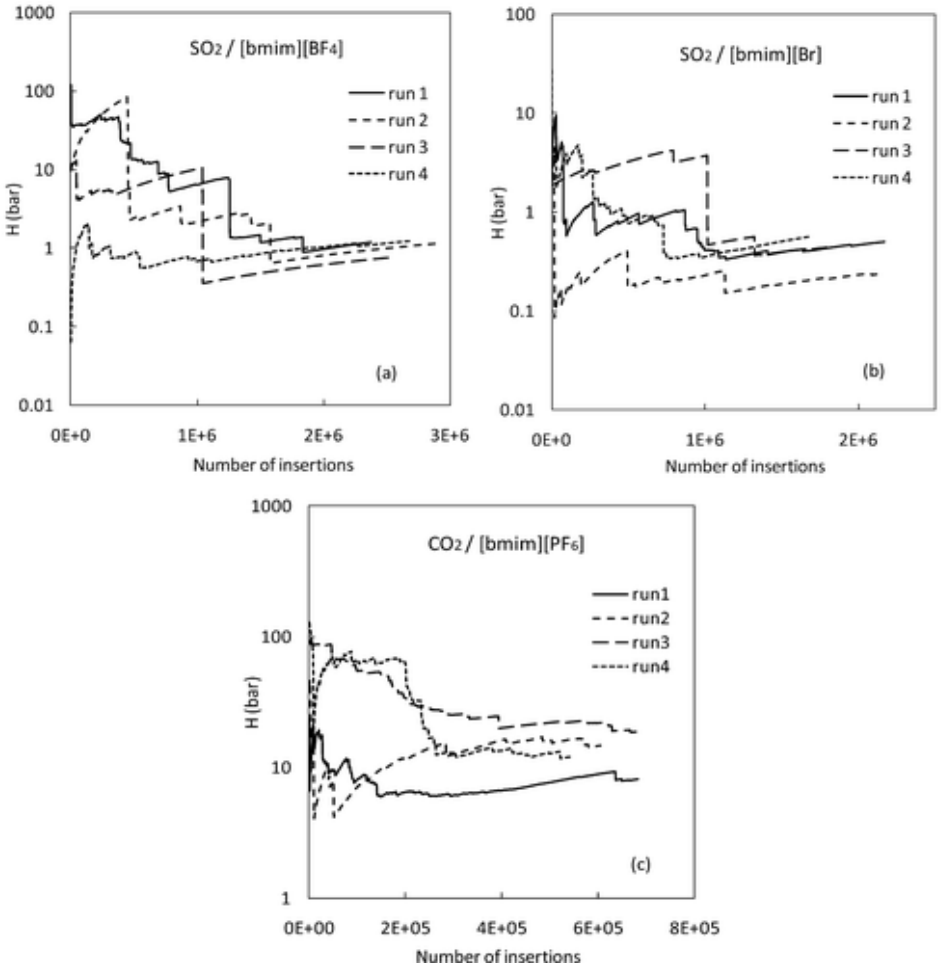


Fig. 1.18. Convergence of Henry's Constant for SO_2 and CO_2 in Imidazolium-Based Ionic Liquids (ILs): The figure shows the convergence of the calculated Henry's constant (H) for SO_2 and CO_2 as a function of the number of insertions during Monte Carlo simulations. The x-axis represents the number of particle insertions, and the y-axis indicates the value of Henry's constant (H) in bar. Henry's constant quantifies gas solubility, and its convergence indicates a stable calculation. The simulations use the Widom particle insertion method to estimate H, as defined by:

$$H = \rho RT \exp(\beta \mu_s^{\infty,ex}),$$

where ρ is the density of the solvent, R is the gas constant, T is the temperature, and $\mu_s^{\infty,ex}$ is the excess chemical potential at infinite dilution. The figure shows that as the number of insertions increases, the Henry's constant stabilises, confirming the accuracy of the solubility calculations for both gases. SO_2 achieves lower H values (higher solubility) compared to CO_2, particularly in ILs with higher charge density anions such as [bmim][Br] and [bmim][NO_3].Figure adapted from Ref.[153]

These examples illustrate how the choice between *NVT* and *NPT* ensembles in Monte Carlo simulations depends on the specific goals of the study. The *NVT* ensemble is typically used when the focus is on understanding the structure and configuration at a fixed temperature and volume, while the *NPT* ensemble is more appropriate when the study involves pressure-dependent phenomena, such as phase transitions or solubility studies. By selecting the appropriate ensemble, researchers can ensure that their simulations accurately reflect the physical conditions of interest and provide reliable predictions for real-world applications.[154,155]

The effectiveness of Monte Carlo stems from its ability to thoroughly scan the configurational landscapes and identify an extensive range of stable structures.[156] One of the key advantages of Monte Carlo over molecular dynamics is its ability to avoid becoming trapped in local minima due to its random sampling nature. By allowing the system to make occasional energetically unfavorable moves (with a probability determined by the Boltzmann factor), Monte Carlo simulations can explore the energy landscape more broadly and efficiently. This is particularly useful in systems where the energy landscape is rugged, with many local minima separated by high energy barriers, such as in the self-assembly of colloidal particles or the folding of complex molecules.[157]

Monte Carlo simulations are generally employed over molecular dynamics regarding the modeling of colloidal self-assembly. In colloidal systems, the primary interest often lies in the equilibrium structures rather than the dynamic pathways leading to these structures. Monte Carlo simulations, with their focus on sampling equilibrium configurations, are well-suited to this task. The stochastic nature of Monte Carlo, operating through random moves, aptly simulates the random walks and fluctuations characteristic of colloidal particles in a solvent. Additionally, because Monte Carlo simulations do not require the calculation of forces and their derivatives at each step (as is required in molecular dynamics), they can be computationally less demanding. This allows for the exploration of larger systems and longer timescales, providing insights into the long-term behavior and stability of colloidal assemblies.[158]

HMonte Carlo simulations have limitations as well. One significant limitation is the lack of explicit time evolution in these simulations. Since the method is based on random sampling rather than integrating equations of motion, it does not provide direct information about the kinetics of the processes being simulated. As a result, while Monte Carlo can accurately predict the final equilibrium state of a system, it cannot describe how quickly or through what specific pathways the system reaches that state.[159] Monte Carlo simulations can struggle with systems where long-range interactions are important, as these require careful treatment to

avoid artifacts in the simulation results. To address this, advanced techniques such as kinetic Monte Carlo (kMC) are sometimes employed, which can provide insights into the temporal evolution of systems by simulating the dynamics of rare events over extended timescales.[160]

Monte Carlo simulations are particularly valuable in the study of colloidal self-assembly due to their ability to model systems at equilibrium. For instance, they have been used to predict the formation of complex crystal structures in colloidal systems by exploring different assembly pathways and identifying the conditions that favour certain morphologies. Examples include the assembly of patchy particles into well-defined lattices[144] or the formation of micelles[161] and vesicles from amphiphilic molecules.[162] These simulations help in understanding the fundamental principles governing self-assembly and guide the design of new materials with desired properties.[156]

In the context of Monte Carlo simulations, it is important to note that these simulations are inherently stochastic. They do not incorporate an explicit measure of time; instead, simulation duration is often represented by the number of Monte Carlo cycles. Unlike molecular dynamics simulations, the physical realism of the moves is not a prerequisite in Monte Carlo methods. However, if restricted to physically meaningful moves, such as small particle displacements, Monte Carlo simulations can mimic realistic dynamics akin to Brownian dynamics.[158] This approach has been successfully applied to model systems where diffusion and random motion play a crucial role, such as in the aggregation of colloidal particles or the diffusion of molecules within a confined space. These simulations can provide valuable information on the diffusion coefficients, rotational diffusion, and other dynamical properties of particles in a system, despite the lack of explicit time tracking.[163]

1.11 | Monte Carlo Simulations in Patchy Particle Systems

The simulation of patchy particle systems, characterized by their specific, directional interactions, presents unique challenges and opportunities. The dynamic behavior of these systems, which leads to diverse self-assembled structures, necessitates the development and application of efficient simulation techniques. One of the primary methods employed in this realm is the Monte Carlo (MC) simulation.For example, Patti et al. (2007) used lattice Monte Carlo simulations to study the self-assembly of hybrid organic-inorganic materials (Figure 1.19).[160] Their work demonstrated how different patchy particles, modeled as surfactants with head

and tail segments, could self-assemble into ordered structures such as micelles and hexagonal phases. The simulations revealed that the specific interactions between the organic and inorganic components, driven by the strength and directionality of the patches, were crucial in determining the final structure of the material. These results provided valuable insights into the design of functional materials with controlled porosity and surface properties.

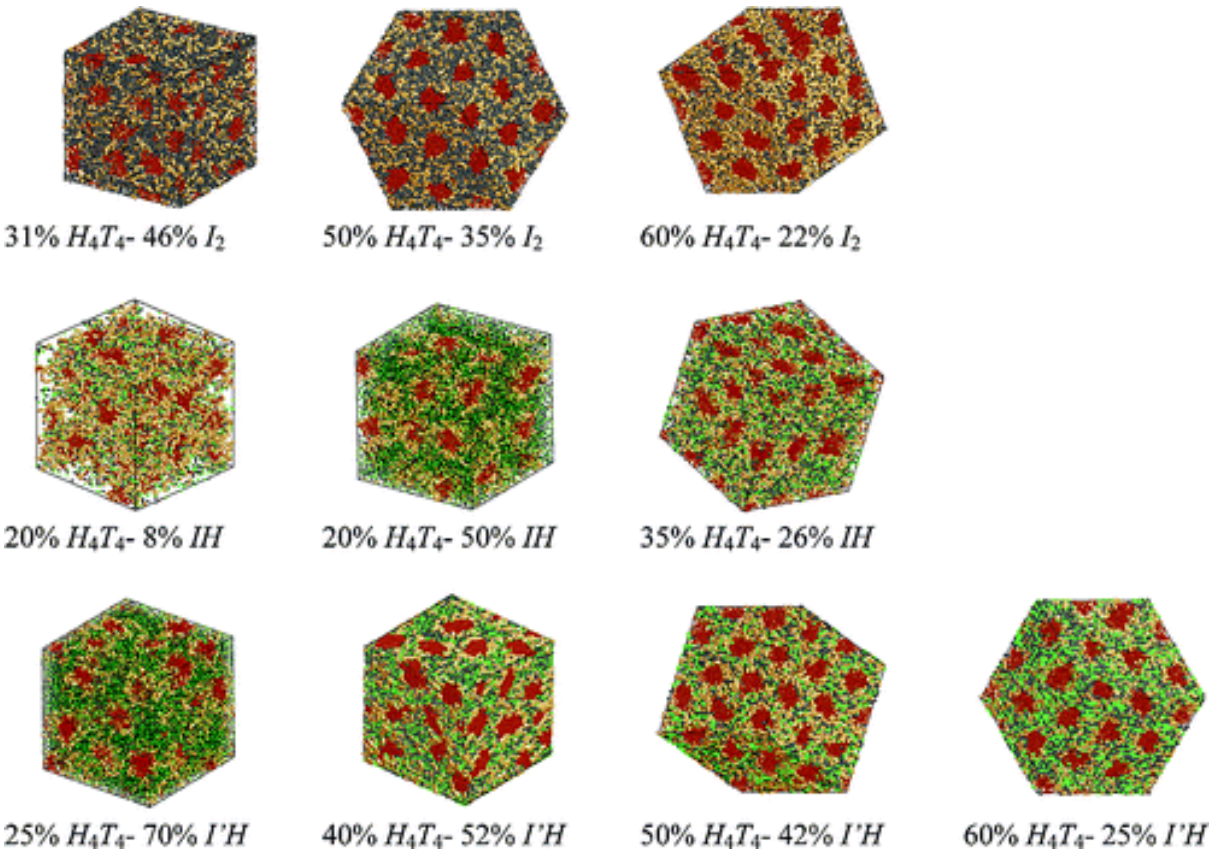


Fig. 1.19. Illustrates the self-assembled structures formed by patchy particles, representing surfactants (H4T4) and hybrid precursors (I2). The surfactant heads (yellow), tails (red), and the inorganic (gray) and organic (green) segments of the precursor organize into distinct phases. The percentages (e.g., 35% H4T4 and 26% I2) indicate the volume fractions of surfactant and inorganic precursor used in the simulation. Figure is adapted from Ref.[160]

In another example, Rao et al. (2020) used Monte Carlo simulations to investigate the hierarchical self-assembly of triblock patchy particles into photonic crystals (Figure 1.20).[164] By tuning the interactions between the patches, they were able to guide the particles through a two-stage self-assembly process, first forming tetrahedral clusters, which subsequently assembled into cubic and hexagonal tetrastack crystals. This approach allowed them to bypass amorphous phases and metastable structures, promoting the formation of photonic crystals with minimal defects. Their findings revealed that the cubic tetrastack structure exhibits a

complete photonic band gap (PBG) at lower dielectric filling fractions, making it a promising candidate for photonic applications.

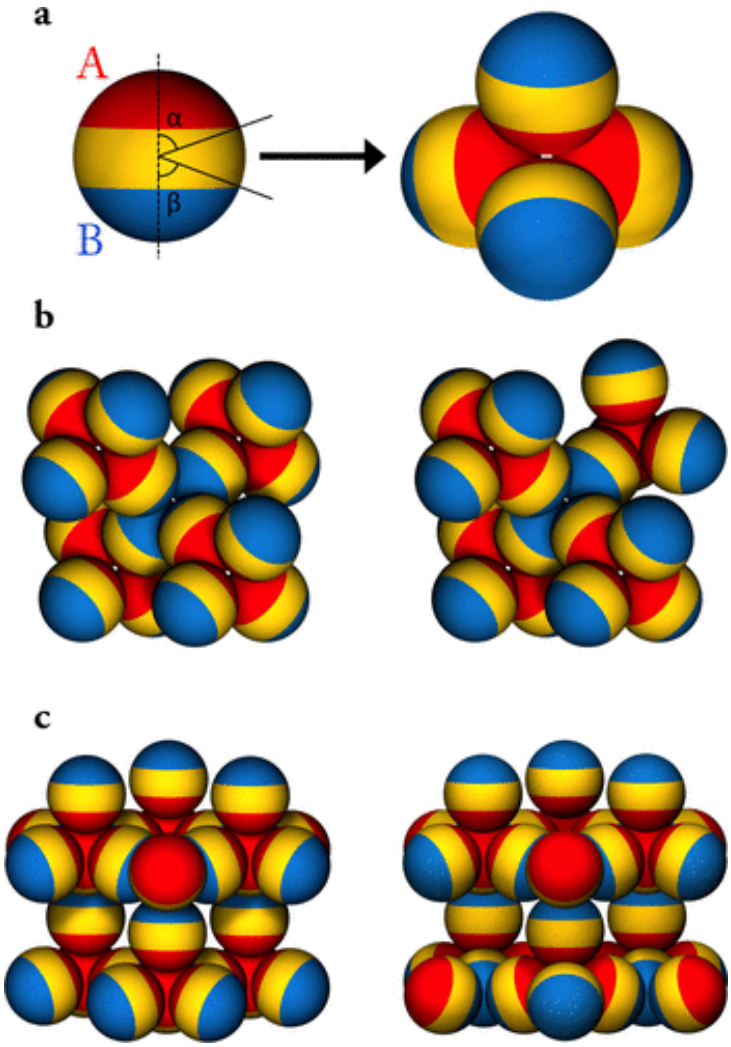


Fig. 1.20. Tetrastack Crystal Structures Formed by Triblock Patchy Particles: This figure shows the assembly of triblock patchy particles into cubic (left) and hexagonal (right) tetrastack structures. Panel (a) illustrates a triblock particle, with two distinct patches at opposite poles (red and blue) forming tetrahedra. Panel (b) highlights 16-particle motifs in cubic and hexagonal polymorphs, where the arrangement of tetrahedra varies between staggered and eclipsed configurations. Panel (c) depicts the layered nature of these crystals, consisting of alternating kagome and triangular planes. These structural motifs are crucial in determining the photonic properties of the tetrastack lattices, with the cubic polymorph supporting a complete photonic band gap at lower dielectric filling fractions. Figure adapated from Ref.[164]

These examples underscore the versatility of Monte Carlo simulations in studying patchy particle systems. By accurately modeling the specific, directional interactions that define these systems, MC simulations provide a powerful tool for predicting and controlling the self-assembly of complex materials. The insights gained from these simulations can guide experimental efforts to synthesize new materials with tailored properties for a wide range of applications.

Monte Carlo methods rely on statistical sampling to explore the configurational space of the system under study. In the realm of patchy particles, this involves the selection of a particle and proposing a new configuration. The decision to accept or reject the proposed configuration is based on the Metropolis criterion.[165]:

$$P_{\text{accept}} = \exp\left(-\frac{\Delta U}{k_B T}\right) \tag{1.3}$$

where ΔU is the change in energy between the new and old configurations, T represents a fictitious temperature, and k_B is Boltzmann's constant.

1.11.1 Sampling Moves: Rototranslations

Properly sampling the configurational space requires the system to undergo both rotations and translations, especially when dealing with anisotropic particles such as patchy ones. Rototranslations, a combination of both rotational and translational moves, are among the simplest techniques to ensure a well-balanced exploration. During a rototranslation, a particle is selected, followed by the determination of random angular and radial displacements. The particle undergoes these displacements, resulting in a change in energy. This energy difference, combined with the Metropolis criterion, guides the move's acceptance or rejection. It's crucial to fine-tune the maximum displacements, both radial and angular, to maintain an optimal acceptance ratio, usually in the range of 0.2-0.4.

1.12 | Thesis Outline

This chapter laid the groundwork for a comprehensive exploration of DNA, encompassing its structural properties and diverse applications, with a discussion on DNA in ribosomal synthesis, highlighting its role as a programmable, autonomous, and sequence-specific method for the synthesis of a wide range of proteins. Techniques using DNA as a tool and vehicle for synthesis outside the body were discussed. This chapter went on to explore the capabilities of computational techniques like molecular dynamics in modelling and understanding binding conformations. Explaining how this insight is crucial in deciphering the complex interactions at the molecular level. Finally, the topic of colloidal self-assembly was brought to light, highlighting DNA's application in such with an emphasis on hierarchical self-assembly. The applications of Monte Carlo sampling to model such events were elaborated.

The following chapters will further explore DNA Templated Synthesis (DTS), discussing its limitations with experimental work attempting to improve its efficacy. Specifically, the subsequent chapter will propose and quantify a novel protection mechanism. This mechanism results in a conformation which shields labile components from hydrolysis. This line of inquiry is complemented by Chapter 3, which undertakes a molecular dynamics study to understand this unique protection mechanism, specifically what conditions and conformations enable this to manifest.

Chapter 4 reports on a computational study of octahedral patchy colloids, where different types of patch-patch interactions are exploited to program distinct self-assembly pathways to yield simple cubic crystals in Monte Carlo simulations. The objective of this study was to assess the quality of the crystals thus self-assembled and characterise the crystallisation pathways in the different scenarios considered. The functionalisation of the patches with DNA can encode different energetics of the interactions considered in this study.

References

1. Schmidt, M. F. in *Chemical Biology: and Drug Discovery* 33–47 (Springer, 2022).
2. Webster, M. W. & Weixlbaumer, A. The intricate relationship between transcription and translation. *Proceedings of the National Academy of Sciences* **118,** e2106284118 (2021).
3. Takahashi, S. & Sugimoto, N. Watson–Crick versus Hoogsteen base pairs: chemical strategy to encode and express genetic information in life. *Accounts of Chemical Research* **54,** 2110–2120 (2021).
4. Cannarozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G. & Barral, Y. A role for codon order in translation dynamics. *Cell* **141,** 355–367 (2010).
5. Küpfer, P. A. & Leumann, C. J. Synthesis, base pairing properties and trans-lesion synthesis by reverse transcriptases of oligoribonucleotides containing the oxidatively damaged base 5-hydroxycytidine. *Nucleic acids research* **39,** 9422–9432 (2011).
6. Jash, B. & Kool, E. T. Conjugation of RNA via 2-OH acylation: Mechanisms determining nucleotide reactivity. *Chemical Communications* **58,** 3693–3696 (2022).
7. Laurentin Táriba, H. E. in *Agricultural Genetics: From the DNA Molecule to Population Management* 19–36 (Springer, 2023).
8. Bercy, M. & Bockelmann, U. Hairpins under tension: RNA versus DNA. *Nucleic acids research* **43,** 9928–9936 (2015).
9. Caudron-Herger, M. & Rippe, K. Nuclear architecture by RNA. *Current opinion in genetics & development* **22,** 179–187 (2012).
10. Xue, Y. Architecture of RNA–RNA interactions. *Current opinion in genetics & development* **72,** 138–144 (2022).
11. Moldave, K. Eukaryotic protein synthesis. *Annual Review of Biochemistry* **54,** 1109–1149 (1985).
12. Pál, C., Papp, B. & Lercher, M. J. An integrated view of protein evolution. *Nature Reviews Genetics* **7,** 337–348 (2006).
13. Brogna, S., Sato, T.-A. & Rosbash, M. Ribosome components are associated with sites of transcription. *Molecular Cell* **10,** 93–104 (2002).

14. Stevenson-Jones, F., Woodgate, J., Castro-Roa, D. & Zenkin, N. Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. *Proceedings of the National Academy of Sciences* **117,** 8462–8467 (2020).

15. Blanchard, S. C., Kim, H. D., Gonzalez Jr, R. L., Puglisi, J. D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proceedings of the National Academy of Sciences* **101,** 12893–12898 (2004).

16. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108,** 557–572 (2002).

17. Travers, A. & Muskhelishvili, G. DNA structure and function. *The FEBS journal* **282,** 2279–2295 (2015).

18. Bendich, A. J. & Drlica, K. Prokaryotic and eukaryotic chromosomes: what's the difference? *Bioessays* **22,** 481–486 (2000).

19. Carrivain, P., Cournac, A., Lavelle, C., Lesne, A., Mozziconacci, J., Paillusson, F., Signon, L., Victor, J.-M. & Barbi, M. Electrostatics of DNA compaction in viruses, bacteria and eukaryotes: functional insights and evolutionary perspective. *Soft Matter* **8,** 9285–9301 (2012).

20. Latchman, D. S. Transcription factors: an overview. *International Journal of Experimental Pathology* **74,** 417 (1993).

21. Lemon, B. & Tjian, R. Orchestrated response: a symphony of transcription factors for gene control. *Genes & Development* **14,** 2551–2569 (2000).

22. Turner, P. Gene regulation: Controlling roles for snurps. *Nature* **316,** 105–106 (1985).

23. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* **3,** a003707 (2011).

24. Beyer, D., Skripkin, E., Wadzack, J. & Nierhaus, K. H. How the ribosome moves along the mRNA during protein synthesis. *Journal of Biological Chemistry* **269,** 30713–30717 (1994).

25. Söll, D. & RajBhandary, U. L. tRNA: structure, biosynthesis, and function (1995).

26. Bartee, L. & Brook, J. *MHCC Biology 112: Biology for Health Professions* (Open Oregon, 2019).

27. Amort, M., Wotzel, B., Bakowska-Zywicka, K., Erlacher, M. D., Micura, R. & Polacek, N. An intact ribose moiety at A2602 of 23S rRNA is key to trigger peptidyl-tRNA hydrolysis during translation termination. *Nucleic acids research* **35,** 5130–5140 (2007).

28. Kazemi, M., Sočan, J., Himo, F. & Åqvist, J. Mechanistic alternatives for peptide bond formation on the ribosome. *Nucleic Acids Research* **46,** 5345–5354 (2018).

29. Sun, Q., Zhu, X., Qi, J., An, W., Lan, P., Tan, D., Chen, R., Wang, B., Zheng, S., Zhang, C., *et al.* Molecular architecture of the 90S small subunit pre-ribosome. *Elife* **6,** e22086 (2017).

30. Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nature Reviews Molecular Cell Biology* **9,** 242–253 (2008).

31. Rodnina, M. V., Beringer, M. & Wintermeyer, W. How ribosomes make peptide bonds. *Trends in Biochemical Sciences* **32,** 20–26 (2007).

32. Anderson, W. F. The effect of tRNA concentration on the rate of protein synthesis. *Proceedings of the National Academy of Sciences* **62,** 566–573 (1969).

33. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proceedings of the National Academy of Sciences* **89,** 5381–5383 (1992).

34. Goodnow Jr, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nature Reviews Drug Discovery* **16,** 131–147 (2017).

35. McGregor, L. M., Gorin, D. J., Dumelin, C. E. & Liu, D. R. Interaction-dependent PCR: Identification of ligand- target pairs from libraries of ligands and libraries of targets in a single solution-phase experiment. *Journal of the American Chemical Society* **132,** 15522–15524 (2010).

36. Flood, D. T., Kingston, C., Vantourout, J. C., Dawson, P. E. & Baran, P. S. DNA encoded libraries: a visitor's guide. *Israel Journal of Chemistry* **60,** 268–280 (2020).

37. Halpin, D. R. & Harbury, P. B. DNA display I. Sequence-encoded routing of DNA populations. *PLoS Biology* **2,** e173 (2004).

38. Halpin, D. R. & Harbury, P. B. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biology* **2,** e174 (2004).

39. Halpin, D. R., Lee, J. A., Wrenn, S. J. & Harbury, P. B. DNA display III. Solid-phase organic synthesis on unprotected DNA. *PLoS Biology* **2,** e175 (2004).

40. Hedberg, A. & Johansen, S. D. Nuclear group I introns in self-splicing and beyond. *Mobile DNA* **4,** 1–10 (2013).

41. De la Peña, M., García-Robles, I. & Cervera, A. The hammerhead ribozyme: a long history for a short RNA. *Molecules* **22,** 78 (2017).

42. Petkovic, S. & Müller, S. RNA circularization strategies in vivo and in vitro. *Nucleic acids research* **43,** 2454–2465 (2015).

43. Klumpp, K., Hang, J. Q., Rajendran, S., Yang, Y., Derosier, A., Wong Kai In, P., Overton, H., Parkes, K. E., Cammack, N. & Martin, J. A. Two-metal ion mechanism of RNA cleavage by HIV RNase H and mechanism-based design of selective HIV RNase H inhibitors. *Nucleic acids research* **31,** 6852–6859 (2003).

44. Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *Journal of the American Chemical Society* **123,** 6961–6963 (2001).

45. Percivalle, C., Bartolo, J.-F. & Ladame, S. Oligonucleotide-templated chemical reactions: pushing the boundaries of a nature-inspired process. *Organic & Biomolecular Chemistry* **11,** 16–26 (2013).

46. Qian, L., Wang, Y., Zhang, Z., Zhao, J., Pan, D., Zhang, Y., Liu, Q., Fan, C., Hu, J. & He, L. Analogic China map constructed by DNA. *Chinese Science Bulletin* **51,** 2973–2976 (2006).

47. Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Lind-Thomsen, A., Mamdouh, W., Gothelf, K. V., Besenbacher, F. & Kjems, J. DNA origami design of dolphin-shaped structures with flexible tails. *ACS Nano* **2,** 1213–1218 (2008).

48. Ke, Y., Sharma, J., Liu, M., Jahn, K., Liu, Y. & Yan, H. Scaffolded DNA origami of a DNA tetrahedron molecular container. *Nano Letters* **9,** 2445–2447 (2009).

49. Castro, C. E., Kilchherr, F., Kim, D.-N., Shiao, E. L., Wauer, T., Wortmann, P., Bathe, M. & Dietz, H. A primer to scaffolded DNA origami. *Nature Methods* **8,** 221–229 (2011).

50. Karabıyık, H., Sevinçek, R. & Karabıyık, H. π-Cooperativity effect on the base stacking interactions in DNA: is there a novel stabilization factor coupled with base pairing H-bonds? *Physical Chemistry Chemical Physics* **16,** 15527–15538 (2014).

51. Kuzyk, A., Jungmann, R., Acuna, G. P. & Liu, N. DNA origami route for nanophotonics. *ACS photonics* **5,** 1151–1163 (2018).

52. Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).

53. Saccà, B., Meyer, R., Erkelenz, M., Kiko, K., Arndt, A., Schroeder, H., Rabe, K. S. & Niemeyer, C. M. Orthogonal protein decoration of DNA origami. *Angewandte Chemie* **122,** 9568–9573 (2010).

54. Shih, W. M. & Lin, C. Knitting complex weaves with DNA origami. *Current Opinion in Structural Biology* **20,** 276–282 (2010).

55. Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M. & Shih, W. M. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Research* **37,** 5001–5006 (2009).

56. Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. M., Sander, B., Stark, H., Oliveira, C. L., *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459,** 73–76 (2009).

57. Liu, J., Geng, Y., Pound, E., Gyawali, S., Ashton, J. R., Hickey, J., Woolley, A. T. & Harb, J. N. Metallization of branched DNA origami for nanoelectronic circuit fabrication. *Acs Nano* **5,** 2240–2247 (2011).

58. Kuzuya, A. & Ohya, Y. Nanomechanical molecular devices made of DNA origami. *Accounts of chemical research* **47,** 1742–1749 (2014).

59. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325,** 725–730 (2009).

60. Linko, V., Shen, B., Tapio, K., Toppari, J. J., Kostiainen, M. A. & Tuukkanen, S. One-step large-scale deposition of salt-free DNA origami nanostructures. *Scientific Reports* **5,** 15634 (2015).

61. Saccà, B. & Niemeyer, C. M. Functionalization of DNA nanostructures with proteins. *Chemical Society Reviews* **40,** 5910–5921 (2011).

62. Heuer-Jungemann, A., Kirkwood, R., El-Sagheer, A. H., Brown, T. & Kanaras, A. G. Copper-free click chemistry as an emerging tool for the programmed ligation of DNA-functionalised gold nanoparticles. *Nanoscale* **5,** 7209–7212 (2013).

63. Koo, K. M., Sina, A. A., Carrascosa, L. G., Shiddiky, M. J. & Trau, M. DNA–bare gold affinity interactions: mechanism and applications in biosensing. *Analytical Methods* **7,** 7042–7054 (2015).

64. Hu, Q., Li, H., Wang, L., Gu, H. & Fan, C. DNA nanotechnology-enabled drug delivery systems. *Chemical Reviews* **119,** 6459–6506 (2018).

65. Li, L., Xing, H., Zhang, J. & Lu, Y. Functional DNA molecules enable selective and stimuli-responsive nanoparticles for biomedical applications. *Accounts of Chemical Research* **52,** 2415–2426 (2019).

66. Li, J., Fan, C., Pei, H., Shi, J. & Huang, Q. Smart drug delivery nanocarriers with self-assembled DNA nanostructures. *Advanced Materials* **25,** 4386–4396 (2013).

67. Genheden, S., Reymer, A., Saenz-Méndez, P. & Eriksson, L. A. Computational chemistry and molecular modelling basics (2017).

68. Chen, X., Guan, N.-N., Sun, Y.-Z., Li, J.-Q. & Qu, J. MicroRNA-small molecule association identification: from experimental results to computational models. *Briefings in Bioinformatics* **21,** 47–61 (2020).

69. Protasov, V., Serenko, A., Nenashev, V., Kulikov, I. & Chernykh, I. *High-performance computing in astrophysical simulations* in *Journal of Physics: Conference Series* **681** (2016), 012022.

70. Van Elteren, A., Pelupessy, I. & Zwart, S. P. Multi-scale and multi-domain computational astrophysics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **372,** 20130385 (2014).

71. Berendsen, H. J. Bio-molecular dynamics comes of age. *Science* **271,** 954–954 (1996).

72. Markwick, P. R. & McCammon, J. A. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Physical Chemistry Chemical Physics* **13,** 20053–20065 (2011).

73. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences* **102,** 6679–6685 (2005).

74. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99,** 1129–1143 (2018).

75. Comer, J., Chen, R., Poblete, H., Vergara-Jaque, A. & Riviere, J. E. Predicting adsorption affinities of small molecules on carbon nanotubes using molecular dynamics simulation. *ACS Nano* **9,** 11761–11774 (2015).

76. Im, W. & Brooks III, C. L. Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **102,** 6771–6776 (2005).

77. Śledź, P. & Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology* **48,** 93–102 (2018).

78. Alder, B. J. & Wainwright, T. E. Phase transition for a hard sphere system. *The Journal of Chemical Physics* **27,** 1208–1209 (1957).

79. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267,** 585–590 (1977).

80. Dawe, G. B., Musgaard, M., Aurousseau, M. R., Nayeem, N., Green, T., Biggin, P. C. & Bowie, D. Distinct structural pathways coordinate the activation of AMPA receptor-auxiliary subunit complexes. *Neuron* **89,** 1264–1276 (2016).

81. Lindahl, E. R. Molecular dynamics simulations. *Molecular Modeling of Proteins,* 3–23 (2008).

82. González, M. A. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **12,** 169–200 (2011).

83. Schneider, J., Hamaekers, J., Chill, S. T., Smidstrup, S., Bulin, J., Thesen, R., Blom, A. & Stokbro, K. ATK-ForceField: a new generation molecular dynamics software package. *Modelling and Simulation in Materials Science and Engineering* **25,** 085007 (2017).

84. Guvench, O. & MacKerell, A. D. Comparison of protein force fields for molecular dynamics simulations. *Molecular Modeling of Proteins,* 63–88 (2008).

85. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *Journal of Chemical Theory and Computation* **11,** 3499–3509 (2015).

86. Klauda, J. B., Monje, V., Kim, T. & Im, W. Improving the CHARMM force field for polyunsaturated fatty acid chains. *The Journal of Physical Chemistry B* **116,** 9424–9431 (2012).

87. Tobias, D. J., Martyna, G. J. & Klein, M. L. Molecular dynamics simulations of a protein in the canonical ensemble. *The Journal of Physical Chemistry* **97,** 12959–12966 (1993).

88. Boltzmann, L. *Über die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie:(vorgelegt in der Sitzung am 8. Februar 1866)* (Staatsdruckerei, 1866).

89. Nielsen, S. O. Nested sampling in the canonical ensemble: Direct calculation of the partition function from NVT trajectories. *The Journal of Chemical Physics* **139** (2013).

90. Fraternali, F. Restrained and unrestrained molecular dynamics simulations in the NVT ensemble of alamethicin. *Biopolymers: Original Research on Biomolecules* **30,** 1083–1099 (1990).

91. Toxvaerd, S. Molecular dynamics at constant temperature and pressure. *Physical Review E* **47,** 343 (1993).

92. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347,** 631–639 (1990).

93. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **9,** 646–652 (2002).

94. Adcock, S. A. & McCammon, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews* **106,** 1589–1615 (2006).

95. Rudling, A., Orro, A. & Carlsson, J. Prediction of ordered water molecules in protein binding sites from molecular dynamics simulations: the impact of ligand binding on hydration networks. *Journal of Chemical Information and Modeling* **58,** 350–361 (2018).

96. Dror, R. O., Green, H. F., Valant, C., Borhani, D. W., Valcourt, J. R., Pan, A. C., Arlow, D. H., Canals, M., Lane, J. R., Rahmani, R., *et al.* Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **503,** 295–299 (2013).

97. Cordero-Morales, J. F., Jogini, V., Lewis, A., Vásquez, V., Cortes, D. M., Roux, B. & Perozo, E. Molecular driving forces determining potassium channel slow inactivation. *Nature structural & Molecular Biology* **14,** 1062–1069 (2007).

98. Delemotte, L., Tarek, M., Klein, M. L., Amaral, C. & Treptow, W. Intermediate states of the Kv1. 2 voltage sensor from atomistic molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **108,** 6109–6114 (2011).

99. Bond, P. J., Holyoake, J., Ivetac, A., Khalid, S. & Sansom, M. S. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *Journal of Structural Biology* **157,** 593–605 (2007).

100. Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Laughton, C. A. Cooperativity in drug- DNA recognition: a molecular dynamics study. *Journal of the American Chemical Society* **123,** 12658–12663 (2001).

101. Spiegel, K. & Magistrato, A. Modeling anticancer drug–DNA interactions via mixed QM/MM molecular dynamics simulations. *Organic & Biomolecular Chemistry* **4,** 2507–2517 (2006).

102. Lei, H., Wang, X. & Wu, C. Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations. *Journal of Molecular Graphics and Modelling* **38,** 279–289 (2012).

103. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC biology* **9,** 1–9 (2011).

104. Duan, L. L., Feng, G. Q. & Zhang, Q. G. Large-scale molecular dynamics simulation: Effect of polarization on thrombin-ligand binding energy. *Scientific Reports* **6,** 31488 (2016).

105. Pal, S. K., Zhao, L. & Zewail, A. H. Water at DNA surfaces: ultrafast dynamics in minor groove recognition. *Proceedings of the National Academy of Sciences* **100,** 8113–8118 (2003).

106. De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry* **59,** 4035–4061 (2016).

107. Alivisatos, A. P., Johnsson, K. P., Peng, X., Wilson, T. E., Loweth, C. J., Bruchez Jr, M. P. & Schultz, P. G. Organization of'nanocrystal molecules' using DNA. *Nature* **382,** 609–611 (1996).

108. Wang, Y., Wang, Y., Zheng, X., Ducrot, É., Yodh, J. S., Weck, M. & Pine, D. J. Crystallization of DNA-coated colloids. *Nature Communications* **6,** 7253 (2015).

109. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. in *Spherical Nucleic Acids* 3–11 (Jenny Stanford Publishing, 2020).

110. Biancaniello, P. L., Kim, A. J. & Crocker, J. C. Colloidal interactions and self-assembly using DNA hybridization. *Physical Review Letters* **94,** 058302 (2005).

111. Kim, A. J., Biancaniello, P. L. & Crocker, J. C. Engineering DNA-mediated colloidal crystallization. *Langmuir* **22,** 1991–2001 (2006).

112. Di Michele, L. & Eiser, E. Developments in understanding and controlling self assembly of DNA-functionalized colloids. *Physical Chemistry Chemical Physics* **15,** 3115–3129 (2013).

113. Li, W., Palis, H., Mérindol, R., Majimel, J., Ravaine, S. & Duguet, E. Colloidal molecules and patchy particles: Complementary concepts, synthesis and self-assembly. *Chemical Society Reviews* **49,** 1955–1976 (2020).

114. Aldaye, F. A. & Sleiman, H. F. Dynamic DNA templates for discrete gold nanoparticle assemblies: control of geometry, modularity, write/erase and structural switching. *Journal of the American Chemical Society* **129,** 4130–4131 (2007).

115. Rogers, W. B., Shih, W. M. & Manoharan, V. N. Using DNA to program the self-assembly of colloidal nanoparticles and microparticles. *Nature Reviews Materials* **1,** 1–14 (2016).

116. Li, Z., Fan, Q. & Yin, Y. Colloidal self-assembly approaches to smart nanostructured materials. *Chemical reviews* **122,** 4976–5067 (2021).

117. Cai, Z., Li, Z., Ravaine, S., He, M., Song, Y., Yin, Y., Zheng, H., Teng, J. & Zhang, A. From colloidal particles to photonic crystals: Advances in self-assembly and their emerging applications. *Chemical Society Reviews* **50,** 5898–5951 (2021).

118. Van Blaaderen, A. Chemistry: Colloidal molecules and beyond. *Science* **301,** 470–471 (2003).

119. Wang, Y., Wang, Y., Breed, D. R., Manoharan, V. N., Feng, L., Hollingsworth, A. D., Weck, M. & Pine, D. J. Colloids with valence and specific directional bonding. *Nature* **491,** 51–55 (2012).

120. Li, Z. & Yin, Y. Stimuli-responsive optical nanomaterials. *Advanced Materials* **31,** 1807061 (2019).

121. Duguet, E., Désert, A., Perro, A. & Ravaine, S. Design and elaboration of colloidal molecules: an overview. *Chemical Society Reviews* **40,** 941–960 (2011).

122. Hou, S., Bai, L., Lu, D. & Duan, H. Interfacial Colloidal Self-Assembly for Functional Materials. *Accounts of Chemical Research* **56,** 740–751 (2023).

123. Stark, H. Physics of colloidal dispersions in nematic liquid crystals. *Physics Reports* **351,** 387–474 (2001).

124. Khan, Z. A., Kumar, R., Mohammed, W. S., Hornyak, G. L. & Dutta, J. Optical thin film filters of colloidal gold and silica nanoparticles prepared by a layer-by-layer self-assembly method. *Journal of Materials Science* **46,** 6877–6882 (2011).

125. Lin, I.-H., Miller, D. S., Bertics, P. J., Murphy, C. J., de Pablo, J. J. & Abbott, N. L. Endotoxin-induced structural transformations in liquid crystalline droplets. *Science* **332,** 1297–1300 (2011).

126. Gröschel, A. H., Schacher, F. H., Schmalz, H., Borisov, O. V., Zhulina, E. B., Walther, A. & Müller, A. H. Precise hierarchical self-assembly of multicompartment micelles. *Nature Communications* **3,** 710 (2012).

127. Zhang, W., Liao, S. & Cui, F. Hierarchical self-assembly of nano-fibrils in mineralized collagen. *Chemistry of Materials* **15,** 3221–3226 (2003).

128. Zhang, J., Kankala, R. K., Ma, J., Zhou, Y., Wang, S.-B. & Chen, A.-Z. Hollow Tobacco Mosaic Virus Coat Protein Assisted Self-Assembly of One-Dimensional Nanoarchitectures. *Biomacromolecules* **22,** 540–545 (2020).

129. Chen, L.-J. & Yang, H.-B. Construction of stimuli-responsive functional materials via hierarchical self-assembly involving coordination interactions. *Accounts of Chemical Research* **51,** 2699–2710 (2018).

130. Sang, Y. & Liu, M. Hierarchical self-assembly into chiral nanostructures. *Chemical Science* **13,** 633–656 (2022).

131. Grunwald, M. & Geissler, P. L. Patterns without patches: Hierarchical self-assembly of complex structures from simple building blocks. *ACS Nano* **8,** 5891–5897 (2014).

132. Diao, Y. Y., Liu, X. Y., Toh, G. W., Shi, L. & Zi, J. Multiple structural coloring of silk-fibroin photonic crystals and humidity-responsive color sensing. *Advanced Functional Materials* **23,** 5373–5380 (2013).

133. Ratna, D. & Karger-Kocsis, J. Recent advances in shape memory polymers and composites: a review. *Journal of Materials Science* **43,** 254–269 (2008).

134. Wu, X., Hao, C., Kumar, J., Kuang, H., Kotov, N. A., Liz-Marzán, L. M. & Xu, C. Environmentally responsive plasmonic nanoassemblies for biosensing. *Chemical Society Reviews* **47,** 4677–4696 (2018).

135. Vadapalli, D. R. P. & Roy, S. Analyzing two-dimensional tunable photonic crystal waveguides for communication band optical filter applications. *Journal of Optics* **50,** 346–353 (2021).

136. Grzelczak, M., Liz-Marzán, L. M. & Klajn, R. Stimuli-responsive self-assembly of nanoparticles. *Chemical Society Reviews* **48,** 1342–1361 (2019).

137. Yadav, S., Sharma, A. K. & Kumar, P. Nanoscale self-assembly for therapeutic delivery. *Frontiers in Bioengineering and Biotechnology* **8,** 127 (2020).

138. Dou, Q., Wang, S., Zhang, Z., Wang, Y., Zhao, Z., Guo, H., Liu, H. & Dai, Q. A highly sensitive quartz crystal microbalance sensor modified with antifouling microgels for saliva glucose monitoring. *Nanoscale* **12,** 19317–19324 (2020).

139. Maduraiveeran, G. Nanomaterials-based portable electrochemical sensing and biosensing systems for clinical and biomedical applications. *Journal of Analytical Science and Technology* **13,** 35 (2022).

140. Kalsin, A. M., Fialkowski, M., Paszewski, M., Smoukov, S. K., Bishop, K. J. & Grzybowski, B. A. Electrostatic self-assembly of binary nanoparticle crystals with a diamond-like lattice. *science* **312,** 420–424 (2006).

141. Noya, E. G., Vega, C., Doye, J. P. & Louis, A. A. The stability of a crystal with diamond structure for patchy particles with tetrahedral symmetry. *The Journal of Chemical Physics* **132** (2010).

142. Romano, F., Sanz, E. & Sciortino, F. Crystallization of tetrahedral patchy particles in silico. *The Journal of Chemical Physics* **134** (2011).

143. Morphew, D., Shaw, J., Avins, C. & Chakrabarti, D. Programming hierarchical self-assembly of patchy particles into colloidal crystals via colloidal molecules. *ACS Nano* **12,** 2355–2364 (2018).

144. Karner, C., Dellago, C. & Bianchi, E. Hierarchical self-assembly of patchy colloidal platelets. *Soft Matter* **16,** 2774–2785 (2020).

145. He, X., Song, M., Liang, H. & Pan, C. Self-assembly of the symmetric diblock copolymer in a confined state: Monte Carlo simulation. *The Journal of Chemical Physics* **114,** 10510–10513 (2001).

146. Suh, D., Radak, B. K., Chipot, C. & Roux, B. Enhanced configurational sampling with hybrid non-equilibrium molecular dynamics–Monte Carlo propagator. *The Journal of chemical physics* **148** (2018).

147. Neyts, E. C. & Bogaerts, A. Combining molecular dynamics with Monte Carlo simulations: implementations and applications. *Theoretical Chemistry in Belgium: A Topical Collection from Theoretical Chemistry Accounts*, 277–288 (2014).

148. Ržička, Š. & Allen, M. P. Collective translational and rotational Monte Carlo moves for attractive particles. *Physical Review E* **89,** 033307 (2014).

149. Huang, L. & Wang, L. Accelerated Monte Carlo simulations with restricted Boltzmann machines. *Physical Review B* **95,** 035105 (2017).

150. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the national academy of sciences* **99,** 12562–12566 (2002).

151. Martsinovich, N. & Troisi, A. Modeling the self-assembly of benzenedicarboxylic acids using monte carlo and molecular dynamics simulations. *The Journal of Physical Chemistry C* **114,** 4376–4388 (2010).

152. Whitelam, S. & Geissler, P. L. Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles. *The Journal of Chemical Physics* **127** (2007).

153. Ghobadi, A. F., Taghikhani, V. & Elliott, J. R. Investigation on the solubility of SO2 and CO2 in imidazolium-based ionic liquids using NPT Monte Carlo simulation. *The Journal of Physical Chemistry B* **115,** 13599–13607 (2011).

154. Ströker, P., Hellmann, R. & Meier, K. Systematic formulation of thermodynamic properties in the N p T ensemble. *Physical Review E* **103,** 023305 (2021).

155. Rosales-Pelaez, P., Sanchez-Burgos, I., Valeriani, C., Vega, C. & Sanz, E. Seeding approach to nucleation in the NVT ensemble: The case of bubble cavitation in overstretched Lennard Jones fluids. *Physical review E* **101,** 022611 (2020).

156. Hansmann, U. H. & Okamoto, Y. New Monte Carlo algorithms for protein folding. *Current Opinion in Structural Biology* **9,** 177–183 (1999).

157. Garcia, A. E. & Sanbonmatsu, K. Y. Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins: Structure, Function, and Bioinformatics* **42,** 345–354 (2001).

158. Chen, J. C. & Kim, A. S. Brownian dynamics, molecular dynamics, and Monte Carlo modeling of colloidal systems. *Advances in Colloid and Interface Science* **112,** 159–173 (2004).

159. Neyts, E. C. On the time scale associated with Monte Carlo simulations. *Journal of Chemical Physics* **141** (2014).

160. Patti, A., Mackie, A. D. & Siperstein, F. R. Monte Carlo simulation of self-assembled ordered hybrid materials. *Langmuir* **23,** 6771–6780 (2007).

161. Nakagawa, N., Maeda, S., Ishii, S. & Ohno, K. A monte carlo simulation of the formation of micelles in a ternary system of water, oil and amphiphilic polymers. *Materials transactions* **48,** 653–657 (2007).

162. Lewandowski, K. & Banaszak, M. Collapse-driven self-assembly of multiblock chains: A Monte Carlooff-lattice study. *Journal of non-crystalline solids* **355,** 1289–1294 (2009).

163. Patti, A. & Cuetos, A. Brownian dynamics and dynamic Monte Carlo simulations of isotropic and liquid crystal phases of anisotropic colloidal particles: A comparative study. *Physical Review E* **86,** 011403 (2012).

164. Rao, A. B., Shaw, J., Neophytou, A., Morphew, D., Sciortino, F., Johnston, R. L. & Chakrabarti, D. Leveraging hierarchical self-assembly pathways for realizing colloidal photonic crystals. *ACS nano* **14,** 5348–5359 (2020).

165. Frenkel, D., Smit, B., Tobochnik, J., McKay, S. R. & Christian, W. Understanding molecular simulation. *Computers in Physics* **11,** 351–354 (1997).

CHAPTER 2

The Protective Effect in the Context of DTS

DNA as a scaffold for synthesis facilitates the production of sequence-defined polymers in a programmable, autonomous, and sequence-specific manner. DTS can facilitate molecular evolution, which can greatly enhance material and drug development. However, DTS occasionally encounters truncation of polymers, arising from the hydrolysis of the labile linkers, essential for transfer reactions. This limitation significantly restricts the efficacy of DTS in synthesising long-chain polymers. This chapter discusses methods of sequence-specific synthesis, outlining the efforts made to devise and assess a strategy to shield labile components from hydrolysis within the context of DTS.

2.1 | Natures Approach to Synthesis

Nature is proficient in the design and production of a plethora of sequence-specific biomolecules.[1] Nature performs this task with unparalleled accuracy and efficiency. The intricate biomolecules produced possess a broad spectrum of properties and functions, perfected over millions of years of evolution.[2–4] Natures arsenal for the synthesis of these compounds consists of proteins, nucleic acids, and enzymes.[5,6] These biomolecules play distinct and complementary roles in the synthesis and regulation of various compounds. Proteins, particularly those that function as enzymes, act as catalysts, speeding up biochemical reactions that would otherwise occur too slowly to sustain life. Enzymes are a specialized subset of proteins, designed to facilitate specific chemical reactions with high precision, often requiring cofactors or coenzymes to function optimally.[6]

Nucleic acids, including DNA and RNA, are the molecules responsible for storing and transmitting genetic information. DNA serves as the blueprint for the synthesis of proteins, while RNA translates this genetic information into functional proteins through processes like transcription

and translation.[5] Unlike proteins, nucleic acids are not directly involved in catalysis but instead guide the assembly of amino acids into polypeptides, which then fold into functional proteins. The key differences between these molecules lie in their functions and mechanisms of action: proteins and enzymes directly catalyze biochemical reactions, while nucleic acids provide the genetic instructions necessary for the synthesis of these catalysts. Together, they form a highly integrated system where nucleic acids encode the information required to build proteins and enzymes, which in turn carry out the chemical processes essential for life.[5,6]

With these synthetic tools, nature can fabricate a myriad of proteins in a sequence-specific manner. It does this through the ribosome, catalysing the sequential anchoring of t-RNA molecules to long templating mRNA strands.[1] The sequence specificity of this process arises from the precise nature of nucleobase hybridisation. Achieving such mastery in molecular design has been a pursuit of science.

2.2 | Approaches to Sequence-Defined Polymer Synthesis

2.2.1 Ribosomal Synthesis of Modified Amino Acids

Chemists have attempted to adapt the ribosomal synthesis methodology through the incorporation of chemically modified amino acids into the ribosome.[7] However, this method is confined to amino acids, severely limiting its versatility. Research has incorporated 'unnatural' amino acids in an adapted ribosomal synthesis,

Chemists have attempted to adapt the ribosomal synthesis methodology through the incorporation of chemically modified amino acids into the ribosome.[7] However, this method is confined to amino acids, severely limiting its versatility. Research has incorporated 'unnatural' amino acids in an adapted ribosomal synthesis; however, contrary to the claim that the modifications are minimal, there are numerous examples where significant deviations from natural amino acids have been achieved. For instance, unnatural amino acids (UAAs) have been synthesized and incorporated into proteins to alter their functionality and properties drastically. UAAs such as azides, alkynes, and other reactive groups have been successfully incorporated, allowing for the introduction of new chemical functionalities that are not present in natural amino acids. These UAAs have been used in various applications, including the creation of novel

biocatalysts, drug development, and protein engineering, showcasing the vast potential beyond minor modifications like the addition of a methyl group.[8].

However, despite these advancements, the adaptation of ribosomal synthesis for UAAs remains primarily targeted at modifying protein structures rather than creating entirely new types of polymers. This approach is limited by the ribosome's natural affinity for proteinogenic amino acids and its role in translating genetic code into proteins. As a result, the range of monomer units that can be utilized in ribosomal synthesis is restricted to those that closely resemble natural amino acids. Additionally, the process can be low-yielding and expensive, particularly when attempting to incorporate UAAs that significantly deviate from the structure of natural amino acids. Consequently, the programmable synthesis of non-protein-based sequence-defined polymers has predominantly been produced *via* other DNA approaches like split and pool combinatorial approaches.[9]

2.2.2 Non-Templated Methodologies

Sequence-defined polymers are pivotal in diverse fields such as drug delivery systems, pharmaceuticals, high-performance materials,[10] and information storage.[11] Traditionally, their synthesis has been achieved through solid phase methods[12] or reversible deactivation radical polymerization (RDRP) and step-growth techniques.[10]

Sequence-specific multi-step synthesis has been traditionally a laborious task requiring multiple purification steps. The construction of sequence-defined polymers on the nanoscale remains a challenge for non-templated methods.[13] However, various strategies of autonomous sequence-specific multi-step synthesis have shown great promise.[14,15]

Orthogonal and iterative growth strategies are limited in performing only one coupling step per reaction, often requiring protective/de-protective strategies.[2] In orthogonal strategies, specific functional groups on monomers are selectively activated to react with complementary groups on a growing polymer chain. This requires the use of protecting groups to block reactive sites that are not intended to participate in the reaction, which adds complexity and necessitates additional steps to remove these groups afterward. The use of different catalysts for each reaction step, such as transition metal catalysts or acid/base catalysts, ensures that only the desired functional groups react, allowing for the precise incorporation of each monomer. However, this process is often slow and can be expensive due to the need for multiple purification steps and the careful control of reaction conditions.[10]

Controlled radical polymerization techniques like reversible addition-fragmentation chain-transfer (RAFT) polymerization are capable of incorporating multiple monomer units per reaction setup.[11] RAFT polymerization operates through a reversible chain transfer mechanism where a chain transfer agent (typically a thiocarbonylthio compound) mediates the growth of the polymer chain. The process involves the formation of a dormant species that can be reactivated to continue polymer growth, allowing for controlled addition of monomer units. This method leverages functional groups like acrylates or methacrylates, which can be polymerized under mild conditions using radical initiators. The versatility of RAFT lies in its ability to control molecular weight and incorporate diverse monomer units. However, achieving sequence-specific control can be challenging due to side reactions and the need for a precise balance between the reactivity of the monomer, the radical initiator, and the chain transfer agent.[11]

In non-templated methods like step-growth polymerization, the polymer grows by the reaction of bifunctional or multifunctional monomers that form bonds such as ester, amide, or urethane linkages. The process often requires catalysts like tin(II) octoate for polyesters or isocyanate-based catalysts for polyurethanes, which drive the condensation reaction between the functional groups of the monomers. These reactions are typically carried out under conditions that promote the removal of small molecules like water or alcohol, which can drive the polymerization towards higher molecular weights. While these methods are effective for producing polymers with specific properties, they generally lack the precision of sequence-defined methods and often result in polymers with a broader distribution of chain lengths and sequences.[10]

Programmable synthesis, on the other hand, offers superior control over unit sequence and dispersity. One method of programmable synthesis attempts to employ molecular machines to template multi-component, sequence-specific polymerisations, particularly utilising rotaxane on an axle.[16] Rotaxane-based molecular machines represent a significant advancement in the field of programmable synthesis. In this approach, a macrocycle is threaded onto a molecular axle, and it travels along the axle, picking up and transferring specific monomers to form a polymer chain. The sequence specificity is derived from the pre-arranged order of the monomers on the axle. For example, in the study by Lewandowski et al., the artificial molecular machine demonstrated the ability to synthesize a peptide in a sequence-specific manner, mimicking the ribosome's function but with entirely synthetic components. Despite its potential, this method has limitations, including the relatively low yield of 30% and the lengthy reaction time of 35 hours to complete a single synthesis cycle. Furthermore, the sequence specificity is entirely

dependent on the synthetic pre-organization of the structure, making it less versatile compared to natural systems like the ribosome, which can read genetic information directly.[16]

Another approach in programmable synthesis involves DNA-templated synthesis (DTS). DTS uses the specificity of DNA base pairing to bring reactants into close proximity, thereby facilitating the formation of sequence-specific polymers. This method allows for the construction of complex molecules by encoding the desired sequence into a DNA strand, which then guides the synthesis process. DTS has the potential to create a much more chemically diverse set of products compared to rotaxane-based systems, as it can incorporate a wide range of chemical functionalities. However, DTS also faces challenges, such as the need for careful design to prevent non-specific interactions and the potential limitations in the types of chemical reactions that can be templated by DNA.[17]

2.3 | DNA-Templated Synthesis

DNA-templated synthesis (DTS) is a programmable, autonomous, sequence-specific method of synthesis inspired by the ribosomal synthesis of proteins.[17] DTS utilises modified DNA strands which self-assemble to pre-designed architectures, selectively enhancing the proximity of DNA-bound monomers.[18] Low concentrations of modified oligomers are used in order to discourage off-template polymerisations. Reaction concentrations as such are typically kept in the micromolar to nanomolar range.[19]

Despite RNA being utilised *via* the ribosome in the translation process, DNA is preferred for DTS. This preference stems from the presence of extra hydroxyl in RNA, which has the capability of acting as a nucleophile, interfacing with templated reaction.[20] DNA is an extremely versatile molecule, capable of self-assembling into a variety of structures documented to host DTS reactions.[21] Its structural versatility stems from its flexible phosphate backbone and specific hydrogen bonding.[22] Each nucleotide undergoes Watson-Crick base pairing with its complementary base pair. On its own, hydrogen bonds are relatively weak compared to the bond enthalpies of covalent and ionic bonds; however, the cumulative effect of these bond enthalpies can be substantial. For instance, a human chromosome typically ranges in size from $50 - 300 \times 10^6$ base pairs.[23] The lengths of the modified DNA strands used in DTS, however, are not of this scale. The lengths and sequences of DNA are meticulously chosen to enable reaction progression *via* means of toehold-mediated strand displacement.[17] After a modified oligomer has transferred its attached monomer to the growing polymer chain, an exchange

of the spent adapter is required. Toehold-mediated strand displacement operates *via* the incoming strand forming a greater number of base pair interactions with the templating strand. This increased base-pair interaction leads to a thermodynamic driving force for displacement.[24] This toehold can be introduced *via* several ways: it could be present on the growing polymer DNA, or could be introduced through an enzyme-mediated cleavage (Figure 2.1).[25,26]

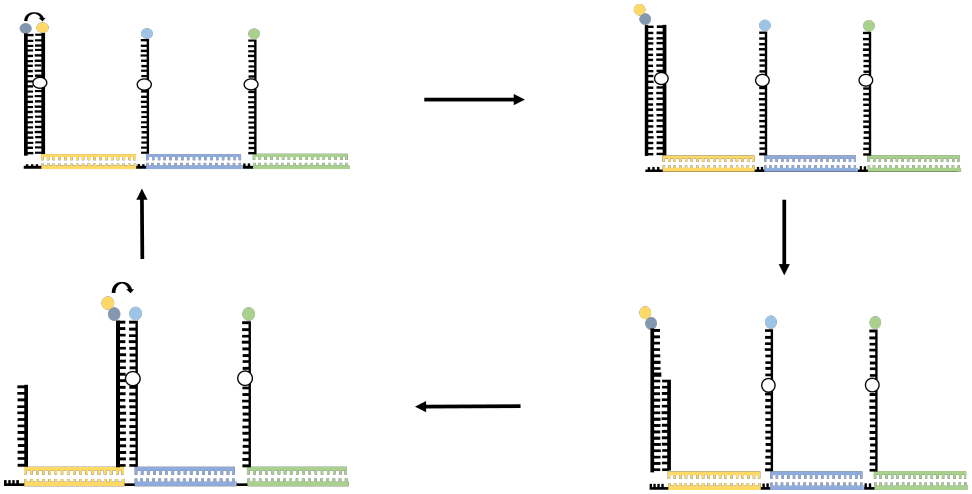


Fig. 2.1. Fig. 2.1. An example of a DTS walker where toe-hold-mediated strand displacement is enabled through enzymatic cleavage of a spent DTS adapter.[42] In this mechanism, the toe-hold is a short, single-stranded DNA sequence that initiates the displacement of a longer, complementary strand. When the enzyme cleaves the adapter, the newly exposed toe-hold binds to the incoming strand, facilitating the displacement of the original strand, thus driving the walker forward.

Molecular Evolution

Unlike the molecular machine method of synthesis, DNA enables the possibility for molecular evolution.[17,27] For molecular evolution to occur, two criteria must be met. The first requires a unique nucleotide sequence for each monomer used in the synthesis.[21] The second required for the templating DNA contains the sequence of all nucleobases corresponding to each monomer comprising the product in a specific order.[28] Fulfilling both criteria enables the identification of the product purely by the sequence of nucleobases. For example, such a mechanism enabling both these criteria could involve a long templating strand with discrete binding sequences of nucleobases corresponding to monomer-carrying, t-RNA-like adapters (Figure 2.2).

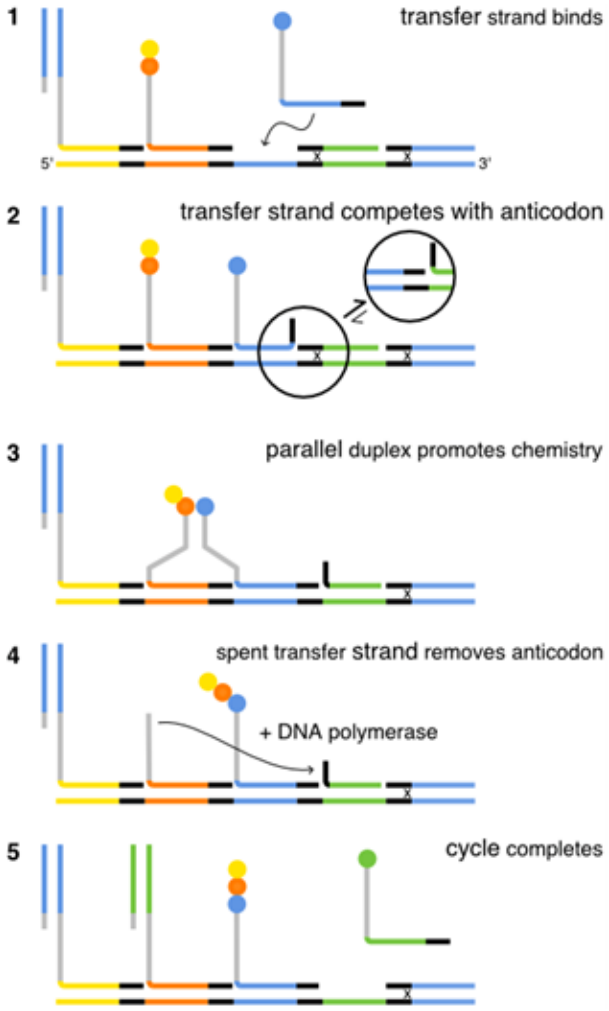


Fig. 2.2. An example of a DTS scheme hosted on a long template where the corresponding nucleobases are colour-coded, and the transfer reaction of the corresponding monomers unveils a DNA sequence that can be extended by a polymerase. This DNA chain extension builds a complementary DNA strand with respect to the next reaction site, enabling the next modified DNA adapter to join, thus allowing the reaction to progress.

Molecular evolution is directed through screening and identifying desirable characteristics *via* a screening process. This process involves subjecting the resultant polymers to an appropriate assay diagnostic for clear analysis of its efficacy concerning its desired application.[29] Through comparison of a variety of products produced through various permutations and combinations of monomers, one can identify desirable monomer sequences with respect to its desired application.[17] After identifying the most effective strand(s) followed by permutations of splicing, mutating and recombing the successful coding DNA strands, a directed evolution takes place (Figure 2.3-2.4).[30]

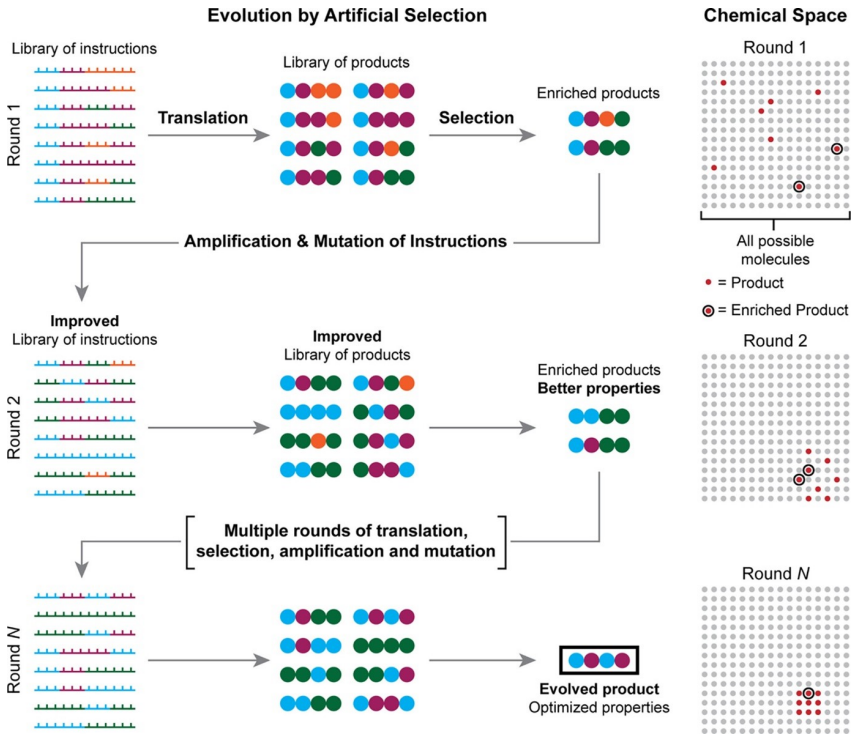


Fig. 2.3. Artificial selection of a desired compound synthesised *via* DNA templated synthesis. A library of DTS instructions is translated via DTS to their complementary products. These products are then subjected to a target selection process. The corresponding instructions of the selected products are then amplified, mutated, and translated to generate a library of products with more refined properties. This process iterates N times. This is a process of artificial selection. Figure taken with permission from Ref.[17].

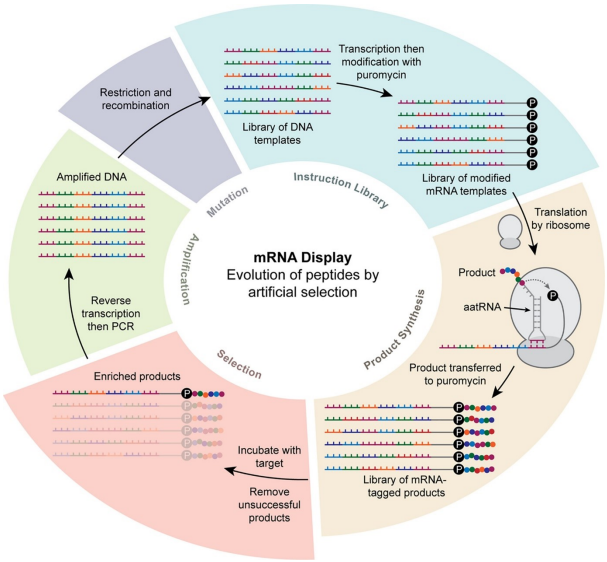


Fig. 2.4. an analogous artificial selection cycle utilising mRNA display hosting an evolution of peptide where the desirable peptides are isolated through a binding assay with the intended target. Figure taken with permission from Ref.[17]

2.3.1 Limitations in DTS

One notable limitation of using DTS for polymer synthesis is the decline in the efficiency of transfer reaction between the building blocks as polymer chain length increases.[19] The reduced efficiency stems from the method of addition being one of extension as opposed to insertion. As the chain length of the growing polymer increases the effective concentration of the electrophile and nucleophile for the attaching monomer and growing polymer decreases, thus decreasing reaction efficiency.

Owing to DNA's negative phosphate backbone, 3' & 5' hydroxyl groups, and the hydrogen bond-forming potential of its nucleobases, it requires a polar solvent for its dissolution. DNA has demonstrated solubility in solutions of glycerol, ethylene glycol, formamide, methanol, and DMSO, with a small amount of water. However, in these solvents hydrogen bonding between the nucleobases is, to an extent, disrupted.[31,32] For undisturbed duplex formation, the solvent is not only required to screen the negative charges of the phosphate backbone but also not disturb the Watson-Crick base pairing.[33]

This requirement for an aqueous solvent poses some complications in the context of DTS, specifically hydrolysis of labile linkers.[34,35] To address this issue a method of protection is required. This method should be incorporated in a way that does not disrupt the unique architecture of the DNA complex and has minimal to no impact concerning their hybridisation.

Intercalators

DNA interacts with various natural ligands during the cell cycle; these interactions can modulate the function of DNA consequently altering the proteins synthesised through ribosomal synthesis.[36] These ligands can bind within the major and minor grooves or intercalate between the bases of DNA. These interactions are mediated by van der Waals, electrostatic forces, and hydrogen bonding.[37] The primary enthalpic force governing the interactions of intercalators with DNA is $\pi - \pi$ stacking.[36,38] Once a molecule intercalates, it's inserted into the hydrophobic core of the DNA duplex. This hydrophobic environment could result in protection from hydrolysis. Perhaps the insertion of labile components into this hydrophobic core could result in protection in the context of DTS.

However, intercalation into the hydrophobic core of the double helix could potentially protect a hydro-sensitive component from hydrolysis. A protective mechanism like this could be implemented into DTS to limit the truncation of hydro-sensitive building blocks.[39,40]

Abasic Binders

An abasic site is a region in a DNA strand where a nucleobase is absent. The presence of an abasic site reduces the activation energy required for the insertion of small molecules into the duplex *via* the abasic site. This lowering of the activation energy also specifies the site of insertion. This absence of a nucleobase also enables the inserting molecule to potentially engage with hydrogen bonding to the opposing nucleobase of the abasic site of the hybridising strand. These advantages make abasic binders an attractive option for the controlled insertion of molecules into the DNA core. Abasic binders therefore could enable the controlled insertion of labile components into the hydrophobic core of the DNA, shielding from hydrolysis and thus preventing the premature truncation of polymers. The Turberfield Group has pioneered this research, experimentally demonstrating the protection of a labile thioester at pH 11 conditions through the incorporation of a minor groove abasic site.[41] This experimental outcome has the potential to be developed into a protection mechanism that could be utilised in the context of DTS.

Turberfield Group's Approach to Thioester Protection in DNA-Templated Synthesis

The research in this chapter involves investigating the stability and efficiency of acyl transfer in a DNA-templated synthesis (DTS) system, specifically focusing on the effects of incorporating an abasic site on one DNA strand under alkaline conditions. Thioesters readily hydrolyse at pH 11 conditions, therefore any reduction in this rate of hydrolysis is evidence of protection.[42] An experiment was set up to determine the efficacy of acyl transfer in the context of DTS with the inclusion of an abasic site on one strand. This abasic site was varied in relative position concerning the modified linker-thioester-TAMRA component. TAMRA was used as a fluorescent marker to determine the relative yield of the acyl transfer. The hybridising DNA strand therefore was host to a linker-amine and an abasic site (Figure 2.5). Both modified DNA strands were hybridised and subjected to pH 11 conditions for 48 h, where the extent of acyl transfer between the two strands was then determined *via* denaturing PAGE.[41]

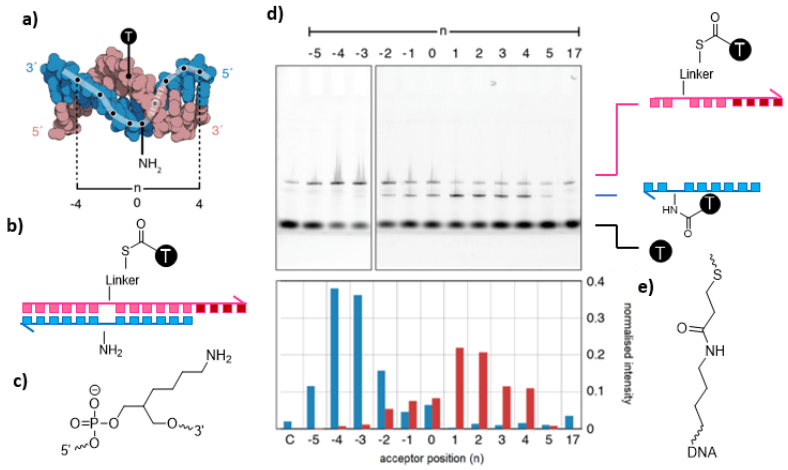


Fig. 2.5. Position-dependent DTS with reactants tethered at internal DNA modifications. a) Space-filling 3D model of the central region of the DNA duplex employed in this study, illustrating the relative positions of the donor TAMRA thioester (T) and acceptor amine. b) Schematic depiction of the DNA-templated acyl transfer reaction, with squares representing nucleotides and missing squares denoting abasic sites. The overhang on the donor strand (blue) enables strand distinction using PAGE. Polyacrylamide Gel Electrophoresis (PAGE) is a technique used to separate DNA strands based on their size and charge. In this context, PAGE allows for the clear differentiation between donor and acceptor strands due to the mobility differences stemming from the overhang.
c) Structure of the abasic site complemented with amine modification. d) PAGE gels represented using TAMRA fluorescence (top) and reaction yields progression, quantified by densitometry (lower), for DTS reactions between the thioester and the amine at relative positions $-5 \leq n \leq 5$, $n = -17$. Control (C) lacks abasic acceptor modification. e) Chemical Structure of the 'linker', delineating its molecular composition and connection between the abasic site and the modifying group. Figure adapted from ref.[43].

The relative yield of DTS was determined *via* densitometry which compared the TAMRA fluorescence in a denaturing polyacrylamide gel electrophoresis gel (PAGE gel) between the two modified DNA strands (Figure 2.5). Denaturing PAGE separated the individual DNA strands *via* their intrinsic electrophoretic mobility through the polyacrylamide matrix. The appearance of these bands were detected *via* their TAMRA fluorescence.

This experiment revealed varying yields of DTS when the abasic amine was positioned over the major groove. The classification of an abasic site as being in the major or minor groove is not absolute and is only defined relative to a fixed perspective along the double helix. In DNA, the major and minor grooves refer to the spaces between the backbones of the double helix, which alternate depending on the perspective of observation.[44] When we refer to an abasic site as being in the "major groove," we mean that, from the perspective of the attached linker, the site is positioned on the more accessible and spacious side of the helix. Conversely, when it is

described as being in the "minor groove," it is situated on the narrower, less accessible side relative to the linker.

The specific placement of the TAMRA-linker determines which groove the abasic site falls into as it relates to the directionality of the DNA strand. In this experiment, higher DTS yields were observed when the abasic site was positioned in the major groove relative to the linker, as opposed to the minor. It was discovered that the protective arose when the abasic site was positioned over the minor groove, conversely no protection arose when the abasic site was positioned along the major groove.

Between the -2 and -5 positions, a large quantity of unreacted thioester TAMRA remained. As this thioester TAMRA did not react with amine *via* acyl transfer nor degrade from hydrolysis, some protective interaction must have manifested, shielding the thioester from both the amine and hydrolysis.

There exists a few modes of binding in which the TAMRA moiety could engage with the duplex, one of those being interaction with the minor groove mediated through van der Waals and electrostatic interactions.[45] The current hypothesis for why such a protective effect manifested involves the idea that TAMRA inserts into the abasic site enthalpically driven by $\pi - \pi$ stacking interactions with neighbouring nucleobases and entropically by the release of water from the abasic site. As for why this protective effect only arises when the abasic site is situated over the minor groove could stem from molecules' preference for minor groove binding.[46] Perhaps this effect lowers the activation energy barrier for insertion into the abasic site.

Interestingly, the experiment showed that the protection was higher when the abasic site was positioned over the minor groove. The linker used in this system is predominantly an alkyl chain, which typically provides a significant degree of flexibility.[47] This flexibility should, in theory, allow the linker to interact with the abasic site regardless of whether it is positioned in the major or minor groove. However, the results suggest that the linker's flexibility alone is not the sole determining factor.

One possible explanation is that the stereochemical orientation of the linker plays a crucial role. The enantiomeric identity of the linker—its specific three-dimensional arrangement—may cause it to preferentially interact with sites in the groove that align with its stereochemistry. When the abasic site is positioned in the minor groove, it may be more favorably aligned with the linker's stereochemical orientation, facilitating better interaction and thus providing higher protection. Conversely, if the abasic site is in the major groove, the linker's orientation might prevent it from effectively interacting with the site, leading to reduced protection. This

interaction specificity could explain the observed differences in protection levels depending on the groove positioning of the abasic site. The supplier of the amine-DNA linker - used to construct the DNA-thioester-TAMRA - does not know the enantiomeric identity of their oligomer. Therefore this line of enquiry cold not be addressed in this scope of this PHD. In-spite of this theory it's important to underline that the minor groove is located in closer proximity to the linker.
(Figure 2.6).

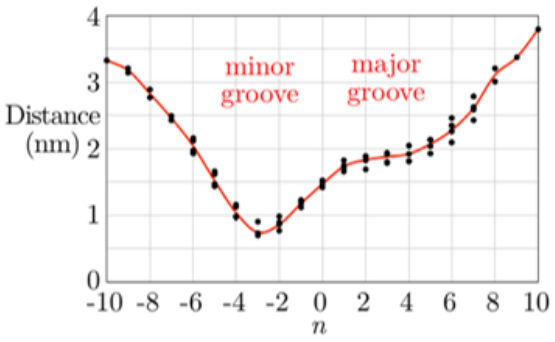


Fig. 2.6. The distance between the 4' carbons for the synthetic linkers was measured using the protein database (PDB) crystal structures of the DNA duplexes using computational software by former PhD student Robert Oppenheimer.

2.3.2 Implementation of Molecular Dynamics

Throughout history, both scientific models and equations have illuminated the intricacies and mechanisms of physics and nature.[48] Our understanding of such phenomena is therefore mirrored by the accuracy and precision of our mathematical models. A true model should accurately describe the individual components of a system as well as their cumulative interactions. Computation has significantly expanded and deepened our ability to model and understand the complexities of physics and nature.[49,50] In particular, computational research has facilitated a deeper, more detailed understanding of the molecular world, where interactions occur beyond the scope of direct visual inspection. Simulation and computation are imperative in forming such a detailed understanding. In this chapter, molecular dynamics simulations will be integrated with experimental techniques to comprehensively probe the DNA-TAMRA abasic protective effect discovered by Oppenheimer.[41] The combination of empirical experiments with computational models offers a powerful approach to visualizing and quantifying nanoscopic

phenomena, providing an atomistic understanding of molecular interactions that neither method could achieve alone.[51]

Molecular dynamics was employed to elucidate the structure-activity relationship to this selective protective effect. Molecular dynamics has been successful in probing other analogous molecular interactions such as drug intercalation.[52–55] Intercalation entails the insertion of a small molecule in between a pair of nucleobases. As discussed in chapter two, this sort of insertion of the TAMRA to the abasic site is hypothesised to be the cause of this unique protective effect. With molecular dynamics, we aim to discover what conditions lead to the insertion of the TAMRA unit into the abasic site and how this interaction may result in the protection of the thioester. The following section explores the principles and operations of molecular dynamics.

2.4 | Integration of Newtonian Mechanics in Molecular Dynamics

Molecular dynamics (MD) is a robust simulation technique that elucidates the behaviour of a multi-particle system.[56–58] MD results in a trajectory of all atoms in the system over the given time frame. The principal operation of MD is classical mechanics.[59] Utilising specific potential functions, the motion of an N-body system is iteratively derived using Newton's second law of motion.[60]

Mathematically expressed as:

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2} \tag{2.1}$$

F_i denotes the force exerted on the ith particle, m_i represents the corresponding mass. The second derivative of the particle position r_i concerning time t is the acceleration. This derivative captures the motions exhibited as a response to force.

The force and consequent motion of every particle emanates from the interactions of potential energies between them. Therefore potential energy and force are intimately intertwined, resulting in an expression of force as:

$$F_i = -\nabla_i V(r_1, \ldots, r_N) \tag{2.2}$$

The force vector F_i acts in the opposing direction of the gradient operator ∇_i concerned with the i th particle acting on the potential energy function $V(r_1, \ldots, r_N)$. The potential energy of the system is dependent on the positions of all N particles, represented by position vector r_i. This potential energy function therefore encapsulates all interactions within the system. This expression introduces a layer of complexity, as all the individual potentials are interlinked, resulting in a multi-body problem. Consequently, this multi-body problem creates a complex matrix of equations. The solving of which necessitates a computational solution.[61] To approach the problem, a temporal segmentation method is applied: the finite difference method. This operates by splitting time into finite chunks δt, allowing for step-wise resolution enabling iterative solving of the equations of motion. The forces on each particle are calculated at each time step τ with the associated motion applied to each respective particle for every δt. This iterative mechanism operates throughout the process, recording particle behaviour in a temporal continuum.

As the particle motion is ultimately dictated by the potential energy exerted on the particle, the potential energy must be accurate. For an accurate description of potential, the correct force field must be implemented.[62,63] Forcefields describe the interaction between particles in a system, through the consideration of bond stretching, angle bending, torsional potentials, electrostatic and van der Waals potentials.[63,64]

2.5 | Force Field

As described, the force field delineates the interaction between nuclei which govern their motion. Utilising the *Born-Oppenheimer approximation*, the motions of electrons are decoupled from the dynamics of the nuclei. However, the electronic and nuclear wave functions both dictate the resulting charge distribution. The parameters for these forcefields are obtained *via* experimental data, theoretical insight, and high-level quantum mechanical computation.[65] The two main contributors to the potential are intramolecular interactions and intermolecular interactions.[63]

$$E_{\text{eff}} = E_{\text{bonded}} + E_{\text{non-bonded}} \tag{2.3}$$

E_{eff} is the total potential energy, E_{bonded} represents the energy contribution from bonded interactions within a molecule in the system. $E_{\text{non-bonded}}$: the energy contribution from non-bonded interactions between molecules in the system.

Bonded and non-bonded energy components can be deconstructed into a series of discrete interactions. The bonded component comprises bond-streching, angle-bending and dihedrlas. Non-bonded: electrostatics and van-der-walls Summing these interactions gives the total energy of the molecule in the system:

$$E_{\text{bonded}} = \sum E_{\text{bonds}} + \sum E_{\text{angles}} + \sum E_{\text{dihedrals}} \tag{2.4}$$

$$E_{\text{non-bonded}} = \sum E_{\text{elec}} + \sum E_{\text{vdw}} \tag{2.5}$$

$$E_{\text{eff}} = \sum E_{\text{bonds}} + \sum E_{\text{angles}} + \sum E_{\text{dihedrals}} + \sum E_{\text{elec}} + \sum E_{\text{vdw}} \tag{2.6}$$

2.6 | Bond Stretching and the Harmonic Approximation

The Bond stretching term describes the variation in energy between two atoms with respect to bond length. To further deconstruct this relation, the Taylor expansion is employed. By truncating the expansion at the second-order term, the harmonic approximation is obtained:[66]

$$E(r) = \frac{1}{2}k(r - r_0)^2 \tag{2.7}$$

r_0 denotes reference bond length, or the length at which the bond energy is minimised. r describes the actual bond length. k: the harmonic force constant. k can be determined by:

$$k_r = \left.\frac{d^2E}{dr^2}\right|_{r=r_0} \tag{2.8}$$

The energy is minimised at $r = r_0$. As the distance between nuclei decreases past this equilibrium bond length, the potential energy spike to the exponential. The repulsive force stems from the repelling electron densities. An attracting force is also applied when the bond angle exceeds its equilibrium, pulling the nuclei back to minimum potential.

Within the MD paradigm, every covalent bond remains intact. Therefore, all vibrational amplitudes exhibited by the covalent bonds exist within the bounds of the harmonic approximation.

2.7 | Dihedral Angles in Molecular Modeling

The dihedral, or torsional, angle is an essential parameter for understanding molecular conformation and flexibility.[67] The dihedral angle describes the angle between two planes formed *via* four sequentially bonded atoms, denoted as A-B-C-D. Specifically, the dihedral angle,

represented by ω, describes the angle between the plane A-B-C and the plane B-C-D. This value can range from -180° to $+180^\circ$.

2.7.1 Energy Function of Dihedral Angles

Associated energy change with respect to the dihedral angle is often characterized using a Fourier series expansion:

$$E(\omega) = \sum_{n=1}^{N} \frac{1}{2} V_n \cos(n\omega - \delta_n) \tag{2.9}$$

V_n quantifies the rotational barrier or force constant for dihedral rotation. n reflects the barrier's periodicity, with $n = 1$ and $n = 2$ corresponding to single and double barriers, respectively. Phase offset δ_n, adjustable or zero, influences the potential energy landscape and is dictated by molecular conformational preference.

2.7.2 Deviations in Dihedral Angles

Large deviations in dihedral angles are more common compared to bond stretching or angle bending. This primarily is because bonds and angles are dictated by strong covalent interactions with stiff force constants, leading to substantial energy penalties for deviations. Dihedral angles however, are mainly governed by non-bonded interactions, such as the avoidance of steric repulsions and the pursuit of favorable conjugation.

The precise description of dihedral potentials is crucial in molecular simulations and modeling. Inaccuracies in representing these potentials can result in unrealistic molecular conformations and potentially incorrect predictions of molecular behaviour.

2.8 | Electrostatic Forces in Molecular Dynamics

Electrostatic forces have a significant effect on the motion of molecules. The distribution of electron density within a system dictates conformational change, resulting in atomic motion. The electro-negativity of a nuclei dictates the distribution of elections; electro-negativity is defined as an atom's capacity to attract electrons.[68] Atoms with greater electro-negativity draw more electron density towards themselves compared to those with lower electro-negativity. This variation in electro-negativity consequently leads to uneven electron distribution, which results in the generation of electrostatic forces in a system.

2.8.1 Modeling Electrostatic Interactions

To model the electrostatic interactions, point charges are assigned to each atomic site within the molecule. With these point charges represent the partial charges derived from the atom's electronegativity. The corresponding interaction energy arising from the resultant charge is described by the Colomb pontntial:[69]

$$E_{\text{elec}} = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \tag{2.10}$$

ε_0 represents the permittivity of free space. This is a constant which relates the electric field and electric displacement in a vacuum. q_i and q_j denote atomic charges of atoms i and j. r_{ij} represents the respective distance between atoms i and j.

2.8.2 Importance in Molecular Dynamics

Comprehensive and accurate implementation of electrostatic forces is crucial in determining the accuracy of molecular dynamics. Electrostatic interactions are pivotal in molecular dynamics, these forces dictate intermolecular interactions, which influence the solvation process, molecular recognition patterns and binding affinities. Not only this, but electrostatic interactions have major contributions to determining the stability of conformations in a molecular system. Even the very motion of molecules: translation, rotation and vibration are directed by the electrostatic force. Therefore, the accurate description of partial charge is imperative in molecular dynamics.[56,70]

2.9 | Long-range Electrostatic Attraction in Molecular Dynamics

A fundamental challenge in the incorporation of electrostatic interaction in molecular dynamics stems from their long-range nature. Due to the periodic boundary conditions, long-range electrostatic interactions have infinite range.[71] To handle this phenomenon, the potential is partitioned into long and short-ranged interactions. The short-ranged interactions can be truncated beyond a specific cut-off (r_c) by rendering the potential to 0 beyond such distance. The long-ranged interactions are handled by the inclusion of the lattice vector in the coulombic potential:

$$E = \frac{1}{4\pi\varepsilon_0}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{q_i q_j}{r_{ij}} + n_L\right) \tag{2.11}$$

This equation, however, converges at an exceedingly slow rate is dependent on the sequence in which terms are evaluated.

To address this challenge, the Ewald summation technique is employed which transforms the summation into two distinct series that converge far more rapidly.[72,73] This summation technique oppeates by addressing each point charge with a charge distribution, equal in magnitude and opposite in sign neutralising the primary charge in real space. Secondly, the technique counteracts the charge in reciprocal space.

A more economically viable method is the Particle-Particle-Particle-Mesh PPPM. Where, instead of indiual point charges a grid based charge distribution is utalised. The corresponding pointial and forces are then derived *via* addressing the discrete Poisson equations of this grid.[74]

2.10 | Van der Waals Interactions in Molecular Dynamics

Another key force in molecular dynamics stems from Van der Waals interaction. The Van der Waals interaction is comprised by several components: dispersion, repulsion and induction. Dispersion results from correlated spontaneous distribution of two neighbouring atoms or molecules which result in an attractive force.[75] Repulsion stems from the overlapping of electron densities resulting in a corresponding repulsive force.[76] Induction occurs due to the distortion of a neighbouring atom/molecule's electron cloud originating from a polarity present in the other atomic system. This distortion results in a temporary dipole consequently mirroring the distorting atoms' polarity causing an attractive force between the two systems.[77]

2.10.1 Modeling Van der Waals Interactions: The Lennard-Jones Potential

The Lenard-Jones (LJ) model is utilised in order to represent the Van der Waals interactions from a molecular system:[78]

$$V_{LJ}(r) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right) \tag{2.12}$$

r represents the distance between the centres of two interacting atoms. ε_{ij} describes the depth of the potential well. Finally, σ_{ij} represents the distance at which the inter-atomic potential equals zero.

The Lennard-Jones potential can accurately describe the nature of the van der Waals forces *via* the r^{-12} term representing steep repulsion at short distances and r^{-6} capturing the longer-ranged dispersion interactions.[79]

2.11 | Choice of Force Field in Molecular Dynamics Simulations

With consideration of all the equations of motion stemming from parameters defined by the force field, the choice of the force field has a great impact on the system. Considering the intricate nature of DNA, all-atom molecular dynamics often encounter difficulties in accurately modelling it's behavior.[80] DNA's unique properties and complexity necessitate the use of a specialised forceifeld.[81]

Specific atomistic forceifields designed for modelling nucleic acids emerged during the latter region of the 20th century, those being AMBER and CHARM. The AMBER forcefields emerged in the 1980s, which were consequently transformed through a series of refinements dictated through experimental observations of small olgiomernucleotides. This led to the creation of the parm94 parameter set.[82] The simultaneous development and testing of the CHARMM family of forcefields resulted in the birth of CHARMM27 and CHARMM36.[83] All force fields are meticulously refined and validated again experimental and high-level quantum mechanics, ensuring their accuracy in appropriate molecular simulations.

Recent research has validated the proficiency of AMBER forcefields, specifically the BSC0OL15 and BSC1 branches which yielded consistent results with direct NMR observabeles.[84] Additionally, the general AMBER Force Field (GAFF), has been extensively employed in accurately modelling a variety of small molecules. Research incorporating the synergistic combination of GAFF and AMBERbsc models the intercaltion into DNA with good accuracy.[53,55,85]

The complex interaction stemming from the non-covalent binding of a small molecule and DNA, present in the intercalation process, requires appropriate forcefield selection. For the forcefield to operate effectively the correct partial charges from each of the constituent atoms are required.[86–88] Due to the extensive research into DNA, the partial charges of the various nucleotides have been accurately determined and obtained from the appropriate database. For

the small molecule and abasic sites present in the duplex, quantum mechanical calculations are required.[89]

2.12 | Gaussian Software for Quantum Calculations

For quantum calculation, the Gaussian software suite is widely utilised in this field of research. The suite has proven to provide insight into electronic structure for a wide variety of molecular systems through the iterative solving of the Schrödinger equation.

2.12.1 Solving the Schrödinger Equation

The time-independent Schrödinger equation, given by:

$$\mathcal{H}\Psi = E\Psi \tag{2.13}$$

$\mathcal{H}$ is the Hamiltonian operator, Ψ is the wavefunction, and E is the energy. This equation represents a many-body problem, and therefore, challenging to solve directly for systems encapsulating multiple electrons.[90,91]

As described by Schrödinger's equation the Hamiltonian operator $\hat{H}$ is defined as:

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}} \tag{2.14}$$

m_e is the mass of an electron, m_k is the mass of nucleus k, i and j correspond to all electrons while k and l correspond to all nuclei, r_{ab} is the distance between particles a and b, Z_k describes atomic number of nucleus k. e represents the charge of an electron (1.602×10^{-19} C), and ∇^2 is the Laplacian operator.

The Gaussian suit employs a series of approximations and numerical techniques to address the challenges in solving this multi-body problem.

Born-Oppenheimer Approximation

To reduce the complexity of the Schrödinger equation, the Born-Oppenheimer approximation is applied.[92,93] Through the assumption that the nuclei remain stationary relative to the motion of the electrons, due to substantial relative mass of the nuclei by comparison. This enables the wavefucntion to be separated into respective electronic and nuclear components. This

greatly reduces the computational load, enabling Gaussian to solve the electronic Schrödinger equation with fixed nuclear coordinates

Basis Sets and Expansion

Another partitioning of the system complexity is achieved through the basis set approach. This method, uses a linear combination of atomic wave function (or atomic orbitals AOs) to describe a molecular wave function (or molecular orbitals MOs).[94] The molecular wave function ψ can be expressed as:

$$\psi = \sum_{\mu} c_{\mu} \chi_{\mu} \tag{2.15}$$

c_{μ} represents the respective coefficients to χ_{μ}: the atomic wave functions. The summation of these terms is called the linear combination of atomic orbitals (LCAO).[95]

The appropriate choice of the atomic wave functions has a great impact on the efficiency of the calculations. There exist two mathematical functions that have been frequently employed to map an atomic wave function: Slater-type orbitals (STOs) and Gaussian-type orbitals (GTOs).[96] STOs present a function which more closely resembles an accurate solution for the hydrogen atom Schrödinger equation.[97] GTOs however, do not offer the same accuracy, instead they provide a significant computational advantage. This stems from the product of multiple GTO functions remaining gaussian in nature which greatly increases the speed of computational integration. In order to mimic the behaviour of STOs multiple contracted GTOs are combined, resulting in common basis set notation like 3-21G or 6-31G.[98]

Electron Correlation

To accurately model multi-body quantum systems, electron correlation needs to be considered. Methods are implied in the Gaussian software such as Configuration Interaction (CI), Møller-Plesset (MP) perturbation theory, and Coupled Cluster (CC) theory, among others.[99]

Density Functional Theory (DFT)

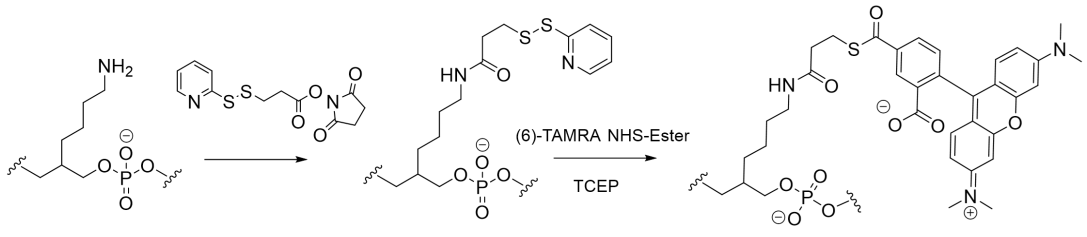
DFT is quantum mechanical approach which simplifies the Schrödinger to a three-dimensional function in order to extract properties of the system. It achieve this by using election density as its primary variable as opposed to the many-body wavefunction. Stemming from its various exchange-correlation functionals, DFT offers a balance between computational efficiency and accuracy.[100]

2.13 | Project Aim

The overall aim of this research into DTS is to develop a system capable of autonomous, sequence-specific polymerization that incorporates a wide range of building blocks. To achieve this, it is crucial to limit the hydrolysis of the monomers involved in the synthetic process. Building on the previous work by the Turberfield Group,[41] this study aims to further investigate and quantify the protective effect provided by the DNA duplex. Specifically, the focus is on understanding how the thioester-TAMRA modification and the positioning of an abasic site across the minor groove contribute to shielding the labile thioester within the DNA structure. By gaining insights into these protective dynamics, the research seeks to enhance DTS methodologies and expand the range of reactions and building blocks that can be effectively utilized in this templated synthesis approach. For example, understanding and improving the protective effect could allow the incorporation of more electrophilic substituents that are typically prone to hydrolysis, such as thioesters, selenium esters, and other such labile functional groups.[42] These groups are highly reactive but often degrade under standard conditions. Enhanced protection within the DNA duplex could stabilise these reactive intermediates, enabling their use in a broader array of chemical reactions within the DTS framework.

2.14 | Synthesis of TAMRA-Modified DNA

The initial goal was to repeat the previous experiments using a HPLC time course for more accurate quantification. Prior to the development of such an assay, the modified thioester TAMRA had to be synthesised. The general synthetic procedure involves the ligation of 3-(2-Pyridyldithio)propionic acid *N*-hydroxysuccinimide ester (SPDP) to an internally modified amine DNA, following a reduction of the disulfide bond, then another ligation to TAMRA-NHS ester (scheme 2.1).



Scheme 2.1: The synthesis of TAMRA-Modified DNA.

The internally modified amine strand was used due to its commercial availability and the large scope of reactions which document amines function as a nucleophile. Similarly, SPDP was chosen due to its commercial availability. SPDP contains an N-Hydroxysuccinimide (NHS) ester and a disulfide functionality which can be reduced at a later stage to reveal the free-thiol. The disulfide functionality was introduced with a pyridine-protecting group to prevent side reactions.

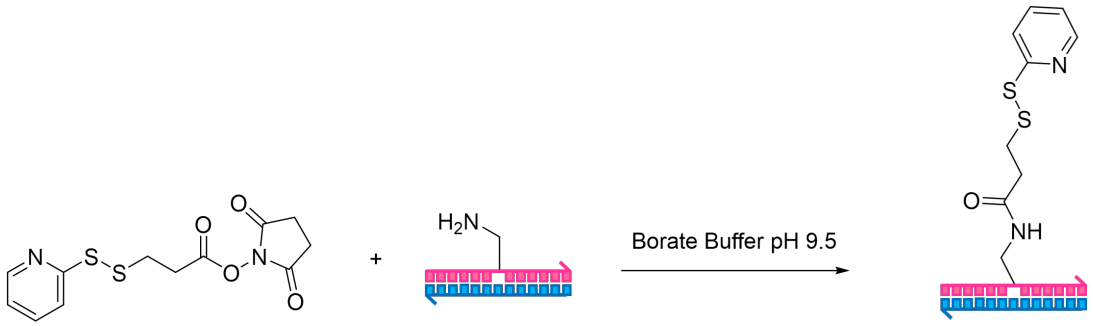
2.14.1 Disulphide Ligation to Amine-Modified DNA

The disulfide-modified DNA was synthesised in a reaction mixture containing 10% borate buffer in DMF, resulting in a final DNA concentration of 100 µM. The desired product was achieved under both pH 8.5 and pH 9.5 conditions, yielding superior results. Following the reaction, the SPDP-DNA was isolated through fractional collection. A comparison of the integrals of the relative signals from SPDP-modified-DNA and amine-DNA revealed a higher conversion using the borate buffer at pH 9.5. Specifically, the buffer at pH 9.5 achieved an 80% conversion, while the buffer at pH 8.5 reached a 72% conversion.

The identity of the SPDP-DNA was confirmed *via* LCMS (m/z: 11106.45, calculated: 11106.95).

Borate buffer is pertinent for DNA modification reactions due to its capability to maintain stable pH levels, and it is specifically conducive for reactions involving nucleophilic substitution. The stability of the pH provided by borate buffer is crucial. Borate buffer's mildly alkaline pH range (typically 8.0-9.0) enhances the nucleophilicity of reactants like amines and thiols, which are often involved in nucleophilic substitution reactions.[101] This environment not only facilitates the reaction but also prevents the degradation of DNA and other sensitive biomolecules.[102]

Borate's ability to form reversible complexes with nucleophiles, such as hydroxyl groups, can stabilise reaction intermediates, improving both the efficiency and selectivity of the modification process. This stabilization is particularly beneficial in reactions where intermediate stability is critical for achieving high yields and specificity. Borate buffer is also chemically inert towards the DNA backbone and bases, ensuring that the DNA's structural integrity is preserved throughout the reaction.[103] This compatibility is essential for maintaining the biological function of the DNA post-modification.



Scheme 2.2: The synthesis of SPDP-Modified-DNA, from amine-modified-DNA in borate buffer pH 9.5 (DNA duplex represented as a truncated sequence)

2.14.2 Reduction of SPDP modified DNA

For the reduction, previously published protocols were adapted using TCEP as the reducing agent thiol.[104–107] A ratio of 1:50 disulfide modified DNA: TCEP was used and after 1 hour at room temperature, both pH 9.5 and pH 8.5 borate buffers resulted in full conversion. TCEP was chosen due to its water solubility and specificity for disulfide reduction. The reduced product was subsequently analysed by HPLC which exhibited a clear shift in elution time, verifying the reaction success (Figure 2.7).

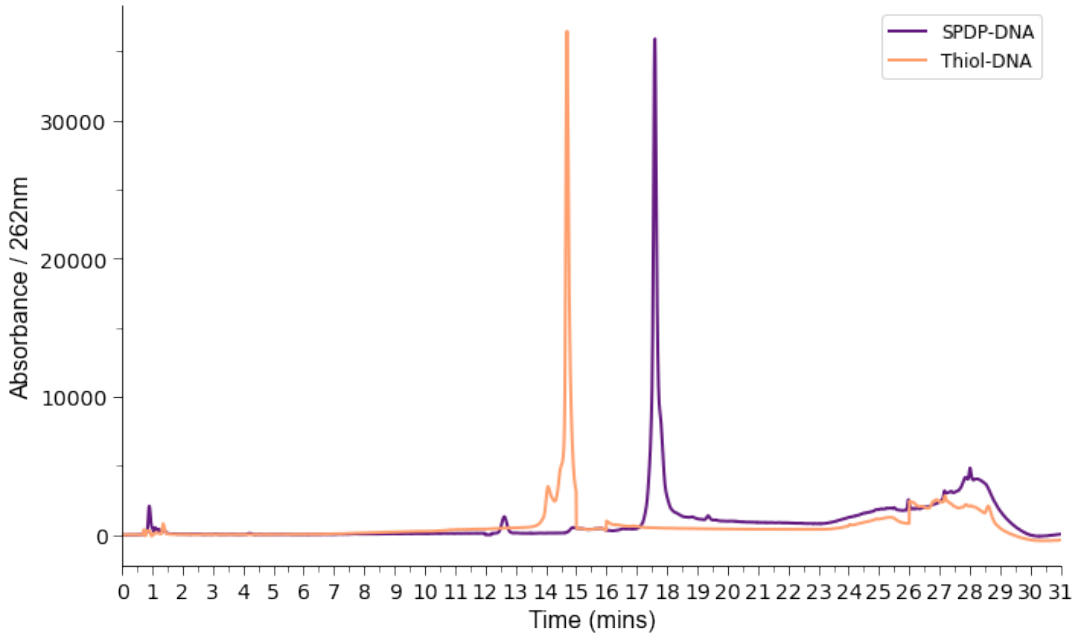


Fig. 2.7. HPLC trace showing the 262 nm absorbance channel of DNA-SPDP and DNA-thiol.

2.14.3 TAMRA Modification

With a procedure established for the instalment of the disulfide moiety onto amine-DNA, the addition of *6)-carboxy tetramethylrhodamine, succinimidyl ester* (TAMRA) was attempted

Scheme 2.1. TAMRA was used in the synthesis to confirm the results from the protection assay produced by the Turberfield group. The TAMRA-NHS was initially dissolved in DMF. The synthetic procedure was established *via* a modification of a previously documented procedure. A buffer of pH 12 resulted in the saponification of the TAMRA-NHS ester. A change in buffer conditions was therefore required. Borate Buffers of 0.1 M pH 8.5 and 9.5 were utilised for the TAMRA ligation, with the pH 9.5 buffer resulting in a higher yield of modified DNA-TAMRA (97% conversion to 85% conversion respectively).

The TAMRA-DNA formation reaction yielded three major peaks, observed through a quantitative absorbance at 262 nm *via* HPLC (Figure 2.8). Through a comparison of retention times and the relative absorbance profile of the three signals, it was determined that the peaks correspond to DNA-thiol, free-TAMRA, and TAMRA-DNA. TAMRA exhibits a strong absorption at 550 nm, while DNA exhibits its strong absorption at 262 nm. Through observation of the relative absorbance profile produced by each signal combined with control HPLC runs of each DNA-thiol and free-TAMRA, the identity of each signal was confirmed. With the LCMS of the single-strand-DNA-TAMRA corresponding to m/z: 11381.51, calculated mass: 11382.02

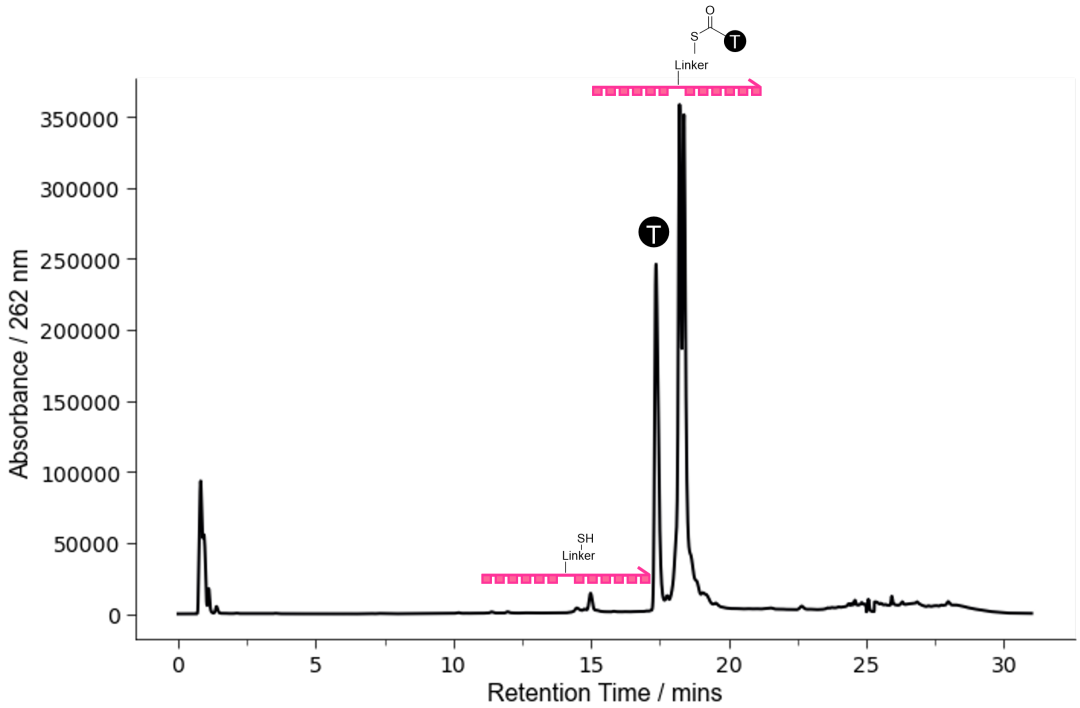


Fig. 2.8. HPLC trace showing the 262 nm absorbance channel of the reaction mixture from the TAMRA-DNA formation utilising borate buffer pH 9.5

The emergence of two signals with analogous retention times hints at the existence of two distinct conformations of the DNA-linker-TAMRA complex. It is well-established in literature

that TAMRA can exist in equilibrium between two isomeric forms: the lactone and the zwitterion.

TAMRA may undergo tautomerization to form a lactone (Scheme 2.3).[108] The transition to the zwitterionic form has been identified to be contingent on hydrogen bonding, thereby facilitating the lactone formation predominantly in aprotic solvents. However, the lactone form is non-fluorescent, and the duplication of peaks remains observable in the fluorescent spectra of TAMRA. Thus, it can be inferred that the lactone formation is not accountable for the appearance of this double peak.

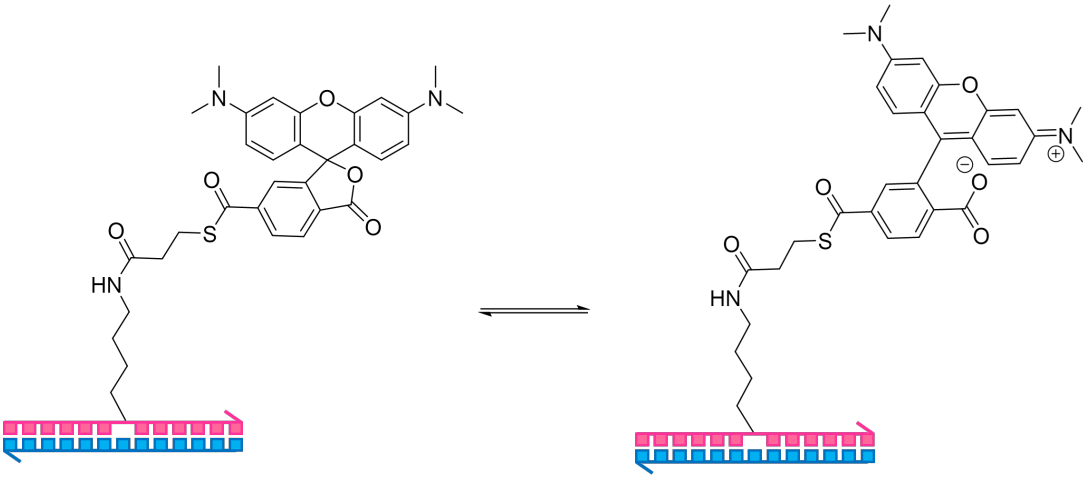
The prevailing hypothesis posits that two favourable conformations of the *ss*DNA-TAMRA may exist with differing polarities. It is plausible that an equilibrium exists between these two conformations, as single-stranded DNA (ssDNA) is known for its flexibility and ability to adopt multiple stable structures depending on its sequence, environment, and interactions with attached molecules like TAMRA.[109,110] The differing polarities of these conformations could arise from variations in the spatial orientation of the TAMRA moiety relative to the ssDNA backbone. This would likely result in differentiated elution times when passing through a reverse phase HPLC column, as each conformation interacts differently with the column matrix.

To further investigate the likelihood and nature of these conformations, kinetic experiments can be conducted by subjecting the ssDNA-TAMRA samples to HPLC at varying temperatures. By observing the relative proportion of the two signals at different temperatures, one could gain insights into the thermodynamics and stability of these conformations.[111] If the equilibrium shifts with temperature, this would indicate the presence of distinct conformations with different energy profiles.[112]

Additionally, the extraction of fractions corresponding to each signal during HPLC separation could be performed. These isolated fractions could then be subjected to circular dichroism (CD) spectroscopy or nuclear magnetic resonance (NMR) spectroscopy to determine the secondary structure and overall shape of each conformation.[113,114] CD spectroscopy, in particular, is well-suited to analyzing the helical content and conformational changes of nucleic acids, while NMR can provide detailed information about the molecular structure.[115]

In terms of literature precedence, studies on fluorescently labeled ssDNA and other oligonucleotides have demonstrated that the attachment of bulky fluorescent groups like TAMRA can indeed induce multiple conformations,[116] These conformations are influenced by factors such as the length of the ssDNA, sequence, and the presence of complementary or interacting

molecules. For instance, the introduction of fluorescent nucleobases can lead to varied stacking interactions and steric hindrances, which, in turn, create distinct conformations within the ssDNA strand.[117] The equilibrium between these conformations is likely due to the flexibility and adaptability of ssDNA, which allows it to adopt different shapes depending on the environmental conditions.[118]



Scheme 2.3: The tautomerization of TAMRA-modified DNA.

To provide a precise evaluation of the protection capabilities of the TAMRA-thioester, a time-dependent HPLC degradation assay was undertaken. The methodology for this assay drew inspiration and was adapted from the DTS protocol previously established within the Turberfield group by PhD candidate Robert Oppenheimer.[41] The revised protocol necessitated monitoring the degradation of the duplex under pH 11 conditions. The formation of this duplex was carried out before the introduction of the alkaline buffer. The duplex was formed with addition of magnesium chloride to ensure its stable formation.[119,120] Following the buffer's addition, the duplex's final concentration was adjusted to one micromolar. This duplex was then subjected to 48-hour HPLC regime, with samples extracted at various intervals for analysis *via* the TAMRA fluorescence channel (λ_{ex}: 530nm, λ_{em}: 580nm)(Figure 2.9).

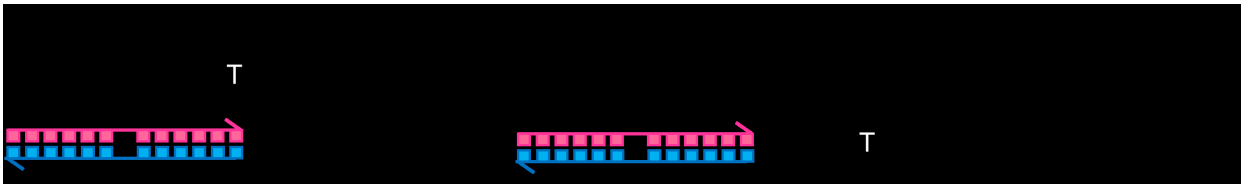


Fig. 2.9. Schematic illustration of the protection assay. The duplex, modified with TAMRA-thioester, features a variable abasic site on its complementary strand. When exposed to pH 11 conditions, the result is thiol-modified DNA along with free TAMRA. In this depiction, the position 0 protective strand is highlighted. Squares symbolize nucleotides, while absent squares indicate abasic sites.

Various HPLC protocols were explored to achieve a clear time course illustrating the degradation of the thioester TAMRA duplex. The objective was to ensure the distinct separation of signals without any observed shifts in their elution. The calculation for the degree of protection was determined by comparing the relevant integrals of the signals visible in the TAMRA fluorescent channel (λ_{ex}: 530nm, λ_{em}: 580nm), as shown in Equation 2.16.

$$\text{Degree of unhydrolysed TAMRA-DNA} = \frac{\text{Signal of TAMRA-Duplex}}{\text{Total TAMRA signals}} \tag{2.16}$$

Degree of unhydrolysed TAMRA-DNA

Equation 2.16. In this equation, the "Signal" refers to the integral of the respective HPLC peaks. The total TAMRA signals are constituted from both the TAMRA DNA duplex and free TAMRA signals.

The resulting relative degree of protection as calculated *via* Equation 2.16 aligned with the trends observed by Robert Oppenheimer. With the general trend demonstrates increased

protection with the abasic site position over the minor groove. The extent of protection is evident when observing the resulting chromatograms resulting from the protection assay concerning the -3 abasic site in comparison to the fully complementary strand (Figures 2.11 to 2.15).

Discrepancies in the elution time of the DNA-thioester-TAMRA/ -3 duplex and the DNA-thioester-TAMRA/ fully complementary duplex were noted. Such variations can be attributed to the TAMRA-linker assuming different conformations within distinct duplexes. These potentially differing binding conformations will result in varying polarities of the duplexes.

From the analysis of the results, it was evident that an enhanced protective effect existed corresponding to abasic strands with positions $-2 \leq x \leq -4$. Apart from the position 0 strands, protection for all other positions approximated 35% (Figure 2.15). The error bars presented in the figures represent the standard deviation of results from experiments performed in triplicate, providing a measure of the reproducibility of the observed effects. Fortunately, the error associated with the measurement of DNA-TAMRA fluorescence intensity is mitigated by the method of calculating the degree of protection as a relative comparison. As the fluorescence intensity is calculated *via* a relative comparison between intact DNA-TAMRA with that of the degradation product (free TAMRA), any variations in the total amount of material, such as differences in injection volume or concentration, are effectively normalised. As a result, the conclusions drawn about the protective efficacy of different abasic site positions are robust, and the standard deviation accurately reflects the variability in the protective effect rather than measurement inconsistencies.

Interestingly, the abasic 0 strands showcased the least protection, suggesting that the extensive cavity in the DNA exposes the TAMRA thioester to the aqueous environment. This result seems to contradict the general hypothesis, which posits that the distance and accessibility to the abasic site are proportional to the degree of protection; therefore, the 0 position should theoretically offer some level of protection. However, it is important to consider that when the abasic site is at the 0 position, it is directly opposite the abasic site where the linker is attached. This positioning likely creates a larger cavity, further exposing the TAMRA thioester to hydrolysis by H_2O.

Two abasic sites directly opposite each other may reduce the overall stability of the DNA structure at this location. The lack of stabilizing interactions such as van der Waals forces and hydrogen bonding in this region could leave the TAMRA thioester more vulnerable to the aqueous environment, leading to a higher rate of hydrolysis. This could explain why the abasic

0 strands offer the least protection, despite the expectation that closer proximity to the abasic site would enhance protection.

However, it should be noted that no repeats were completed for the 0 abasic position, and this result was not confirmed by Dr. Jennifer Frommer(Figure 2.16) The results presented in the graph are based on the average of three independent experiments.[43] In light of such further experimentation the results from the protection assay concerning position 0 were removed from the figure (Figures 2.15).

The graph in Figure2.16 of Frommer's extension of my study confirms my findings regarding the protective effect of abasic sites, with the exception of the 0 position.Her data illustrate a clear relationship between the presence of a minor groove abasic site and enhanced protection, particularly confirming the superiority of the abasic site at position -3 in providing maximal protection.

The results indicate little difference in protection between the ribose and purely amine spacers (Figure 2.16 **a**), as observed in the graph, supporting the hypothesis that the internal abasic site, rather than the nature of the linker, plays the critical role in shielding the thioester from hydrolysis. These findings highlight the robustness of the protective effect across various configurations, reinforcing the potential for using abasic sites to stabilise reactive intermediates in DNA-templated synthesis.

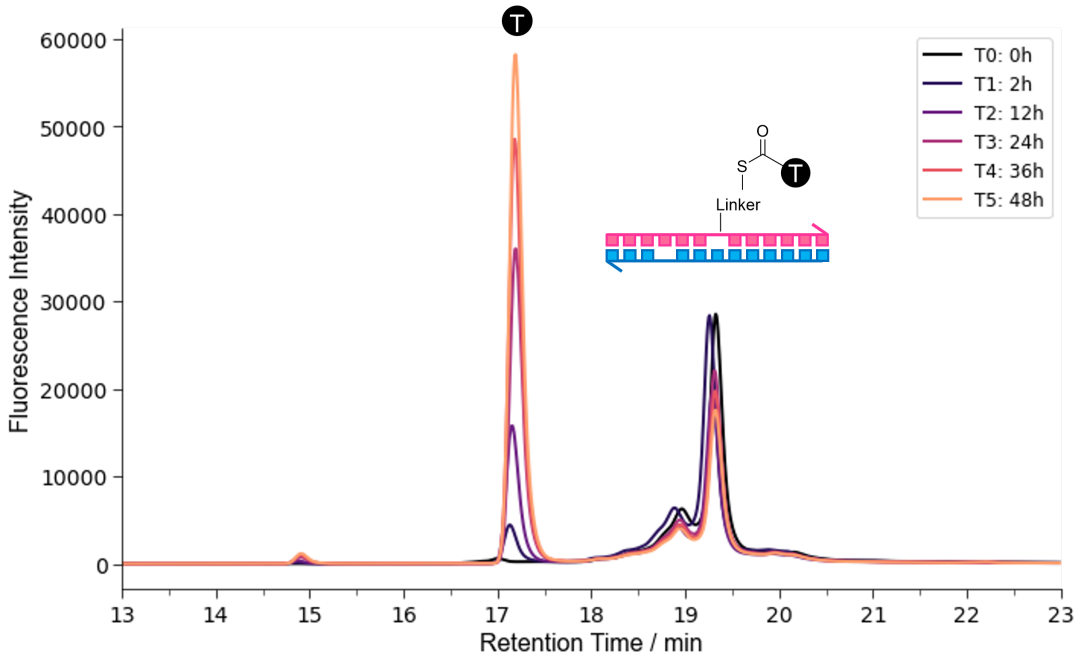


Fig. 2.11. Time course from the protection assay for the fully complementary duplex, as visualized in the TAMRA fluorescence channel. The respective signals correspond to free TAMRA and the TAMRA-modified duplex, each illustrated with an overhead cartoon representation. A 2X excess of the complementary strand was present in the protection assay; however, it does not appear in the HPLC data since the analysis was conducted using the TAMRA fluorescence channel, to which the unlabeled complementary strand does not contribute.

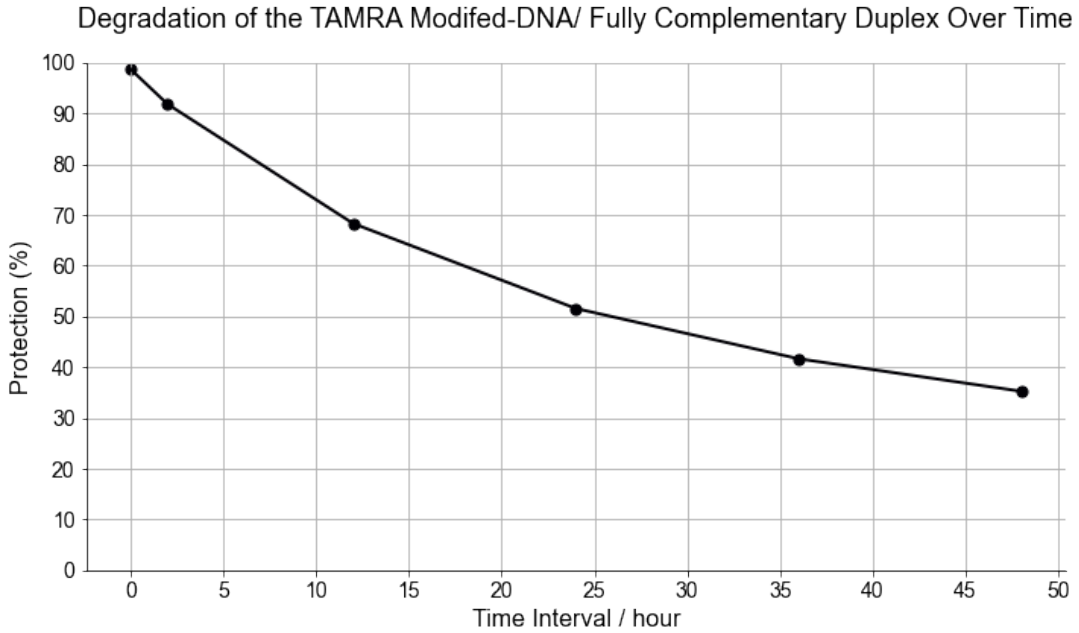


Fig. 2.12. Evolution of protection degree for the fully complementary duplex with TAMRA-modified-DNA over the time course, determined using Equation 2.16. The protection level at T_5 is measured at 35%.

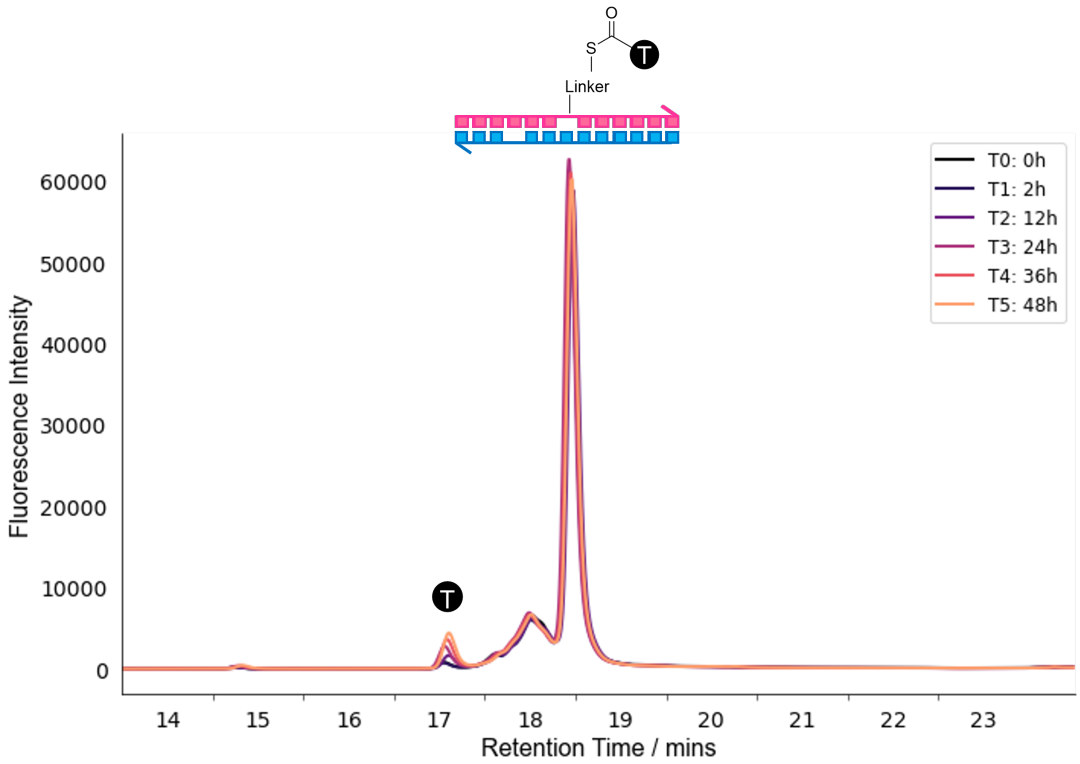


Fig. 2.13. Time course from the protection assay for the -3 abasic duplex, as visualized in the TAMRA fluorescence channel. The respective signals correspond to free TAMRA and the TAMRA-modified duplex, each illustrated with an overhead cartoon representation. A 2X excess of the abasic DNA strand was present in the protection assay; however, it does not appear in the HPLC data as the analysis was conducted using the TAMRA fluorescence channel, to which the unlabeled abasic strand does not contribute.

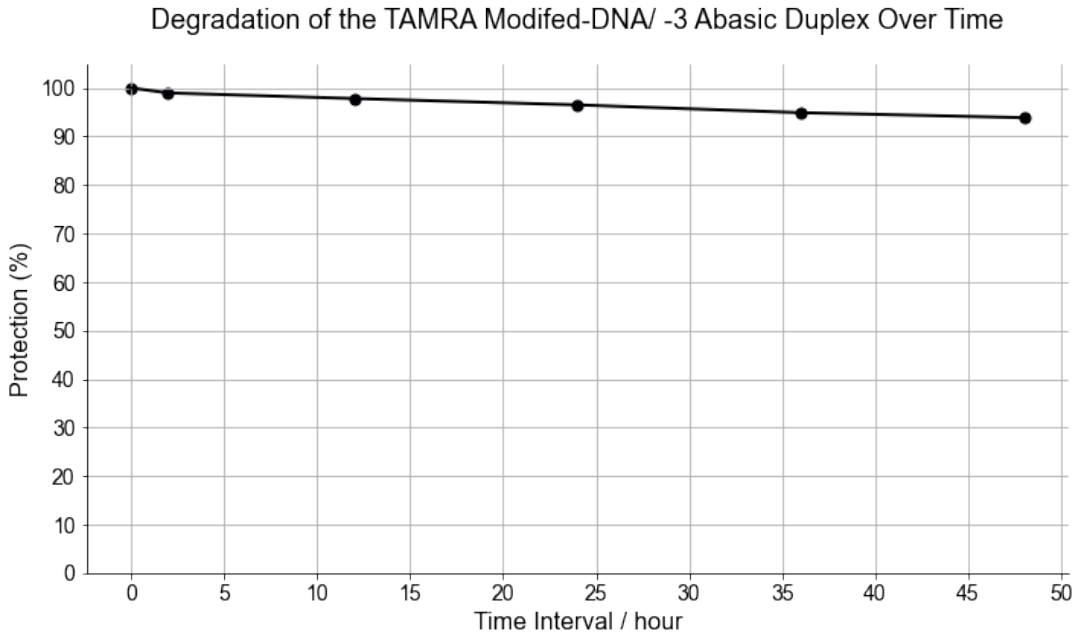


Fig. 2.14. Evolution of protection degree for the -3 abasic duplex over the time course, determined using Equation 2.16. The protection level at T_5 is measured at 35%.

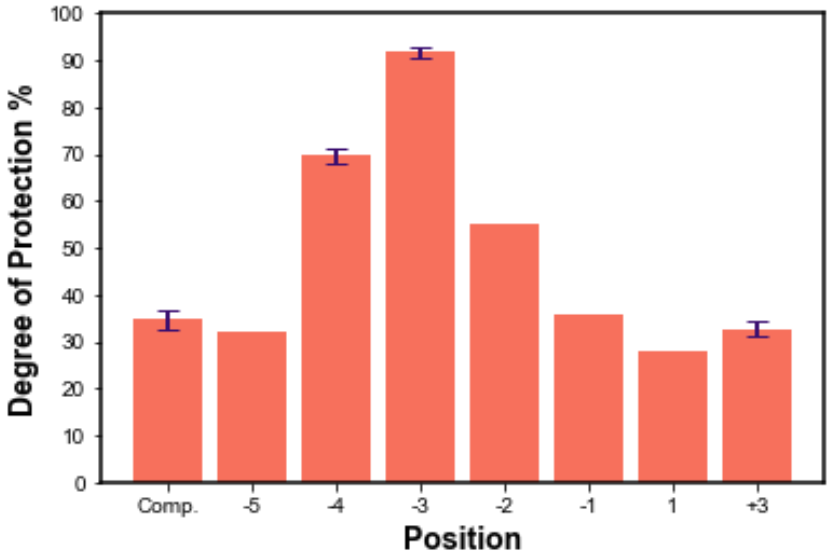


Fig. 2.15. Degree of Protection for all assays completed at the 48-hour time point (T5), determined using Equation 2.16. Concerning the positions -3, +3, +4 and the fully complementary region ('Comp.'), data were averaged from three independent repeats of each assay, with standard deviations (SD) calculated to reflect the variability in the results.

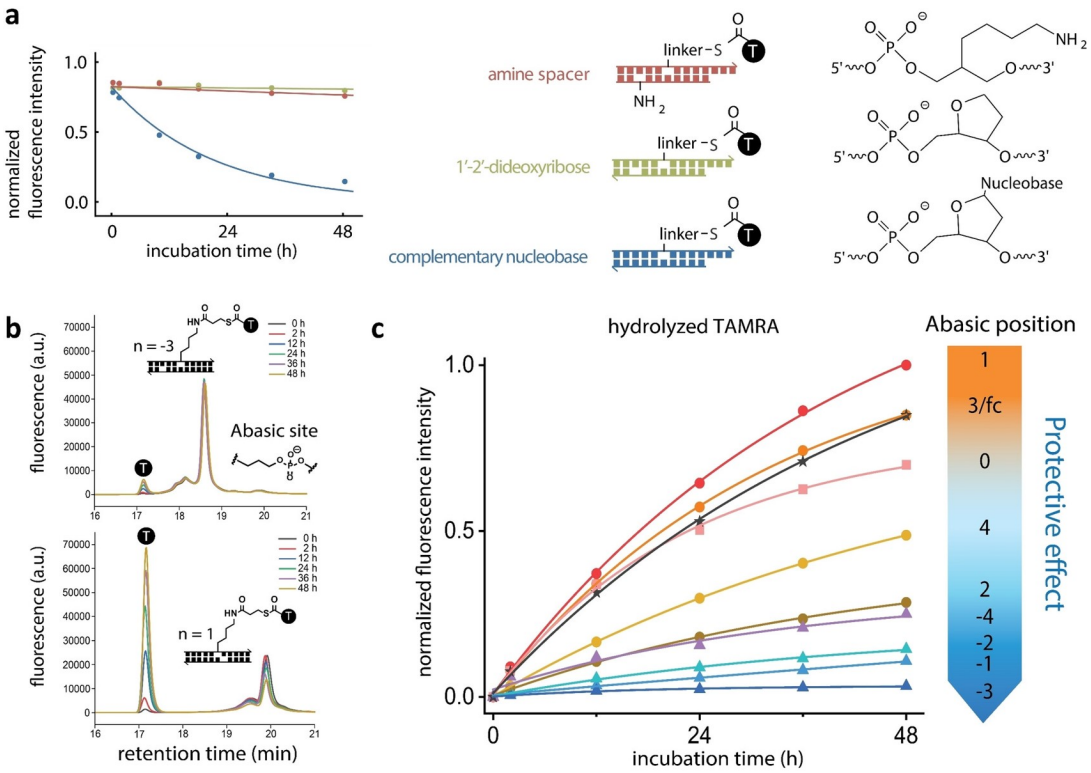


Fig. 2.16. Protection of a TAMRA thioester by an abasic site in the complementary strand. Hydrolysis of the thioester was measured by RP-HPLC after incubation at 5°C, pH11 for 0–48h. TAMRA fluorescence was measured with ex:550nm and em:580nm. a) Protective effects of each of the three modifications shown at the n=_3 position in the complementary strand. The proportion of unhydrolyzed thioester was calculated by comparing the areas under the RP-HPLC peaks corresponding to the DNA-thioester and free TAMRA. b) Comparison of protection afforded by complementary strands with the abasic site (C3-spacer) located in the _3 position (maximum protection, top) and in the 1 position (minimum protection, bottom). c) Position-dependence (n= 4 to _4) of the protective effect of an abasic site (C3-spacer) monitored by RP-HPLC time course using the integrated fluorescence intensity of hydrolyzed TAMRA as an indicator. A fully complementary opposite strand (fc) was used as a control.

2.15 | Results and Discussion

The results from the time course paralleled the protective effects previously observed by Robert Oppenheimer. A distinctive protective effect emerged when the abasic protecting strand had an abasic site within the region of the minor grooves. This can be discerned by evaluating the signal intensity of the free TAMRA in chromatograms from protection assays with -3 and fully complementary hybridising strands (Figures 2.11 to 2.14) . The differences in signal intensity and retention times between the TAMRA-duplexes of -3 and the fully complementary duplexes suggest varied conformations of the TAMRA-linker among duplexes.

These different binding conformations both affect the retention time in the HPLC and the quantum yield of TAMRA.

The quantum yield of TAMRA is known to vary significantly depending on its interaction with nearby nucleobases, which can quench or enhance fluorescence through mechanisms such as Förster resonance energy transfer (FRET) or static quenching.[121] This phenomenon has been observed in other studies where the fluorescence intensity of similar dyes is directly influenced by the DNA sequence and structural context in which they are situated.[122]

Enhanced protection was evident for abasic strands between positions -2 and -4. However, protections for other positions, excluding the 0 strand, hovered around 35% (see Table 2.15). Intriguingly, the abasic 0 strand showed the least protection, potentially due to conformational changes from the two opposing abasic sites, exposing the TAMRA-thioester more to the aqueous environment, however, further investigation and repeats are required. This protective effect, therefore, must stem from the TAMRA's interaction with an abasic site over the minor groove. This observation aligns with the known affinity of smaller molecules for the minor groove, coupled with the proximity to the -4 and -3 positions. It is possible that TAMRA's preferential minor groove binding lowers the activation energy required for insertion into abasic sites. This then enables the unique protection of the thioester to manifest.

2.15.1 The Influence of the Abasic Site Dependent TAMRA Interaction on the DNA Duplex's Hydrophobicity

Variations in retention times of TAMRA-modified duplexes were noted during degradation assays. Given that a reverse-phase column was used for these assays, species with increased hydrophobicity had therefore extended elution times. Given TAMRA's hydrophobic nature, one would expect decreased hydrophobicity if TAMRA were integrated into the duplex. Following this reasoning, a protective interaction between TAMRA and a minor groove abasic site should result in a duplex with reduced hydrophobicity and quicker elution. This was indeed confirmed during the degradation assays, where the -3 abasic duplex displayed a notably earlier elution compared to the +3 and fully complementary duplexes (Figure 2.17).

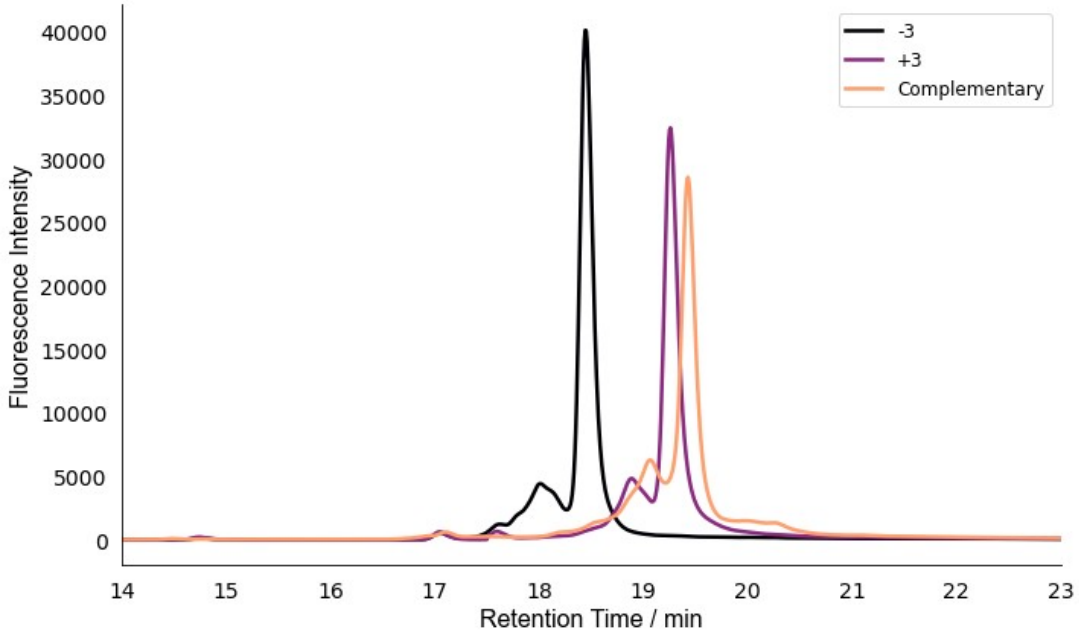
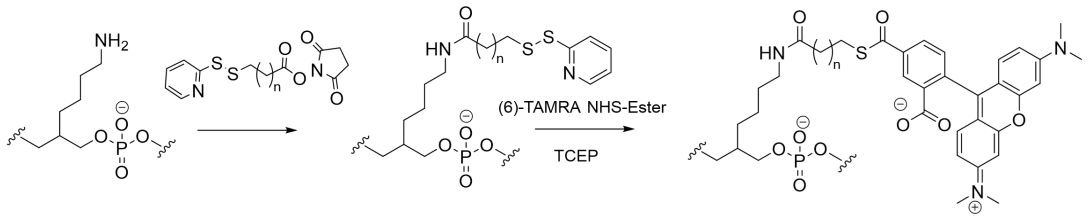


Fig. 2.17. Superimposed HPLC traces from the protection assay for the -3, +3, and fully complementary duplexes with TAMRA-modified DNA showing each duplex at T0: 0h.

In order to confirm this change in elution time stems from the interaction of TAMRA with the abasic site, a control was executed. This control involved analysis of the elution time produced by the -3, +3, abasic strands and fully complementary strands hybridised to the DNA-linker-thiol. This consequently formed the same duplexes without the TAMRA component. The result of this was that all duplexes from this control occurred at the same elution time, indicating the distinct varied effect the TAMRA had on the polarity for the various duplexes. Specifically, the -3 abasic duplex caused a shift from the thiol -3 basic duplex elution time by +0.4 mins, +3 abasic resulted in a shift of +1.2 mins, and the complementary duplex complementary showed a shift of +1.4 mins.

To investigate whether this observed protective effect is specific to the minor groove or merely a consequence of the linker length, we synthesised and ligated a longer SPDP-linker to the amine-modified DNA strand. This linker extended the reach of the TAMRA unit (Scheme 2.4). Through this modification, we aimed to discern if the protective effect was genuinely a function of the minor groove's inherent properties or if it was simply a result of the accessibility of the abasic site.



Scheme 2.4: The synthesis of TAMRA-Modified DNA, where n =4

With the invaluable assistance of Dr. Jennifer Frommer in synthesising the elongated SPDP-NHS ester, the extended TAMRA-DNA was produced. The synthetic steps used were analogous to the previous stems utilised to install the shorter TAMRA linker. Following the successful synthesis and purification of the extended TAMRA-DNA (m/z = 11423.501, calculated: 11424.07), these samples underwent the degradation assays, as shown in Figure 2.18 to 2.21 and summarised in 2.22. The protection assay results for the lengthier linker indicate that the protective effect manifests irrespective of the existence or position of an abasic site.

In tests involving the +3, -3, and fully complementary strands, the degree of protection (determined using Equation 2.16) consistently remained above 77% across all instances. The fact that protection arose with the fully complementary strand suggests that a protective conformation was enabled without the need for an abasic site. However, It's imperative to highlight that the greatest degree of protection was observed in the duplex featuring the -3 abasic site. However, it is important to consider the positioning of the abasic sites in relation to the thioester-TAMRA complex. As illustrated in Fig. 2.6, the -3 abasic site is closest to the thioester-TAMRA complex, while the +3 site is three nucleobases farther away, with the fully complementary strand serving as a control without an abasic site. This range of distances provides insight into how proximity may influence the degree of protection.

The selection of the -3 and +3 positions was deliberate to compare the abasic site that offered the best protection (-3) with a site the same number of base pairs away in the opposite direction (+3). This choice allows us to determine if the differential protective effect between the minor and major groove is prevalent with a linker of longer length. The observation that the -3 position offers the greatest protection suggests that the minor grove character of the abasic site enhances protection, even when the linker has greater flexibility and length. For a clearer understanding, the inclusion of more abasic positions would be necessary for a comprehensive study to fully understand how groove orientation and distance together influence protection. Unfortunately, material constraints and the complexity of synthesising these modified strands limit the ability to test every possible position. The synthesis of these

materials is not only expensive but also time-consuming, making it challenging to explore every abasic site thoroughly. Despite these limitations, the results obtained from the selected positions provide valuable insights into the protective dynamics at play, even if they are not exhaustive.

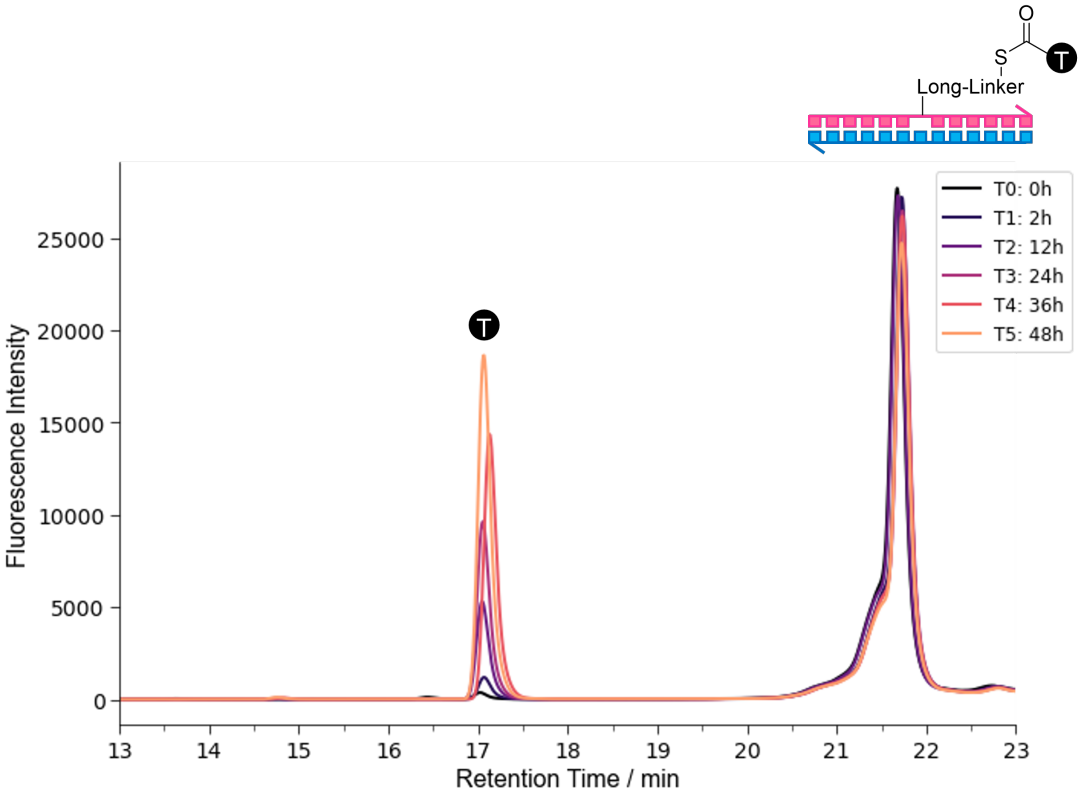


Fig. 2.18. Time course from the protection assay for the fully complementary duplex with the Long-TAMRA modified DNA, as visualised in the TAMRA fluorescence channel. The respective signals correspond to free TAMRA and the TAMRA-modified duplex, each illustrated with an overhead cartoon representation.

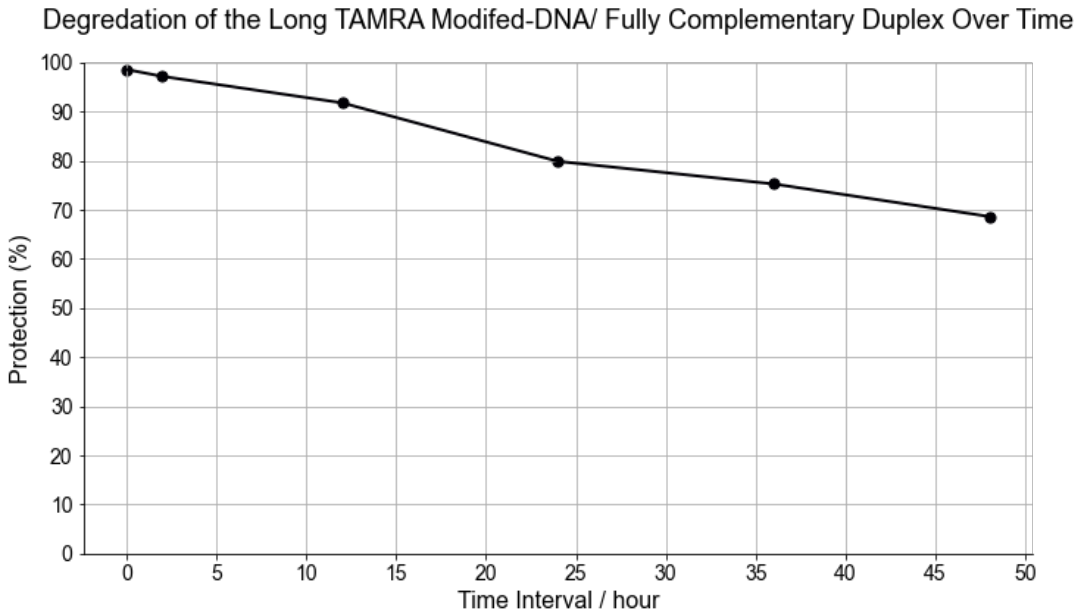


Fig. 2.19. Evolution of protection degree for the fully complementary duplex with the Long-TAMRA modified DNA over the time course, determined using Equation 2.16. The protection level at T_5 is measured at 35%.

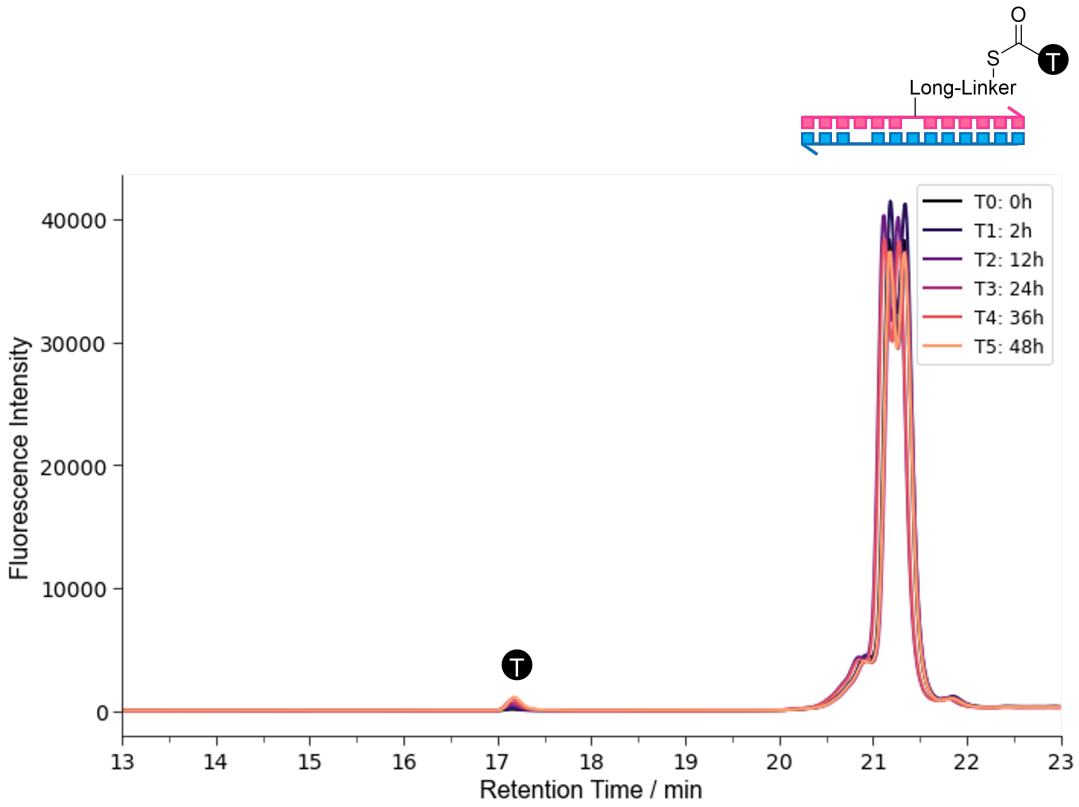


Fig. 2.20. Time course from the protection assay for the -3 abasic duplex with the Long-TAMRA modified DNA, as visualised in the TAMRA fluorescence channel. The respective signals correspond to free TAMRA and the TAMRA-modified duplex, each illustrated with an overhead cartoon representation.

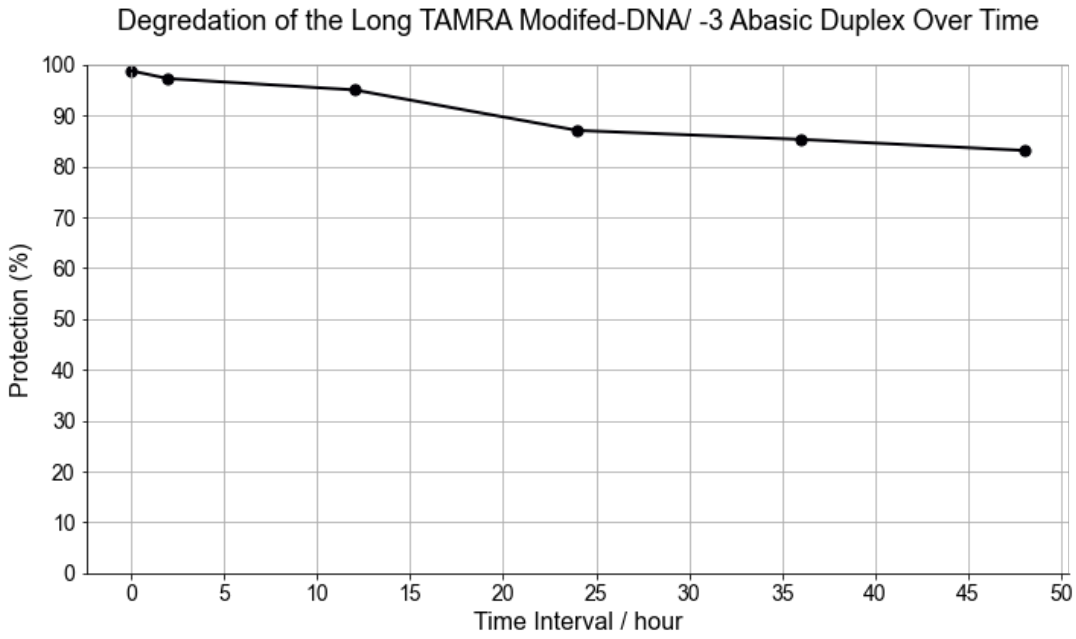


Fig. 2.21. Evolution of protection degree for the -3 abasic duplex with the Long-TAMRA modified DNA over the time course, determined using Equation 2.16. The protection level at T_5 is measured at 35%.

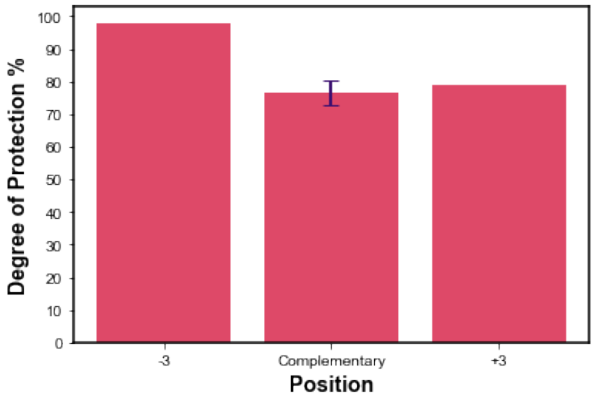


Fig. 2.22. Degree of Protection for all assays completed with the longer Linker for time point at 48h T5, determined using Equation 2.16.Standard deviation derived from 3 independent results concerning the assay involving the complementary hybridising strand.

Through observations of the resultant elution times stemming from the duplexes with the longer linker, it is evident that the difference in elution between duplexes is less than that of the difference observed with the smaller linker. This may be suggestive of a similar binding configuration of the long-TAMRA DNA for the various duplexes. It was observed that two signals emerged from the chromatogram involving the long-TAMRA-DNA -3 duplex. These two signals intersected due to their close retention time. It is again possible that these two signals are symptomatic of two different binding configurations present in the long-TAMRA DNA/-3 duplex. As discussed earlier with the shorter linker, if these signals represent different

conformations, they are likely in equilibrium and interchangeable. To further investigate this, similar thermodynamic experiments could be conducted by isolating each signal and reinjecting the samples into the HPLC. This would help determine if the signals are indeed due to distinct conformations and if they are capable of interconverting. Additionally, structural determination techniques such as circular dichroism (CD) spectroscopy or nuclear magnetic resonance (NMR) could be employed to better understand the nature of each conformation.

2.15.2 Implementation of the protective effect into DTS

The strategy of protecting a labile thioester from hydrolysis by incorporating a minor groove -3 abasic site was applied to a DNA-templated synthesis (DTS) reaction by Dr. Jennifer Frommer (Figure 2.26). Two reaction systems were compared: both utilized a DNA-thioester-TAMRA duplex, with one system including the -3 abasic protective site and the other lacking it. Both systems were incubated for 48 hours before initiating a toehold-mediated strand displacement with a DNA strand containing a reactive amine, thereby enabling DTS. The relative yields of the DTS reactions were then compared between the two systems. To accurately quantify the rate of hydrolysis, fluorescent calibration curves for all TAMRA-containing species were obtained. We collaborated on applying the equations derived from these calibration curves to the experimental results.

It is essential to understand that the fluorescence response of TAMRA is highly dependent on its molecular environment. Several studies have observed that the quantum yield of fluorescent dyes tends to diminish when integrated into a DNA system.[123,124] This effect was evident in our degradation assay, where the fluorescence intensity of free TAMRA was considerably higher than that of the TAMRA-modified duplex. Such disparities in the quantum yields between free and duplex-bound TAMRA can lead to inaccuracies if fluorescence intensities are compared directly, potentially exaggerating differences in the degree of protection across the tested duplexes.

To address this issue and achieve quantitatively accurate measurements of the rate of degradation, we generated calibration curves specific to each TAMRA species in the system (Figures: 2.23 - 2.25). By measuring the fluorescence intensities of known concentrations of free TAMRA and TAMRA-modified duplexes, we established calibration curves that allowed us to convert fluorescence readings into accurate concentrations for each species. This normalisation accounted for the differing quantum yields and ensured that the calculated rates of hydrolysis and protection were precise.

The results, depicted in Figure 2.26, clearly demonstrate that the presence of the abasic site significantly improved the yield of the desired amide product. Specifically, with the protective site, the yield of the DTS reaction became largely independent of the incubation time, while without the protection, hydrolysis of the thioester reduced the overall yield. This highlights the critical role of the abasic site in maintaining the integrity of the thioester, thus facilitating a more efficient and productive reaction.

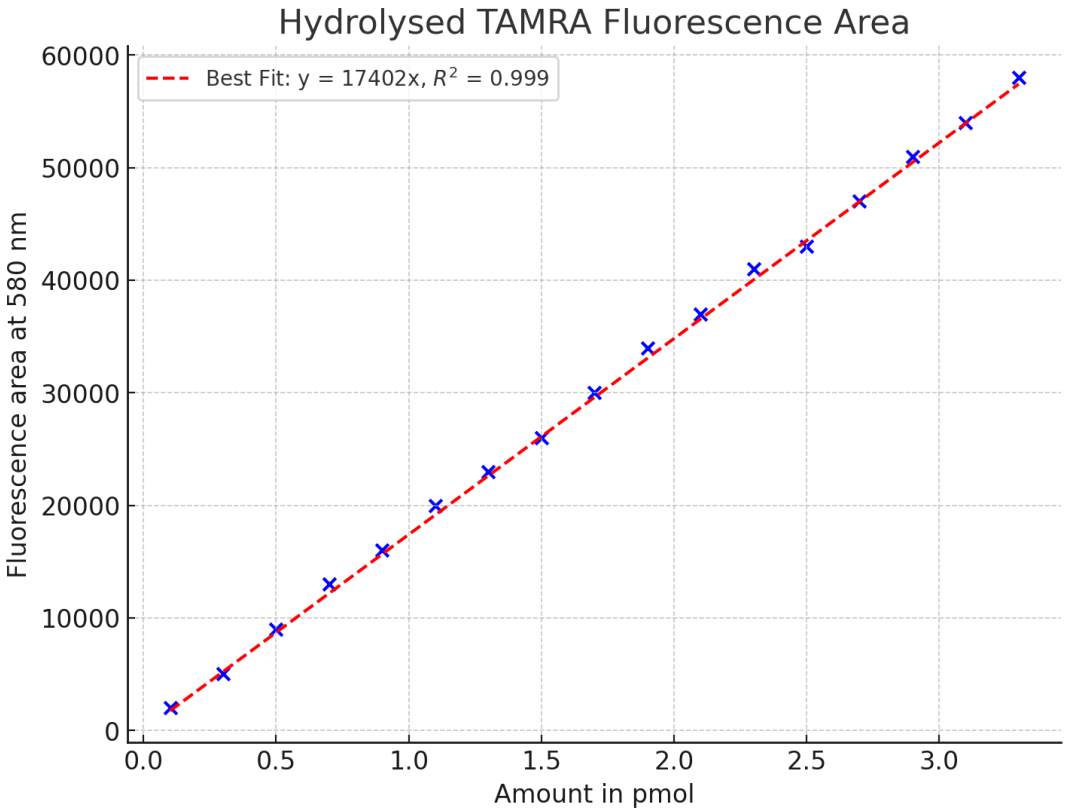


Fig. 2.23. The result of the TAMRA fluorescence calibration as determined by integrating the signal intensity produced at each concentration of free TAMRA by HPLC.

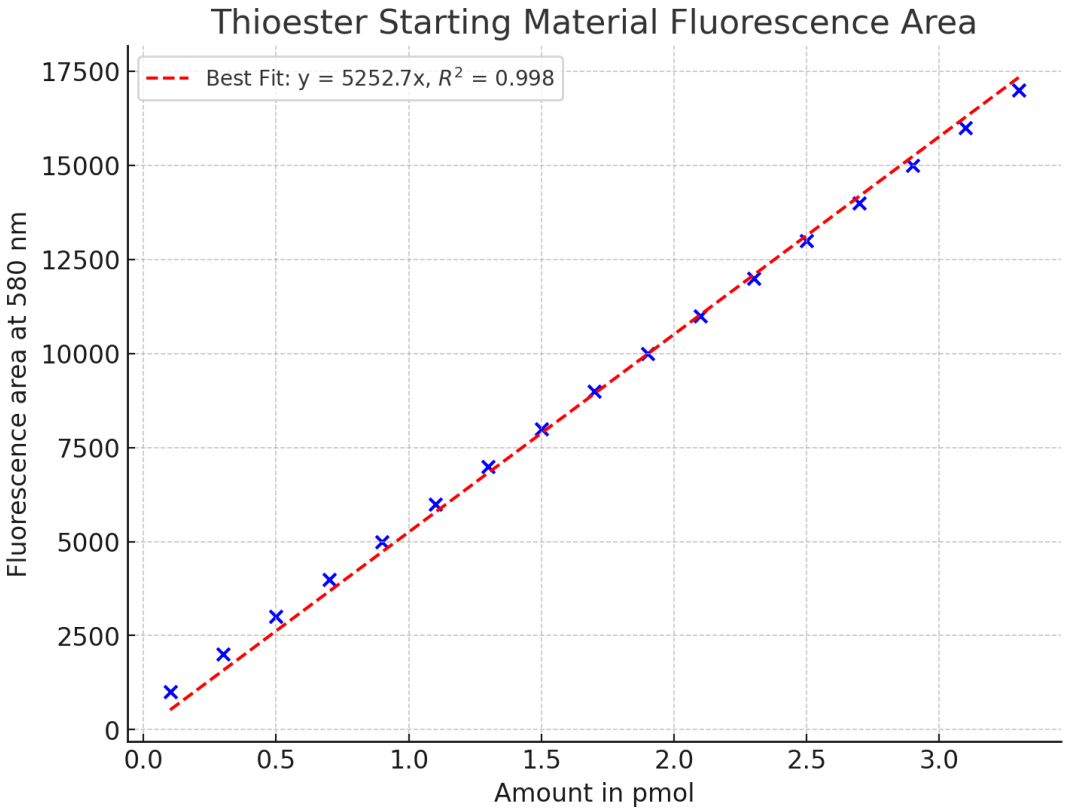


Fig. 2.24. The result of the TAMRA flourescence calibration as determined by integrating the signal intensity produced at each concentration of DNA-thioester-TAMRA by HPLC.

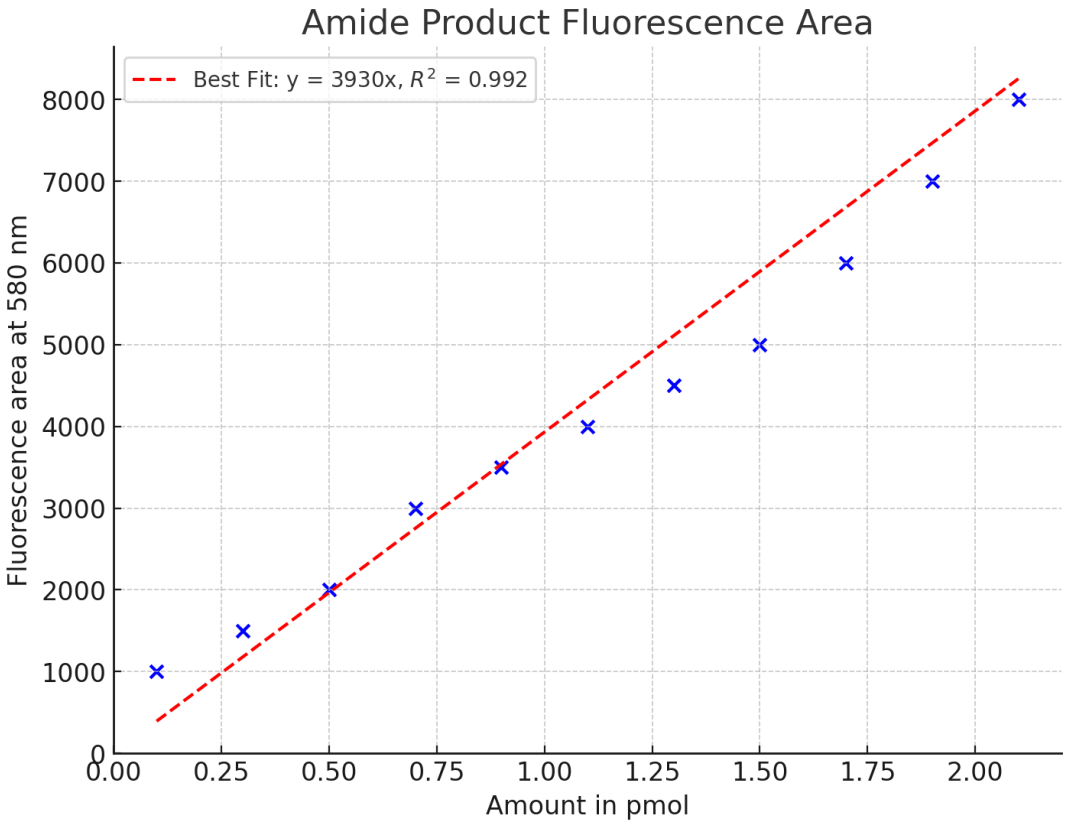


Fig. 2.25. The result of the TAMRA calibration as determined by integrating the signal intensity produced at each concentration of the amide product by HPLC.

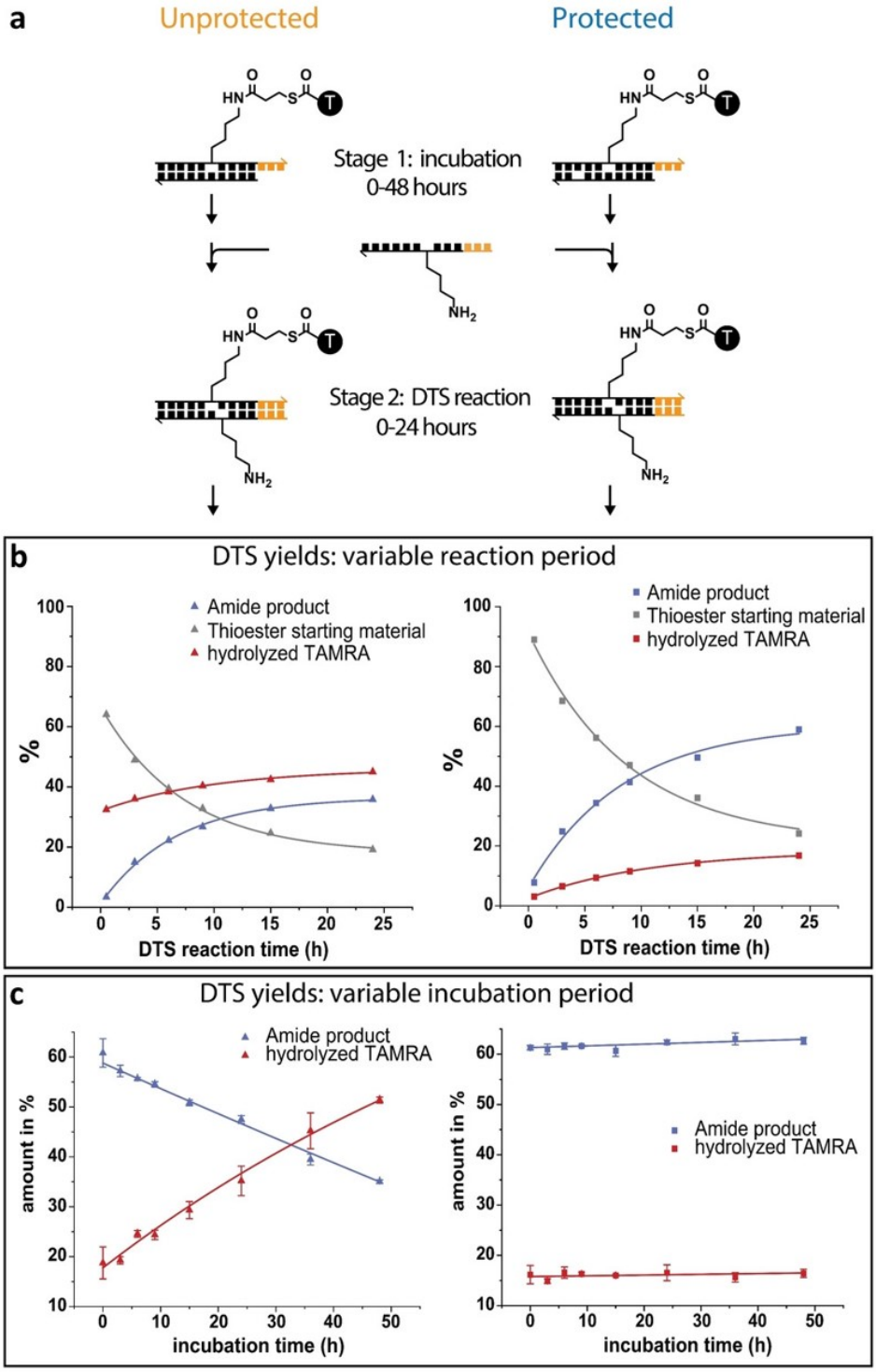


Fig. 2.26. a) Stage 1 (incubation). Donor adapters, hybridized to a complementary keeper strand with or without an abasic protective site at position n=-3, were incubated at pH11 at 5°C for between 0 and 48h. Stage 2 (reaction). Reactions were initiated using toehold-mediated strand exchange to replace the keeper strand with an acceptor strand forming a duplex with an amine acceptor at the n=1 position. DTS was performed at 15°C for 24hours. b) DTS reaction progress for the unprotected and protected systems as functions of reaction time after 48h incubation at stage 1. DTS reaction progression was assessed by RP-HPLC using the TAMRA fluorescence channel (ex:550nm and em:580nm) (FigureS9). c) As b) but with fixed 24h reaction time and with variable incubation time. DTS reactions were performed in triplicate: error bars represent the standard deviation of the yield.

Implementation of the Protective Effect into DTS

In the degradation assay with an extended linker, it was observed that significant protection was consistently achieved regardless of the abasic site's position or existence within the sequence. This consistency presents potential challenges concerning its integration into a DTS mechanism. Robbie's research demonstrates a fundamental trade-off between protection and reactivity in DTS systems. His findings show that while the protective mechanism is active, no acyl transfer can occur, making protection and reactivity mutually exclusive. For the protective effect to function as intended, the protected labile linker must be selectively exposed to the reactant when needed. As well as designing a DTS mechanism to selectively expose the reacting electrophile when needed, a bi-functional protecting group is required to render this mechanism viable. Such bi-functionality ensures that while the group offers protection, the preceding acyl transfer won't result in the protecting group attaching to the growing polymer. This protecting component must then be exposed from the abasic site to facilitate the transfer reaction (Fig. 2.27).

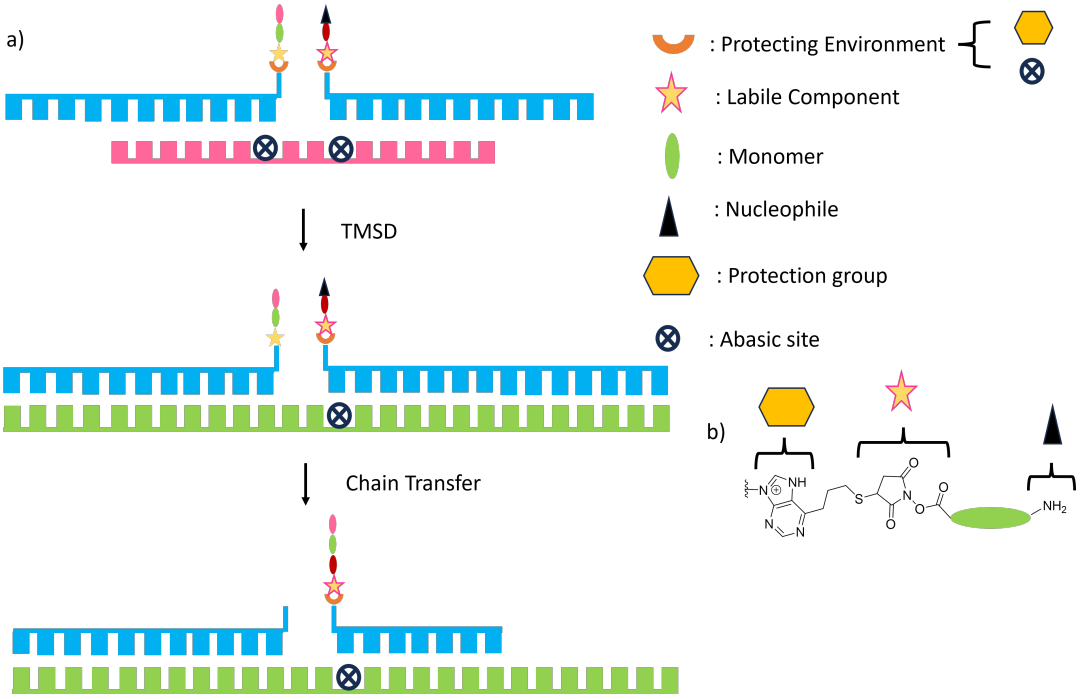


Fig. 2.27. This cartoon illustrates the incorporation of a protective effect within a dynamic template strand displacement (TMSD) mechanism. The figure shows both linkers extending from nucleobases that oppose abasic sites, which are hosting a protective environment—realized by a protecting group and an abasic site. The mechanism involves labile linkers, monomer(s), and a linker containing a nucleophile. In the simplified cartoon (a), TMSD is initiated by a longer templating strand. The incoming strand contains a nucleobase opposing the longer polymer chain, which forces the electrophilic polymer chain out into the solution, enabling the nucleophile to engage in chain transfer. (b) depicts a chemical schematic of a bifunctional protecting group that can be implemented at the forefront of the chain, allowing the terminal nucleophile to participate in chain transfer.

2.16 | Methods: Molecular Dynamics Simulation Preparation

In order to elucidate the structure-activity relationship discovered from the experimental strand, molecular dynamics simulations were employed to further study this system.

2.16.1 TAMRA Linker Topology Generation

The topology file for the TAMRA linker was created through the use of the ACYPYPE tool, in conjunction with the GAFF force field. Formal charges were assigned to the atomistic components of the TAMRA linker through a high-level DFT calculation in Gaussian, using the 6-31++G** basis set and the B3LYP functional.

2.16.2 DNA Duplex Topology Generation

The formation of the DNA duplex topology file was completed through the PDB2GMX command from the GROMACS software suite. For its operation, the program required input of co-ordinates and an appropriate force field. Co-ordinates were derived from a DNA database, which described a truncated 30 base-pair sequence. This truncation from 45 base pairs was strategically chosen to reduce computational demand through the exclusion of 'extra' base pairs which were deemed to have negligible impact to the TAMRA-linker abasic site interaction. For the forcefield, the parmBSC2 was chosen.

2.16.3 Introduction of Abasic Site

After the creation of the DNA and TAMRA topology files, two abasic sites were introduced into the dulex. One abasic site for the attachment of the TAMRA-linker and one for the +3 and -3 abasic positions. These sites were created by editing the topology file in order to remove the necessary nucleobases. A custom python script was developed to handle the deletion of appropriate atoms and rendering of others to hydrogens in order to describe the appropriate abasic site and the appropriate positions. The script also had to handle the subsequent renumbering of the atoms across the various topology sections: atoms, bonds, pairs, angles and dihedrals to ensure an accurate description of the DNA system. Appropriate atomic force field parameters were used for atoms which were rendered to hydrogens during the process.

2.16.4 Topology Concatenation and Bond Addition

The relevant connecting bond between the TAMRA-linker and the DNA was manually described in the concatenated topology file describing both the DNA and TAMRA-linker with their respective force fields. Bonding constants for this bond were derived from an additional topology file created *via* ACYPEPE calculation with the GAFF forcefield. Subsequently, the appropriate co-ordinate file of the DNA-linker-TAMRA duplex was created.

The addition topology file was also used to extract all relevant pair, angle, and dihedral interactions present in the linker-DNA connection. All appropriate interactions were input into the contacted topology file calling upon the parmbBSC forcefield for the DNA and the gaff force-field describing the linker-TAMRA components.

2.16.5 Gaussian Scans for Abasic Sites

For an accurate portrait of the spatial charges present in the abasic sites, additional Gaussian scans were performed entailing the carbon and hydrogen atoms comprising the abasic site. This was done to ensure the charges of the abasic accurately portrayed the electronic environment of a purely basic site. The charges of the flanking phosphate groups remained unaltered, as the charges of the phosphate group were consistent throughout the duplex as described by the DNA database.

2.16.6 Dihedral Scans of the TAMRA Unit

It's hypothesised that the pivotal interaction in the protective effect originates from TAMRA and the abasic site. Research has shown that the enthalpic driving force driving the interaction of small molecules into abasic site(s) is π-stacking .[125] Considering the TAMRA molecule, the dihhedral angle present between the two-ring system will have a considerable effect on its pi-conjugation and therefore needs to be correctly modelled. To achieve this a dihedral scan of the TAMRA unit was executed *via* the Gaussian software. Because of the inherent symmetry within the molecule, the dihedral angle was varied within a range of $0°$ to $90°$. The resulting potential curve was then mapped to the proper dihedral equation with the corresponding co-efficient used to describe the four dihedral concerned with the twist between the two ring systems (Figure 2.28).

$$V_{\text{rb}}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n(\cos(\psi))^n \tag{2.18}$$

$$\psi = \phi - 180 \tag{2.19}$$

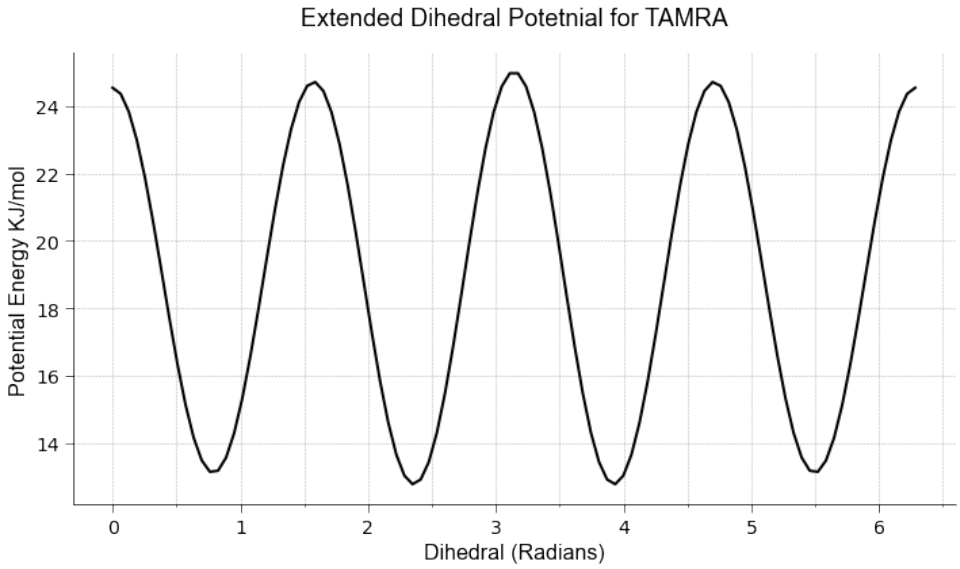


Fig. 2.28. This figure illustrates the extended potential profile for a four-atom dihedral bonding configuration in the TAMRA molecule, utilising constants derived from fitting the proper dihedral potential. The potential is mapped across a full range of 2π.

2.16.7 Ressonance Forms

The final intricacy to contemplate for the modelling is the resonance forms of TAMRA. It is well documented how TAMRA undergoes resonance between its zwitterion and lactone forms.[108,126,127] Considering that Gaussian does not differentiate the formal bonding configurations of the molecule and only seeks to minimise the electronic wave function concerning the centre of mass of the constituent nuclei, this did not raise a concern regarding the partial charges. Regardless, the formal binding configuration present in the system will have a significant impact on the conformation freedom exhibited by the TAMRA. To capture both resonant forms of the TAMRA both binding configurations of the lactone and zwitterion were modelled.

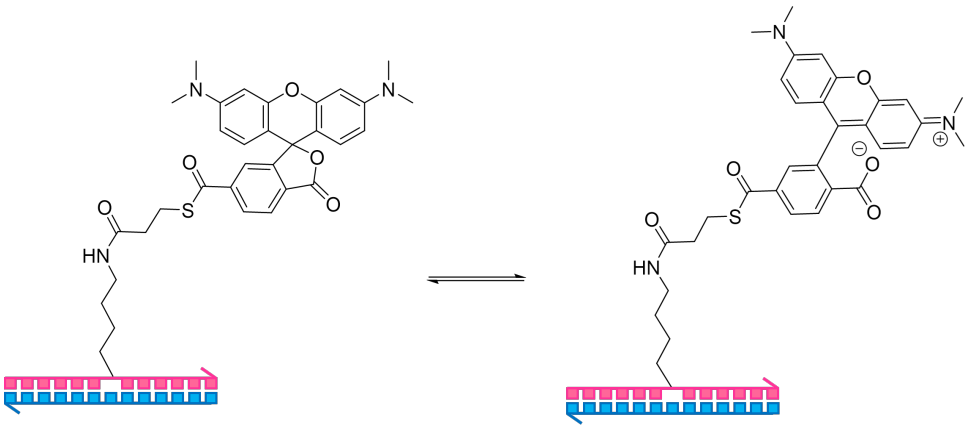


Fig. 2.29. Resonance equilibrium between TAMRA and lactone forms. Equilibrium between the resonant zwitterionic and lactone forms of the TAMRA molecule.

2.16.8 Simulation Protocol

Simulations were performed with periodic boundary conditions within a cubic box. Initially, the DNA duplex was positioned at the centre of the box. The simulation box was then solvented with the TIP3P water model.[128] Preceding this solvation step, two distinct simulation regimes were constructed: The first regime involved the addition of enough sodium ions to neutralise the charge of the negative phosphate backbone which equated to 28 sodium ions, this will be referred to as neutral salt concentration. The second regime, involved adding additional sodium and chloride ions until physiological concentrations of $0.125\,\mathrm{mM}$ were reached, this will be referred to as physiological salt concentration.

Preceding the addition of water and ions to format the environment a brief energy minimisation was conducted. This step called upon the AMBER parm99/bsc for the DNA and the GAFF for the TAMRA linker, in line with earlier discussions. With the positions of the modified-DNA-duplex harmonically restrained, the environment was equilibrated through *NVT* and *NPT* simulations. Regarding *NVT*, the system's temperature was increased from $0\,\mathrm{K}$ to $278.15\,\mathrm{K}$ to comply with experimental conditions. For the *NPT* (isothermal-isobaric) ensemble, the pressure was equibrated to 1 bar, again matching experimental conditions. Preceding the harmonisation of the environment to experimental conditions, a brief molecular dynamic simulation of 1 ns was executed. This was completed to allow some equilibration between the modified-DNA system and its environment to take place. Following this initial run, an extensive molecular dynamic simulation of 400 ns was completed.

Pair distance plots between the frontal three rings of the TAMRA molecule and the carbon back bone of abasic site (either -3 or $+3$) were utilised as quantitative descriptors for TAMRA's binding state (Figure 3.1, Table 2.1). This specific choice of atom groups was motivated by several key factors.

Firstly, TAMRA, being a large and rigid molecule with an extended conjugated ring system - specifically the 3 frontal conjugated 6 members rings -, exhibited minimal rotation during molecular dynamics simulations. This stability indicates that the positions of the atoms within these three rings remain relatively fixed, making them ideal for accurately representing the molecule's overall orientation and proximity to the abasic site.

Secondly, focusing on the carbon atoms within these rings ensures that the measured distances reflect the core structure of TAMRA, minimizing the influence of peripheral fluctuations from flexible side chains or solvent interactions.

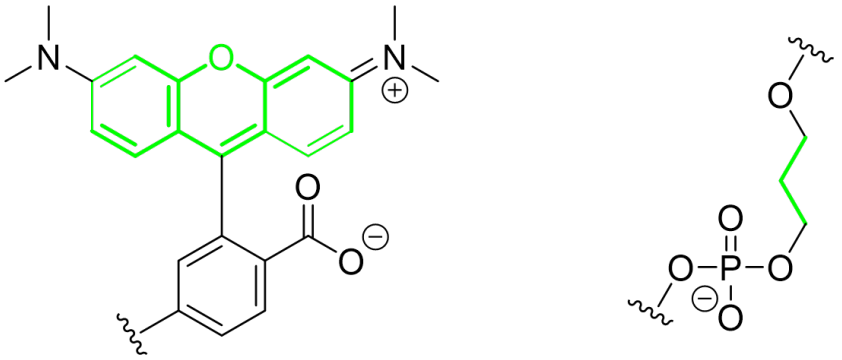


Fig. 2.30: The frontal three rings of the conjugated ring system in TAMRA and the alkyl chain of the abasic site were selected for the calculation of Root Mean Square Deviation (RMSD) plots in the molecular dynamics simulations. The bonds highlighted in green on the skeletal structures indicate the specific regions used for RMSD analysis.

Visual analysis of molecular dynamics trajectories revealed TAMRA's initial insertion into the abasic site involved these frontal ring systems. These atoms play a crucial role in stabilizing the binding through -stacking and other non-covalent interactions. Thus, monitoring the distances between these rings and the abasic site offers a reliable descriptor of the binding state, largely unaffected by rotational or conformational changes in other parts of the molecule.

The selection of the frontal three rings of TAMRA for analysis effectively captures the molecule's binding interactions with the abasic site, providing a robust and invariant measure that aligns with the observed behavior in the simulations.(Table 2.1).

Metrics constituting the plot were extracted from the simulation preceding the equilibration of the DNA-TAMRA complex, specifically when a stable binding conformation was reached. The determination of the point of equilibrium was informed by a comprehensive analysis of the pair distance plots, coupled with careful observation of the simulation trajectories.

2.17 | Manual Pre-positioning of TAMRA within the Abasic Site: Implications for Steric Interactions and Conformational Stability

Preliminary experiments involved manually placing the latone from of the TAMRA linker into the abasic site, resulted in significant steric clashes with the surrounding DNA residues due to the rigid, bulky nature of TAMRA compared to the flexible DNA backbone (Figures 2.34-2.35). These clashes, characterised by repulsive interactions, prevented TAMRA from remaining in

the abasic site under both salt conditions, indicating that steric and conformational challenges outweigh any stabilising effects from the ionic environment.

Additionally, manual insertion distorted the bond angles and torsional angles within both the DNA structure and the TAMRA molecule, introducing strain that the system sought to relieve during MD simulations. This strain led to the dissociation of TAMRA as the system moved towards a more stable conformation. These observations highlight the importance of carefully considering steric and conformational factors when setting up initial configurations for MD simulations, particularly in systems where these factors are critical to molecular stability and binding interactions.

This situation is exacerbated by the fact that abasic sites in DNA introduce additional flexibility and structural perturbations, as demonstrated by Barsky et al. in their study.[44] The study highlights that abasic sites do not maintain a simple gap but rather induce flexibility and local structural changes in the DNA, which can contribute to the challenges of manually placing a large molecule like TAMRA within the site.

This instability is further evidenced by the Root Mean Square Deviation (RMSD) plots, which showed large fluctuations in RMSD for both salt conditions, indicating significant instability in the pre-positioned TAMRA. Specifically, observations from the MD trajectories and the RMSD analysis revealed that under physiological salt conditions, TAMRA remained in proximity to the abasic site longer than under neutral salt conditions. However, in both cases, TAMRA ultimately did not remain within the abasic site. Notably, during the simulation at neutral salt concentration, TAMRA was not found in the abasic site by the time MD simulations were initiated, suggesting that it was likely forced out during the NPT equilibration phase.

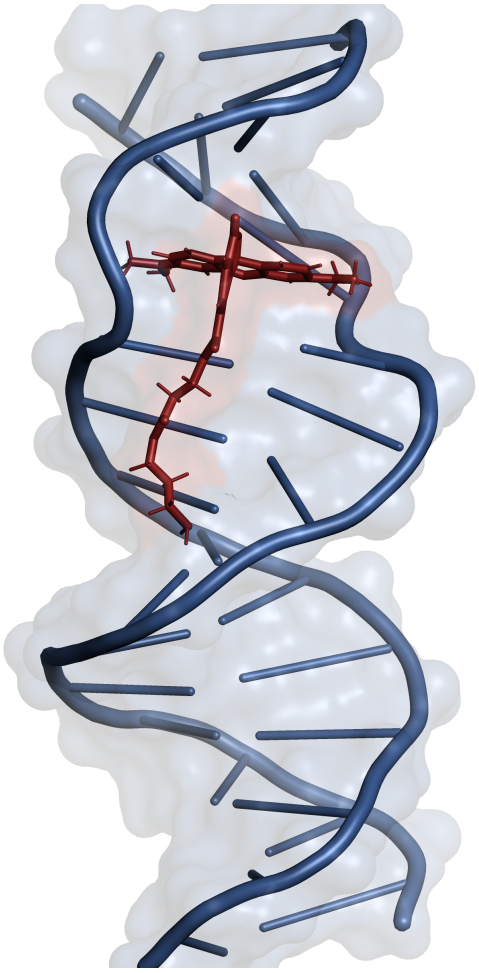


Fig. 2.31. Starting configuration of manually inserted lactone TAMRA into the -3 abasic site

.

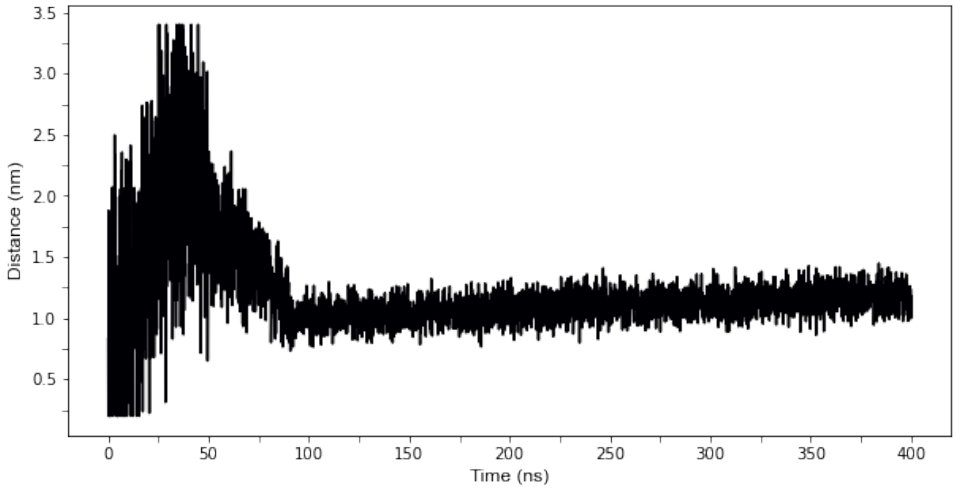


Fig. 2.32. Pair distance of Manually inserted TAMRA with -3 abasic site with physiological salt conditions.

.

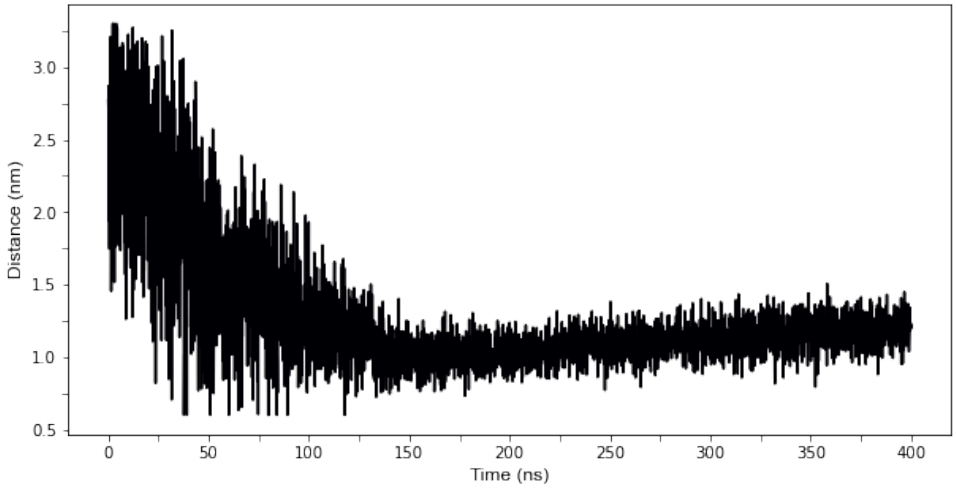


Fig. 2.33. Pair distance of the manually inserted TAMRA with -3 abasic site with neutral salt conditions

.

Fig. 2.34. Final configuration of the manual inserted lactone TAMRA physiological salt concentration

.

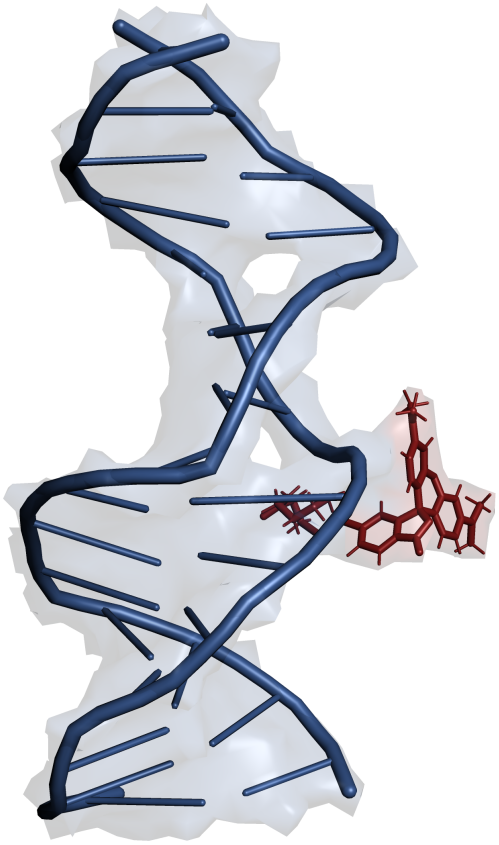


Fig. 2.35. Final configuration of the manual inserted lactone TAMRA neutral salt concentration

.

2.17.1 Molecular Dynamics Results and Interpretation

The only simulation to result in the insertion of TAMRA into its corresponding abasic site involved the lactone TAMRA form, the minus three abasic site and physiological salt concentrations.

In contrast to this, other simulations with the minus three abasic site, including the lactone at neutral salt concentration of the zwitterion form at physiological salt concentration did not result in the insertion of the TAMRA unit into the minus three abasic site. The zwitterion binding configuration resulted in insertion in its abasic site (the abasic introduced for the attachment of the linker ligated to the thioester-TAMRA). Regarding the simulation entailing the lactone bonding configuration at neutral salt concentration, no insertion of the TAMRA was observed. The simulation involving the plus three abasic site demonstrated intriguing behaviour where the TAMRA and linker threaded through the abasic site, seemingly in an attempt to access the plus three abasic site.

Table 2.1
Summary of simulation results, the key outcomes of the molecular dynamics simulations, highlighting the minimum, maximum, and average pair distances between the TAMRA unit and the respective abasic site (either -3 or +3) targeted for insertion. The table categorises results based on the salt condition of each simulation, the resonance bonding configuration of the TAMRA molecule, the specific abasic site in the DNA duplex, and a qualitative analysis of the TAMRA unit's behaviour observed during the simulation. This comprehensive summary provides valuable insights into the interaction dynamics between the TAMRA unit and abasic sites under varying conditions.

Abasic Position	TAMRA Form	Conditions	Final Configuration	Average Displacement / nm	Max. Displacement / nm	Min. Displacement / nm
-3	Lactone	Physiological	Inserted -3 abasic site	0.414	0.706	0.312
-3	Zwitterion	Physiological	Inserted to own abasic site	1.070	1.487	0.566
+3	Lactone	Physiological	No insertion (linker threaded through own abasic site)	2.019	2.551	1.492
-3	Lactone	Neutral	No insertion	1.049	1.472	0.326

2.17.2 Minus Three Abasic Duplex with the Lactone TAMRA form

Molecular dynamics concerning the lactone bonding TAMRA at physiological salt concentration, with the 3 abasic site yielded an insertion conformation aligning with our initial hypothesis

(Figure 2.36). This insertion was confirmed through the evaluation of the relative pair distance plots between the frontal three rings of the TAMRA molecule and the carbon backbone of the -3 abasic pocket (Figure 2.37). For this specific scenario, equilibrium was determined to occur at 29 ns, therefore metrics concerning this parameter were extracted after this point.

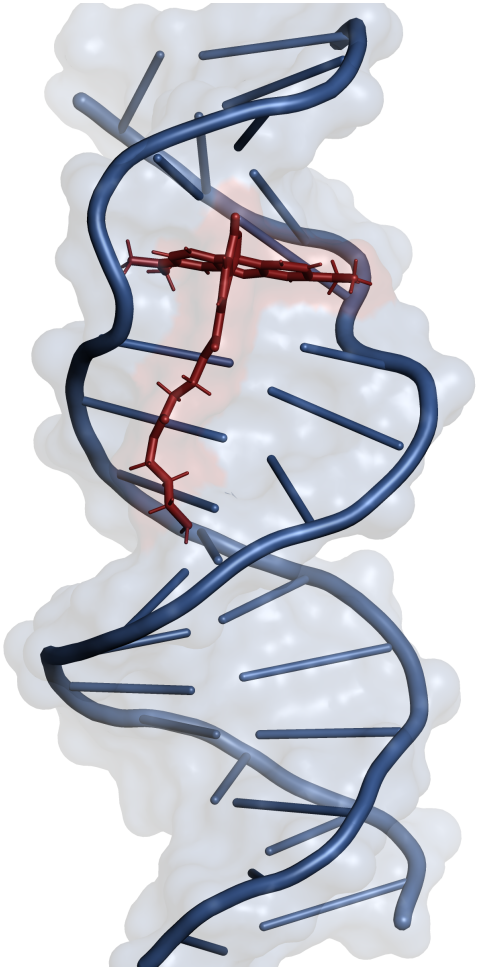


Fig. 2.36. Final Configuration of Lactone TAMRA in -3 Abasic Site at physiological conditions

.

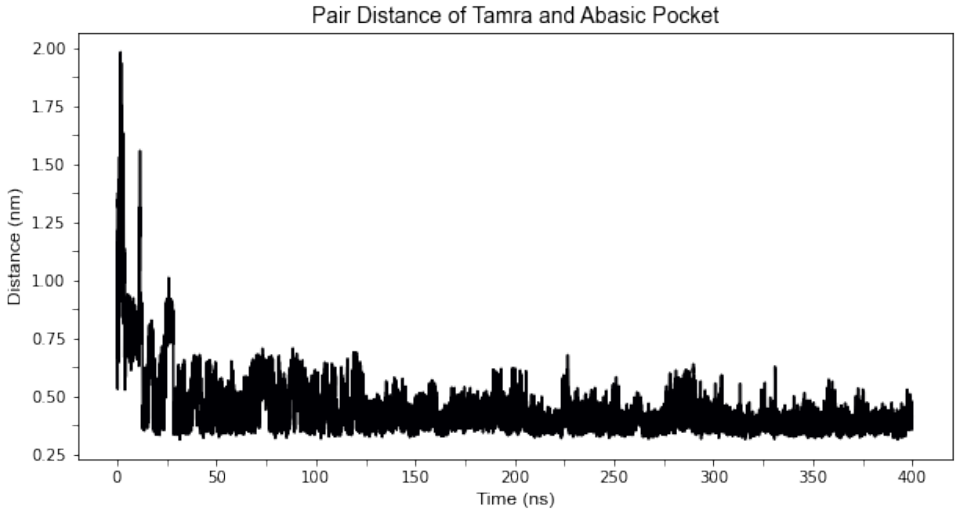


Fig. 2.37. Pair distance of TAMRA with -3 abasic site. Equilibrium in the simulation was identified at the 29 ns mark, indicating the point of stable interaction between TAMRA and the abasic site.

.

2.17.3 Behavioral Difference of Lactone TAMRA at Varying Salt Concentrations

The MD simulation concerning the lactone form of TAMRA in neutral salt concentration resulted in no insertion (Figure 2.38). Distinct from the physiological salt concentration regime the TAMRA unit did not interact with the -3 abasic site. This differential behaviour concerning the two salt concentrations demonstrates the profound influence that charge screening has on interaction dynamics.[53] This simulation involving the lactone form and neutral salt concentration reached some sort of equilibrium after 170 ns (Figure 2.39). This is in contrast to the short equilibrium time exhibited with the physiological salt concentration. This again is evidence of how electrostatic interactions influence the conformational stability of the DNA-linker-TAMRA complex. Specifically, the neutral salt concentration regime is subject to a relative lack of charge screening as compared to the physiological salt concentration regime. This lack of charge screening consequently results in the linker-TAMRA remaining in solution. This behaviour is constant with research concerned with the effect of electrostatic interaction on biomolecular stability and orientations.[129,130] Visual inspection of the trajectory revealed that the TAMRA unit never fully inserts into the -3 abasic site but accesses a meta-stable conformation that lies in close proximity to the abasic sites. The linker appears to curl around itself seemingly in an attempt to minimise solvent interaction. This characteristic interplay

between the linker and the solvent demonstrates how the surrounding environment affects the conformation of the complex.

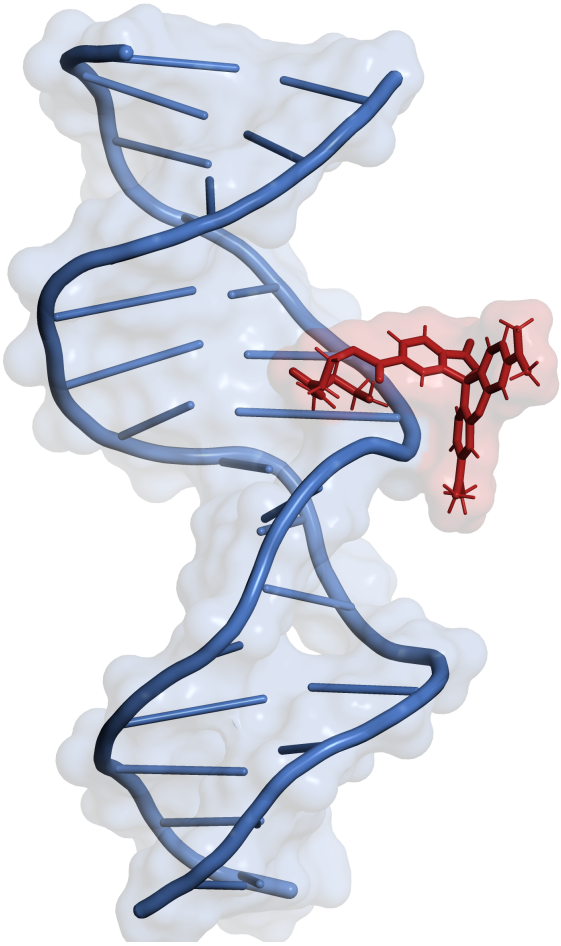


Fig. 2.38. Final configuration of lactone TAMRA at neutral salt conditions showing the TAMRA unit not inserted in any abasic site.

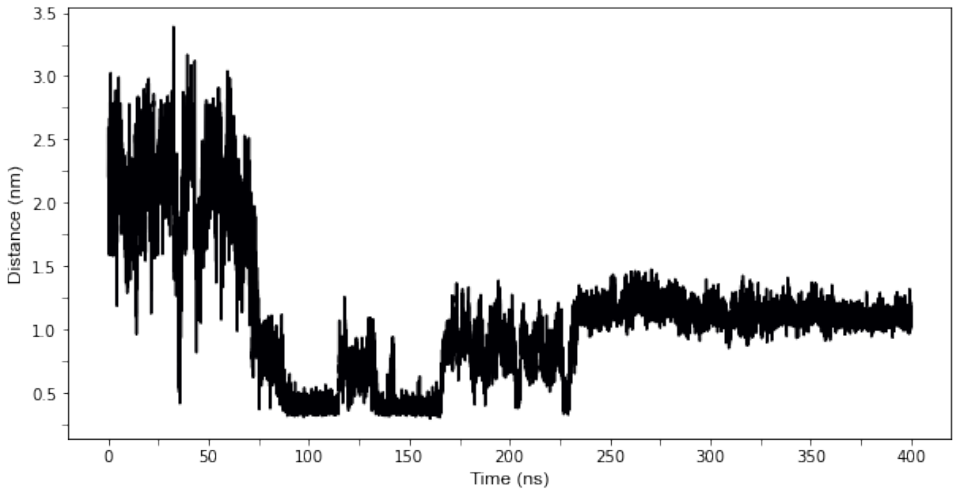


Fig. 2.39. Pair distance plot of TAMRA and -3 abasic site under neutral salt conditions, with the system reaching equilibrium at the 170 ns mark.

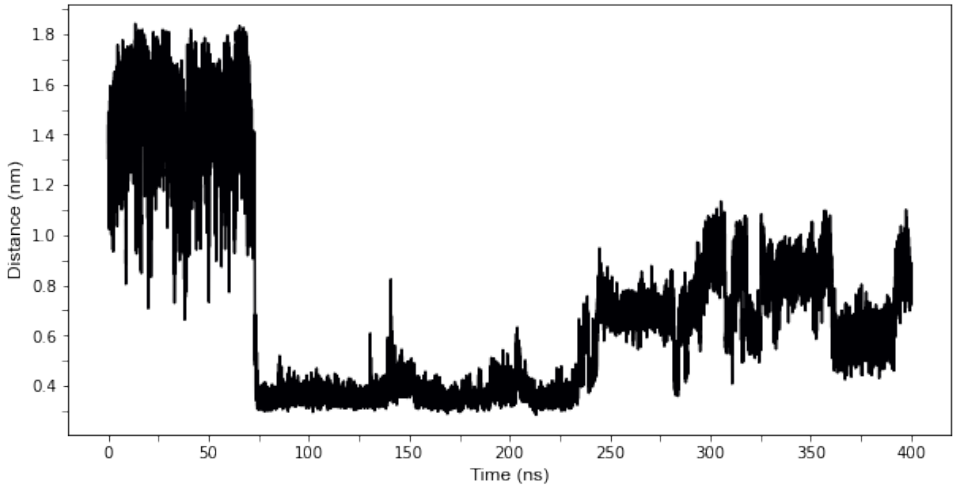


Fig. 2.40. Pair Distance plot of TAMRA with its own abasic site under neutral salt conditions. Extended time periods at various pair distances are indicative of a range of meta-stable states.

This starkly contrasting behaviour of the lactone bonding form of TAMRA with the varying salt concentration underlines the necessity of simulating a variety of conditions to obtain a holistic understanding of the interaction dynamic concerned with bio-molecular complexes.

2.17.4 Behavioral Difference Between The Resonant Bonding Configurations

With the same physiological salt concentration, the resonance zwitterion bonding form of the TAMRA was simulated. With this regime, no insertion into the -3 abasic site was observed. Interestingly in this scenario, the linker-TAMRA seemed to curl around it's self and insert into its own abasic site (Figure 2.41). The system seemed to reach a stable configuration at 24 ns, therefore equilibrium was determined to be reached at this point (Figure 2.42 and 2.43).

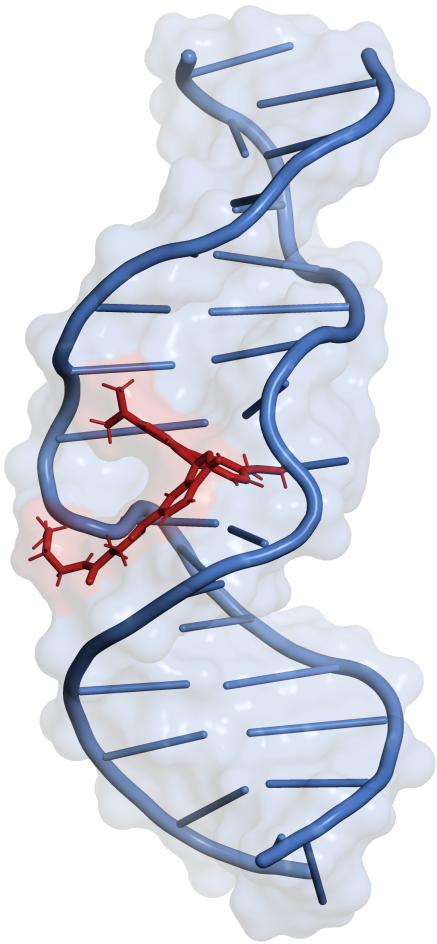


Fig. 2.41. Resonance form of TAMRA in its own abasic site minimising solvent interactions.

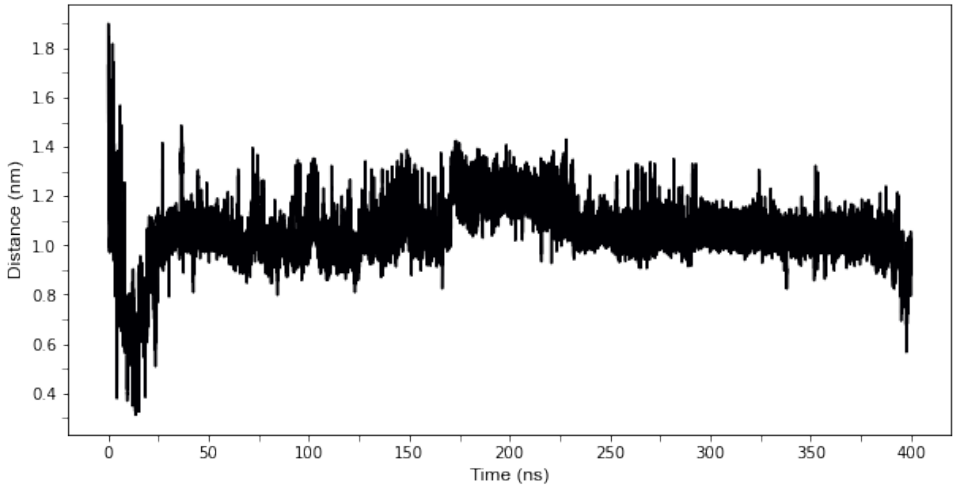


Fig. 2.42. Pair distance plot of resonance TAMRA with -3 abasic site. Equilibrium is established at the 24 ns mark.

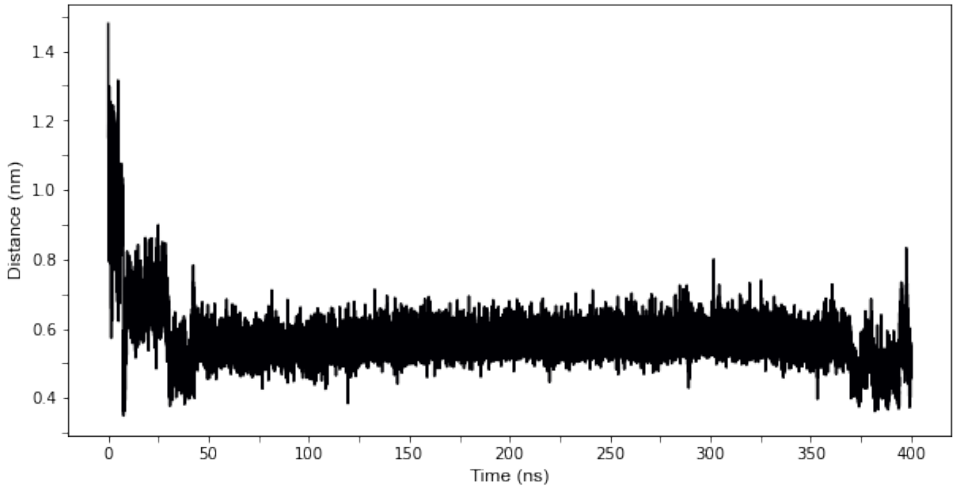


Fig. 2.43. Pair Distance Plot of Resonance TAMRA with its own Abasic Site.

2.17.5 Simulation of the +3 Abasic Duplex

The final scenario tested was the +3 abasic site with the lactone resonance form of TAMRA and physiological salt concentrations. Over a span of 200 ns, the conformation of the TAMRA-linker seemed to result in a kinetic entrapment involving the head of the TAMRA unit inserting through its abasic site (Figure 2.44). This phenomenon may be attributed to the inherent chirality of DNA in combination with the chiral orientation of the linker attachment to the DNA backbone. The current combination of both results in the linker oriented towards the minor groove of the DNA system, potentially influencing its behaviour in the simulation.

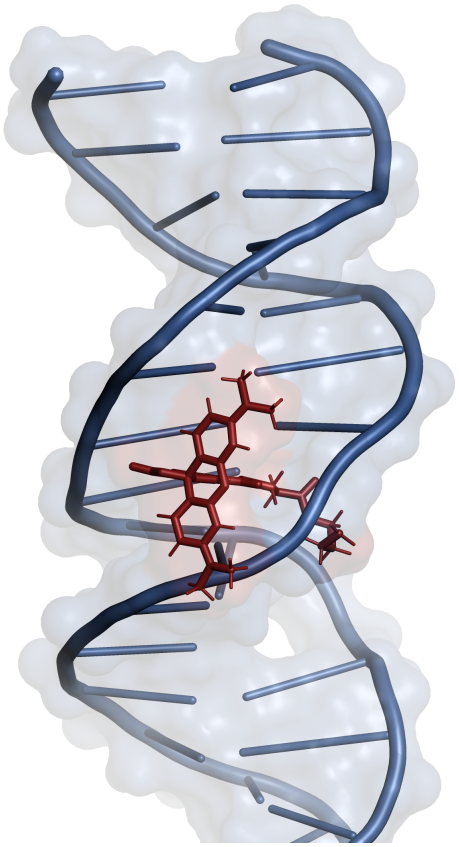


Fig. 2.44. Lactone TAMRA at physiological conditions with the +3 abasic site the linker is observed to thread through its own abasic site in the presence of the plus abasic site.

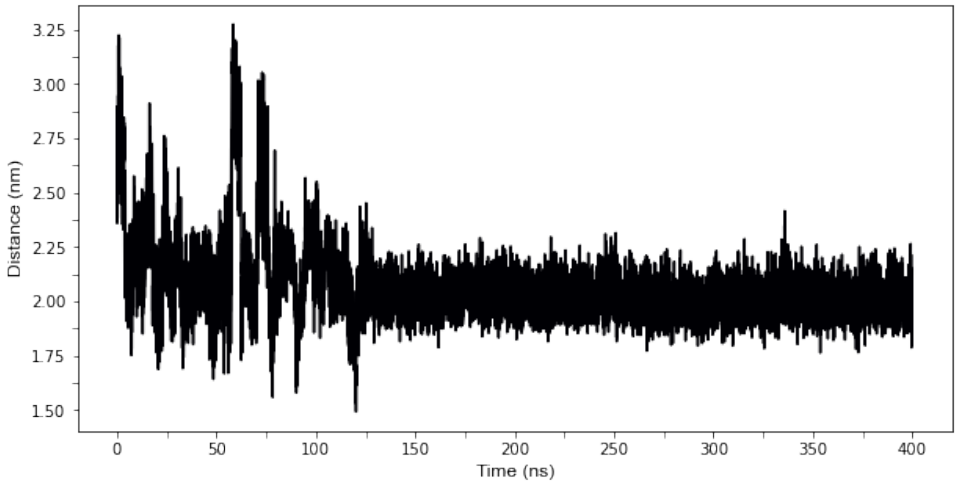


Fig. 2.45. Pair distance plot of TAMRA with +3 abasic site. equilibrium was reached at the 100 ns mark.

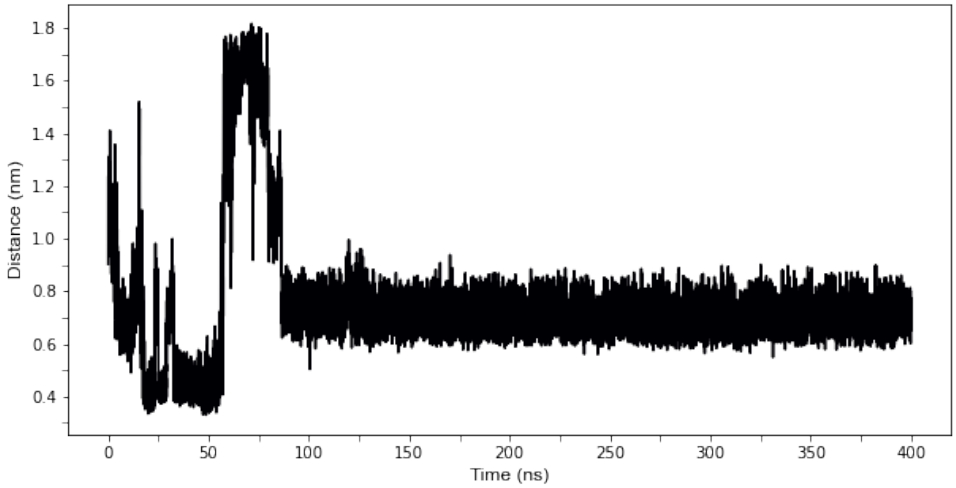


Fig. 2.46. Pair distance plot of TAMRA with its own abasic site for the scenario with +3 abasic position equilibrium was achieved at 40 ns.

2.17.6 Chirality Inherent to The System

Inherent chirality within the DNA duplex results in a right-handed helix under physiological DNA (B-DNA). The chirality arises from the intrinsic chirality present in D-deoxyribose sugars, constituting the ribose of the DNA backbone. The D-deoxyribose sugar is host to three chiral centres in their composition.[131] Stemming from the preferred chirality and angle of the glycosidic bond, the DNA structure results in major and minor grooves. In combination with this chirality, the attachment of the linker to the abasic site can occur in two chiral forms (*R* and *S* enantiomers, Figure 2.47). The provider of the modified oligomers: IDT, is uncertain about the chiral identity of their product. Considering the time constraint of my PhD only the *S* enantiomer was modeled. Considering the length and flexibility of the linker, it was argued that the enantiomeric identity of the attachment will have little effect on the conformational preference exhibited in the model. However, this would be an intriguing avenue to explore for future investigations.

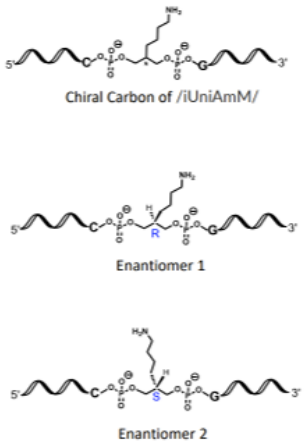


Fig. 2.47. Enantiomeric configurations of amino-modified abasic site depicting the R and S enantiomeric configurations of the chiral carbon in the amino-modified abasic site used for the linker in the modified DNA duplex. The illustration differentiates the two spatial arrangements, highlighting their distinct chiral characteristics and potential implications for the molecular interactions within the duplex.

2.18 | Discussion

This computational study represents a preliminary experimentation investigation aimed at understanding the mechanism of this for the unique protection in the context of DNA-templated synthesis (DTS). While the initial findings are promising, several critical approaches should be considered to further validate and expand upon these results.

To ensure the reproducibility and statistical significance of the observed phenomena, it is imperative to run multiple independent simulations starting from different initial conditions, such as varying initial velocities velocity distributions corresponding to the temperature of interest and initial conformations. This approach helps build a statistically significant ensemble of trajectories, ensuring that the results are robust and not merely artefacts of particular initial conditions in a given simulation run.[132]

Introducing small perturbations to the starting structures—such as slight positional adjustments—is essential to confirm that the observed behaviour is not an artifact of the specific initial setup. This strategy broadens the exploration of conformational space and captures a more diverse range of behaviours, providing a comprehensive understanding of the system.[133]

Extending the simulation time significantly beyond 400 ns is crucial to allow for the system to have ample opportunity to explore different conformational states, especially when transitions between these states occur slowly. Longer simulations are vital for capturing long-timescale processes and rare events that shorter simulations might miss capture, offering a more complete picture of the system's dynamics.[134]

To better explore the conformational landscape and capture rare events, employing enhanced sampling methods such as Replica Exchange Molecular Dynamics (REMD), Metadynamics, or Accelerated MD is highly recommended. These techniques help overcome energy barriers and facilitate the sampling of otherwise inaccessible states, providing a more exhaustive view of the system's behaviour.[135,136]

Additionally, careful selection and validation of force fields and parameterisations used in the simulations are essential to ensure the accuracy of the results. Validating computational models against experimental data, when available, can increase confidence in the predictive power of the simulations and help identify any discrepancies that may need further investigation.[66]

By analysing various metrics such as dihedral angles or free potential energy distributions over time, convergence can be confirmed, indicating that the simulation has adequately sampled the system's phase space and that the results are independent of the simulation time window.

This step is critical for verifying that the observed behaviour properties are representative of the system's true dynamics.[137]

By implementing these approaches, future studies can build upon the preliminary findings presented here, providing a more robust and comprehensive understanding of the mechanisms underlying the protective effects observed in DTS. This, in turn, could significantly advance the field and contribute to the development of more effective DNA-templated synthesis strategies.

2.19 | Conclusion

Through experimentation, it was uncovered that the protective effect is far more prominent when the abasic site resides in the minor groove. A possible rationale for this minor groove preference might stem from the propensity of small molecules to bind to the minor groove, thus reducing the activation energy needed for its insertion. Another such rationale could be due to proximity, as the minor groove is located closer than the major. It was observed that the longer linker provided more consistent protection irrespective of the existence or position of an abasic site. This is evident as the protection observed for the fully complementary and $+3$ abasic strand lies in the range $77 < x < 79\%$, whereas, with the shorter linker, the protection across all positions, excluding $-2 \leq x \leq -4$, was approximately 35% (calculated using Equation 2.16 without calibration). It is possible that this exemplified protective effect stemming from the longer linker is enabled by greater conformational freedom. As the linker is longer and more flexible, the TAMRA can bind to previously inaccessible regions of the duplex.

The protective mechanism observed in these experiments has significant implications for DNA-Templated Synthesis (DTS). In the work of Frommer et al.,[41] this protective strategy was successfully implemented within a DTS framework to safeguard a thioester group, which is prone to hydrolysis. Their results showed that incorporating an abasic site adjacent to the thioester, particularly within the minor groove, drastically enhanced the yield of the desired amide product. This underscores the minor groove's role in facilitating a more effective interaction between the protective site and the reactive group, further supported by the consistency of protection provided by the longer linker, which allowed TAMRA to access previously inaccessible regions within the DNA duplex. These findings suggest that leveraging the flexibility and strategic positioning of linkers and protective sites can significantly improve the efficiency and yield of DTS reactions, particularly for protecting labile functional groups from degradation.

This study revealed distinct dynamics across various modeling conditions, particularly regarding the bonding identity of TAMRA and the surrounding salt environment. These factors significantly influenced the resulting conformation and binding preferences of the system. Under physiological salt concentrations, TAMRA in its lactone resonance form inserted into the -3 abasic site, a behavior not observed at neutral salt concentrations, where no insertion occurred. Conversely, TAMRA in its zwitterion resonance form demonstrated a preference for inserting into its own abasic site under the same physiological conditions.

The final scenario examined the +3 abasic site, where the lactone bonding configuration and physiological salt concentration similarly led to TAMRA inserting into its own abasic site. Experimental investigations demonstarted substantial protection was afforded when the abasic site was situated between the -2 and -4 positions, whereas the fully complementary duplex showed no such protection. These findings suggest that the protective effect does not arise from TAMRA's insertion into its own abasic site but rather from the specific positioning within the DNA duplex, particularly in the minor groove where smaller molecules are more likely to bind. Through combining the streams of computational and experimental work it is reeled that for protection to arise the lactone from of TAMRA inserts into the -3 abasic site, and the computational work highlights the nessecity for physiological salt conditions.

As a fluorescence signal of the DNA-linker-TAMRA duplex was observed for the -3 abasic site, the state of TAMRA must be in its zwitterionic resonance form. The observed protection of the labile thioester from hydrolysis at the -3 abasic site, coupled with the detectable shift in elution of the DNA-TAMRA duplex relative to the DNA-thiol duplex, strongly suggests that a binding reaction occurred. From previous work in this area, we know that the protective effect between the TAMRA molecule and the abasic site occurred without any reaction between the thioester and amine nucleophile, indicating that no DNA-templated synthesis (DTS) can occur under these conditions.[43]

Regarding the stability of the inserted form, it is likely that an equilibrium exists between the inserted and non-inserted forms. The stability of this inserted form could be further examined by employing umbrella sampling, a technique used to calculate the potential of mean force (PMF) along a reaction coordinate by introducing a series of biased simulations that allow for the efficient sampling of rare events.[52] Umbrella sampling is particularly well-suited to this scenario, as it would enable the determination of the free energy landscape associated with the insertion and non-insertion states of the TAMRA molecule. By analysing this landscape,

we could gain insights into how stable the inserted form is relative to the non-inserted form, providing a quantitative measure of the equilibrium between these states.

In future work on the implementation of a protective effect in DNA-Templated Synthesis (DTS), it will be crucial to develop protective groups with bifunctionality to ensure that reactions can progress effectively. Molecular dynamics (MD) simulations combined with umbrella sampling could be employed to identify and optimise suitable protecting groups. These computational techniques would allow for an in-depth analysis of the free energy landscapes associated with different potential protecting groups, helping to predict their stability and effectiveness in shielding reactive intermediates. This approach is particularly beneficial as it can systematically explore the relevant regions of the configurational space, identifying the most energetically favorable structures.[138]

Once suitable protecting groups have been identified through MD and umbrella sampling, their practical applicability can be confirmed by subjecting these groups to the protection assay described in the experimental section of this chapter. This will involve assessing their ability to shield labile intermediates from hydrolysis and determining whether they can be effectively incorporated into the DTS framework. This iterative process of computational prediction followed by experimental validation will enhance the design and application of protective groups in DTS.

General Information

Materials & Equipment

All chemicals and solvents were ordered from Sigma-Aldrich or Fisher Scientific and used without further purification, unless otherwise stated; DNA was purchased from IDT and all solvents were reagent grade.

DNA strands:

Table 2.2
DNA Sequences with Modifications

Name	Sequence
Amine modified DNA	GCCGAGCCAGCAGTCAGCGC/iUniAmM/GTCCTAATCTACCTG
Abasic -5	CAGGTAGATTAGGACAGCGC /iSpC3/GACTGCTGGCTCGGC
Abasic -4	CAGGTAGATTAGGACAGCG/iSpC3/TGACTGCTGGCTCGGC
Abasic -3	CAGGTAGATTAGGACAGC/iSpC3/CTGACTGCTGGCTCGGC
Abasic -2	CAGGTAGATTAGGACAG/iSpC3/GCTGACTGCTGGCTCGGC
Abasic -1	CAGGTAGATTAGGACA/iSpC3/CGCTGACTGCTGGCTCGGC
Abasic 0	CAGGTAGATTAGGAC/iSpC3/GCGCTGACTGCTGGCTCGGC
Abasic 1	CAGGTAGATTAGGA/iSpC3/AGCGCTGACTGCTGGCTCGGC
Abasic 3	CAGGTAGATTAG/iSpC3/ACAGCGCTGACTGCTGGCTCGGC
Abasic 5	CAGGTAGATT/iSpC3/GGACAGCGCTGACTGCTGGCTCGGC

Equipment

HPLC

High-Performance Liquid Chromatography (HPLC) analysis, and purification of oligonucleotides were performed on a modular Shimadzu instrument with the following modules: CBM-20A system controller, LC-20AD solvent deliver module, SIL-20AC HT autosampler, CTO-20AC column oven, SPD-M20A photodiode array UV-Vis detector, RF-20A spectrofluorometric detector and an FRC-10 fraction collector. Chromatography was performed on a Waters XBridge Oligonucleotide BEH C_{18} column, 130 Å, 2.5 µm, 4.6×50mm.

Buffers

A: 5% ACN, 0.1 M triethylammonium acetate (TEAA)

B: 70% ACN, 0.1 M triethylammoniumacetate.

HPLC methods

Method 1: Flow rate: 0.8 mL/min; column temperature: 60 °C, buffer gradient varied with respect to buffer B in the following manner: 0% 0-10 min, 0-50% over 18 min, 50-95% over 4 min, 95-0% over 3 min, 0% over 5 min.

Method 2: Flow rate: 0.8 mL/min; column temperature: 60 °C, buffer gradient varied with respect to buffer B in the following manner: 1% over 5min 1%-30% over 15 min, 30%-95% 5 min, 95%-1% 1 min, 1% over 3 min.

Method 3: Flow rate: 0.8 mL/min; column temperature: 35 °C, buffer gradient varied with respect to buffer B in the following manner: 1% over 3 min, 1-15% over 17 min, 15%-95% over 5min, 95%-1% over 2 min, 1 % over 4 min.

Oligonucleotide Synthetic Equipment

Termormixer: Eppendorf Thermomixer C

Nanodrop: Thermoscientific nano-drop one

Centrifuge: Eppendorf Centrifuge 5424 R

Spin Vacuum: Eppendorf Concentrator plus

NMR

Proton (^{1}H) and Carbon (^{13}C): Resonance Spectroscopy ^{1}H NMR and ^{13}C NMR spectra were recorded on a Bruker DPX-300, DPX-400 or HD500 spectrometer. Chemical shifts are reported in ppm. Solvent signals were used as reference (Trimethylsilane, ^{1}H: $\delta = 0$ ppm; chloroform ^{13}C: $\delta = 77.16$ ppm; DMSO, ^{1}H: $\delta = 2.50$, ^{13}C: 39.52 ppm. The signals are characterized as s = singlet, d = doublet, dd = doublet in doublet, ddd = doublet in doublet in doublet, t = triplet, q = quartet and m = multiplet. The coupling constant J is recorded in Hertz (Hz).

LCMS

Liquid chromatography-mass spectrometry (LC-MS) analysis of oligonucleotides was performed on a Bruker Amazon^{TM} X. Chromatography was carried out on a Waters XBridge^{TM} OST C_{18} 2.5 M column at a flow rate of 0.8 mL min^{-1} using a gradient of buffers A and B: buffer A, 10 mM NH_4OAc in a 95:5 mixture of H_2O and MeOH; buffer B, 10 mM NH_4OAc in 30:70

mixture of H_2O and MeOH. The eluent was directly injected into the mass spectrometer, and the data was acquired in the negative-ion mode.

Spin filtration

Procedure: Load of DNA ≤ 2 nmol per amicon 3 k ultra-centrifugal spin filter. The sample is diluted to 500μL with nano pure water. The sample is subjected to spin filtration in a centrifuge for 12 min at 14000 rcf at $15°C$.

Relative Comparison of the Absorbance Signals of Chemically Modified DNA Strands in the Context of HPLC

The degree of conversion of a reaction involving chemically modified DNA strands was determined by a comparison from the area of the 262 nm absorbance from their signal produced in the HPLC chromatogram.

$x = 262\,nm$ signal produced by a DNA strand on the HPLC

$$x = \int_{t_1}^{t_2} A\,dt$$

$$\text{conversion} = \frac{x_n}{\sum x}$$

Buffers

DTS Buffer: *T*he composition of the DTS buffer used by Robbie.

Table 2.3
The composition of the DTS buffer used by Robbie

	mMol	Equivalents
TAPS	12.5	0.25
CAPS	12.5	0.25
NA_2HPO_4	12.5	0.25
NaCl	12.5	0.25

The buffer was made to the proportions described in the table and the pH of the DTS was adjusted by using hydrochloric acid (HCl) and sodium hydroxide (NaOH).

Borate Buffers

Borate buffers were prepared by solubilizing sodium tetraborate decahydrate to a 1 M solution. An aliquot of the solution was diluted to 0.1 M and the pH was adjusted with glacial acetic acid or NaOH.

Synthesis of Disulphide DNA

Previous Synthesis for Disulphide-Modified DNA

Robbie's synthesis of dilsulphide modified DNA: $10\,\mu$L, 1 mM, of Internal amine-modified DNA in H_2O was mixed with $540\,\mu$L DMF, $100\,\mu$L 100 mM succinimidyl $3-(2$-pyridyldithiopropionate (SPDP, Cambridge Bioscience) in DMF and $53\,\mu$L 250 mM DTS buffer pH 12. The reaction was incubated for 15 hours at RT. The reaction was ethanol precipitated and subjected to $3\times$ spin filtration.

Developed Synthesis for Disulphide-Modified DNA

- Synthesis of disulphide modified DNA: $2\,\mu$L, 1 M SPDP dissolved in DMF was added to Internally modified amine DNA ($2\,\mu$L, 2 nmol, 1 mM) with $8\,\mu$L borate buffer 0.1 M pH 9.8.
- The solution was left shaking overnight in a thermomixer at 1400 rpm for 16 hours.
- The solution was then subjected to spin filtration ($6\times$).
- The SPDP modified DNA was separated from the reaction via HPLC fractional collection using HPLC method 2. Conversion (80%). The SPDP modified containing fractions were reduced by vacuum and subjected to spin filtration ($6\times$). Conversion (86%), Average recovered yield ($40-50\%$).

Synthesis of Protecting Group Modified DNA

- $1\,\mu$L of 0.05 M of TCEP in borate buffer 0.1 M, pH 9.5 was added to $1\,\mu$L, 1 mM, of desalted SPDP DNA and the solution was left mixing at 1400 rpm in a thermomixer for 1 h.

- $1\,\mu$L of 100 nmol of activated NHS ester dissolved in DMF was added along with $7\,\mu$L of Borate buffer 0.1 M pH 9.5. The reaction was left to stir for 16h.
- The solution was then subjected to spin filtration at 4°C ($6\times$).
- The protecting group modified DNA was separated from the reaction via HPLC fractional collection using HPLC method 3. The SPDP modified containing fractions were reduced by vacuum and subjected to spin filtration at 19°C ($6\times$).
- Conversion from benzene-DNA formation: 55%
- Conversion from TAMRA-DNA formation: 97%

Protection Assay

HPLC Time Course: $7\,\mu$L, $4.29\,\mu$M thioester-modified duplex was hybridized to a $2\times$ molar excess of the hybridizing strand ($7\,\mu$L, $8.57\,\mu$M) with $0.6\,\mu$L 0.5 M $MgCl_2$ and $0.4\,\mu$L of deionized water to result in a $2\,\mu$M duplex solution of $15\,\mu$L total volume. The duplex was left in a thermomixer for 15 minutes, 400 rpm, 21°C. $15\,\mu$L of DTS buffer 0.25 M, pH 11 was added and the sample was subjected to an HPLC time course with time points taken at 0h, 2h, 12h, 24h, 36h, 48h using HPLC method 3. This resulted in a final solution of $30\,\mu$L of $1\,\mu$M with respect to the concentration of duplex. $2\,\mu$L of the sample was injected per timepoint resulting in 2 pmol of duplex per timepoint. The time course was automated with the batch processing in lab solutions.

References

1. Brar, G. A. & Weissman, J. S. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nature Reviews Molecular Cell Biology* **16,** 651–664 (2015).
2. Clarke, B. Darwinian evolution of proteins. *Science* **168,** 1009–1011 (1970).
3. Goh, C.-S., Bogan, A. A., Joachimiak, M., Walther, D. & Cohen, F. E. Co-evolution of proteins with their interaction partners. *Journal of molecular biology* **299,** 283–293 (2000).
4. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **16,** 379–394 (2015).
5. Beyer, D., Skripkin, E., Wadzack, J. & Nierhaus, K. H. How the ribosome moves along the mRNA during protein synthesis. *Journal of Biological Chemistry* **269,** 30713–30717 (1994).
6. Bhavsar, R. B., Makley, L. N. & Tsonis, P. A. The other lives of ribosomal proteins. *Human Genomics* **4,** 1–18 (2010).
7. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annual review of biochemistry* **83,** 379–408 (2014).
8. Adhikari, A., Bhattarai, B. R., Aryal, A., Thapa, N., Puja, K., Adhikari, A., Maharjan, S., Chanda, P. B., Regmi, B. P. & Parajuli, N. Reprogramming natural proteins using unnatural amino acids. *RSC advances* **11,** 38126–38145 (2021).
9. Xuan, S. & Zuckermann, R. N. Engineering the atomic structure of sequence-defined peptoid polymers and their assemblies. *Polymer* **202,** 122691 (2020).
10. De Neve, J., Haven, J. J., Maes, L. & Junkers, T. Sequence-definition from controlled polymerization: the next generation of materials. *Polymer Chemistry* **9,** 4692–4705 (2018).
11. Haven, J. J., De Neve, J. A. & Junkers, T. Versatile approach for the synthesis of sequence-defined monodisperse 18-and 20-mer oligoacrylates. *ACS Macro Letters* **6,** 743–747 (2017).
12. Holloway, J. O., Wetzel, K. S., Martens, S., Du Prez, F. E. & Meier, M. A. Direct comparison of solution and solid phase synthesis of sequence-defined macromolecules. *Polymer Chemistry* **10,** 3859–3867 (2019).

13. Ten Brummelhuis, N. Controlling monomer-sequence using supramolecular templates. *Polymer Chemistry* **6,** 654–667 (2015).

14. Tokura, Y., Harvey, S., Chen, C., Wu, Y., Ng, D. Y. & Weil, T. Fabrication of defined polydopamine nanostructures by DNA origami-templated polymerization. *Angewandte Chemie International Edition* **57,** 1587–1591 (2018).

15. Jonathan, B., Eugen, S., O'Reilly Rachel, K., *et al.* Programmable One-Pot Multistep Organic Synthesis Using DNA Junctions (2012).

16. Lewandowski, B., De Bo, G., Ward, J. W., Papmeyer, M., Kuschel, S., Aldegunde, M. J., Gramlich, P. M., Heckmann, D., Goldup, S. M., D'Souza, D. M., *et al.* Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science* **339,** 189–193 (2013).

17. O'Reilly, R. K., Turberfield, A. J. & Wilks, T. R. The evolution of DNA-templated synthesis as a tool for materials discovery. *Accounts of Chemical Research* **50,** 2496–2509 (2017).

18. Kanan, M. W., Rozenman, M. M., Sakurai, K., Snyder, T. M. & Liu, D. R. Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* **431,** 545–549 (2004).

19. Li, X. & Liu, D. R. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angewandte Chemie International Edition* **43,** 4848–4870 (2004).

20. Forster, A. C. & Symons, R. H. Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site. *Cell* **50,** 9–16 (1987).

21. Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *Journal of the American Chemical Society* **123,** 6961–6963 (2001).

22. Kuzuya, A. & Komiyama, M. DNA origami: Fold, stick, and beyond. *Nanoscale* **2,** 309–321 (2010).

23. Matthews, H. Dna structure prerequisite information. *Biology Concept and Connections. Benjamin Cummings, California* (1997).

24. Li, M.-X., Xu, C.-H., Zhang, N., Qian, G.-S., Zhao, W., Xu, J.-J. & Chen, H.-Y. Exploration of the kinetics of toehold-mediated strand displacement via plasmon rulers. *ACS Nano* **12,** 3341–3350 (2018).

25. Tian, Y., He, Y., Chen, Y., Yin, P. & Mao, C. A DNAzyme that walks processively and autonomously along a one-dimensional track. *Angewandte Chemie* **117,** 4429–4432 (2005).

26. He, Y. & Liu, D. R. Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker. *Nature nanotechnology* **5,** 778–782 (2010).

27. Liu, D. R. Translating DNA into synthetic molecules. *PLoS Biology* **2,** e223 (2004).

28. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448,** 828–831 (2007).

29. Ma, F., Chung, M. T., Yao, Y., Nidetz, R., Lee, L. M., Liu, A. P., Feng, Y., Kurabayashi, K. & Yang, G.-Y. Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. *Nature Communications* **9,** 1030 (2018).

30. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences* **94,** 12297–12302 (1997).

31. Nakano, S.-i. & Sugimoto, N. The structural stability and catalytic activity of DNA and RNA oligonucleotides in the presence of organic solvents. *Biophysical Reviews* **8,** 11–23 (2016).

32. Privalov, P. L. & Crane-Robinson, C. Forces maintaining the DNA double helix. *European Biophysics Journal* **49,** 315–321 (2020).

33. Arcella, A., Portella, G., Collepardo-Guevara, R., Chakraborty, D., Wales, D. J. & Orozco, M. Structure and properties of DNA in apolar solvents. *The Journal of Physical Chemistry B* **118,** 8540–8548 (2014).

34. El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chemical Society Reviews* **39,** 1388–1405 (2010).

35. Meng, W., Muscat, R. A., McKee, M. L., Milnes, P. J., El-Sagheer, A. H., Bath, J., Davis, B. G., Brown, T., O'Reilly, R. K. & Turberfield, A. J. An autonomous molecular assembler for programmable chemical synthesis. *Nature Chemistry* **8,** 542–548 (2016).

36. Bhaduri, S., Ranjan, N. & Arya, D. P. An overview of recent advances in duplex DNA recognition by small molecules. *Beilstein Journal of Organic Chemistry* **14,** 1051–1086 (2018).

37. Almaqwashi, A. A., Paramanathan, T., Rouzina, I. & Williams, M. C. Mechanisms of small molecule–DNA interactions probed by single-molecule force spectroscopy. *Nucleic Acids Research* **44,** 3971–3988 (2016).

38. Waller, M. P., Robertazzi, A., Platts, J. A., Hibbs, D. E. & Williams, P. A. Hybrid density functional theory for π-stacking interactions: Application to benzenes, pyridines, and DNA bases. *Journal of Computational Chemistry* **27,** 491–504 (2006).

39. Sato, Y., Nishizawa, S., Yoshimoto, K., Seino, T., Ichihashi, T., Morita, K. & Teramae, N. Influence of substituent modifications on the binding of 2-amino-1, 8-naphthyridines to cytosine opposite an AP site in DNA duplexes: thermodynamic characterization. *Nucleic acids research* **37,** 1411–1422 (2009).

40. Jana, B., Pal, S., Maiti, P. K., Lin, S.-T., Hynes, J. T. & Bagchi, B. Entropy of water in the hydration layer of major and minor grooves of DNA. *The Journal of Physical Chemistry B* **110,** 19611–19618 (2006).

41. Oppenheimer, R. *Artificial Programmable Poylmerases* Trinity 2019. PhD thesis (University of Oxford, St Catherine's College, 2019).

42. Bracher, P. J., Snyder, P. W., Bohall, B. R. & Whitesides, G. M. The relative rates of thiol–thioester exchange and hydrolysis for alkyl and aryl thioalkanoates in water. *Origins of Life and Evolution of Biospheres* **41,** 399–412 (2011).

43. Frommer, J., Oppenheimer, R., Allott, B. M., Núñez-Pertíñez, S., Wilks, T. R., Cox, L. R., Bath, J., O'Reilly, R. K. & Turberfield, A. J. A New Architecture for DNA-Templated Synthesis in Which Abasic Sites Protect Reactants from Degradation. *Angewandte Chemie* **136,** e202317482 (2024).

44. Barsky, D., Foloppe, N., Ahmadia, S., Wilson III, D. M. & MacKerell Jr, A. D. New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic acids research* **28,** 2613–2626 (2000).

45. Unruh, J. R., Gokulrangan, G., Lushington, G., Johnson, C. K. & Wilson, G. S. Orientational dynamics and dye-DNA interactions in a dye-labeled DNA aptamer. *Biophysical journal* **88,** 3455–3465 (2005).

46. Cai, X., Gray Jr, P. J. & Von Hoff, D. D. DNA minor groove binders: back in the groove. *Cancer treatment reviews* **35,** 437–450 (2009).

47. Kanarskaya, M. A., Pyshnyi, D. V. & Lomzov, A. A. Diversity of Self-Assembled RNA Complexes: From Nanoarchitecture to Nanomachines. *Molecules* **29,** 10 (2023).

48. Chen, Z., Liu, Y. & Sun, H. Physics-informed learning of governing equations from scarce data. *Nature Communications* **12,** 6136 (2021).

49. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99,** 1129–1143 (2018).

50. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347,** 631–639 (1990).

51. Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W. F. & Mark, A. E. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition* **38,** 236–240 (1999).

52. Mukherjee, A., Lavery, R., Bagchi, B. & Hynes, J. T. On the molecular mechanism of drug intercalation into DNA: a simulation study of the intercalation pathway, free energy, and DNA structural changes. *Journal of the American Chemical Society* **130,** 9747–9755 (2008).

53. Wilhelm, M., Mukherjee, A., Bouvier, B., Zakrzewska, K., Hynes, J. T. & Lavery, R. Multistep drug intercalation: molecular dynamics and free energy studies of the binding of daunomycin to DNA. *Journal of the American Chemical Society* **134,** 8588–8596 (2012).

54. Lei, H., Wang, X. & Wu, C. Early stage intercalation of doxorubicin to DNA fragments observed in molecular dynamics binding simulations. *Journal of Molecular Graphics and Modelling* **38,** 279–289 (2012).

55. Galindo-Murillo, R. & Cheatham III, T. E. Ethidium bromide interactions with DNA: an exploration of a classic DNA–ligand complex with unbiased molecular dynamics simulations. *Nucleic Acids Research* **49,** 3735–3747 (2021).

56. Hansson, T., Oostenbrink, C. & van Gunsteren, W. Molecular dynamics simulations. *Current opinion in structural biology* **12,** 190–196 (2002).

57. Alder, B. J. & Wainwright, T. E. Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics* **31,** 459–466 (1959).

58. Allen, M. P. *et al.* Introduction to molecular dynamics simulation. *Computational Soft Matter: from Synthetic Polymers to Proteins* **23,** 1–28 (2004).

59. Binder, K., Horbach, J., Kob, W., Paul, W. & Varnik, F. Molecular dynamics simulations. *Journal of Physics: Condensed Matter* **16,** S429 (2004).

60. Leach, A. R. *Molecular modelling: principles and applications* (Pearson education, 2001).

61. Van Gunsteren, W. F. & Berendsen, H. J. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English* **29,** 992–1023 (1990).

62. González, M. A. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **12,** 169–200 (2011).

63. Monticelli, L. & Tieleman, D. P. Force fields for classical molecular dynamics. *Biomolecular Simulations: Methods and Protocols,* 197–213 (2013).

64. Liu, Y.-P., Kim, K., Berne, B., Friesner, R. A. & Rick, S. W. Constructing ab initio force fields for molecular dynamics simulations. *The Journal of Chemical Physics* **108,** 4739–4755 (1998).

65. Wang, X., Li, J., Yang, L., Chen, F., Wang, Y., Chang, J., Chen, J., Feng, W., Zhang, L. & Yu, K. Dmff: an open-source automatic differentiable platform for molecular force field development and molecular dynamics simulation. *Journal of Chemical Theory and Computation* **19,** 5897–5909 (2023).

66. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25,** 1157–1174 (2004).

67. Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M. & MacKerell Jr, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of Chemical Theory and Computation* **8,** 3257–3273 (2012).

68. Allred, A. L. & Rochow, E. G. A scale of electronegativity based on electrostatic force. *Journal of Inorganic and Nuclear Chemistry* **5,** 264–268 (1958).

69. George, A., Mondal, S., Purnaprajna, M. & Athri, P. Review of Electrostatic Force Calculation Methods and Their Acceleration in Molecular Dynamics Packages Using Graphics Processors. *ACS Omega* **7,** 32877–32896 (2022).

70. Schröder, C. & Steinhauser, O. The influence of electrostatic forces on the structure and dynamics of molecular ionic liquids. *The Journal of Chemical Physics* **128** (2008).

71. Sagui, C. & Darden, T. A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual Review of Biophysics and Biomolecular Structure* **28,** 155–179 (1999).

72. Belhadj, M., Alper, H. E. & Levy, R. M. Molecular dynamics simulations of water with Ewald summation for the long range electrostatic interactions. *Chemical Physics Letters* **179,** 13–20 (1991).

73. Wells, B. A. & Chaffee, A. L. Ewald summation for molecular simulations. *Journal of Chemical Theory and Computation* **11,** 3684–3695 (2015).

74. Luty, B. A. & van Gunsteren, W. F. Calculating electrostatic interactions using the particle- particle particle- mesh method with nonperiodic long-range interactions. *The Journal of Physical Chemistry* **100,** 2581–2587 (1996).

75. Dzyaloshinskii, I. E., Lifshitz, E. M. & Pitaevskii, L. P. The general theory of van der Waals forces. *Advances in Physics* **10,** 165–209 (1961).

76. Margenau, H. Van der Waals forces. *Reviews of Modern Physics* **11,** 1 (1939).

77. Slater, J. C. & Kirkwood, J. G. The van der Waals forces in gases. *Physical Review* **37,** 682 (1931).

78. Tee, L. S., Gotoh, S. & Stewart, W. E. Molecular parameters for normal fluids. Lennard-Jones 12-6 Potential. *Industrial & Engineering Chemistry Fundamentals* **5,** 356–363 (1966).

79. Cuadros, F., Cachadiña, I. & Ahumada, W. Determination of Lennard-Jones interaction parameters using a new procedure. *Molecular Engineering* **6,** 319–325 (1996).

80. Ricci, C. G., de Andrade, A. S., Mottin, M. & Netz, P. A. Molecular dynamics of DNA: comparison of force fields and terminal nucleotide definitions. *The Journal of Physical Chemistry B* **114,** 9882–9893 (2010).

81. Pérez, A., Luque, F. J. & Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Accounts of Chemical Research* **45,** 196–205 (2012).

82. Galindo-Murillo, R., Robertson, J. C., Zgarbova, M., Sponer, J., Otyepka, M., Jurecka, P. & Cheatham III, T. E. Assessing the current state of amber force field modifications for DNA. *Journal of Chemical Theory and Computation* **12,** 4114–4127 (2016).

83. Hart, K., Foloppe, N., Baker, C. M., Denning, E. J., Nilsson, L. & MacKerell Jr, A. D. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *Journal of Chemical Theory and Computation* **8,** 348–362 (2012).

84. Dans, P. D., Ivani, I., Hospital, A., Portella, G., González, C. & Orozco, M. How accurate are accurate force-fields for B-DNA? *Nucleic Acids Research* **45,** 4217–4230 (2017).

85. Bignon, E., Chan, C.-H., Morell, C., Monari, A., Ravanat, J.-L. & Dumont, E. Molecular Dynamics Insights into Polyamine–DNA Binding Modes: Implications for Cross-Link Selectivity. *Chemistry–A European Journal* **23,** 12845–12852 (2017).

86. Palanisamy, K., Prakash, M. & Rajapandian, V. Combined DFT and MD simulation studies of protein stability on imidazolium–water (ImH+ W n) clusters with aromatic amino acids. *New Journal of Chemistry* **44,** 17912–17923 (2020).

87. Qin, W., Li, X., Bian, W.-W., Fan, X.-J. & Qi, J.-Y. Density functional theory calculations and molecular dynamics simulations of the adsorption of biomolecules on graphene surfaces. *Biomaterials* **31,** 1007–1016 (2010).

88. Spinello, A., Terenzi, A. & Barone, G. Metal complex–DNA binding: Insights from molecular dynamics and DFT/MM calculations. *Journal of Inorganic Biochemistry* **124,** 63–69 (2013).

89. Mori, M., Dietrich, U., Manetti, F. & Botta, M. Molecular dynamics and DFT study on HIV-1 nucleocapsid protein-7 in complex with viral genome. *Journal of Chemical Information and Modeling* **50,** 638–650 (2010).

90. Winter, N. W., Laferrière, A. & McKoy, V. Numerical solution of the two-electron Schrödinger equation. *Physical Review A* **2,** 49 (1970).

91. Sinanoğlu, O. Many-Electron Theory of Atoms and Molecules. II. *The Journal of Chemical Physics* **36,** 3198–3208 (1962).

92. Sherrill, C. D. The born-oppenheimer approximation. *School of Chemistry and Biochemistry, Georgia Institute of Technology* **242** (2005).

93. Baer, M., Lin, S. H., Alijah, A., Adhikari, S. & Billing, G. D. Extended approximated Born-Oppenheimer equation. I. Theory. *Physical Review A* **62,** 032506 (2000).

94. Davidson, E. R. & Feller, D. Basis set selection for molecular calculations. *Chemical Reviews* **86,** 681–696 (1986).

95. Froyen, S. & Harrison, W. A. Elementary prediction of linear combination of atomic orbitals matrix elements. *Physical Review B* **20,** 2420 (1979).

96. Magalhaes, A. L. Gaussian-type orbitals versus slater-type orbitals: a comparison. *Journal of Chemical Education* **91,** 2124–2127 (2014).

97. Forster, A. & Visscher, L. GW100: A Slater-Type Orbital Perspective. *Journal of Chemical Theory and Computation* **17,** 5080–5097 (2021).

98. Scuseria, G. E. Linear scaling density functional calculations with Gaussian orbitals. *The Journal of Physical Chemistry A* **103,** 4782–4790 (1999).

99. Raghavachari, K. & Anderson, J. B. Electron correlation effects in molecules. *The Journal of Physical Chemistry* **100,** 12960–12973 (1996).

100. Cohen, A. J., Mori-Sánchez, P. & Yang, W. Challenges for density functional theory. *Chemical reviews* **112,** 289–320 (2012).

101. Koniev, O. & Wagner, A. Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chemical Society Reviews* **44,** 5495–5551 (2015).

102. DeSanctis, M. L., Soranno, E. A., Messner, E., Wang, Z., Turner, E. M., Falco, R., Appiah-Madson, H. J. & Distel, D. L. Greater than pH 8: The pH dependence of EDTA as a preservative of high molecular weight DNA in biological samples. *PLoS One* **18,** e0280807 (2023).

103. Jang, E. K., Son, R. G. & Pack, S. P. Novel enzymatic single-nucleotide modification of DNA oligomer: prevention of incessant incorporation of nucleotidyl transferase by ribonucleotide-borate complex. *Nucleic Acids Research* **47,** e102–e102 (2019).

104. Srivastava, P., Schito, M., Fattah, R. J., Hara, T., Hartman, T., Buckheit Jr, R. W., Turpin, J. A., Inman, J. K. & Appella, E. Optimization of unique, uncharged thioesters as inhibitors of HIV replication. *Bioorganic & medicinal chemistry* **12,** 6437–6450 (2004).

105. Maret, B., Regnier, T., Rossi, J.-C., Garrelly, L., Vial, L. & Pascal, R. Reduction with tris (2-carboxyethyl) phosphine (TCEP) enables the use of an S-sulphonate protecting group for thiol-mediated bioconjugation. *Rsc Advances* **4,** 7725–7728 (2014).

106. Dmitrenko, O., Thorpe, C. & Bach, R. D. Mechanism of SN2 disulfide bond cleavage by phosphorus nucleophiles. Implications for biochemical disulfide reducing agents. *The Journal of organic chemistry* **72,** 8298–8307 (2007).

107. Han, J. C. & Han, G. Y. A procedure for quantitative determination of tris (2-carboxyethyl) phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Analytical Biochemistry* **220,** 5–10 (1994).

108. Grimm, J. B., Tkachuk, A. N., Xie, L., Choi, H., Mohar, B., Falco, N., Schaefer, K., Patel, R., Zheng, Q., Liu, Z., *et al.* A general method to optimize and functionalize red-shifted rhodamine dyes. *Nature Methods* **17,** 815–821 (2020).

109. Plumridge, A., Meisburger, S. P., Andresen, K. & Pollack, L. The impact of base stacking on the conformations and electrostatics of single-stranded DNA. *Nucleic Acids Research* **45,** 3932–3943 (2017).

110. Li, Q., Liang, X., Mu, X., Tan, L., Lu, J., Hu, K., Zhao, S. & Tian, J. Ratiometric fluorescent 3D DNA walker and catalyzed hairpin assembly for determination of microRNA. *Microchimica Acta* **187,** 1–8 (2020).

111. Schmitt, T. J. & Knotts, T. A. Thermodynamics of DNA hybridization on surfaces. *The Journal of chemical physics* **134** (2011).

112. Blake, R. & Delcourt, S. G. Thermal stability of DNA. *Nucleic acids research* **26,** 3323–3332 (1998).

113. Bishop, G. R. & Chaires, J. B. Characterization of DNA structures by circular dichroism. *Current protocols in nucleic acid chemistry* **11,** 7–11 (2002).

114. Wijmenga, S. S. & Van Buuren, B. N. The use of NMR methods for conformational studies of nucleic acids. *Progress in nuclear magnetic resonance spectroscopy* **32,** 287–387 (1998).

115. Lukin, M. & de Los Santos, C. NMR structures of damaged DNA. *Chemical reviews* **106,** 607–686 (2006).

116. Liu, Q., Liu, M., Jin, Y. & Li, B. Ratiometric fluorescent probe: a sensitive and reliable reporter for the CRISPR/Cas12a-based biosensing platform. *Analyst* **147,** 2567–2574 (2022).

117. Xu, W., Chan, K. M. & Kool, E. T. Fluorescent nucleobases as tools for studying DNA and RNA. *Nature chemistry* **9,** 1043–1055 (2017).

118. Marin-Gonzalez, A., Vilhena, J., Perez, R. & Moreno-Herrero, F. A molecular view of DNA flexibility. *Quarterly Reviews of Biophysics* **54,** e8 (2021).

119. Howorka, S., Movileanu, L., Braha, O. & Bayley, H. Kinetics of duplex formation for individual DNA strands within a single protein nanopore. *Proceedings of the National Academy of Sciences* **98,** 12996–13001 (2001).

120. Brittany, R., Elizabeth, M., Rachel, C., Kara, O., Elisabeth, W., Rachel, S., *et al.* Kinetics and Thermodynamics of DNA, RNA, and Hybrid Duplex Formation (2013).

121. Kolbeck, P. J., Vanderlinden, W., Gemmecker, G., Gebhardt, C., Lehmann, M., Lak, A., Nicolaus, T., Cordes, T. & Lipfert, J. Molecular structure, DNA binding mode, photophysical properties and recommendations for use of SYBR Gold. *Nucleic acids research* **49,** 5143–5158 (2021).

122. Agbavwe, C. & Somoza, M. M. Sequence-dependent fluorescence of cyanine dyes on microarrays. *PloS one* **6,** e22177 (2011).

123. Seio, K., Kanamori, T. & Masaki, Y. Solvent-and environment-dependent fluorescence of modified nucleobases. *Tetrahedron Letters* **59,** 1977–1985 (2018).

124. Massey, M., Algar, W. R. & Krull, U. J. Fluorescence resonance energy transfer (FRET) for DNA biosensors: FRET pairs and Förster distances for various dye–DNA conjugates. *Analytica chimica acta* **568,** 181–189 (2006).

125. Shankaraiah, N., Siraj, K., Nekkanti, S., Srinivasulu, V., Sharma, P., Senwar, K. R., Sathish, M., Vishnuvardhan, M., Ramakrishna, S., Jadala, C., *et al.* DNA-binding affinity and anticancer activity of β-carboline–chalcone conjugates as potential DNA intercalators: Molecular modelling and synthesis. *Bioorganic chemistry* **59,** 130–139 (2015).

126. Stephenson, C. J. & Shimizu, K. D. A fluorescent diastereoselective molecular sensor for 1, 2-aminoalcohols based on the rhodamine B lactone–zwitterion equilibrium. *Organic & Biomolecular Chemistry* **8,** 1027–1032 (2010).

127. Tyson, J., Hu, K., Zheng, S., Kidd, P., Dadina, N., Chu, L., Toomre, D., Bewersdorf, J. & Schepartz, A. Extremely bright, near-IR emitting spontaneously blinking fluorophores enable ratiometric multicolor nanoscopy in live cells. *ACS Central Science* **7,** 1419–1426 (2021).

128. Price, D. J. & Brooks III, C. L. A modified TIP3P water potential for simulation with Ewald summation. *The Journal of Chemical Physics* **121,** 10096–10103 (2004).

129. Zhou, H.-X. & Pang, X. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical Reviews* **118,** 1691–1741 (2018).

130. Cherstvy, A. Electrostatic interactions in biological DNA-related systems. *Physical Chemistry Chemical Physics* **13,** 9942–9968 (2011).

131. Winogradoff, D., Li, P.-Y., Joshi, H., Quednau, L., Maffeo, C. & Aksimentiev, A. Chiral systems made from DNA. *Advanced Science* **8,** 2003113 (2021).

132. Tuckerman, M. E., Alejandre, J., López-Rendón, R., Jochim, A. L. & Martyna, G. J. A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal–isobaric ensemble. *Journal of Physics A: Mathematical and General* **39,** 5629 (2006).

133. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **4,** 435–447 (2008).

134. Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science* **330,** 341–346 (2010).

135. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the national academy of sciences* **99,** 12562–12566 (2002).

136. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **314,** 141–151 (1999).

137. Ormeño, F. & General, I. J. Convergence and equilibrium in molecular dynamics simulations. *Communications Chemistry* **7,** 26 (2024).

138. Wojtas-Niziurski, W., Meng, Y., Roux, B. & Bernèche, S. Self-learning adaptive umbrella sampling method for the determination of free energy landscapes in multiple dimensions. *Journal of chemical theory and computation* **9,** 1885–1895 (2013).

CHAPTER 3

Programming colloidal self-assembly pathways for simple cubic crystals

3.1 | Functionalisation and Implementation of Colloids as Building Blocks

Many desirable functional materials have a structure defined by intricate nano-architectures. These materials have applications in photonics,[1,2] biomedicine,[3] and energy conversion domains.[4–6] Nature has been discovered to employ bottom-up hierarchical self-assembly in order to construct three-dimensional nano-architectures. This design process underpins the construction of sophisticated systems such as muscle tissue[7], amyloid fibrils[8], and collagen networks[9]. Sceintists have employed colloidal particles to manifest hierarchical self-assembly.[10–13]

Colloidal particles, as nanometre-to-micrometre-sized entities dispersed in a liquid medium, are attractive building blocks for programming self-assembly.[14] The interplay between repulsive and attractive interactions typically governs the self-assembly of colloidal building blocks, though entropy-driven self-assembly in the presence of excluded volume interactions alone is also known.[15] A variety of colloidal building blocks are now synthetically available, often offering directional and/or specific interaction to encode the self-assembly information.[16–18]. The choice of DNA-mediated interactions between colloidal particles to drive colloidal self-assembly has gained a lot of traction.[19,20] Due to the high specificity of Watson and Crick base pairing and the thermal reversibility of the interactions,[21,22] it has been successfully utilised in programming colloidal self-assembly. Recent studies have shown the versatility of using DNA as a cross-linker in colloidal crystals, leading to varied structural outcomes.[23–25]

Different strategies employ different DNA configurations to drive the assembly of these colloids. Previous research emphasizes factors like DNA chain length, nucleobase sequence, and

structural ratio of single-stranded DNA segments in determining the self-assembly dynamics.[11] Additionally, other interactions between DNA chains, especially steric and electrostatic repulsion of the phosphate backbone, play a role in the formation of colloidal crystals, leading to a wide range of lattice symmetries.[26,27]

Colloidal self-assembly encompasses the arrangement of particles into intricate structures. Such superstructures, influenced by individual nanoparticle properties and their collective attributes, have shown tremendous sensitivity to environmental changes.[28,29] This paves the way for creating stimuli-responsive materials and intelligent devices.[30,31] As advancements continue to flourish, we observe a refined control over the dynamic reactions of these materials to various stimuli, revealing the untapped potential of colloidal self-assembly, especially when intertwined with DNA functionalisation.[19,32,33]

Colloidal particles with short-ranged isotropic interactions tend to close-pack. Colloidal open crystals, having densities significantly lower than what is achieved at close packing, have proved challenging targets for programming colloidal self-assembly. In particular, diamond-structured colloidal crystals are targeted for self-assembly because of their ability to support a complete photonic band gap (PBG), and thus applications as photonic crystals at optical frequencies.[34] A simple cubic structure is also known to support a complete PBG, albeit with a smaller maximum gap to midgap ratio at a higher refractive index contrast than what is offered by the much sough-after cubic diamond.[2,35] Additionally, the simple cubic crystal structure possesses useful mechanical properties, such as those of auxetic metamaterials,[36] making it an attractive target for self-assembly.

Patchy colloidal particles with sticky parts on their surface provide highly directional interactions, which can be made specific as well *via* DNA functionalisation of the patches.[37–39] Owing to their highly directional interactions, patchy colloidal particles are promising building blocks for low-coordinated colloidal open crystals. In fact, colloidal spheres with six attractive patches in an octahedral arrangement – referred to as octahedral patchy particles hereafter – are intuitively obvious building blocks to stabilise a simple cubic crystal. In a computational study mapping the phase diagram of a model octahedral patchy particles, the simple cubic crystal was indeed confirmed to be most stable at low pressures for a relatively narrow value of the patch width.[40]

In recent years, there has been significant progress in programming hierarchical self-assembly of colloidal building blocks, which may open up alternative pathways to improve assembly kinetics. The Chakrabarti group has devised a design framework to encode hierarchical self-

assembly pathways in triblock patchy particles, having attractive patches on opposite poles, *via* a hierarchy of strengths for patch-patch interactions, and exploited the design framework for the self-assembly of diamond-structured colloidal crystals in a series of computational studies.[10,41,42] The two-stage self-assembly pathways via tetrahedra are shown to promote crystallisation via even-member ring selection – a strategy that is subsequently implemented *via* specificity of interactions to achieve facile self-assembly of colloidal diamond from a binary mixture of tetrahedral patchy particles for a wider range of patch widths than what is possible for a one-component tetrahedral patchy particles.[43] In this context, we sought to program different self-assembly pathways in octahedral patchy particles by distinguishing its four patches on a plane, referred to as *equatorial patches* hereafter and labelled E, from the remaining two *axial* patches, labelled A, across the plane, depending upon how they interact amongst themselves. We considered two scenarios: one where E-E bonds are stronger than A-A bonds, which we call equatorial bias, and the other where A-A bonds are stronger than E-E bonds corresponding to axis bias. In both case, E-A bonds are allowed, but are of an intermediate strength. These two scenarios were compared with the so-called control scenario, where the patches are not distinguished in the absence of any bias. Our objective was to assess whether biasing improves the quality of the crystal thus self-assembled as opposed to the control scenario and investigate the crystallisation pathways in those distinct scenarios.

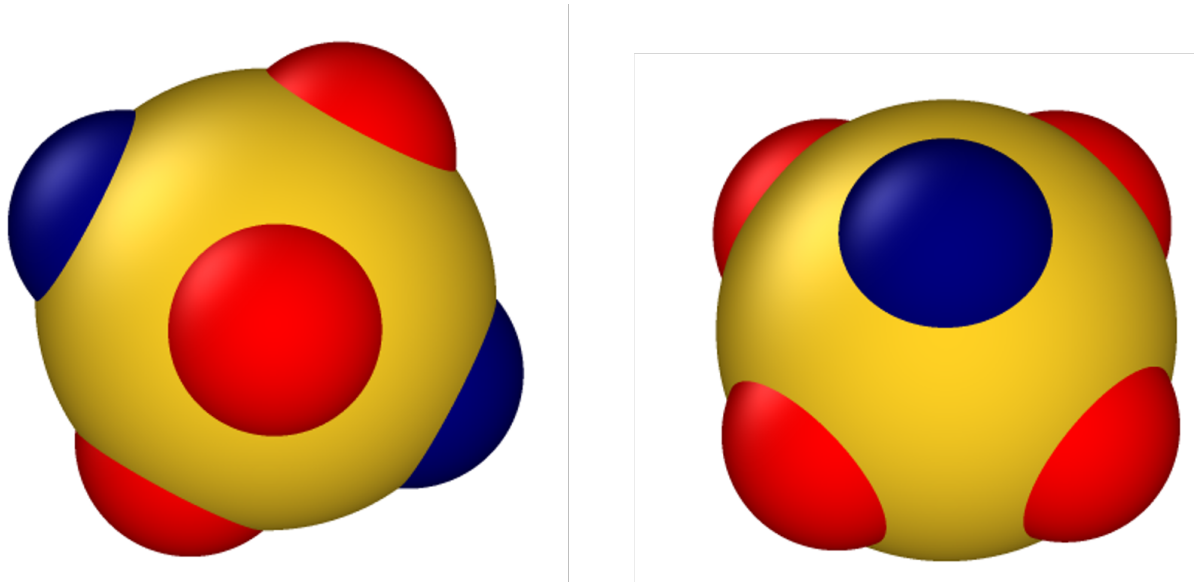


Fig. 3.1. This figure shows two perspectives of the theoretical octahedral patchy particle used in these simulations. The blue patches represent axial (A) patches, the red represent equatorial (e) patches.

3.2 | Modelling Octahedral Patchy Particle

As previously stated, octahedral patchy particles were employed in this research. With interactions between particles conducted via the Kern-Frenkel potential scheme to model the self-assembly of octahedral patchy particles. This approach combines two mathematical frameworks: the first treats the particles as hard spheres, while the second captures the patch interaction between the patchy particles as a square-well potential with an angular condition. For bonding interactions to occur in the Kern-Frenkel model, two conditions must be met: The center-to-center distance between the particles is less than the distance δ. The angles between a pair of patches lie within the half-angle θ of each particle. The potential energy U_{ij}^{kf} is given by:

$$U_{ij}^{kf}(r_{ij}, \Omega_j, \Omega_i) = \sum_{\alpha=1}^{M} \sum_{\beta=1}^{M} U_{\alpha\beta}^{SW}(r_{ij}) \cdot f_{\alpha\beta}(r_{ij}, \Omega_j, \Omega_i)$$

where $r_{ij} = |r_{ij}|$ is the centre-to-centre distance between particles i and j, and $d\Omega_i$ and Ω_j describe the orientations of particles i and j, respectively.

$$U_{ab}^{SW}(r_{ij}) = \begin{cases} \infty & \text{if } r_{ij} < \sigma \\ -\epsilon & \text{if } \sigma \leq r_{ij} < \sigma + \delta \\ 0 & \text{otherwise} \end{cases}$$

In the present study, we used reduced units: the length in units of σ, the energy in units of ε, and the temperature in units of $\frac{\varepsilon}{k_B}$.

$$f(r_{ij}, \Omega_i, \Omega_j) = \begin{cases} 1 & \text{if } \Omega_i \cdot \hat{r}_{ij} > \cos\theta \text{ and } \Omega_j \cdot \hat{r}_{ji} > \cos\theta, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

The interaction energy bias set between axial and equatorial patches was at a ratio of 5:1. The investigated energetic regimes included equatorial bias, axial bias, and a neutral control scenario, each at particle densities of 0.1, 0.3, and 0.5. For the density of 0.5, we further delineated the simulations into scenarios with active E-A interactions and those with A-B interactions switched off, where E-A denotes the interplay between dissimilar patches—equatorial and axial.

The patches were defined as being either axial or equatorial. With a bias set at a ratio of 1:5 between the two groups of patches for the different regimes, i.e., for the axial bias regime the patch strengths would equal $\epsilon_{AA} = 1, \quad \epsilon_{EE} = 0.2$. An additional control scenario was employed where the patch strengths were set to equal $\epsilon_{AA} = 1, \quad \epsilon_{EE} = 1$. The interaction range of the patches set to a spherical cutoff of 1.2σ, and half angle of $20°$.

NVT Monte Carlo simulation were carried out employing N $= 2000$ octahedral patchy articles, in a cubic box applying periodic boundary conditions adhering to the minimum image convention. The acceptance ratio was set to 0.3 with the step size adjusting accordingly to achieve such acceptance. The starting configuration of the system at T*=1 pre-organised as simple cubic. This simple cubic structure was melted during the simulation of T*=1, with the resulting liquid progressively cooled. The reduced temperature was reduced with simulations spanning from 1 down to 0.2, across which 10×10^6 Monte Carlo steps were conducted, inclusive of 5×10^6 steps for system equilibration. Analysis of the system's properties was facilitated by block averaging after every 5×10^4 Monte Carlo steps. At reduced temperatures below 0.2, the simulation length was extended to 2.5×10^7 Monte Carlo steps, with a corresponding increase in equilibration steps to 2×10^7, while maintaining the same block averaging interval. Potential energy was monitored for divergence every 1×10^5 steps, with a set divergence threshold of 1×10^{-10}. The system was confirmed to be equilibrated, provided the potential energy demonstrated no significant fluctuations in energy.

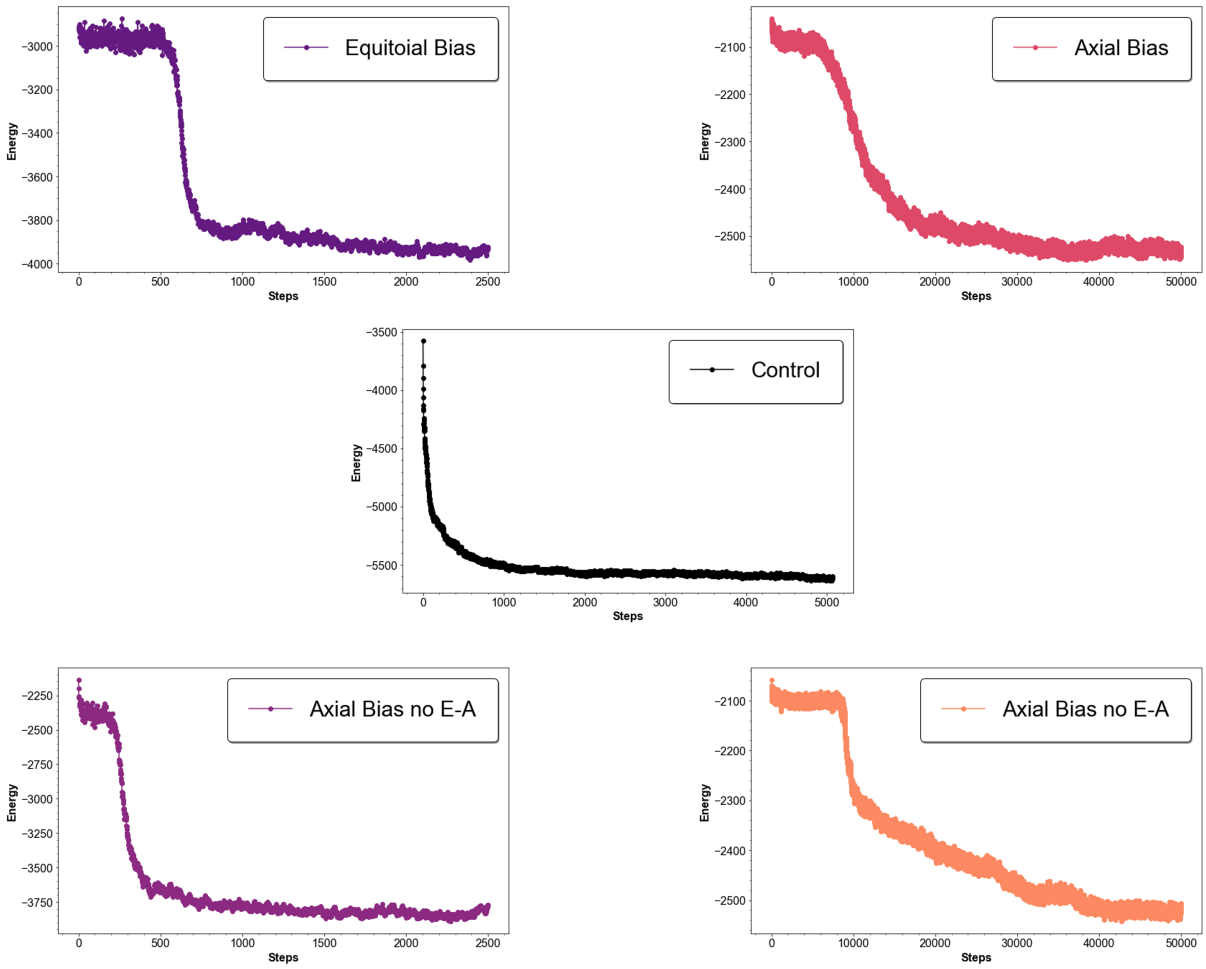


Fig. 3.2. The instantaneous energy of the system corresponding to the reduced temperature of crystallisation over monte carlo steps for the various biasing regimes. Equatorial bias T* = 0.16, Axial bias T*=0.08, Control T* = 0.2, Equatorial bias no E-A T* = 0.15, Axial bias no E-A T* = 0.07

The analysis of the resulting structure, using the global Steinhardt bond order parameter Q_4,[44,45] confirmed the self-assembly of a simple cubic crystal. Note that $Q_4 = 0.764$ and $Q_6 = 0.354$ for a perfect simple cubic lattice.[45] The global order parameter Q_4 provides information on the overall particle arrangement in the crystal, reflecting the degree of long-range order inherent in the system.

The order parameter utilises spherical harmonics to capture the angular alignment of particles within a spherical cutoff, this yields an indication of their symmetrical distribution. In the case of the simple cubic structure, the quantum number l was set to 4, representing a 4-fold, this degree of symmetry best captures the angular distribution of particles in the simple cubic lattice.

For a perfect simple cubic structure, the expected global Q_4 value is close to 0.763, which is the theoretical value for a system with complete simple cubic order.

The global Q_4 parameter values reveal that simple cubic crystals are self-assembled in a better quality at density $\rho^\star = 0.5$ compared to $\rho^\star = 0.3$ and 0.1. Comparing the various biasing scenarios with E-A allowed in terms of the global Q_4 parameter values, we infer that the equatorial bias and control scenarios form the better-quality crystals. We note that the global Q_4 parameter values in the axial bias scenario are considerably smaller. concerning the density at 0.1 and 0.3, no crystal structure resulted, however, the interconnected fluid formed at density 0.1 had a higher Q4 than the fluid at density 0.3 (0.46 as opposed to 0.33). For each scenario, five independent runs were carried out each initiated from a distinct equilibrium configuration well separated along the Monte Carlo trajectory obtained at a reduced temperature just above the temperature where crystallisation was observed. The median Q_4 values from these simulations are presented in Table 3.1.

Table 3.1
Median calculated global Q_4 parameter for the crystal resulting from density 0.5 for each regime, encompassing axial bias, equatorial bias, control case (no bias), axial bias without E-A interactions, and equatorial bias without E-A interactions. The global Q_4 is computed by averaging over 1,000 frames post-equilibration for each run with the standard deviation displayed beside.

Regime	$\langle Q_4 \rangle$	σ
Axial Bias	0.583	4.43×10^{-3}
Equatorial Bias	0.698	2.59×10^{-3}
Control	0.689	1.73×10^{-3}
Axial Bias (E-A off)	0.548	6.32×10^{-3}
Equatorial Bias (E-A off)	0.584	4.97×10^{-3}

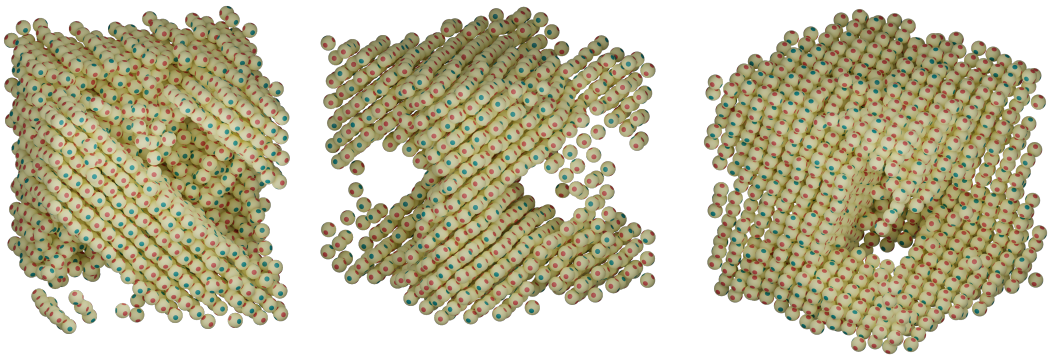


Fig. 3.3. Final frame configurations ($\rho = 0.5$) for the equatorial biasing regime ($T^* = 0.16$), control regime ($T^* = 0.2$), and axial biasing regime ($T^* = 0.07$) from left to right

.

The simulations demonstrate a clear discrepancy in the global Q_4 values with axial bias (0.583), equatorial bias (0.698), and the control case (0.689), while the E-A interaction switched-off cases show lower global Q_4 values.

3.4 | Structured Fluids Prior to Crystallisation

The different biasing scenarios as opposed to the control scenario are anticipated to result in enhanced structuring in the fluid upon cooling before the crystallisation takes place. To characterise the short-range ordering in these structured fluids and compare, we show in Figure 3.4 to 3.5 the radial distribution functions (RDFs) at temperatures just above the temperatures, at which the crystallisation was observed in our simulations. As anticipated, the peaks in the RDFs persist at larger separations in the equatorial- and axial-bias scenarios, compared to the control scenario. The enhanced structuring is also reflected in slightly more pronounced first and subsequent peaks in the two biasing scenarios. Interestingly, an additional broader peak is observed for the fluid in the equatorial-bias scenario at a separation around $r \approx 1.5$, corresponding to a sphere of nearest neighbours. Also, in this case, the third peak develops a shoulder at a distance $r \approx 2.5$ corresponding to the secondary co-ordination shell and its nearest neighbour. We therefore infer that in the equatorial-bias scenario, the octahedral patchy particles form a more structured fluid, with a tendency for the formation of 2D sheets. When these biasing scenarios are considered with E-A ruled out, the fluids have comparable structuring, though the peaks beyond the first ones appear to be slightly less pronounced in the equatorial-bias scenario compared to the axial-bias scenario. However, we note that the signatures for sheet-like coordination in a layer are absent in the equatorial bias scenario

when E-A is ruled out. This is believed to be due to the absence of any E-A interaction either catalysing or engaging in the addition of nearest neighbours.

Given that the ordering in the structured fluids, from which the crystallisation into a simple cubic crystal takes place, varies depending upon the different scenarios considered here, it is likely that the crystallisation pathways are distinct in these cases. We therefore investigated the pathways for crystallisation in those different scenarios.

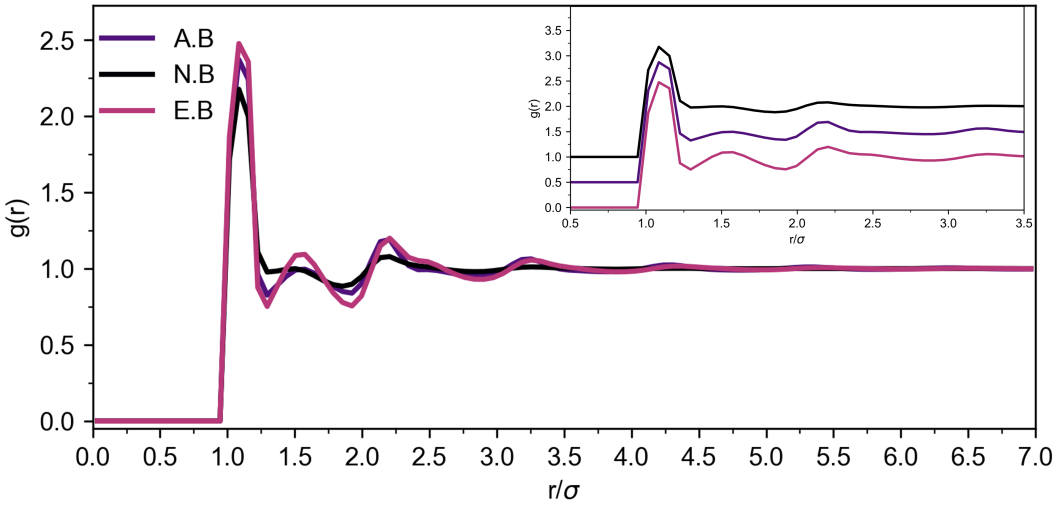


Fig. 3.4. The radial distribution function $g(r)$ of the liquid phase at the reduced temperature (T*) before crystallisation, showing the control (black line, T* = 0.3), axial bias (purple line, T* = 0.09), and equatorial bias (pink line, T* = 0.17) regimes, each distinctively separated by an increment of 0.5 for clarity.

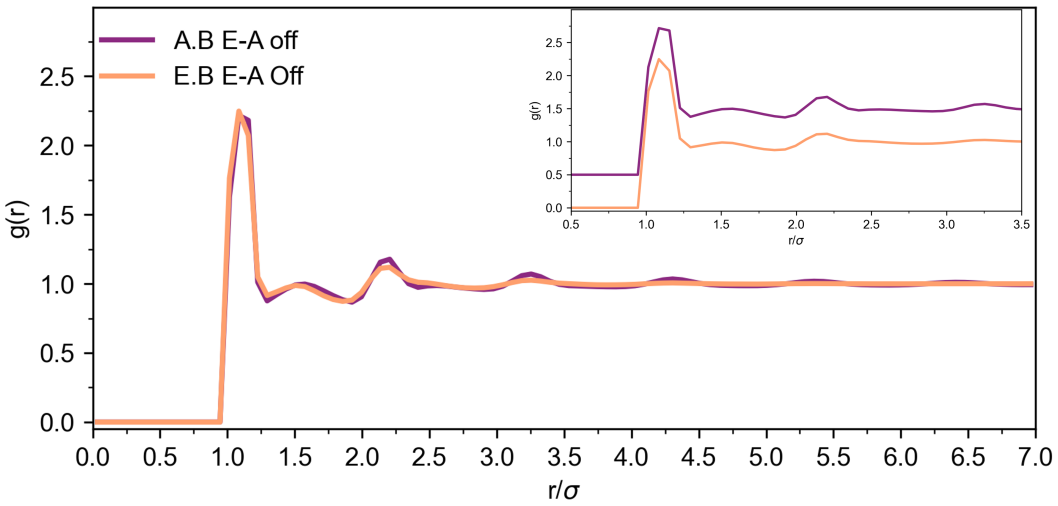


Fig. 3.5. Radial distribution function $g(r)$ for the liquid phase at the reduced temperature (T*) before crystallisation, comparing the axial bias with E-A interactions off (purple line, T* = 0.07) and the equatorial bias with E-A interactions off (orange line, T* = 0.16).

The pair distribution function and cluster analysis revealed differing structural arrangements of the colloidal particles before crystallisation. Each resulting crystal from the regimes produced simple cubic crystals of varying quality as determined by the global Q_4 and local $\bar{q}_4$ distribution. To elucidate the discrepancy in the resulting global Q_4 parameter from the density 0.5 regimes, nucleation pathway analysis was employed. This was executed using the translation bond order parameter $d_4(ij)$. It was discovered that this local bond order parameter exhibited a strongly defined signal for the simple cubic structure (Figure 3.6). Two criteria were used to determine the state of the cluster: for a cluster to be considered a crystal in this context it required $d_4(ij)$ value greater than 0.92 with each particle in that cluster forming 5-6 bonds.

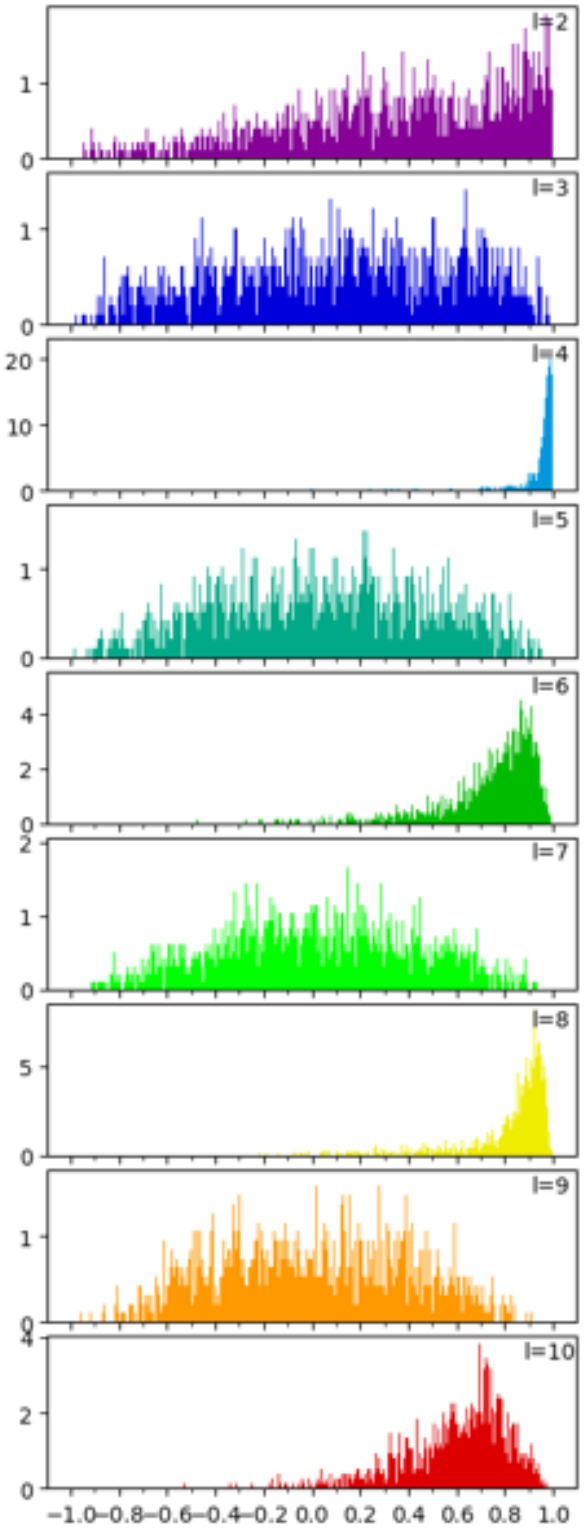


Fig. 3.6. Translation Bond Order Parameter $d_l(ij)$ for a Simple Cubic Lattice, with quantum number l values in $\{2, 3, \ldots, 10\}$.

3.5 | Calculating the Gyration Tensor

The gyration tensor was implemented to obtain insight into the shape of the cluster. The gyration tensor yields the second moments of positions for a cluster of particles. For a cluster described in 3 spatial dimensions. The gyration tensor was calculated using the central positions of each constituent colloidal particle of the largest crystalline cluster for each frame of the simulation.

$$S_{xy} = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)(y_i - y_j) \tag{3.2}$$

$$\frac{1}{2N^2} \begin{pmatrix} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)(x_i - x_j) & \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)(y_i - y_j) & \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)(z_i - z_j) \\ \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i - y_j)(x_i - x_j) & \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i - y_j)(y_i - y_j) & \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i - y_j)(z_i - z_j) \\ \sum_{i=1}^{N} \sum_{j=1}^{N} (z_i - z_j)(x_i - x_j) & \sum_{i=1}^{N} \sum_{j=1}^{N} (z_i - z_j)(y_i - y_j) & \sum_{i=1}^{N} \sum_{j=1}^{N} (z_i - z_j)(z_i - z_j) \end{pmatrix} \tag{3.3}$$

$$\frac{1}{2N^2} \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \tag{3.4}$$

where

$$S_{xy} = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)(y_i - y_j)$$

$$S_{xz} = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)(z_i - z_j)$$

$$S_{yy} = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i - y_j)(y_i - y_j)$$

This computation of the gyration tensor results in a 3X3 matrix displaying a spatial distribution of inertia, with the diagonal elements of the gyration tensor representing the moments of inertia along the principal axes of the body, while the off-diagonal elements represent the products of inertia. The diagonal elements are known as the principal moments of the gyration

tensor($\lambda_x^2 \leq \lambda_y^2 \leq \lambda_z^2$) . Using these moments parameters such as the radius of gyration, asphericity, acylindricity and anisotropy are obtained. All such values were calculated for the largest crystalline cluster for each frame.

$$R_g^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2 \tag{3.6}$$

$$b \overset{\text{def}}{=} \lambda_z^2 - \frac{1}{2}(\lambda_x^2 + \lambda_y^2) = \frac{3}{2}\lambda_z^2 - \frac{R_g^2}{2} \tag{3.7}$$

$$c \overset{\text{def}}{=} \lambda_y^2 - \lambda_x^2 \tag{3.8}$$

$$\kappa^2 \overset{\text{def}}{=} \frac{b^2 + (3/4)c^2}{R_g^4} = \frac{3}{2}\frac{\lambda_x^4 + \lambda_y^4 + \lambda_z^4}{(\lambda_x^2 + \lambda_y^2 + \lambda_z^2)^2} - \frac{1}{2} \tag{3.9}$$

3.6 | Nucleation Pathway Analysis

For each of the different scenarios, five independent Monte Carlo trajectories, which all resulted in crystallisation, were analysed to characterise the crystallisation pathways. From the analyses, a few features emerge as general trends. It is noted that the initial fluctuations in the principal moments corresponding to small cluster size, do not yield significant meaning. This is due to the small cluster size giving rise to large gyration tensors as compared to larger clusters due to the $\frac{1}{N^2}$ factor. Additionally, the accumulation of particles in these small clusters has a significant impact on the cluster's geometry, as each particle yields a relatively large contribution to the total mass distribution in the cluster. This consequently leads to significant variation in the principal moments during this phase of the simulation. The initial nucleation phase occurs around 0-1600 steps, with corresponding cluster size 0-200, and the crystal growth phase occurs after 1600. General trends :

Upon cooling to the relative crystallisation temperature of the axial and equatorial biasing scenarios, it is noted that the relative rate of crystallisation concerning Monte-Carlo steps is slower in the axial biasing case (Figure 3.7-3.9). This is attributed to the relative difference in binding energy. As the equatorial biasing features 4 patches of binding strength $(\epsilon = 1)$ as opposed to two of strength $\epsilon = 1$ in the axial scenario, consequently resulting in a stronger mean attractive force between colloids in the equatorial case. These stronger attractions between colloids perhaps facilitate a faster rate of crystallisation. The increased tendency of binding in the equatorial scenario is further supported by the higher number of average bonds formed per colloid in the liquid above crystallisation. In the equatorial scenario, an average of three bonds are formed per colloidal particle as compared to the axial scenario with two bonds. Following an initial period of rapid fluctuations, when the size of the largest crystalline cluster is rather small, both the asphericity and acylindricity of the largest crystalline cluster steadily increase before falling off in the equatorial-bias scenario (Figure 3.8). The broad maxima appear to correlate and coincide with the period when the size of the largest crystalline cluster starts to rise rapidly following an initial period of slow growth. This peak asphericity and acylindiricity are accompanied by the sharp rise in λ_z^2, λ_x^2 and λ_y^2. In contrast, in the axial-bias scenario, the acylindricity continues to rise even when the asphericity is on a slow decline or practically not changing (Figure 3.9). With this trend accompanied by the continuous rise by the λ_x^2 and λ_y^2. The consistently low simultaneous value of the asphericity and acylindircity in the equatorial scenario can be justified by the near-constant simultaneous values of the

principal moments, signifying a near three-dimensional symmetric growth. The slight increase in both asphericity and acylindricity in correspondence to the rapid increase in cluster size can be the artefact of a non-perfect simultaneous increase in the principal moments. The disparity in the trends of asphericity and acylindricity can be explained by the marginally slower rise in the λ_x^2 and λ_y^2 which can result in more an elongated geometry, distinct from the perfect cylinder.

Intriguingly large initial fluctuation in the principal moments was not observed in all nucleation studies concerning the axial basis regime (Figure 3.10). The asphericity values during this initial crystallisation phase was relatively low compared to other scenarios in this phase with the λ_z^2 moment being slightly higher than λ_y^2 and λ_x^2. Consequently the ascylindricity was near 0 indicating cylindrical symmetry during this phase. Further investigation is required to determine how this simulation involving axial biasing resulted in little fluctuations of the principal moments.

Analyses concerning the axial biasing scenario with no E-A interactions were almost indistinguishable to the axial biasing scenario with E-A interactions (Figure 3.9 and 3.11). The only discernable difference stems from the initial faster rate of growth of the largest crystalline cluster with E-A off. As no E-A interaction is enabled, the resultant binding from this mechanism results in the strongest bonding between particles. Because of this clusters formed in this regime are likely to be more one-dimensional enabling a more seamless crystallisation. Cluster analysis of the liquid phase before crystallisation highlighted differences in bonding patterns between scenarios with E-A interactions turned off and on. Specifically, the ratio of axial-axial to equatorial-equatorial bonds in the E-A off scenario was markedly higher (0.84 to 0.16) compared to the E-A on scenario, which presented a more diverse bond distribution (0.69 to 0.22, with an additional 0.1 being axial-equatorial bonds). A lower ratio of axial-axial bonds consequently implies more branching structures due to the geometric constraints of the bonds formed.

Differences between the equatorial binding scenario concerning E-A on and E-A off were more significant (Figure 3.12). Significant fluctuations in the largest crystalline cluster and all corresponding principal moments and parameters were observed in the E-A off case highlighting a large degree of hierarchical self-assembly. These large fluctuations across all parameters suggest transient three-dimensional symmetry resultant of large sub-cluster aggregation. Due to the absence of E-A interactions large sub-clusters are likely to be bound *via* continuous

stretches of relatively weaker axial bonds resulting in equilibrium concerning the bonding of the sub-clusters.

In the control scenario, proceeding with the initial sharp increase in principal moments corresponding to the initial nucleation phase, the principal moments fall and then spike in correspondence with one another as the cluster size increases (Figure 3.7). After this point the principal moments λ_z^2, λ_x^2 and λ_y^2 all plateau at 10 for the rest of the simulation. This trend signifies three-dimensional growth as the cluster size increases with all principal moments remaining approximately equal. Accompanying this, asphericity and acylindricity remain practically close to 0 through the growth of the largest crystalline cluster. In some of the simulations, significant fluctuation in the largest clusters can be observed with corresponding fluctuations in the λ_x^2 component. This is evidence of some degree of hierarchical growth, as the largest cluster can be identified to be composed of subsidiary smaller clusters. As no enthalpic bias is in place the result of cluster formation and 3-dimensional growth can be attributed to an entropic effect.

3.6.1 Largest Crystalline Cluster Analysis for the Control Scenario

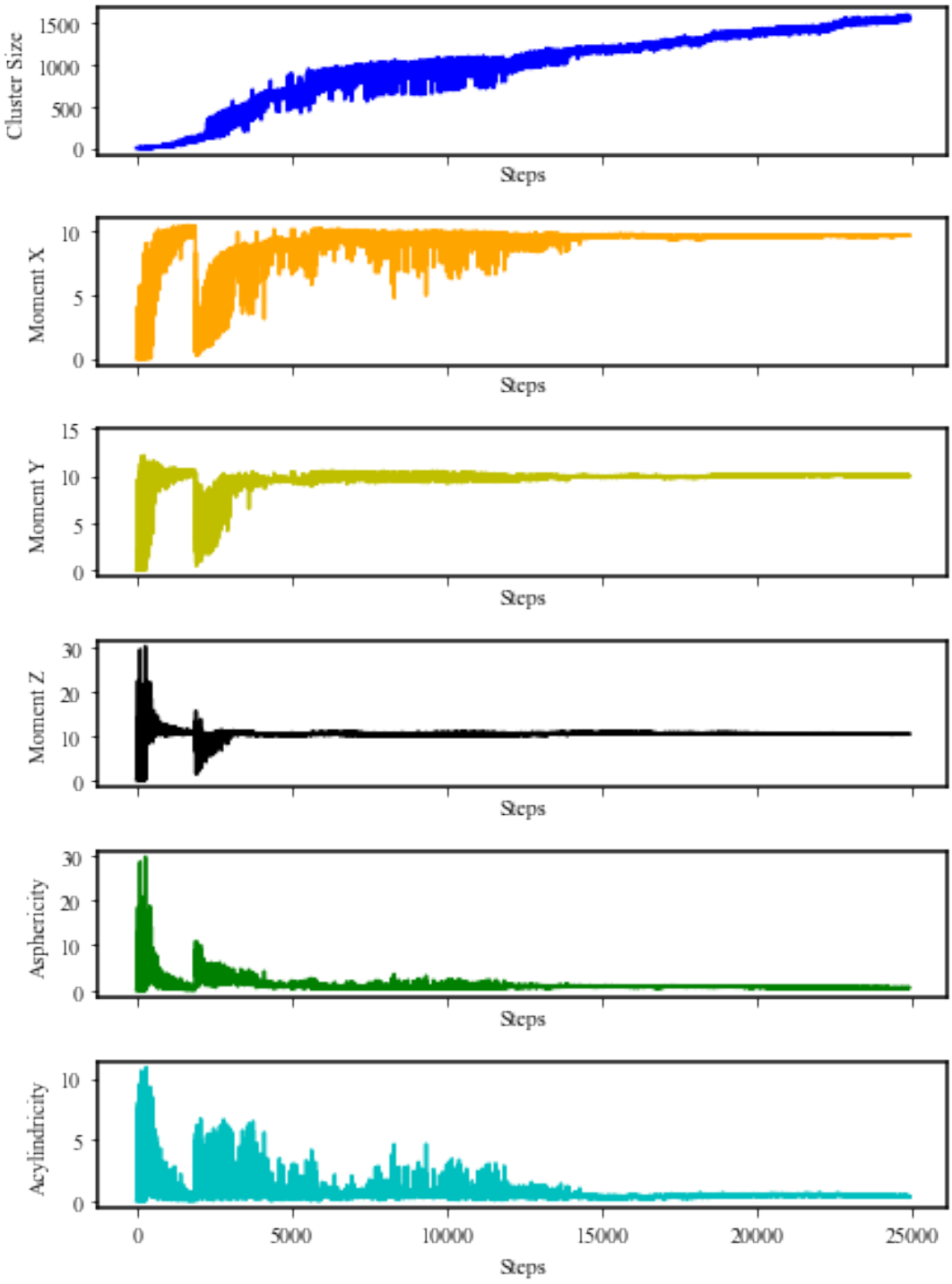


Fig. 3.7. Results of the largest crystalline cluster analysis for the control regime ($\rho = 0.5$, $T^* = 0.2$), with no energetic bias. The figure displays: the size of the largest crystalline cluster, principal moments, asphericity, and acylindricity.

3.6.2 Largest Crystalline Cluster Analysis for the Equatorial Bias Scinario

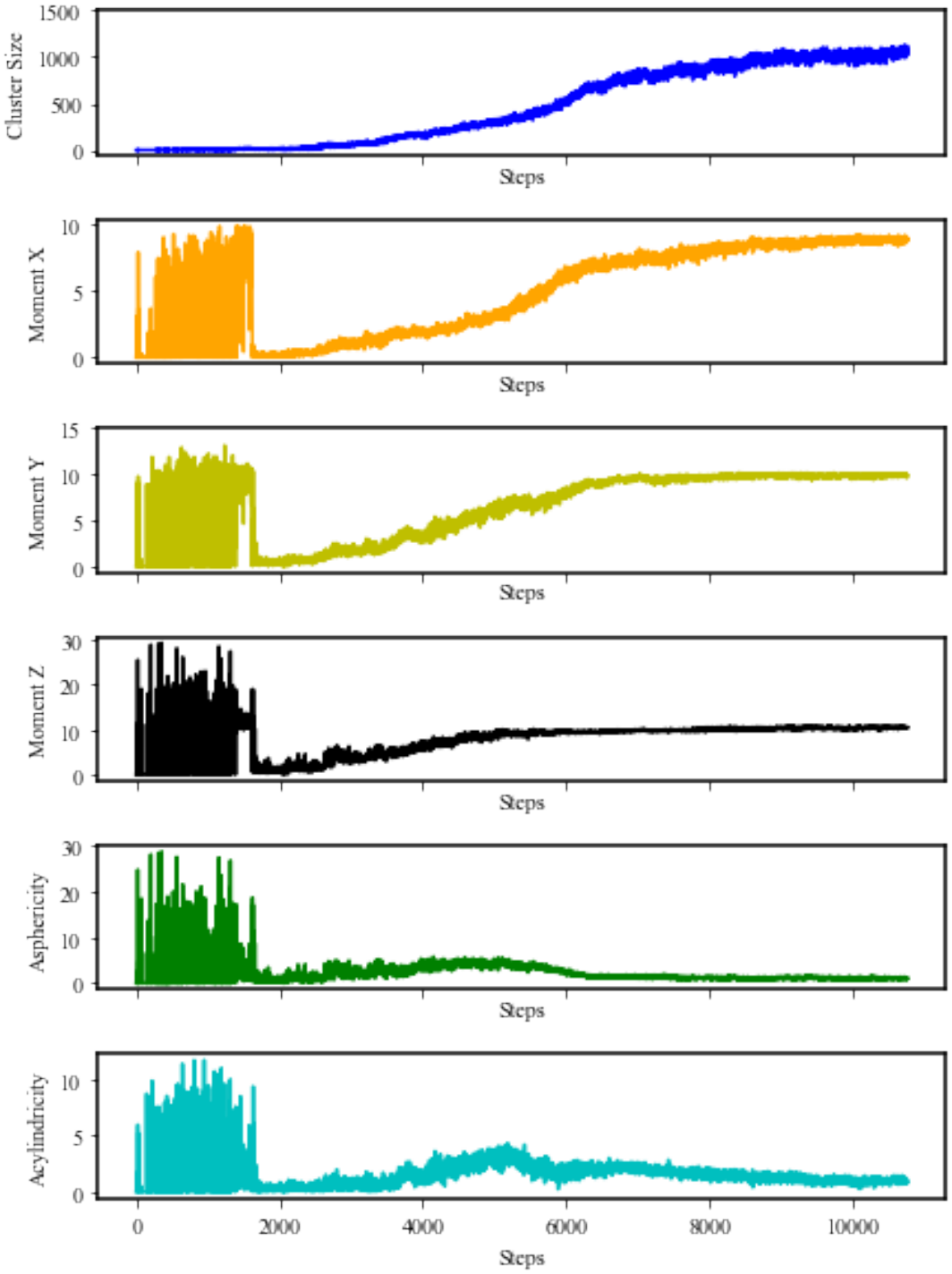


Fig. 3.8. Results of the largest crystalline cluster analysis for the equatorial bias regime at a density of($\rho = 0.5$, $T^* = 0.16$). The figure displays: the size of the largest crystalline cluster, principal moments, asphericity, and acylindricity.

3.6.3 Largest Crystalline Cluster Analysis for the Axial Bias Scenario

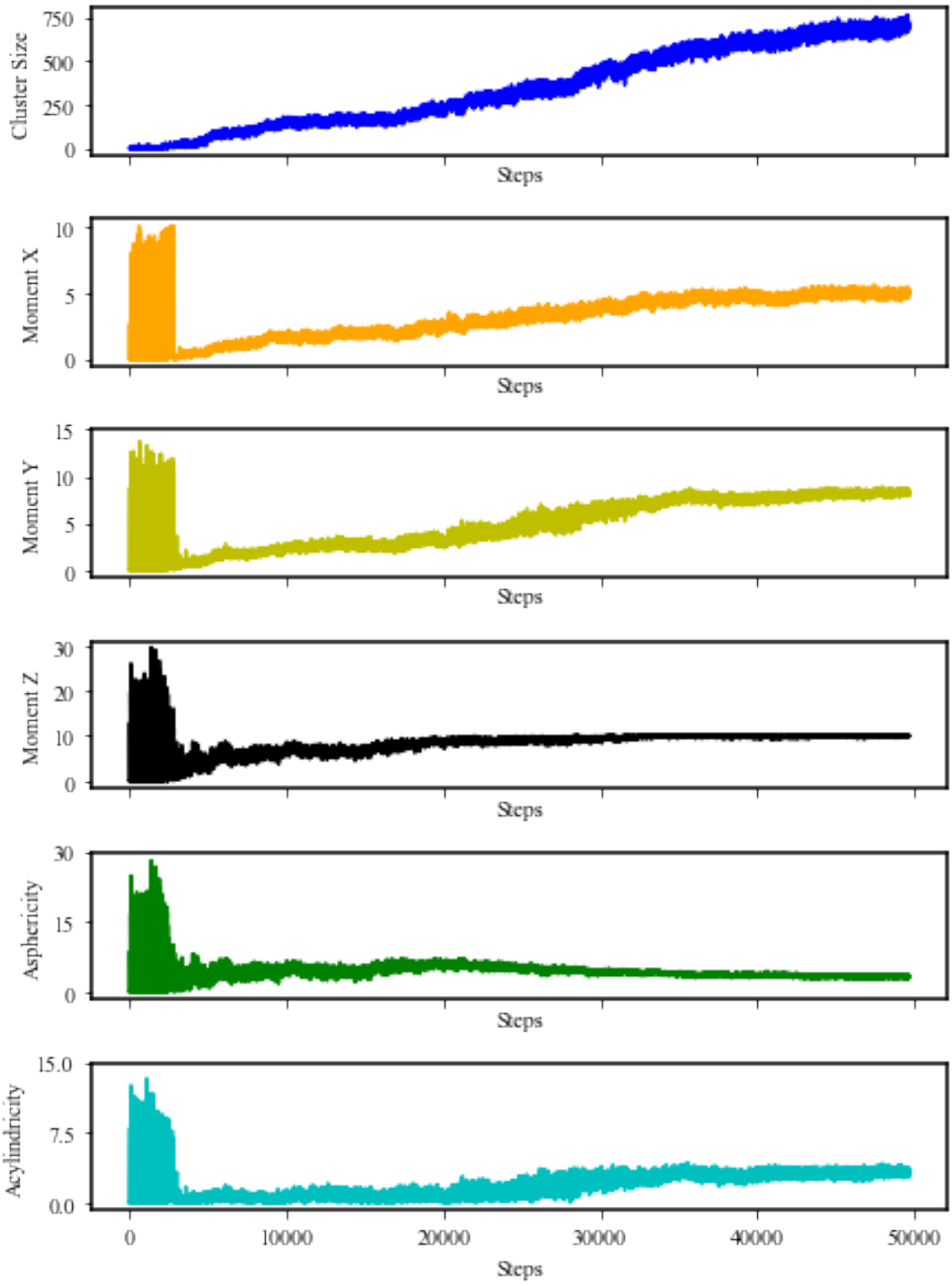


Fig. 3.9. Results of the largest crystalline cluster analysis for the axial bias regime at a density of($\rho = 0.5$, $T^* = 0.08$). The figure displays: the size of the largest crystalline cluster, principal moments, asphericity, and acylindricity.

3.6.4 Largest Crystalline Cluster Analysis for the Axial Bias Scenario no Fluctuations

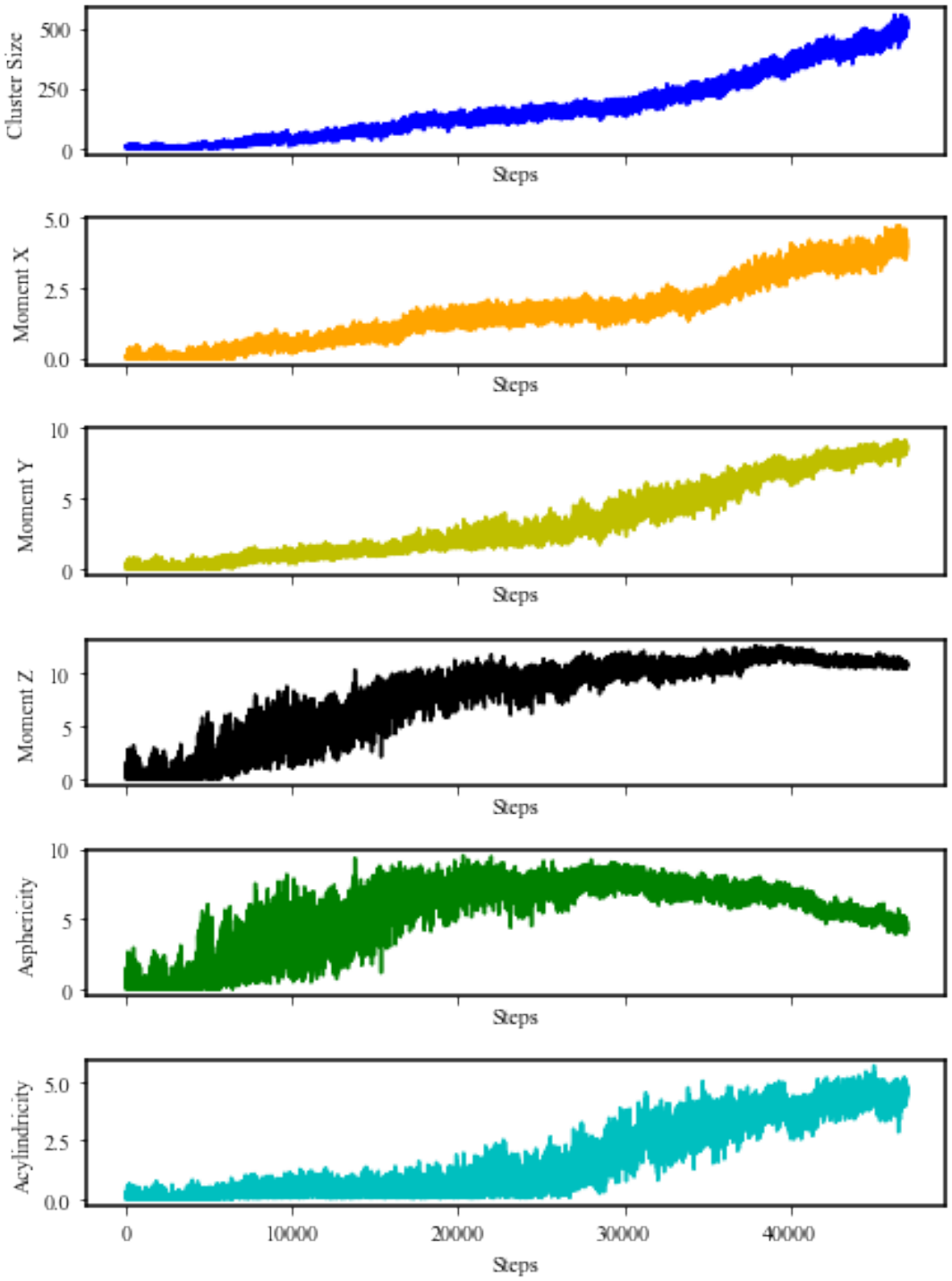


Fig. 3.10. Results of the largest crystalline cluster analysis for the axial bias regime at a density of ($\rho = 0.5$, $T^* = 0.08$) with no fluctuations in the early stage of nucleation. The figure displays: the size of the largest crystalline cluster, principal moments, asphericity, and acylindricity.

3.6.5 Largest Crystalline Cluster Analysis for the Axial Bias no E-A Scenario

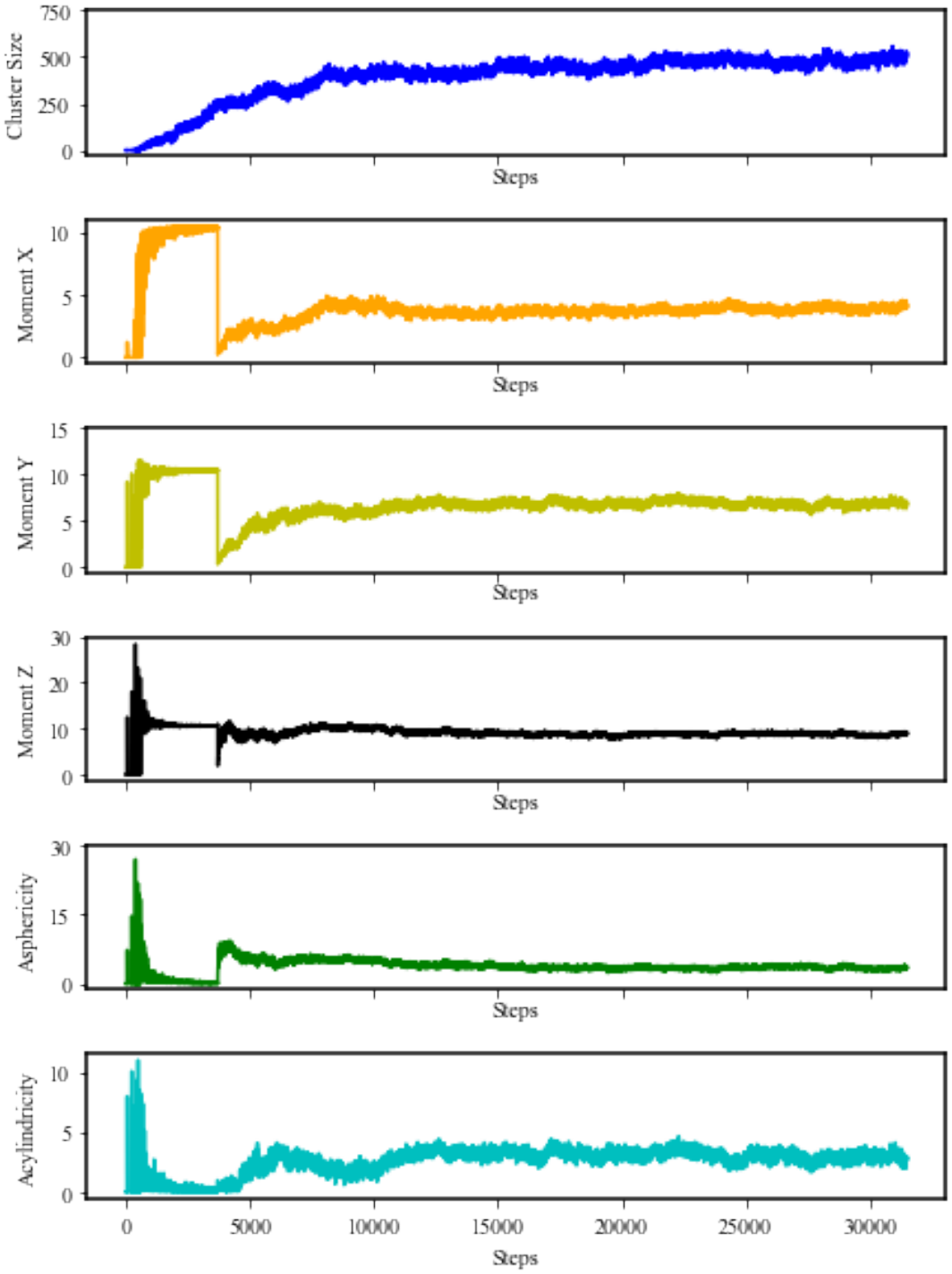


Fig. 3.11. Results of the largest crystalline cluster analysis for the axial bias regime at a density of ($\rho = 0.5$, $T^* = 0.06$) with no E-A interaction. The figure displays: the size of the largest crystalline cluster, principal moments, asphericity, and acylindricity.

3.6.6 Largest Crystalline Cluster Analysis for the Equatorial Bias no E-A Scenario

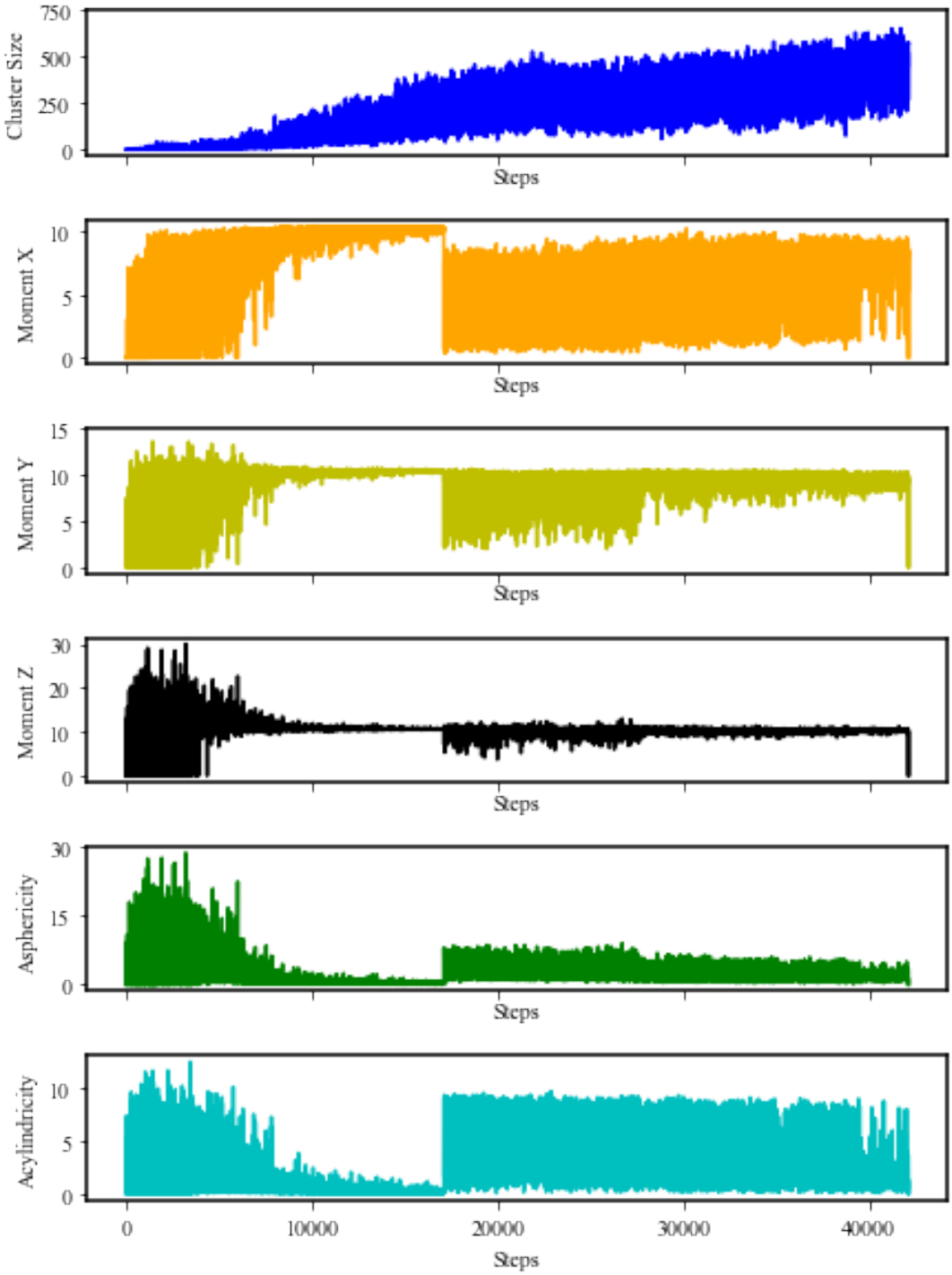


Fig. 3.12. Results of the largest crystalline cluster analysis for the equatorial bias regime at a density of ($\rho = 0.5$, $T^* = 0.15$) with no E-A interaction. The figure displays: the size of the largest crystalline cluster, principal moments, asphericity, and acylindricity.

3.7 | Conclusion

It is of interest to note that a recent study has combined theory, simulation and experiment to investigate the self-assembly of octahedral DNA origami frames with the use of different strengths for vortex-to-vortex interactions in the orthogonal directions.[46] The experimental strand of this work has been informed by Monte Carlo simulations of octahedral hard particles with sticky vertices, where vertex-vertex interactions are modelled by the Kern-Frenkel potential, and theoretical calculations of the second virial co-efficient for hard spheres decorated with six patches conforming to the octahedral symmetry. Interestingly, octahedral hard particles of two types are considered in this study, where each octahedron has four A-type and two B-type vertices. Only the vertices of the same type on particles of different types attract, with a range A_A and B_B vertex interaction energies considered while A_B vertex interaction energy set to zero. The study demonstrates that assemblies of different 3D crystalline morphologies but the same lattice symmetry, such as cube-like, sheet-like, and cylinder-like, are formed based on the relative strength of vertex-to-vertex interactions in orthogonal directions.[46]

Our studies explored how an energetic bias applied to wither axially or equatorial on a non-binary system can influence the self-assembly of colloidal simple cubic crystals from octahedral patchy particles. As compared to the control scenario, in which case no energetic bias is at play concerning the patch-patch interactions, the biasing does not significantly enhance the quality of the simple cubic crystal, thus self-assembled. Notably, the equatorial-bias scenario appears to resulted in a crystal structure of slightly better quality than the control scenario, as indicated by the calculated global Q_4 parameter. The analysis of the fluid phases from which the crystallisation occurs in the different scenarios considered in this study reveals that the crystallisation takes place from a force-structured fluid if an equatorial bias or axial bias is in operation compared to the control scenario with no bias. The equatorial bias results in the most structured fluid, which tends to form 2D sheets. The enhanced structuring in the axial-bias scenario as compared to the control scenario is attributed to the propensity of chain formation through the energetically favourable bonding between the axial patches. The crystallisation from these distinctly structured fluids is observed to follow somewhat different pathways as revealed by the growth of the largest crystalline clusters. It seems that uniform growth of the largest crystalline cluster in all three dimensions favours the formation of good-quality simple cubic crystals. The uniformity of this growth is especially affected, when an axial bias is in operation. In summary, while the biasing scenarios do not appear

to promote the self-assembly of simple cubic crystal from octahedra patchy particles of an improved quality, they give rise to different pathways for the crystal growth from distinctly structure fluids. However, we envisage the colloidal simple cubic crystals thus self-assembled to have distinct mechanical properties, the study of which should be of interest. In the axial biasing regimes, a relatively rapid progression of λ_z^2 and λ_y^2 was observed as compared to the progression of λ_x^2. This phenomenon is symptomatic of 2-dimensional growth. Comparatively, the equatorial biasing regimes demonstrated near-equal progression of all principal moments indicative of more isotropic growth of the largest crystalline cluster. Differences between the regimes were also evident in the asphericity and acylindricity trends. The axial regimes exhibited an initial asphericity increase followed by a rise in acylindiricty, this is suggestive of a cylindrical-like cluster formation in the early simulation stages. The equatorial biasing regime with no E-A interactions demonstrated a simultaneous spike in both asphericity and aclyindricity during the growth of the largest crystalline cluster before returning to 0 when the largest crystalline cluster reached 1000. This is in contrast to the equatorial bias regime with no E-A interaction which exhibited constant non-zero asphericity and acylindricity, indicating deviation from both 3d and cylindrical symmetry. Interestingly, substantial variations in the size of the largest crystalline cluster were observed in the equatorial biasing regime with E-A off, showcasing a dynamic state of binding and unbinding of multiple, large crystalline clusters throughout the simulation. These fluctuations were also present in the principal moments as well as acylindricity, asphericity and all principal moments, demonstrating the geometrical impact the large clusters had on one another. It is of note that the $\bar{q}_4$ distribution curves also demonstrated two distinct crystal environments.

A notable limitation of the A-E off scenario is the absence of cluster moves in the simulations. Cluster moves are crucial in colloidal self-assembly studies as they facilitate the exploration of larger configurational spaces, enhancing the simulation's ability to capture cooperative phenomena essential in hierarchical assembly processes. The lack of these moves potentially undermines the reliability of the results in these scenarios, as it restricts the system's ability to accurately mimic the dynamism and complexity inherent in colloidal self-assembly.

Looking ahead, the implementation of energetically biased hierarchical assembly strategies holds significant promise for constructing more complex crystal structures, such as body-centered cubic (BCC), face-centered cubic (FCC), and hexagonal close-packed (HCP) lattices. Unlike the simple cubic crystal—which is highly symmetric and has uniform bonding orientations—these more intricate structures involve varying coordination environments and complex stacking

sequences. In our studies, the simple cubic structure assembled similarly under both biased and non-biased conditions, largely due to its straightforward bonding requirements and symmetrical nature. However, BCC, FCC, and HCP lattices require precise directional interactions and higher coordination numbers, which can be more effectively controlled through energetic biasing.

By selectively enhancing specific patch interactions via energetic biasing, we could guide particles to preferentially form intermediary structures or arrangements with the correct local symmetries, which are essential for these more complex lattice formations. For example, the correct stacking order in FCC or HCP lattices, where particles must align in very specific ways, may be encouraged by energetic biasing that differentiates between different particle orientations. This would allow the system to avoid disordered intermediate states or incorrect bonding arrangements that would otherwise occur in a non-biased scenario. Additionally, hierarchical assembly with energetic biasing may favor pathways where precursor structures form initially and then evolve into the final target morphology, further improving the accuracy of self-assembly processes.

These strategies could significantly improve the quality and fidelity of assembled crystals, especially for BCC, FCC, and HCP structures, which are more sensitive to the nuances of particle interaction. Ultimately, applying energetic biasing to such systems could open new avenues for the creation of advanced materials with tailored mechanical, optical, or electronic properties. Future research into this direction could help overcome current limitations in colloidal self-assembly methodologies and lead to the precise construction of complex colloidal architectures through controlled hierarchical processes.

References

1. Takahashi, S., Suzuki, K., Okano, M., Imada, M., Nakamori, T., Ota, Y., Ishizaki, K. & Noda, S. Direct creation of three-dimensional photonic crystals by a top-down approach. *Nature Materials* **8,** 721–725 (2009).

2. Ullal, C. K., Maldovan, M., Thomas, E. L., Chen, G., Han, Y.-J. & Yang, S. Photonic crystals through holographic lithography: Simple cubic, diamond-like, and gyroid-like structures. *Applied Physics Letters* **84,** 5434–5436 (2004).

3. Leong, T. G., Randall, C. L., Benson, B. R., Bassik, N., Stern, G. M. & Gracias, D. H. Tetherless thermobiochemically actuated microgrippers. *Proceedings of the National Academy of Sciences* **106,** 703–708 (2009).

4. Klimov, V. I. Mechanisms for photogeneration and recombination of multiexcitons in semiconductor nanocrystals: Implications for lasing and solar energy conversion. *The Journal of Physical Chemistry B* **110,** 16827–16845 (2006).

5. Liang, J., Ma, F., Hwang, S., Wang, X., Sokolowski, J., Li, Q., Wu, G. & Su, D. Atomic arrangement engineering of metallic nanocrystals for energy-conversion electrocatalysis. *Joule* **3,** 956–991 (2019).

6. Liu, J., Wu, Z., Tian, Q., Wu, W. & Xiao, X. Shape-controlled iron oxide nanocrystals: synthesis, magnetic properties and energy conversion applications. *CrystEngComm* **18,** 6303–6326 (2016).

7. Gotti, C., Sensini, A., Zucchelli, A., Carloni, R. & Focarete, M. L. Hierarchical fibrous structures for muscle-inspired soft-actuators: A review. *Applied Materials Today* **20,** 100772 (2020).

8. Fitzpatrick, A. W., Debelouchina, G. T., Bayro, M. J., Clare, D. K., Caporini, M. A., Bajaj, V. S., Jaroniec, C. P., Wang, L., Ladizhansky, V., Müller, S. A., *et al.* Atomic structure and hierarchical assembly of a cross-β amyloid fibril. *Proceedings of the National Academy of Sciences* **110,** 5468–5473 (2013).

9. Gale, M., Pollanen, M. S., Markiewicz, P. & Goh, M. C. Sequential assembly of collagen revealed by atomic force microscopy. *Biophysical Journal* **68,** 2124–2128 (1995).

10. Morphew, D., Shaw, J., Avins, C. & Chakrabarti, D. Programming hierarchical self-assembly of patchy particles into colloidal crystals via colloidal molecules. *ACS Nano* **12,** 2355–2364 (2018).

11. Li, Z., Fan, Q. & Yin, Y. Colloidal self-assembly approaches to smart nanostructured materials. *Chemical reviews* **122,** 4976–5067 (2021).

12. Miszta, K., De Graaf, J., Bertoni, G., Dorfs, D., Brescia, R., Marras, S., Ceseracciu, L., Cingolani, R., Van Roij, R., Dijkstra, M., *et al.* Hierarchical self-assembly of suspended branched colloidal nanocrystals into superlattice structures. *Nature Materials* **10,** 872–876 (2011).

13. Boles, M. A., Engel, M. & Talapin, D. V. Self-assembly of colloidal nanocrystals: From intricate structures to functional materials. *Chemical Reviews* **116,** 11220–11289 (2016).

14. Cademartiri, L. & Bishop, K. J. Programmable self-assembly. *Nature Materials* **14,** 2–9 (2015).

15. Pusey, P. N. & Van Megen, W. Phase behaviour of concentrated suspensions of nearly hard colloidal spheres. *Nature* **320,** 340–342 (1986).

16. Glotzer, S. C. & Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nature Materials* **6,** 557–562 (2007).

17. Kress, R. N. & Jones, M. R. Colloidal interactions get patchy and directional. *Proceedings of the National Academy of Sciences* **117,** 15382–15384 (2020).

18. Wang, Y., Wang, Y., Breed, D. R., Manoharan, V. N., Feng, L., Hollingsworth, A. D., Weck, M. & Pine, D. J. Colloids with valence and specific directional bonding. *Nature* **491,** 51–55 (2012).

19. Di Michele, L., Varrato, F., Kotar, J., Nathan, S. H., Foffi, G. & Eiser, E. Multistep kinetic self-assembly of DNA-coated colloids. *Nature Communications* **4,** 2007 (2013).

20. Rogers, W. B., Shih, W. M. & Manoharan, V. N. Using DNA to program the self-assembly of colloidal nanoparticles and microparticles. *Nature Reviews Materials* **1,** 1–14 (2016).

21. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953).

22. Manghi, M. & Destainville, N. Physics of base-pairing dynamics in DNA. *Physics Reports* **631,** 1–41 (2016).

23. Wang, Y., Wang, Y., Zheng, X., Ducrot, É., Yodh, J. S., Weck, M. & Pine, D. J. Crystallization of DNA-coated colloids. *Nature Communications* **6,** 7253 (2015).

24. Casey, M. T., Scarlett, R. T., Benjamin Rogers, W., Jenkins, I., Sinno, T. & Crocker, J. C. Driving diffusionless transformations in colloidal crystals using DNA handshaking. *Nature Communications* **3,** 1209 (2012).

25. Girard, M., Wang, S., Du, J. S., Das, A., Huang, Z., Dravid, V. P., Lee, B., Mirkin, C. A. & Olvera de la Cruz, M. Particle analogs of electrons in colloidal crystals. *Science* **364,** 1174–1178 (2019).

26. Zhang, Y., Pal, S., Srinivasan, B., Vo, T., Kumar, S. & Gang, O. Selective transformations between nanoparticle superlattices via the reprogramming of DNA-mediated interactions. *Nature Materials* **14,** 840–847 (2015).

27. Thaner, R. V., Eryazici, I., Macfarlane, R. J., Brown, K. A., Lee, B., Nguyen, S. T. & Mirkin, C. A. The significance of multivalent bonding motifs and "bond order" in DNA-directed nanoparticle crystallization. *Journal of the American Chemical Society* **138,** 6119–6122 (2016).

28. Jiang, S., Van Dyk, A., Maurice, A., Bohling, J., Fasano, D. & Brownell, S. Design colloidal particle morphology and self-assembly for coating applications. *Chemical Society Reviews* **46,** 3792–3807 (2017).

29. Grzelczak, M., Liz-Marzán, L. M. & Klajn, R. Stimuli-responsive self-assembly of nanoparticles. *Chemical Society Reviews* **48,** 1342–1361 (2019).

30. Li, Z. & Yin, Y. Stimuli-responsive optical nanomaterials. *Advanced Materials* **31,** 1807061 (2019).

31. Ge, J. & Yin, Y. Responsive photonic crystals. *Angewandte Chemie International Edition* **50,** 1492–1522 (2011).

32. Kuzyk, A., Schreiber, R., Fan, Z., Pardatscher, G., Roller, E.-M., Högele, A., Simmel, F. C., Govorov, A. O. & Liedl, T. DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response. *Nature* **483,** 311–314 (2012).

33. Maune, H. T., Han, S.-p., Barish, R. D., Bockrath, M., Iii, W. A. G., Rothemund, P. W. & Winfree, E. Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nature Nanotechnology* **5,** 61–66 (2010).

34. Biswas, R., Sigalas, M., Subramania, G., Soukoulis, C. & Ho, K.-M. Photonic band gaps of porous solids. *Physical Review B* **61,** 4549 (2000).

35. Zavieh, L. & Mayer, T. S. Demonstration of a three-dimensional simple-cubic infrared photonic crystal. *Applied physics letters* **75,** 2533–2535 (1999).

36. Shen, J., Zhou, S., Huang, X. & Xie, Y. M. Simple cubic three-dimensional auxetic metamaterials. *Physica Status Solidi (b)* **251,** 1515–1522 (2014).

37. Pawar, A. B. & Kretzschmar, I. Fabrication, assembly, and application of patchy particles. *Macromolecular Rapid Communications* **31,** 150–168 (2010).

38. Hueckel, T., Lewis, D. J., Mertiri, A., Carter, D. J. & Macfarlane, R. J. Controlling Colloidal Crystal Nucleation and Growth with Photolithographically Defined Templates. *ACS Nano* (2023).

39. Hueckel, T., Hocky, G. M. & Sacanna, S. Total synthesis of colloidal matter. *Nature Reviews Materials* **6,** 1053–1069 (2021).

40. Noya, E., Vega, C., Doye, J. & Louis, A. Phase diagram of model anisotropic particles with octahedral symmetry. *The Journal of Chemical Physics* **127** (2007).

41. Rao, A. B., Shaw, J., Neophytou, A., Morphew, D., Sciortino, F., Johnston, R. L. & Chakrabarti, D. Leveraging hierarchical self-assembly pathways for realizing colloidal photonic crystals. *ACS nano* **14,** 5348–5359 (2020).

42. Neophytou, A., Manoharan, V. N. & Chakrabarti, D. Self-assembly of patchy colloidal rods into photonic crystals robust to stacking faults. *ACS Nano* **15,** 2668–2678 (2021).

43. Neophytou, A., Chakrabarti, D. & Sciortino, F. Facile self-assembly of colloidal diamond from tetrahedral patchy particles via ring selection. *Proceedings of the National Academy of Sciences* **118,** e2109776118 (2021).

44. Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Physical Review B* **28,** 784 (1983).

45. Rein ten Wolde, P., Ruiz-Montero, M. J. & Frenkel, D. Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. *The Journal of Chemical Physics* **104,** 9932–9947 (1996).

46. Adhikari, S., Minevich, B., Redeker, D., Michelson, A. N., Emamy, H., Shen, E., Gang, O. & Kumar, S. K. Controlling the Self-Assembly of DNA Origami Octahedra via Manipulation of Inter-Vertex Interactions. *Journal of the American Chemical Society* **145,** 19578–19587 (2023).

CHAPTER 4

Conclusion

4.1 | Overview

This thesis has discussed and explored the pivotal role of DNA in templating chemistry and assembly. We have delved into DNA's application in various synthetic strategies, specifically, its role in DNA-templated synthesis. Within this topic, the development and testing of a protection mechanism were carried out, where a unique protection was actualised involving a short-linker-TAMRA-DNA and a minor groove abasic site. In parallel to the central theme of DNA, computational sciences were explored, specifically the application of molecular dynamics and Monte Carlo sampling. Most computations were assisted by bespoke code in Python and C++ enabling the creation of files and analysis of these systems. Molecular dynamics was employed in order to uncover the structure-activity relationship which resulted in this unique protective effect. The final experimental undertaking was concerned with probing how an energetic bias of octahedral patchy particles affects the pathways of assembly, with the bias either being applied to the equatorial or axial patches.

4.2 | DNA as a Vehicle for Chemistry

The synthesis of proteins *via* ribosomes exemplifies a programmable, autonomous, and sequence-specific synthesis approach, operating effectively in a one-pot system. DNA-Templated-Synthesis (DTS) emerges as a compelling method to replicate such synthetic strategies outside the body, utilising building blocks beyond amino acids. A significant challenge in DTS is the hydrolysis of labile linkages. Our collaborative work with the Turbefield group revealed a protective effect when TAMRA interacts with a minor groove abasic site, particularly effective at positions -2 to -4. The HPLC protection assay highlighted the -3 abasic site as providing substantial protection to the labile thioester linkage.

Comparative analysis of fluorescent integrals between free and duplex-bound TAMRA indicated high protection (93%) of TAMRA-modified duplex integrity. The unique TAMRA-abasic interaction was further validated by comparing elution times between different duplexes. Extending the study to a longer linker TAMRA, broader protection was noted, suggesting significant sequestration of the thioester from the hyrodlytic environment irrespective of inclusion or position of abasic site.

However, for the synthetic viability of this protection mechanism, the shielded reactive building block must remain reactive towards the intended nucleophile. The Turberfeild group found that the protection impedes the desired reaction. Therefore necessitating a mechanism for its incorporation into the DTS context. This mechanism must enable the protection of the reactive building block and subsequent deprotection to facilitate the reaction between building blocks. Further work should be done to asses the extent of protection offer by bi-functional protecting groups. As bi-functionality will be required to ensure the protecting group is not transferred to the growing polymer-chain during DNA-templated chain transfer.

This first part of the thesis focused on DNA's utility in DTS, with a particular emphasis on the protection mechanisms involving abasic sites. The enquiry into this unique protective effect was aided by molecular dynamics.

4.3 | Molecular Dynamics of the Protective Effect in DTS

Molecular dynamics were utilised to investigate the unique protective effect of short-linker-TAMRA-DNA interacting with a minor groove abasic site. The research in modelling the intercalation of small molecules with DNA exhibited the utilisation of the GAFF and Amber parm99 bsc force fields for the small molecule and DNA respectively. Owing to this success in analogous studies both forcefields were incorporated in the modelling for the respective parts of the TAMRA-linker and DNA-duplex. It was noted that the TAMRA undergoes resonance between its lactone and zwitterionic forms. Both configurations were modelled under various salt conditions: neutral and physiological. This study focused on the -3 abasic site with the inclusion of the $+3$ site for comparison. It was uncovered that the only scenario to result in insertion in the intended abasic site involved the -3 abasic site, physiological salt concentration and the lactone resonance form. Through combining both strands of research it was concluded that the TAMRA molecule uses its lactone form to insert into the abasic site

and then undergoes resonance back to the zwitterionic form once inserted. This transition was determined due to the fluorescence response of the TAMRA-modified-duplex, as observed in the HPLC protection assay. This work could be extended in several directions. One could consider modelling the R enantiomer to deepen the specific understanding in this protective scenario. Following this umbrella sampling could be implemented to uncover the energetics associated with the binding processes. Expanding this strand with the modelling of other protective groups aside TAMRA, perhaps investigating how the degree of (π)-conjugation affects the energetic profiles of binding. Relating to the TAMRA scenario, however, the long linker could be investigated to gain insight into how this lengthening of the linker introduces a consistent and independent level of protection concerning the identity and location of the abasic site.

4.4 | Programming colloidal self-assembly pathways for simple cubic crystals

The final experimental chapter focused on exploring how an energetic bias can influence the assembly of octahedral patchy particles. Assembly of the particles was influenced either through an axial, equatorial, or no energetic bias, with binding either including or excluding E-A interactions. The attempt to promote hierarchical assembly through an energetic bias did not result in significantly greater crystal quality. The equatorial biasing with E-A interaction resulted in the best quality crystals by a small margin. Both axial biasing regimes in both E-A on and E-A off resulted in poorer quality crystals than the comparative equatorial bias scenario as per the global Q_4 parameter.

With the equatorial biasing regime E-A off resulting in an unusually fluctuating size of the largest crystalline cluster.

The radial distribution function alongside cluster analysis revealed distinct binding configurations in all scenarios. Each liquid in all regimes preceded to result in simple cubic crystal structures to varying degrees of quality. Interestingly the RDF concerning liquid of the equatorial bias revealed more sheet-like binding configuration as compared to the equatorial without E-A interactions.

To probe the mechanism of assembly, nucleation pathway analysis was employed. This indicated the geometries present in the largest crystalline cluster throughout nucleation. The general trend revealed that the axial biasing regimes resulted in more 2-dimensional growth whereas

the equatorial biasing regime resulted in more isotropic growth. This conclusion was reached through analysis of trends in principal moments, specifically, the principal moments exhibited more simultaneous growth in the equatorial biasing regimes, whereas the axial biasing exhibited substantially greater growth for the λ_Z and λ_Y as compared to the λ_X. Intriguingly the nucleation pathway analysis revealed significant fluctuations in the size of the largest crystalline cluster, thought to be because of the equilibrium binding of large sub-clusters, however further work is needed to determine this.

Extensions to this line of work should include a larger-scale simulation of 10,000 particles. The increase in simulation size can result in differing results due to several factors. One of those is enhanced statistical significance. A larger system size also offers a more detailed view of phase transition allowing for the existence of heterogeneities for a prolonged period, this consequently extends the period of phase transition allowing for greater insight.[1] Extension to this study should also incorporate cluster moves to treat clusters as one body as opposed to incorporating individual motion to all constituting particles.

Looking ahead, implementing energetically biased hierarchical assembly strategies holds significant promise for constructing more complex crystal structures such as body-centered cubic (BCC), face-centered cubic (FCC), and hexagonal close-packed (HCP) lattices. Unlike the simple cubic crystal—which assembled similarly under both biased and non-biased conditions due to its straightforward bonding requirements and high symmetry—these intricate structures involve varying coordination environments and complex stacking sequences. They require precise directional interactions and higher coordination numbers, which can be more effectively achieved through energetic biasing.

By selectively enhancing specific particle interactions, energetic biasing can guide particles to preferentially form arrangements with the correct local symmetries essential for these complex lattice formations. For example, encouraging specific stacking orders in FCC or HCP lattices—where particles must align in very particular ways—can help the system avoid disordered intermediate states or incorrect bonding arrangements that might occur without biasing. This approach can improve the accuracy and fidelity of self-assembly processes by promoting pathways where precursor structures form initially and then evolve into the final target morphology.

Applying these strategies could significantly enhance the quality of assembled crystals, especially for BCC, FCC, and HCP structures that are sensitive to the nuances of particle interactions. Ultimately, energetic biasing in hierarchical assembly could open new avenues for creating

advanced materials with tailored mechanical, optical, or electronic properties. Future research in this direction may help overcome current limitations in colloidal self-assembly methodologies, enabling the precise construction of complex colloidal architectures through controlled hierarchical processes.

4.5 | Concluding Remarks and Future Directions

An extensive scope of science was covered throughout this thesis, ranging from bio-chemistry-based wet-lab work concerning the analysis of a unique protection mechanism in the context of DTS, to computational-based enquiry ranging from molecular dynamics to modelling colloidal self-assembly with Monte Carlo simulations. The combination of experimental and computational research highlights the importance of both disciplines. Exhibiting the practicality and real-world insight gained from wet-lab work, whilst drawing attention to the great versatility and analytical capabilities of computational research. Experimental work discovered and to an extent quantified the unique protective effect whilst computational studies were implemented to elucidate the nature of the interaction and to explore pathways of colloidal self-assembly to the simple cubic structure. This thesis has explored the utility of DNA in both synthesis and assembly, through the showcase of DTS and colloidal self-assembly respectively. Furthermore, the discoveries and insights gained in this body of work lay the groundwork for future research enabling the advancement of synthesis and assembly processes.

References

1. Li, B., Zhou, D. & Han, Y. Assembly and phase transitions of colloidal crystals. *Nature Reviews Materials* **1,** 1–13 (2016).