



**POLITECHNIKA  
GDAŃSKA**  
WYDZIAŁ CHEMICZNY



Imię i nazwisko autora rozprawy: **Mateusz Kogut**  
Dyscyplina naukowa: **Chemia**

## **ROZPRAWA DOKTORSKA**

Tytuł rozprawy w języku polskim:

**Termodynamika oddziaływań i przejść strukturalnych DNA w układach wodnych na podstawie symulacji metodą dynamiki molekularnej.**

Tytuł rozprawy w języku angielskim:

**Thermodynamics of DNA interactions and structural transitions in water solutions based on molecular dynamics simulations.**

Promotor: dr hab. inż. Jacek Czub

**Gdańsk, 2019**

*Pragnę podziękować wszystkim, którzy przyczynili się do powstania tej pracy.  
W szczególności dziękuję dr hab. inż. Jackowi Czubowi oraz kolegom i koleżankom z Katedry.*

*Wszystkie obliczenia, których wyniki są przedstawione w tej pracy zostały wykonane  
w centrach obliczeniowych TASK w Gdańsku oraz Cyfronet w Krakowie.*

# Spis treści

<b>1</b>	<b>Wprowadzenie</b>	<b>1</b>
1.1	Różne formy DNA . . . . .	1
1.2	Wpływ osmolitów na DNA . . . . .	3
1.3	Historia badań nad G-kwadrupleksami . . . . .	5
1.4	Podstawy budowy G-kwadrupleksów . . . . .	5
1.5	Różnorodność struktur G-kwadrupleksów . . . . .	7
1.5.1	Topologie nici DNA . . . . .	7
1.5.2	Topologie kąta glikozydowego . . . . .	10
1.5.3	Topologie indukowane jonami metali . . . . .	11
1.5.4	Topologie indukowane stłoczeniem molekularnym . . . . .	11
1.5.5	Wpływ super-zwinięcia DNA na G-kwadrupleksy . . . . .	12
1.6	Stabilność G-kwadrupleksów . . . . .	13
1.7	G-kwadrupleksy w genomie . . . . .	14
1.8	Telomerowe DNA i G-kwadrupleksy . . . . .	15
1.9	Minimalne struktury G-kwadrupleksów . . . . .	18
1.10	Dimeryzacja jednoniciowych G-kwadrupleksów . . . . .	20
<b>2</b>	<b>Założenia i cele pracy</b>	<b>22</b>
<b>3</b>	<b>Podstawy metod badawczych</b>	<b>24</b>
3.1	Podstawy teoretyczne . . . . .	24
3.1.1	Przybliżenie Borna-Oppenheimera . . . . .	24
3.1.2	Empiryczne pola siłowe . . . . .	26
3.2	Metody obliczeniowe . . . . .	29
3.2.1	Minimalizacja energii układu . . . . .	29
3.2.2	Dynamika molekularna . . . . .	30
3.2.3	Algorytmy całkowania równań ruchu . . . . .	31
3.2.4	Krok czasowy . . . . .	32
3.2.5	Długości wiązań i więzy . . . . .	33
3.2.6	Temperatura i ciśnienie . . . . .	33
3.2.7	Periodyczne warunki brzegowe . . . . .	34
3.2.8	Oddziaływania niewiążące . . . . .	35

3.3	Elementy mechaniki statystycznej . . . . .	36
3.3.1	Analiza danych . . . . .	36
3.3.2	Energia swobodna . . . . .	36
3.3.3	Umbrella Sampling . . . . .	37
3.3.4	Adaptive Biasing Force . . . . .	39
<b>4</b>	<b>Procedury symulacyjne</b>	<b>40</b>
4.1	Stabilność G-tetrad w uformowanym G-kwadrupleksie . . . . .	40
4.1.1	Konstrukcja modelu i szczegóły symulacji . . . . .	40
4.1.2	Sterowana dynamika molekularna . . . . .	40
4.1.3	Symulacje typu energii swobodnej dla dysocjacji G-tetrad . . . . .	42
4.1.4	Wymuszona dysocjacja pojedynczej guaniny . . . . .	42
4.2	Rola pętli w stabilności G-kwadrupleksów . . . . .	43
4.2.1	Konstrukcja modeli i szczegóły symulacji . . . . .	43
4.2.2	Procedura zwijania dwupłaszczyznowych G-kwadrupleksów . . . . .	44
4.2.3	Szacowanie stabilności G-kwadrupleksów . . . . .	47
4.3	Dimeryzacja G-kwadrupleksów o konformacji równoległej . . . . .	49
4.3.1	Konstrukcja modeli i szczegóły symulacji . . . . .	49
4.3.2	Symulacje typu energii swobodnej . . . . .	50
4.4	Wpływ osmolitów na DNA . . . . .	51
4.4.1	Konstrukcja modeli i szczegóły symulacji . . . . .	51
4.4.2	Wymuszona denaturacja DNA . . . . .	52
<b>5</b>	<b>Wyniki oraz ich dyskusja</b>	<b>53</b>
5.1	Stabilność G-tetrad w uformowanym G-kwadrupleksie . . . . .	53
5.1.1	Profil energii swobodnej dysocjacji G-tetrad . . . . .	54
5.1.2	Mechanizm dysocjacji G-tetrad . . . . .	56
5.1.3	Energetyczne determinanty stabilności G-tetrad . . . . .	60
5.1.4	Entropia formowania G-tetrad . . . . .	64
5.1.5	Podsumowanie . . . . .	67
5.2	Rola pętli w stabilności G-kwadrupleksów . . . . .	68
5.2.1	Weryfikacja procedury zwijania G-kwadrupleksów . . . . .	68
5.2.2	Stabilność modelowych G-kwadrupleksów . . . . .	70
5.2.3	Podsumowanie . . . . .	72
5.3	Dimeryzacja G-kwadrupleksów o konformacji równoległej . . . . .	74
5.3.1	Preferencje G-kwadrupleksów do dimeryzacji . . . . .	74
5.3.2	Szczegóły struktur dimerów . . . . .	78
5.3.3	Siły napędowe dimeryzacji . . . . .	79
5.3.4	Modulacja równowagi dimeryzacji . . . . .	83



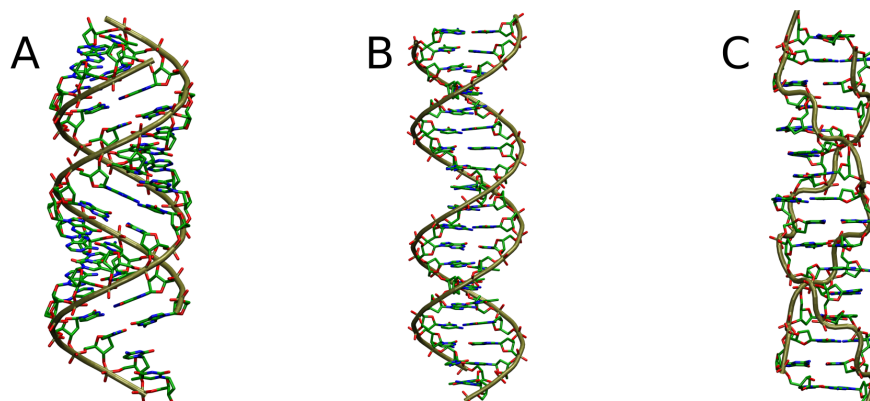
5.3.5	Podsumowanie . . . . .	86
5.4	Wpływ denaturantów na stabilność DNA . . . . .	88
5.4.1	Akumulacja denaturantów i wody wokół DNA . . . . .	88
5.4.2	Miejsca oddziaływań denaturantów w obrębie cząsteczki DNA . . . . .	90
5.4.3	Wpływ denaturantów na rozkład wody wokół DNA . . . . .	92
5.4.4	Wiązania wodorowe między DNA a roztworami denaturantów . . . . .	93
5.4.5	Rozkład przestrzenny cząsteczek denaturantów wokół DNA . . . . .	95
5.4.6	Energie oddziaływania DNA z cząsteczkami roztworu . . . . .	98
5.4.7	Zmiany konformacyjne DNA . . . . .	101
5.4.8	Podsumowanie . . . . .	102
<b>6</b>	<b>Ogólne podsumowanie</b>	<b>104</b>
<b>A</b>	<b>Załącznik</b>	<b>108</b>
A.1	Stabilność G-tetrazy w uformowanym G-kwadrupleksie . . . . .	108
A.1.1	Definicja współrzędnej opisującej dysocjację G-tetrazy ( $N_{hb}$ ) . . . . .	108
A.1.2	Szacowanie średniego czasu pierwszego przejścia (MFTP) . . . . .	109
<b>B</b>	<b>Załącznik</b>	<b>117</b>
B.1	Dimeryzacja G-kwadrupleksów o konformacji równoległej . . . . .	117
<b>C</b>	<b>Załącznik</b>	<b>134</b>
C.1	Wpływ denaturantów na DNA . . . . .	134
C.1.1	Definicja odległości minimalnej . . . . .	135
C.1.2	Współczynnik preferencyjnego oddziaływania . . . . .	135
C.1.3	Oczekiwana liczba cząsteczek wody w pierwszej sferze hydratacji . . . . .	135
<b>D</b>	<b>Udział w innych projektach badawczych</b>	<b>136</b>
<b>E</b>	<b>Dorobek naukowy</b>	<b>137</b>
<b>F</b>	<b>Streszczenie w języku polskim</b>	<b>138</b>
<b>G</b>	<b>Streszczenie w języku angielskim</b>	<b>139</b>
	<b>Bibliography</b>	<b>140</b>



# 1 Wprowadzenie

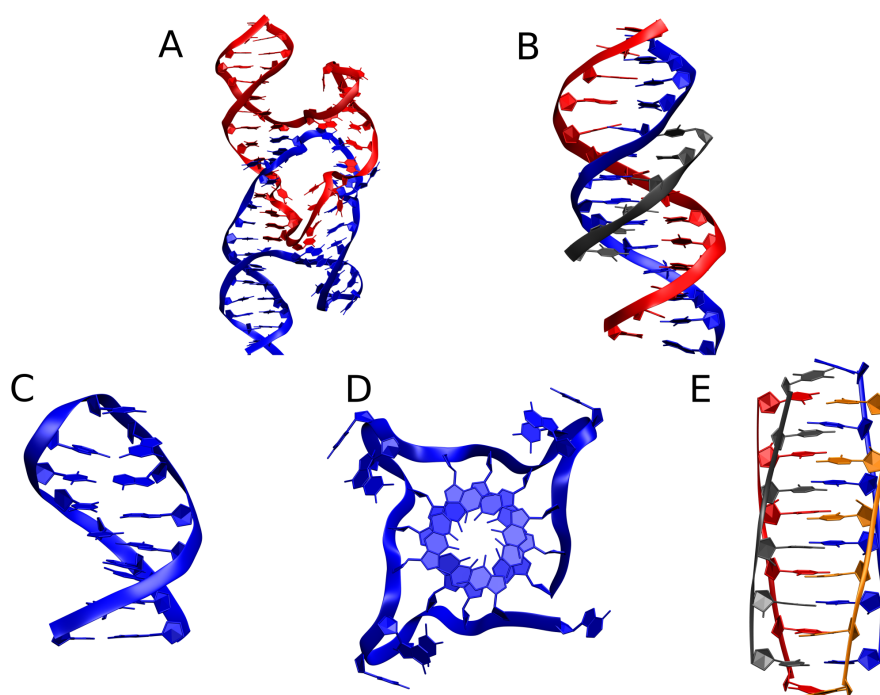
## 1.1 Różne formy DNA

Kwas deoksyrybonukleinowy (DNA) jest podstawowym nośnikiem informacji genetycznej we wszystkich żywych komórkach oraz większości wirusów. Informacja o cechach dziedzicznych każdej jednostki jest zawarta w sekwencji nukleotydów DNA. Reguły parowania zasad azotowych w podwójnej helisie DNA zostały po raz pierwszy opisane w 1953 r. przez Watsona i Cricka [1]. Od tego momentu systematyczne badania wskazały, że forma B-DNA jest dominującą formą biologiczną prawoskrętnej helisy DNA (*in vivo* lub w wysokiej wilgotności 92%) [2].



RYSUNEK 1.1: Trzy najistotniejsze biologicznie konformacje podwójnej helisy DNA. Od lewej, struktura A-DNA, B-DNA, Z-DNA.

W zależności od warunków środowiskowych DNA może również przyjmować inne formy helikalne takie jak A-, B- i Z-DNA (patrz Rys. 1.1) oraz kilka niehelikalnych struktur drugorzędowych (patrz Rys. 1.2), jednak ich formowanie zazwyczaj wymaga niekanonicznego lub/ oraz nieciągłego parowania zasad azotowych. Do takich egzotycznych form DNA zalicza się struktury Hollidaya (ang. Holliday junction), tripleksy DNA (H-DNA), spinki DNA (ang. DNA hairpins), G-kwadrupeksy oraz motyw interkalowany DNA (i-motif DNA) [3, 4]. Co ciekawe, w przypadku G-kwadrupeksów nie występuje klasyczne parowanie zasad azotowych Watsona-Cricka (G–C, A–T), lecz cykliczne parowanie guanin od strony krawędzi Watsona-Cricka oraz krawędzi Hoogsteena (patrz Rys. 1.3). W parowanie guanin zawsze zaangażowane są 4 równoległe fragmenty nici DNA, jednakże niekoniecznie sparowane w kolejności w jakiej występują w sekwencji. Na przejścia strukturalne z prawoskrętnej formy



RYSUNEK 1.2: Różne niekanoniczne formy drugorzędowe DNA: A) struktura Hollidaya, B) tripleks DNA, C) spinka DNA, D) G-kwadrupleks oraz E) motyw interkalowany. Różne kolory oznaczają różne nici DNA.

B-DNA do innych form, w tym tworzenie G-kwadrupleksów, wpływa wiele czynników takich jak sekwencja, siła jonowa, temperatura, stopień hydratacji czy oddziaływanie z białkami. Obecnie badania skupiają się przede wszystkim na ustaleniu w jaki sposób ich tworzenie zależy od warunków *in vitro* oraz jakie role biologiczne mogą pełnić *in vivo* [3, 5].

Badania strukturalne nad formą A-DNA rozpoczęły się od obserwacji, że krystaliczne formy włókien DNA o niskiej wilgotności wytworzyły rentgenowski wzór dyfrakcyjny struktur, które wyraźnie różniły się od natywnej helisy DNA w fazie wodnej [6]. Struktura A-DNA może być formowana zarówno przez dwuniciowe helisy DNA oraz RNA jak również hybrydowe struktury dupleksowe DNA–RNA [3, 6].

Synteza chemiczna odpowiednio zaprojektowanej sekwencji oligonukleotydu, pozwoliła na wyznaczenie pierwszej struktury krystalicznej o wysokiej rozdzielczości, która nieoczekiwanie wykazała możliwość wystąpienia lewoskrętnej struktury helikalnej z naprzemiennymi wiązaniami glikozydowymi w konformacji *anti* i *syn*. Uzyskaną strukturę nazwano formą Z-DNA i ustalono, że jej tworzenie promowane jest przez sekwencje zawierające naprzemiennie pary G–C (szczególnie  $\text{poly}(\text{dGC})_2$ ) oraz wysoką siłę jonową [7, 8]. Następnie pojawiała się hipoteza odnośnie udziału formy Z-DNA w regulacji genów [9]. Tabela 1.1 podsumowuje najistotniejsze różnice między trzema formami podwójnej helisy DNA.

TABELA 1.1: Różnice parametrów geometrycznych pomiędzy trzema głównymi formami podwójnej helisy DNA.

Rodzaj DNA	B-DNA	A-DNA	Z-DNA
Kierunek skręcenia	prawoskrętna	prawoskrętna	lewoskrętna
Duża bruzda	szeroka i głęboka	wąska i bardzo głęboka	płytko
Mała bruzda	wąska i głęboka	bardzo szeroka i płytko	bardzo wąska i głęboka
Średnica helisy (w nm)	2,37	2,55	1,84
Liczba par zasad przypadająca na skręt helisy	10,4	11	12
Kąt skręcenia między sąsiednimi parami zasad	+34,6°	+39,0°	-30,0°
Skok helisy (w nm)	3,54	2,53	4,56

## 1.2 Wpływ osmolitów na DNA

Zdolność dostosowania ciśnienia osmotycznego i stabilności biomakromolekuł do zmieniających się warunków środowiskowych ma kluczowe znaczenie dla przetrwania wielu żywych organizmów. Można to osiągnąć przez wewnątrzkomórkową akumulację małych cząsteczkowych związków organicznych – osmolitów, które mogą jednak zaburzać równowagę konformacyjną białek i kwasów nukleinowych. Na podstawie właściwości chemicznych i strukturalnych naturalnie występujące osmolity można podzielić na trzy grupy: (i) poliole, np. glicerol lub mannitol; (ii) aminokwasy i ich pochodne, np. tauryna,  $\beta$ -alanina i betaina; oraz (iii) mocznik i metyloaminy, np. TMAO, TMG [10–14]. Najczęściej jednak osmolity są dzielone, zgodnie z ich wpływem na stabilność białek, na stabilizatory, które przesuwają równowagę w kierunku postaci natywnej (stabilizacja, np. betaina), i destabilizatory (denaturanty) przesuwające równowagę w kierunku postaci rozwiniętej (destabilizacja, np. mocznik). Choć istnieje wiele teorii wyjaśniających stabilizujące i destabilizujące działanie osmolitów na białka, zwykle nie można ich bezpośrednio przenosić na przypadek kwasów nukleinowych, co podkreśla fakt, że niektóre stabilizujące białka osmolity destabilizują podwójną helisę DNA.

Pozostaje zatem niejasne, czy równowaga związania DNA może być zaburzona przez bezpośrednie oddziaływania między rozpuszczalnikiem a cząsteczką DNA, czy pośrednio, przez oddziaływania osmolit–woda, które równolegle modyfikują właściwości rozpuszczalnika [15–17]. Dotychczasowa wiedza na temat mechanizmów rządzących wpływem rozpuszczalnika na stabilność DNA ma ogromne znaczenie praktyczne ze względu na potencjalne zastosowania w inżynierii genetycznej i biotechnologii.

DNA wykazuje znaczną zmienność konformacyjną, a wśród głównych czynników determinujących względną stabilność poszczególnych struktur drugorzędowych są oddziaływania między rozpuszczalnikiem a cząsteczką DNA. Te oddziaływania również przyczyniają



się do zmienności funkcji DNA, w tym wiązania różnych ligandów [18–23]. Szczególnie interesująca jest podatność do zmiany konformacji z A-DNA na B-DNA w zależności od aktywności wody [24, 25]. Badania z wykorzystaniem spektroskopii w zakresie podczerwieni i UV wykazały, że pełna solwatacja cząsteczki DNA jest niezbędna do utrzymania integralnej struktury DNA, a obniżenie wilgotności względnej skutkuje niewłaściwym ułożeniem par zasad w helisie [26]. Ponadto tendencja do tworzenia formy A- lub B-DNA jest silnie zależna od sekwencji, co sugeruje, że końcowa struktura drugorzędowa jest narzucana przez skomplikowane współoddziaływanie różnych wkładów [27–30].

Podczas gdy woda jest kluczem do stabilizacji biologicznie aktywnych struktur DNA, wiadomo, że cząsteczki osmolitu zaburzają układ przestrzenny cząsteczek wody i ich wzajemne oddziaływania, powodując, że woda przejawia zmienione właściwości fizykochemiczne, takie jak niższa aktywność. W konsekwencji stwierdzono, że zarówno zmniejszona wilgotność względna w badaniach krystalograficznych [25, 31, 32] jak i dodanie współrozpuszczalnika organicznego lub soli [33–36] powoduje podobny korzystny efekt preferencyjnej stabilizacji A-DNA.

Znany jest wpływ kilku małocząsteczkowych związków organicznych o aktywności osmotycznej (osmolitów) na podwójną helisę DNA. Nie istnieją natomiast przekonujące wyniki badań eksperymentalnych, które w pełni wyjaśniałyby mechanizmy przejść strukturalnych DNA wywołanych obecnością takich związków w roztworze. Klasycznym przykładem związku chemicznego mającego istotny wpływ na DNA jest mocznik. Faktem jest, że denaturuje on podwójną helisę DNA, lecz mechanizm takiej denaturacji chemicznej nie jest w pełni zrozumiały [37]. Jednym z sugerowanych mechanizmów jest efektywne osłabienie wiązań wodorowych pomiędzy zasadami. Proces taki jest możliwy, ponieważ mocznik stanowi konkurencję dla wody w tworzeniu wiązań wodorowych z biopolimerami. Inną możliwością jest zaburzenie przez mocznik struktury i sieci oddziaływań samego rozpuszczalnika, co eliminowałoby lub znacząco redukowałoby korzyści energetyczne związane z efektem hydrofobowym. Innym przykładem osmolitu o działaniu denaturującym względem DNA jest betaina. Dodanie betainy do roztworu DNA eliminuje różnice pomiędzy temperaturami topnienia fragmentów DNA wynikającymi z zawartości i kolejności par zasad A-T oraz G-C w DNA [38]. Efekt ten wykorzystuje się, stosując betainę jako dodatek w reakcji łańcuchowej polimerazy (PCR), jednak jego mechanizm do dziś nie został w pełni wyjaśniony. Zupełnie przeciwny efekt na DNA wywiera spermidyna, związek chemiczny należący do poliamin, który jest syntetyzowany przez żywe komórki w celu ochrony informacji genetycznej przed stresem chemicznym i fizycznym. W szczególności, substancja ta odpowiedzialna jest za ochronę DNA plemników przed niesprzyjającym, kwaśnym środowiskiem pochwy. Przy fizjologicznym pH spermidyna jest sprotonowana, przez co staje się podwójnie lub potrójnie naładowaną cząsteczką. Ten dodatni ładunek sprzyja silnym oddziaływaniom z negatywnie naładowanymi grupami fosforanowymi. W przypadku klasycznej struktury helisy DNA

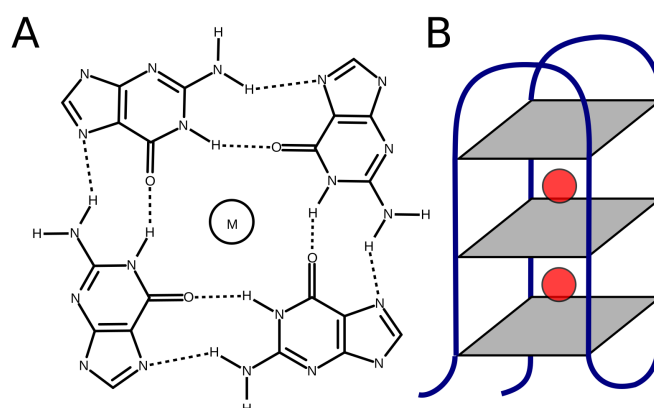
spermidyna obsadza małą bruzdę dzięki czemu skutecznie neutralizuje grupy fosforanowe w obu niciach DNA [39]. Obecnie uważa się, że spermidyna jest jednym ze związków, który najlepiej indukuje przejścia z prawoskrętnej helisy DNA do lewoskrętnej.

### 1.3 Historia badań nad G-kwadrupleksami

Dość nietypowa cecha samoorganizacji kwasu guanylowego została po raz pierwszy odnotowana przez niemieckiego chemika Banga już w 1910 roku. Badacz zaobserwował zjawisko tworzenia się żelu dla wysokich stężeń 5'-monofosforanu guanozyny (5'-GMP) przy określonych wartościach pH [40]. Dużo później badania dyfrakcji rentgenowskiej na włóknach 5'-GMP pokazały, że dokładnie 4 reszty guaninowe układają się w jednej płaszczyźnie, pomiędzy którymi istnieje sieć wiązań wodorowych [41]. W toku badań prowadzonych przez Arnotta i jego współpracowników uzyskano podobny wniosek, dotyczący formowania kompaktowych struktur czteroniciowych o prawoskrętnej strukturze przez sekwencję DNA bogate w guaniny [42]. Następnie w 1978 r. Miles i Fraser pokazali, że stabilność tych struktur zależy od obecności dodatnio naładowanych jonów, które silnie oddziałują z atomami tlenu z G-tetrad. Możliwość tworzenia *in vitro* struktur G-kwadrupleksów przez cyklicznie powtarzające się fragmenty bogate w guaniny związane z telomerami została potwierdzona pod koniec lat 80-tych przez grupę Blackburn oraz Cecha [43, 44]. Pierwsze silne dowody na istnienie struktur G-kwadrupleksów *in vivo* uzyskano przez barwienie makrojąder *Styloichia lemnae* za pomocą swoistych przeciwciał Sty49 o wysokim powinowactwie do G-kwadrupleksów [45]. Później okazało się, że ponad 50% genomu występuje jako cyklicznie powtarzające się sekwencje, przez co należy uwzględnić możliwość formowania niehelikalnych struktur w tych sekwencjach [46, 47]. Badania bioinformatyczne przeprowadzone na ludzkim oraz innych genomach zidentyfikowały bardzo wiele sekwencji bogatych w cyklicznie powtarzające się fragmenty guaninowe, a ludzki genom zawiera ponad 300000 sekwencji potencjalnie mogących tworzyć G-kwadrupleksy [48, 49]. Dla wielu z tych sekwencji wskazano rolę w regulacji procesów komórkowych takich jak transkrypcja, translacja i rekombinacja [50, 51]. Ciągłe prowadzone są badania mające na celu wyznaczenie wszystkich możliwych natywnych stanów strukturalnych tych nietypowych drugorzędowych form DNA oraz mechanizmów odpowiedzialnych za ich stabilność termodynamiczną i kinetyczną w warunkach fizjologicznych.

### 1.4 Podstawy budowy G-kwadrupleksów

W pewnych warunkach fizjologicznych sekwencje DNA bogate w guaniny mogą się specyficznie fałdować tworząc czteroniciową strukturę znaną jako G-kwadrupleksy. Struktury te składają się z dwóch lub więcej tetrad guanozynowych ułożonych jedna nad drugą tworząc

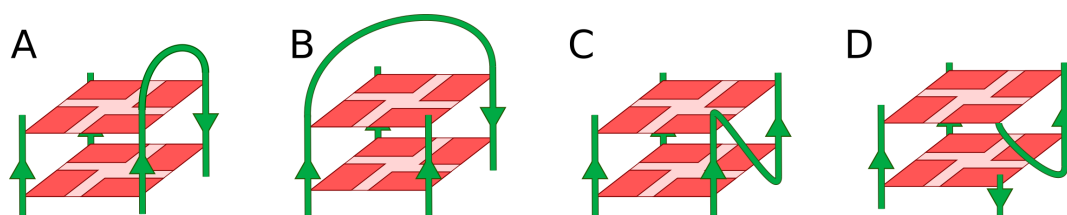


RYSUNEK 1.3: Ogólny schemat struktury G-kwadrupleksów. A) Pojedyncza G-tetradą tworzona przez 4 guaniny z jon metalu w środku. B) Jedna z możliwych konformacji przyjmowanych przez G-kwadrupleksy. Szare kwadraty reprezentują G-tetrady ułożone w stos, niebieska linia oznacza przebieg łańcucha fosfocukrowego, natomiast kolorem czerwonym zaznaczono miejsca wiązania kationów.

tw. rdzeń guaninowy. Pojedyncza tetradą składa się z 4 ganozyn połączonych ze sobą wiązaniami wodorowymi typu Hoogsteena (patrz Rys. 1.3). W takim układzie każda guanina jest donorem oraz akceptorem wiązań wodorowych (wiązanie N1 z O6 oraz N2 z N7). Zaangażowanie atomu azotu N7 w tworzenie wiązań wodorowych chroni guaniny przed metylacją przez siarczan dimetylu. Ta unikalna cecha G-kwadrupleksów pozwala na chemicznie odróżnienie ich od innych drugorzędowych struktur DNA. Jednowartościowe kationy takie jak  $K^+$  i  $Na^+$ , stabilizują całą strukturę rdzenia guaninowego poprzez oddziaływanie z centralnymi elektroujemnymi karbonylowymi atomami tlenu O6. Brak kationów metali w roztworze w znacznym stopniu utrudnia powstawanie struktur kwadrupleksowych [52, 53]. Ustalono również, że największa stabilizacja G-kwadrupleksów występuje podczas interkalacji jonów  $K^+$  dokładnie pomiędzy dwoma kolejnymi G-tetradami [54]. Poszczególne G-tetrady są ułożone w odległości 3,3 Å od siebie, a tak utworzony stos jest połączony za pomocą łańcucha fosfocukrowego DNA. Pozostałe zasady azotowe z nici DNA, czyli nie biorące udziału w tworzeniu G-tetrad, biorą udział w tworzeniu pętli o różnej długości oraz sekwencji. Pętłe te przebiegają wzdłuż krawędzi każdej G-tetrady, a ich różne orientacje prowadzą do powstawania struktur G-kwadrupleksów o różnorodnych topologiach. Jak już łatwo zauważyć, podstawową cechą strukturalną wszystkich G-kwadrupleksów jest rdzeń guaninowy, a główną cechą różnicującą je są pętłe o różnej długości oraz sekwencji. Przeprowadzono wiele badań mających na celu ustalenie, w jaki sposób sekwencja i długość pętli może regulować konformację i stabilność G-kwadrupleksów [55–58].

Do tej pory zidentyfikowano wiele sposobów ułożeń pętli w przestrzeni, uwzględniając w tym możliwy obrót kowalencyjnego wiązania glikozydowego pomiędzy cukrem i zasadą azotową. Jak pokazano na Rys. 1.4 pętłe mogą łączyć dwa równoległe lub antyrównoległe fragmenty guaninowe ( $G_n$ , gdzie  $n$  oznacza dowolną liczbę guanin w ciągu), tworząc jeden

z czterech możliwych typów pętli. Pętla boczna (zwana również pętlą krawędziową) łączy dwa sąsiednie antyrównoległe fragmenty  $G_n$  i zazwyczaj składa się z dwóch lub więcej zasad. Pętla diagonalna łączy dwie przeciwległe i antyrównoległe fragmenty  $G_n$  i składa się z trzech lub więcej zasad. Natomiast pętla śmigłowa (znana również jako pętla z podwójnie zagiętym łańcuchem) łączy dwa sąsiednie równoległe fragmenty  $G_n$  i składa się z jednej zasady (w przypadku krótkiej wersji pętli) lub sześciu lub więcej zasad (w przypadku długiej wersji pętli). Rzadziej spotykane pętle w kształcie litery V łączą trzy krawędzie G-tetrad z dwóch sąsiednich fragmentów  $G_n$  [59].

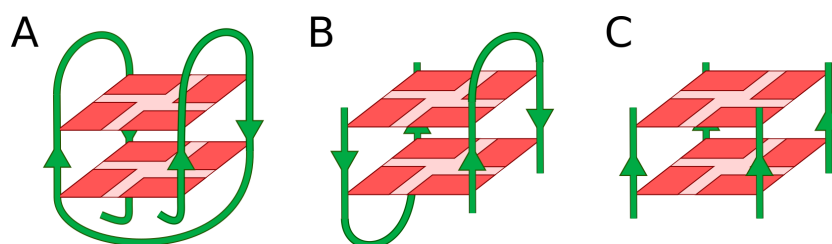


RYSUNEK 1.4: Schemat 4 możliwych typów pętli w G-kwadrupleksach: A) pętla boczna, B) pętla diagonalna, C) pętla śmigłowa i D) pętla w kształcie litery V.

## 1.5 Różnorodność struktur G-kwadrupleksów

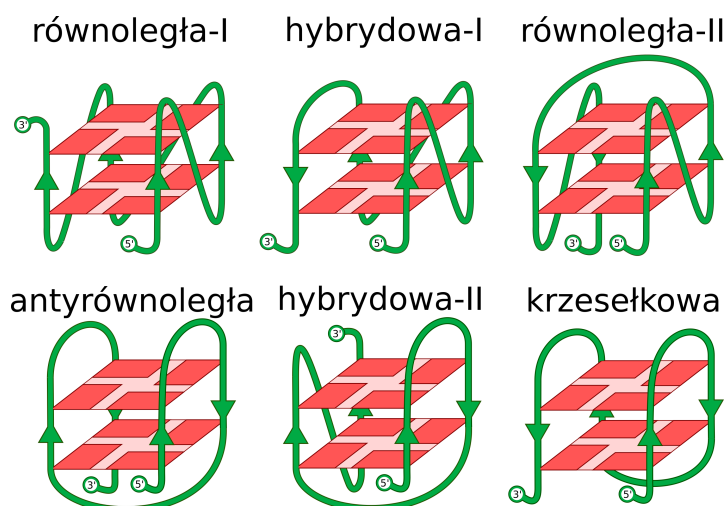
Struktury G-kwadrupleksów są wysoce polimorficzne ze względu na możliwość przyjmowania różnych struktur przez ten sam fragment DNA. Wynika to z wielu możliwych kombinacji różnych czynników takich, jak ułożenie fragmentów nici względem siebie, typy pętli, konformacje wiązań glikozydowych, różnorodność kationów metali mogących wiązać się w rdzeniu guaninowym.

### 1.5.1 Topologie nici DNA



RYSUNEK 1.5: Stechiometria nici G-kwadrupleksów: A) G-kwadrupleks jednoniciowy, B) dwuniciowy i C) czteroniciowy.

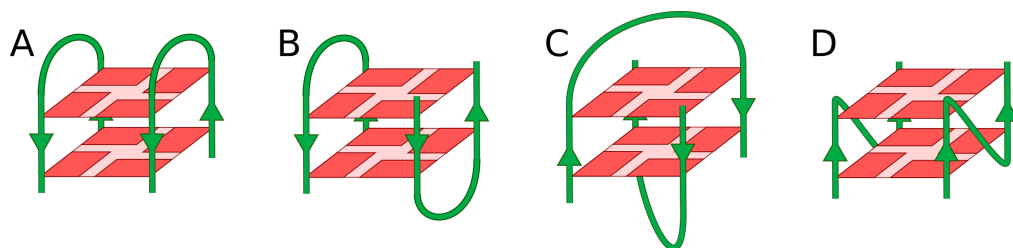
Struktury G-kwadrupleksów mogą tworzyć się przez fałdowanie jednoniciowego DNA bogatego w guaniny (jednocząsteczkowe G-kwadrupleksy) oraz poprzez związanie się ze sobą dwóch lub czterech niezależnych nici DNA (wielocząsteczkowe G-kwadrupleksy). Te trzy możliwości przedstawiono na Rys. 1.5.



RYSUNEK 1.6: Przykładowe topologie przyjmowane przez jednoniciowe G-kwadrupleksy. Kolorem czerwonym oznaczone zostały płaszczyzny guaninowe, kolorem zielonym szkielet nici DNA oraz strzałkami kierunek przebiegu od strony 5'-końca do 3'-końca.

Sekwencje bogate w guaniny tworzące jednoniciowe G-kwadrupleksy zazwyczaj zbudowane są z 4 kolejnych fragmentów guaninowych ( $G_n$ ), które rozdzielone są dokładnie trzema fragmentami pętli o różnych długościach i sekwencjach. Ten konkretny rodzaj G-kwadrupleksów może przyjmować różne topologie (patrz Rys. 1.6), które różnią się ułożeniem fragmentów nici względem siebie oraz wartościami kątów wiązania glikozydowego, przyjmując konformację *syn* lub *anti*. Topologia krzeselkowa jest jednym z prostszych sposobów fałdowania G-kwadrupleksów, a powstaje ona poprzez utworzenie trzech bocznych pętli pomiędzy antyrównoległymi fragmentami  $G_n$ . Drugi sposób fałdowania o kształcie zbliżonym do koszyka (topologia antyrównoległa) powstaje poprzez utworzenie dwóch pętli bocznych oraz jednej środkowej diagonalnej. Topologia równoległa-I, zwana również śmigłową, formowana jest przez 4 równoległe fragmenty  $G_n$  połączone 3 pętlami o podwójnym zagięciu łańcucha. Natomiast obie topologie hybrydowe przedstawione na Rys. 1.6 są połączeniem równoległego i antyrównoległego ułożenia fragmentów  $G_n$  oraz różną kombinacją pętli. Struktury hybrydowe najczęściej mają dwie pętli boczne i jedną pętlę śmigłową [60]. Spośród przebadanych dotąd sekwencji największą różnorodność jednoniciowych struktur G-kwadrupleksów przejawia ludzka sekwencja telomerowa 5'-d(GGGTTA) $_n$ -3' [51, 61]. Zmiany w długości i sekwencji pętli mogą spowodować przyjęcie jednej z 26 teoretycznie możliwych topologii G-kwadrupleksów [60, 62, 63]. Ciekawym przykładem różnorodności topologii G-kwadrupleksów jest sekwencja d[G<sub>3</sub>ATG<sub>3</sub>ACACAG<sub>4</sub>ACG<sub>3</sub>], która związa się do jednoniciowego G-kwadrupleksu z trzema G-tetradami, połączonymi aż trzema różnymi typami pętli (śmigłowa, diagonalna, boczna) każda o różnej długości [64].

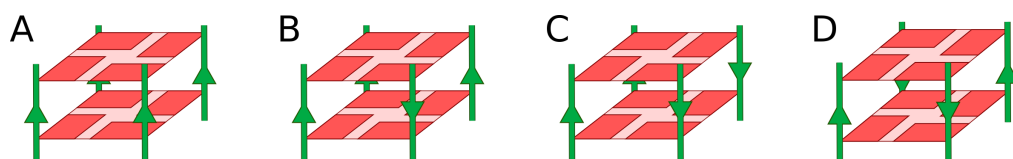
Dwucząsteczkowe G-kwadrupleksy powstające w wyniku połączenia dwóch nici DNA, z których każda zawiera przynajmniej po dwa fragmenty  $G_n$ , również mogą przyjmować wiele różnych konformacji. Jak pokazano na Rys. 1.7, istnieje kilka sposobów ułożenia dwóch



RYSUNEK 1.7: Przykładowe topologie przyjmowane przez dwuniciowe G-kwadrupleksy.

sąsiednich G-spinek tworzących G-kwadrupleks. Możliwe jest ułożenie G-spinek w tym samym kierunku 3'- oraz 5'-końców 5' (równoległe), w kierunku przeciwnym (antyrównoległe) oraz wzajemnie się zazębiających (z dwiema pętlami diagonalnymi). Inną możliwością jest równoległe ułożenie G-spinek z dwiema pętlami śmigłowymi. Jednym z przykładów dwucząsteczkowego G-kwadrupleksu jest struktura promotora MyoD [65].

Schemat formowania czterocząsteczkowych G-kwadrupleksów pozornie wydaje się być znacznie prostszy, ponieważ nie obejmuje pętli (Rys. 1.5C). Przykładem może być rentgenowska struktura czteroniciowego G-kwadrupleksu o sekwencji d(TG<sub>4</sub>T) i wszystkich fragmentach G<sub>n</sub> równoległych względem siebie [66]. Poza tym w roztworach jonów K<sup>+</sup> sekwencje telomerowe takie jak d(TTAGGG) i d(TTAGGGT) również tworzą czteroniciowe G-kwadrupleksy o równoległym ułożeniu nici, a wszystkie konformacje glikozydowe guanin z G-tetrad są typu *anti-anti-anti-anti* [67]. Za pomocą technik wysokorozdzielczego NMR i spektrometrii mas grupa badawcza Balasubramaniana wykazała, że sekwencja d(GGGT), która nie zawiera nukleotydu flankującego na 5'-końcu również tworzy strukturę czteroniciową. Jednakże, uwzględniając możliwość dowolnej kombinacji różnych orientacji każdego z fragmentów G<sub>n</sub> oraz kątów glikozydowych guanin, teoretycznie możliwych jest również wiele struktur czteroniciowych G-kwadrupleksów [68].

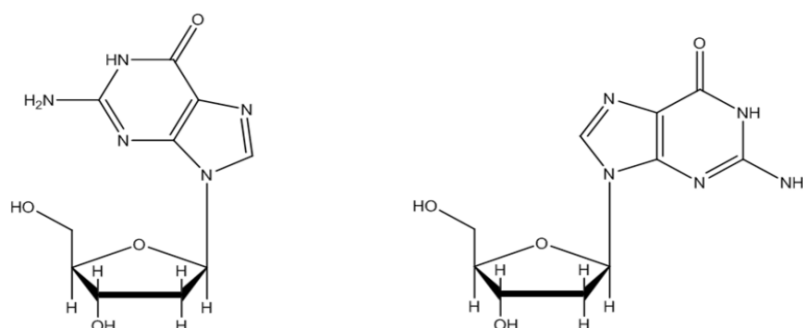
RYSUNEK 1.8: Wszystkie możliwe względne ułożenia fragmentów G<sub>n</sub> występujące w G-kwadrupleksach.

Cztery fragmenty G<sub>n</sub> występujące w obrębie jednocząsteczkowych lub wielocząsteczkowych G-kwadrupleksów mogą przyjmować jeden z czterech możliwych sposobów ułożenia względem siebie (patrz Rys. 1.8). Pierwszy sposób ułożenia składa się z czterech równoległych nici o tym samym kierunku 5'-3' oraz jednego z dwóch możliwych zestawów konformacji kątów glikozydowych (*anti-anti-anti-anti* lub *syn-syn-syn-syn*). Drugi sposób to trzy równoległe ułożone nici oraz jedna w przeciwnym kierunku, a możliwy zestaw konformacji glikozydowych to *anti-syn-syn-syn* lub *syn-anti-anti-anti*. Poza tym istnieją układy hybrydowe tzw. 2+2, które składają się z dwóch naprzemiennie ułożonych nici antyrównoległych

(*syn-syn-anti-anti*) lub dwóch sąsiadujących równoległych (*syn-anti-syn-anti*). Pełny zestaw nici ułożonych równoległe względem siebie jest cechą charakterystyczną dla czterocząsteczkowych struktur G-kwadrupleksów. Natomiast mieszane oraz antyrównoległe ułożenia nici spotyka się zarówno w jednocząsteczkowych jak i dwucząsteczkowych G-kwadrupleksach. Pierwsze badania NMR przeprowadzone dla nici DNA o ludzkiej sekwencji telomerowej DNA pokazały, że w obecności jonów  $\text{Na}^+$  DNA związa do topologii antyrównoległej o kształcie koszyka [69]. Jednak krystaliczna struktura rentgenowska tej samej sekwencji DNA, uzyskana w obecności jonów  $\text{K}^+$  charakteryzuje się konformacją równoległą, w której wszystkie guaniny przyjmują konformację *anti* [61]. Co więcej podobna sekwencja lecz tylko o podwójnym powtórzeniu ludzkiej sekwencji telomerowej d(TAG<sub>3</sub>TTA<sub>3</sub>T) w obecności jonów  $\text{K}^+$  tworzy mieszaninę dwóch struktur hybrydowych [70]. Podobna hybrydowa struktura telomerowego G-kwadrupleksu jest przyjmowana przez fragment bogaty w guaniny powyżej promotora Bcl-2 P1, w której to wiele fragmentów guaninowych jest zaangażowanych w tworzenie jednocząsteczkowej struktury [71]. Struktura ta ma typową dla topologii równoległej jednozasadową trzecią pętlę śmigłową oraz typową dla topologii antyrównoległej trójzasadową (CGC) pierwszą pętlę boczną. Bazując na przytoczonych przykładach można stwierdzić, że krótkie pętle promują lokalne struktury równoległe, podczas gdy długie pętle promują struktury antyrównoległe.

Badania krystalograficzne oraz NMR wykazały, że w obecności jonów potasu jak i sodu sekwencja d(TG<sub>4</sub>T) tworzy jedną dominującą strukturę wielocząsteczkowego G-kwadrupleksu [72]. Natomiast struktury NMR i krystalograficzne wyznaczone przez grupę Haidera dla telomerowej sekwencji *Oxytricha nova* (5'-d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>)-3') mają charakter wielocząsteczkowy z dwoma antyrównoległymi i dwoma równoległymi fragmentami G<sub>n</sub> wraz z czterema resztami tyminowymi tworzącymi pętle diagonalne [73].

### 1.5.2 Topologie kąta glikozydowego



RYSUNEK 1.9: Dwie orientacje wiązania glikozydowego. Konformacja *syn* (z lewej) i *anti* (z prawej).

Różnorodność struktur G-kwadrupleksów jest również związana z elastycznością wiązania glikozydowego. Efektywne przyciąganie lub odpychanie między cukrem a zasadą nukleozydu skutkuje przyjęciem odpowiednio konformacji *syn* lub *anti* (patrz Rys. 1.9). Obie konformacje są obserwowane w strukturach G-kwadrupleksów. W przeciwieństwie do klasycznej podwójnej helisy DNA, struktury G-kwadrupleksów mają aż cztery bruzdy o różnych szerokościach oraz głębokościach, które opisuje się jako szeroka, średnia i wąska. W przypadku czteroniciowych struktur G-kwadrupleksów wszystkie deoksyguanozyny przyjmują taką samą konformację, co powoduje, że wszystkie cztery bruzdy są identyczne. Jednak jednocząsteczkowe i dwucząsteczkowe struktury G-kwadrupleksowe mogą się charakteryzować niemal dowolną kombinacją konformacji *syn* i *anti*.

### 1.5.3 Topologie indukowane jonami metali

W zależności od rodzaju i stężenia jonów obecnych w roztworze konkretna sekwencja DNA tworząca G-kwadrupleksy może formować kilka różniących się między sobą struktur przestrzennych. Część badań NMR i krystalograficznych ludzkiego telomerowego DNA wskazuje, że formowanie antyrównoległych G-kwadrupleksów jest preferowane w obecności jonów  $\text{Na}^+$ , natomiast równoległych i hybrydowych w obecności jonów  $\text{K}^+$  [61, 69]. Podobne przejścia konformacyjne zaobserwowano również dla G-kwadrupleksów *bcl-2*. W obecności potasu widmo dichroizmu kołowego (CD) G-kwadrupleksu przedstawia silne pasmo przy 265 nm, natomiast w obecności sodu w 295 nm [74]. Są to dwa charakterystyczne pasma dla G-kwadrupleksów równoległych (265 nm) oraz antyrównoległych (295 nm) [75]. Co ciekawe, widmo CD G-kwadrupleksu *bcl-2* w wariacie RNA niezależnie od rodzaju jonów posiada dodatni pik na paśmie 260 nm oraz ujemny pik na paśmie 240 nm. Wskazuje to na tworzenie się równoległego G-kwadrupleksu niezależnie od warunków jonowych [75]. W przeciwieństwie do G-kwadrupleksów DNA, tworzenie się struktur G-kwadrupleksów RNA jest zasadniczo niezależne od rodzaju kationu. Co więcej, o ile typ jednowartościowych kationów może regulować topologię jednocząsteczkowych G-kwadrupleksów DNA, to nie wpływa on na topologię dwucząsteczkowych G-kwadrupleksów. Przykładem takiego zachowania jest dwucząsteczkowy G-kwadrupleks formowany z sekwencji  $d(\text{G}_4\text{T}_4\text{G}_4)_2$  przyjmujący formę dwóch zazębionych ze sobą spinek w różnych roztworach jonów  $\text{Na}^+$ ,  $\text{K}^+$  i  $\text{NH}_4^+$  [76].

### 1.5.4 Topologie indukowane stłoczeniem molekularnym

Żywe komórki zawierają wysokie stężenia biomakromolekuł oraz innych rozpuszczalnych i nierozpuszczalnych substancji. Te biomakromolekuły, wliczając w to DNA, RNA, białka i polisacharydy, zajmują około 30–40% objętości komórkowej, a ich całkowite stężenie dochodzi



do 400 g/L [77]. Jak wiadomo, struktura i stabilność a zatem i pełnione funkcje tych biomakromolekuł są regulowane przez warunki środowiskowe, takie jak typ i stężenie jonów metali, pH, rodzaj rozpuszczalnika oraz cząsteczki powodujące stłoczenie molekularne. W celu zbadania przejść konformacyjnych telomerowych G-kwadrupleksów w warunkach stłoczenia molekularnego grupa Miyoshiego wykorzystwała spektroskopię CD w połączeniu z izotermiczną kalorymetrią miareczkową (ITC). Badania pokazały, że w obecności 2 M PEG (glikolu polietylenowego) telomerowe sekwencje DNA *Oxytricha nova* bogate w guaniny oraz cytozyny mają możliwość tworzenia równoległych G-kwadrupleksów oraz struktur o motywie interkalowanym (z ang. i-motif), ale zupełnie nie tworzą klasycznych dupleksów, co ma miejsce przy braku PEG [78]. Podobne badania wykazały, że stłoczenie molekularne obojętnymi polimerami wywoływało przemianę strukturalną G-kwadrupleksu o sekwencji  $d(G_4T_4G_4)$  od struktury antyrównoległej do równoległej [79]. Ponadto, grupa Kana wykazała, że dla sekwencji  $d(T_2AG_3)_4$  i  $d[G_3(T_2AG_3)_3]$  wysokie stężenia PEG 200 sprzyjają tworzeniu się G-kwadrupleksów przy jednoczesnym niedoborze soli [80]. Heddi i Phan wyznaczyli również strukturę NMR G-kwadrupleksu o ludzkiej sekwencji telomerowej w warunkach stłoczenia molekularnego podobnego do tego w żywej komórce [81]. Badanie pokazało również, że glikol polietylenowy (PEG 200) indukował przejście konformacyjne form antyrównoległych oraz hybrydowych do równoległych struktur G-kwadrupleksów. Wyniki te pokazują, że równoległa forma G-kwadrupleksów o pętłach typu śmigłowego jest szczególnie preferowana w warunkach stłoczenia molekularnego. Kilka innych badań wykazało również, że stłoczenie molekularne powoduje przejście od form G-kwadrupleksów do postaci drutów guaninowych (G-wires) [77, 82], jak również regulację równowagi konformacyjnej G-kwadrupleks/dupleks [80, 82, 83]. Efekty stłoczenia molekularnego nie ograniczają się do struktur G-kwadrupleksów, ale mają również wpływ na stabilność form trójniciowych [84].

### 1.5.5 Wpływ super-zwinięcia DNA na G-kwadrupleksy

W genomowym DNA sekwencje bogate w guaniny oraz cytozyny ogólnie są ograniczone do tworzenia długich podwójnych nici DNA o klasycznym schemacie parowania Watsona i Cricka (G–C i A–T). Przejście strukturalne ze struktur B-DNA do innej postaci wymaga zazwyczaj zerwania wiązań wodorowych WC lub lokalnego odwinięcia podwójnej helisy. Lokalne rozwijanie DNA, które towarzyszy procesowi transkrypcji, jest częściowo ułatwiane dzięki elastycznej energii swobodnej związanej z negatywnym super-zwinięciem DNA. Spekuluje się, że podczas transkrypcji genów, kiedy polimeraza RNA porusza się wzdłuż DNA, dochodzi do generowania super-zwinięć, które są dodatnie przed polimerazą RNA, a ujemne za nią. W ten sposób transkrypcja prowadzi do akumulacji energii elastycznej w DNA, dzięki czemu może indukować rozdzielanie nici DNA za sobą i napędzać zmiany konformacyjne w sekwencjach podatnych na super-zwinięcie. Analizy bioinformatyczne przeprowadzone na całym ludzkim genomie zidentyfikowały sekwencje potencjalnie tworzące G-kwadrupleksy

właśnie w regionach bezpośrednio powyżej miejsca startu transkrypcji dla kilku ludzkich genów [47]. Ta obserwacja potwierdza pogląd, że G-kwadrupleksy mogą regulować transkrypcję genów.

## 1.6 Stabilność G-kwadrupleksów

Najbardziej powszechną formą DNA jest podwójna helisa, w której dwie nici są utrzymywane razem przez parowanie zasad azotowych typu Watsona i Cricka. Adeniny i tyminy są połączone ze sobą dwoma wiązaniami wodorowymi, podczas gdy guaniny tworzą trzy wiązania wodorowe z cytozynami. Dla takiego schematu parowania przypada średnio 1 lub 1,5 wiązania wodorowego na zasadę [85]. W odróżnieniu od tego, każda G-tetrada zawiera cykliczny układ ośmiu wiązań wodorowych typu Hoogsteena, dzięki czemu na jedną guaninę przypadają średnio dwa wiązania wodorowe. Sugeruje to, że struktury G-kwadrupleksów mogą być bardziej stabilną formą DNA i mieć wyższą temperaturę topnienia niż klasyczny dupleks DNA. Warto dodać, że G-kwadrupleksy RNA są ogólnie bardziej stabilne niż ich odpowiedniki DNA [50].

Pomimo tego, że struktury G-kwadrupleksów są stabilizowane przez liczne wiązania wodorowe Hoogsteena, do ich uformowania konieczne są jednowartościowe kationy metali, zwłaszcza  $K^+$  i  $Na^+$  [86]. Jony pełnią rolę stabilizującą poprzez zmniejszenie odpychania pomiędzy dipolami grup karbonylowych w centrum każdej G-tetrady. Struktury G-kwadrupleksu są unikalne wśród różnych form DNA pod względem wymagań co do jonów. Wiele publikacji zwraca uwagę na fakt, że jony potasu generują bardziej stabilne struktury G-kwadrupleksów niż jony sodu [87]. Jony potasu posiadają dość duży promień jonowy, dzięki czemu doskonale wpasowują się w przestrzeń pomiędzy dwiema sąsiednimi G-tetradami. Natomiast jony sodu mają mniejszy rozmiar, przez co łatwiej umiejscawiają się w płaszczyźnie G-tetrad, ale są przy tym bardziej ruchliwe [88]. Potas jest również dominującym kationem wewnątrzkomórkowym i powinien być częściej zaangażowany w tworzenie G-kwadrupleksów *in vivo* [46, 89].

Ogólnie jednowartościowe kationy stabilizują G-kwadrupleksy w następującej kolejności:  $K^+ > Na^+ > Rb^+ > NH_4^+ > Cs^+ \gg Li^+$  [53]. Natomiast kationy dwuwartościowe w następującej kolejności:  $Sr^{2+} \gg Ba^{2+} > Ca^{2+} > Mg^{2+}$  [90]. W normalnych warunkach telomerowe sekwencje *Oxytricha* 5'-d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>)-3' spontanicznie asocjują tworząc antyrównoległy G-kwadrupleks, natomiast ich destabilizacja zachodzi już przy milimolowych stężeniach jonów  $Zn^{2+}$ ,  $Co^{2+}$ ,  $Mn^{2+}$ ,  $Mg^{2+}$  oraz  $Ca^{2+}$  [91].

Rozważając stabilność G-kwadrupleksów, należy również uwzględnić efekty związane z długością pętli, sekwencją pętli, jak również długością fragmentów  $G_n$ . Wszystkie te czynniki mogą niezależnie wpływać na przyjmowaną przez G-kwadrupleks topologię. Każdy jednociowy G-kwadrupleks zawiera trzy pętle, których długość i sekwencja odgrywa dużą rolę

w regulacji stabilności. Pokazano, że bardzo długie pętle częściej biorą udział w tworzeniu mało stabilnych G-kwadrupleksów, podczas gdy krótsze pętle mają większą stabilność w warunkach fizjologicznych. Na przykład, sekwencje  $d[G_3(TG_3)_3]$  z obrębu regionu *c-MYC* i inhibitora integrazy HIV  $d[G_3(TG_3)_3]$  mogą tworzyć bardzo stabilne G-kwadrupleksy nawet w obecności komplementarnych do nich nici bogatych w cytozyny [92]. Jednocząsteczkowe G-kwadrupleksy mają najczęściej trzy G-tetrazy ułożone jedna nad drugą i tworzące zwarty układ. Istnieją badania, które wykazały, że zwiększenie lub zmniejszenie liczby G-tetrad może niekorzystnie wpływać na stabilność G-kwadrupleksów [93]. Okazało się, że fragmenty  $G_n$  nici o długości od 3 do 7 zasad mogą tworzyć G-kwadrupleksy, przy czym stabilność wzrasta wraz ze wzrostem długości. Wyjątkiem od tej reguły jest sekwencja  $(G_3T)_4$ , która tworzy bardzo stabilną równoległą strukturę G-kwadrupleksu w obecności roztworu zawierającego  $K^+$  lub  $Na^+$  [93].

## 1.7 G-kwadrupleksy w genomie

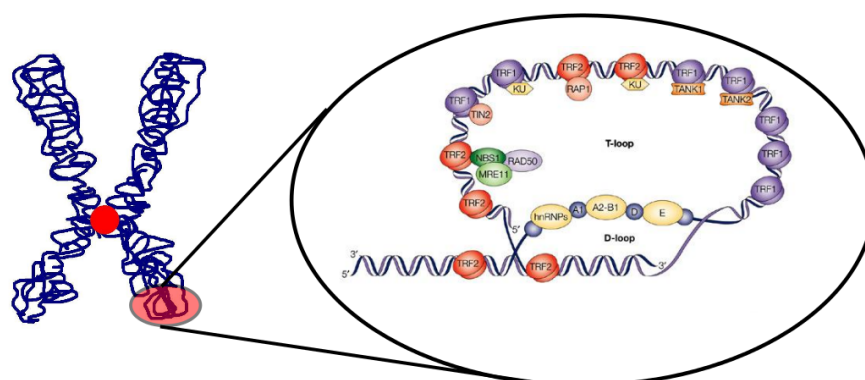
Przeszukanie genomów prokariotycznych pod kątem miejsc potencjalnie tworzących G-kwadrupleksy doprowadziło do zidentyfikowania jako jednego z takich rejonów jednoniciowego naddatka na 3'-końcu liniowych chromosomów u *Streptomyces* [94]. Z kolei od czasu zidentyfikowania różnych sekwencji potencjalnie tworzących G-kwadrupleksy w rejonach promotorowych genów zaangażowanych w transkrypcję, transdukcję i biosyntezę metabolitów, G-kwadrupleksy są również rozpatrywane jako możliwe elementy regulacyjne [95]. Ponadto inne badania bioinformatyczne wskazały na pewne powtarzające się sekwencje bogate w guaniny potencjalnie mogące tworzyć G-kwadrupleksy. Chodzi o biologicznie istotne regiony genomu eukariotycznego, włączając w to telomerowe DNA i geny kodujące białko w onkogennych sekwencjach promotorów takich jak *c-myc* [96], PDGF-A [67], hTERT [97]. Inne rejonny chromosomalne są również bogate w sekwencje potencjalnie tworzące G-kwadrupleksy. Może do nich za wliczyć rybosomalne DNA (kodujące rRNA), rejonny przełącznika ciężkiego łańcucha immunoglobuliny, powtarzające się sekwencje mikro- i mini- satelitów [98], tandemowe powtórzenia sekwencji istniejące w genie podatności na siatkówczaka [99], gen FMR-1, gen  $\beta$ -globiny [100], a także rejon polimorficzny powyżej genu insuliny [101]. Dość znaczna akumulacja sekwencji potencjalnie tworzących G-kwadrupleksy w całym genomie daje silne przesłanki co do tego, że struktury G-kwadrupleksów odgrywają ważną rolę w procesach komórkowych [51, 102].

Na skutek ulepszonej metodologii obliczeniowej dwa niezależne badania bioinformatyczne na całym genomie rzuciły światło na rozkład sekwencji potencjalnie tworzących G-kwadrupleksy, a także na występowanie najpopularniejszych pętli w tych sekwencjach [48, 103]. Dwie grupy badawcze systematycznie przeszukiwały ludzki genom, uwzględniając jedynie sekwencje pasujące do zadanego wzorca  $(G_{3-5}L_{1-7}G_{3-5}L_{1-7}G_{3-5}L_{1-7}G_{3-5})$  z których

potencjalnie może powstać jednocząsteczkowy G-kwadrupleks. Założono, że fragmenty z guaninami ( $G_n$ ) są ograniczone do długości od trzech do pięciu zasad, natomiast fragmenty z pętlami (L) mogą zawierać jakąkolwiek zasadę i ich kombinację, a ich długość jest z zakresu 1–7 [48]. Wyniki tych badań wykazały, że całkowita liczba sekwencji potencjalnie tworzących G-kwadrupleksy w ludzkim genomie wynosi aż 5713000. Jednakże uwzględnia ona wszystkie możliwe sekwencje nakładające się, a po ich usunięciu zmniejsza się około 15-krotnie do 375157 sekwencji [48]. Ponadto grupa Balasubramaniana wykorzystała bazę danych ludzkiego genomu ENSEMBL do zidentyfikowania sekwencji potencjalnie tworzących G-kwadrupleksy w we wszystkich chromosomach, w szczególności w rejonach eksonów DNA. Badania pokazały, że rejony o niskiej częstotliwości występowania fragmentów  $G_n$  znaleziono w niciach kodujących rejony eksonów, co sugeruje, że sekwencje potencjalnie tworzące G-kwadrupleksy są niedostatecznie reprezentowane w sekwencjach mRNA [103].

Wcześniejsze badania wykazały, że stabilność i topologia większości G-kwadrupleksów zależy od długości pętli i ich sekwencji [55, 56]. Dlatego badania bioinformatyczne skupiły się również na analizie długości potencjalnych pętli [103], wyliczeniu najczęstszych sekwencji pętli oraz ustaleniu częstotliwości ich występowania w obrębie sekwencji potencjalnie tworzących G-kwadrupleksy [48]. Pokazano, że najczęściej występującą kombinacją długości pętli jest 1:1:1 (8%). Sugeruje to, że dłuższe pętle generalnie nie są promowane w sekwencjach potencjalnie tworzących G-kwadrupleksy [103]. Co więcej, najczęściej występującą sekwencją pętli jest pojedyncza zasada adeniny lub tyminy. Wprawdzie dłuższe pętle występują rzadziej niż krótsze to jednak kombinacja zasad CCTGT w potencjalnej pierwszej pętli od 5'-końca jest najczęściej występującą sekwencją dla pętli dłuższych niż jedna zasada [48].

## 1.8 Telomerowe DNA i G-kwadrupleksy



RYSUNEK 1.10: Położenie telomerów w chromosomach. Telomery oprócz nici DNA, która tworzy charakterystyczne pętle, zawierają również przyłączone do nich wyspecjalizowane białka TRF1, TRF1 oraz POT1. Schemat zaadaptowany z [104].

Telomery są wyspecjalizowanymi kompleksami kwasów nukleinowych i białek występującymi na końcach wszystkich eukariotycznych chromosomów. Ich występowanie potwierdzono w zielonych algach [105], pierwotniakach orzęsionych [106], drożdżach [107] i kręgowcach [43]. Tandemowo powtarzana sekwencja telomerowa różni się w zależności od organizmu, ale jej ogólna postać jest następująca:  $d[T_{1-3}-(T/A)-G_{3-4}]_n$  [44], np.  $(TTAGGG)_n$  u wszystkich wyższych eukariotów. Telomery stabilizują końce liniowych chromosomów eukariotycznych i chronią je przed degradacją przez egzonukleazy i rekombinacją oraz zapobiegają ich rozpoznawaniu jako podwójnych pęknięć nici. Ludzkie telomerowe DNA składa się z około 10 tysięcy par zasad w postaci dwuniciowego DNA z tandemowymi powtórzeniami sekwencji  $(TTAGGG)_n$ . Dodatkowo na jego 3'-końcu znajduje się jednoniciowy naddatek o długości 100–200 zasad i takiej samej sekwencji. Jak pokazano, te rejony bogate w guaniny są szczególnie prawdopodobnymi miejscami tworzenia się G-kwadrupeksów. U ssaków końce liniowych chromosomów mogą uczestniczyć w tworzeniu bardziej kompaktowych struktur tzw. nasadek telomerów, w celu ochrony całego chromosomu przed uszkodzeniem w wyniku fragmentacji lub łączenia się chromosomów (Rys. 1.10). Zidentyfikowano szereg białek specyficznie wiążących się z telomerowym DNA. Są wśród nich zarówno białka oddziałujące bezpośrednio z DNA (TRF1, TRF2 i POT1) jak i białka stowarzyszone z nimi (RAP1, TPP1 i TIN2). Białka TRF1 i TRF2 wiążą się bezpośrednio z dupleksowym telomerowym DNA tworząc ochronną strukturę tzw. pętlę T (T-loop). Pokazano także, że białko POT1 chętniej wiąże się z jednoniciowym 3'-naddatkiem i jest również związane z kompleksem białek telomerowych (tzw. kompleks szelterynowy). W obrębie tego kompleksu białko TIN2 działa jako łącznik pomiędzy podjednostkami TRF1 i TRF2 i angażuje białko POT1 w kompleks szelterynowy za pośrednictwem podjednostki TPP1. Oddziaływanie pomiędzy RAP1 i TRF2 powoduje pośrednie wiązanie ich z kompleksem szelterynowym [108, 109]. Generalnie pętla T i dodatkowa tzw. pętla D są nierozłącznymi częściami struktury telomeru (patrz Rys. 1.10). Pętla T związana jest z całą strukturą, w której to końce są zapętlone do tyłu, tworząc strukturę kołową. Natomiast pętla D odnosi się do krótkiego odcinka z jednoniciowym fragmentem 3'-końca wplecionego w dupleks [110, 111].

W większości normalnie funkcjonujących komórek somatycznych telomeraza nie jest aktywna, a długość telomerów stopniowo zmniejsza się po każdym cyklu podziału komórkowego z powodu niepełnej replikacji nici opóźnionej. Prowadzi to do zatrzymania wzrostu lub nawet apoptozy po około 40 podziałach w przypadku większości komórek somatycznych [112]. Telomeraza jest jednak aktywna w komórkach płciowych i innych komórkach stale dzielących się (np. komórki nabłonka), a w ponad 85% komórek nowotworowych także występuje wysoki poziom tego enzymu [113]. Enzym ten jest rybonukleoproteiną zawierającą matrycę RNA, z której powtórzenia TTAGGG są kopiowane przez odwrotną transkryptazę i dodawane do jednoniciowych 3'-końców telomerów, w konsekwencji utrzymując długość chromosomu. Zwinięcie do struktury G-kwadrupeksu jednoniciowego fragmentu bogatego

w guaniny hamuje aktywność telomerazy [88].

Mocne dowody na istnienie G-kwadrupeksów *in vivo* pochodzą już z najwcześniejszych obserwacji, że specyficzne względem G-kwadrupeksów przeciwciała Sty49 wiążą się do telomerów z makrojąder pierwotniaków rzęskowych *Stylonychia lemnae* [45]. Wykazano, że jednoniciowy fragment przeciwciała Sty49 wiąże się zarówno z telomerowymi G-kwadrupeksami równoległymi oraz antyrównoległymi. Powinowactwo Sty49 do obu konformacji G-kwadrupeksów jest podobne (w obu przypadkach w zakresie 3–5 nM). Kolejne dowody potwierdzające istnienie G-kwadrupeksów w telomerach są związane z dwoma białkami TEBP,  $\alpha$  i  $\beta$ , które, jak wykazano sprzyjają tworzeniu się G-kwadrupeksów *in vivo* [114].

Stosunkowo niedawno potwierdzono powstawanie G-kwadrupeksów DNA w jądrach komórek ssaków poprzez wizualizację przy pomocy przeciwciał znakowanych fluorescencyjnie [115]. Grupa Balasubramaniana opracowała strukturalnie specyficzne przeciwciało BG4, które preferencyjnie wiąże się do jedno- i wielocząsteczkowych G-kwadrupeksów, jednocześnie nie wykazując powinowactwa do innych struktur kwasu nukleinowego, takich jak spinki RNA, jednoniciowe i dwuniciowe DNA. Dalsze testy wykazały brak swoistości BG4 dla różnych konformacji G-kwadrupeksów. Co ciekawe obserwowano cykliczne zwiększanie i zmniejszone barwienia w różnych fazach komórkowych, co sugeruje, że powstawanie G-kwadrupeksów może być zależne od cyklu komórkowego. W niezależnych badaniach pokazano, że drobnocząsteczkowy ligand pirydozyna może bezpośrednio oddziaływać z DNA, stabilizując G-kwadrupeksy, a następnie promować apoptozę w ludzkich komórkach nowotworowych poprzez indukcję uszkodzeń DNA [116]. W związku z tym pojawiły się propozycje stabilizacji telomerowych G-kwadrupeksów przez małe cząsteczki, co miałyby prowadzić do hamowania wydłużania telomerazy i w konsekwencji do zatrzymania dalszego wzrostu komórek rakowych.

Wiele dotychczasowych badań skupiało się na zrozumieniu źródeł różnorodności strukturalnej ludzkiej telomerowych struktur G-kwadrupeksów (patrz Rys. 1.6). Wczesne badania z grupy Patela wykazały, że fragmenty DNA z pojedynczymi powtórzeniami d(TTAGGGT) ludzkiej sekwencji telomerowej mogą tworzyć równoległy G-kwadrupeks w roztworze zawierającym jony  $K^+$  [67]. Struktura ta składa się z trzech G-tetrad, a kąty glikozydowe wszystkich guanin są w konformacji *anti*. Następnie ustalono topologię oraz strukturę krystaliczną fragmentów DNA o dwukrotnie powtórzonej ludzkiej sekwencji telomerowej d(TAG<sub>3</sub>-TTAG<sub>3</sub>T). Rentgenografia strukturalna pokazała, że ten 12-nukleotydowy fragment tworzy dwucząsteczkowy G-kwadrupeks, a wszystkie fragmenty guaninowe  $G_n$  są równoległe względem siebie i połączone za pomocą pętli śmigłowych [61]. Późniejsze badania NMR dla tej sekwencji potwierdziły występowanie takiej samej topologii w wodnym roztworze kationów  $K^+$  [70]. Na podstawie badań NMR dla potrójnie powtórzonej ludzkiej sekwencji telomerowej 5'-G<sub>3</sub>TTAG<sub>3</sub>TTAG<sub>3</sub>T-3' wyznaczono najbardziej prawdopodobną topologię i strukturę nasadek telomerowych tworzących się na końcach eukariotycznych chromosomów [117].

Ponadto badanie to wskazało na dość specyficzne cechy topologii uformowanego G-kwadrupeksu. Jedną z nich jest tworzenie rdzenia guaninowego składającego się z trzech równoległych fragmentów guaninowych jednej nici oraz czwartego fragmentu o przeciwnym zwrocie i pochodzącego z osobnej nici DNA. Taka struktura G-kwadrupeksu (3+1) ma potencjał do formowania się *in vivo*, przy uwzględnieniu tego, że jednoniciowy 3'-koniec jest wtrącony w fragment dupleksu poprzedzający go [108]. Potencjalnie duże znaczenie telomerowych G-kwadrupeksów DNA dla procesów starzenia się komórek oraz nowotworzenia sprawiło że rozwiązano także istotne biologicznie struktury G-kwadrupeksów o czterech powtórzeniach sekwencji TTAGGG. Wczesne badania NMR dotyczące sekwencji ludzkich telomerów  $d[AG_3(T_2AG_3)_3]$  wykazały, że roztwory zawierające jony sodu zawężają możliwości zwijania G-kwadrupeksów do struktur o konformacji zawierających antyrównoległe ułożenie fragmentów guaninowych połączonych centralną pętlą diagonalną oraz dwoma pętlami bocznymi [67]. Natomiast rentgenowska struktura krystaliczna o tej samej sekwencji jednak uzyskana w obecności  $K^+$  zdecydowanie różni się od tej pochodzącej z NMR. W tym przypadku badana sekwencja utworzyła jednocząsteczkową strukturę G-kwadrupeksu złożoną z trzech płaszczyzn guaninowych z czterema równoległymi fragmentami guaninowymi oraz wszystkimi trzema pętlami śmigłowymi [61]. Ponadto widma CD ludzkiej sekwencji telomerowej w roztworach  $K^+$  charakteryzują się dodatnim sygnałem przy paśmie 260 nm i ujemnym sygnałem przy paśmie 240 nm, co sugeruje formowanie struktury równoległej [118]. Z powodu wysokiego wewnątrzkomórkowego stężenia  $K^+$  (140 mM) w porównaniu do  $Na^+$ , G-kwadrupeks typu równoległego (śmigłowego) wydaje się być fizjologicznie istotną konformacją w obrębie ludzkich telomerów. Jednak systematycznie prowadzone badania NMR na ludzkich sekwencjach telomerowych wykazały, że w obecności jonów potasu tworzą się hybrydowe formy G-kwadrupeksów z równoległymi i antyrównoległymi fragmentami guaninowymi [119].

## 1.9 Minimalne struktury G-kwadrupeksów

Z informacji, które przytoczyłem we wcześniejszych rozdziałach można wywnioskować, że najmniejsza możliwa stabilna struktura DNA, którą można nazwać G-kwadrupeksem musi składać się z przynajmniej dwóch płaszczyzn guaninowych (G-tetrad). W przypadku jednoniciowego G-kwadrupeksu teoretyczna minimalna struktura G-kwadrupeksu musi zawierać zatem 8 nukleotydów guaninowych (po 4 na płaszczyznę) oraz 3 dodatkowe nukleotydy przedzielające 4 fragmenty GG i tworzące pętle. Wynika z tego, że ogólna sekwencja dwupłaszczyznowych G-kwadrupeksów będzie miała postać GGXGGYGGZGG, gdzie G oznacza pozycję guaniny, a X, Y, Z oznacza pozycję sekwencji zasad.

Niestety w literaturze nie znalazłem wielu doniesień dotyczących jednoniciowych G-

kwadrupleksów dwupłaszczyznowych, w szczególności o najkrótszych możliwych sekwencjach. Jest to niekorzystna sytuacja, ponieważ wiedza na temat molekularnych podstaw stabilności małych modelowych struktur powinna przekładać się także na struktury bardziej złożone, które są zdecydowanie trudniejsze w badaniach. Poza tym teoretyczna liczba wszystkich możliwych struktur dwupłaszczyznowych jest zdecydowanie mniejsza niż liczba struktur trójpłaszczyznowych, co powinno być także pewnym ułatwieniem w badaniach.

Podczas dokładnego przeszukania bazy Protein Data Bank pod kątem najkrótszych sekwencji tworzących jednoniciowe G-kwadrupleksy ustaliłem, że do tej pory zaobserwowano jedynie kilka struktur dwupłaszczyznowych G-kwadrupleksów. Szczegóły przedstawiłem w Tabeli 1.2. Część z tych struktur dzieli tę samą topologię zwinięcia oraz jest bardzo podobna strukturalnie do siebie, co jest wynikiem użycia tej samej sekwencji w różnych eksperymentach.

TABELA 1.2: Zestawienie struktur jednoniciowych G-kwadrupleksów dwupłaszczyznowych zdeponowanych w bazie PDB.

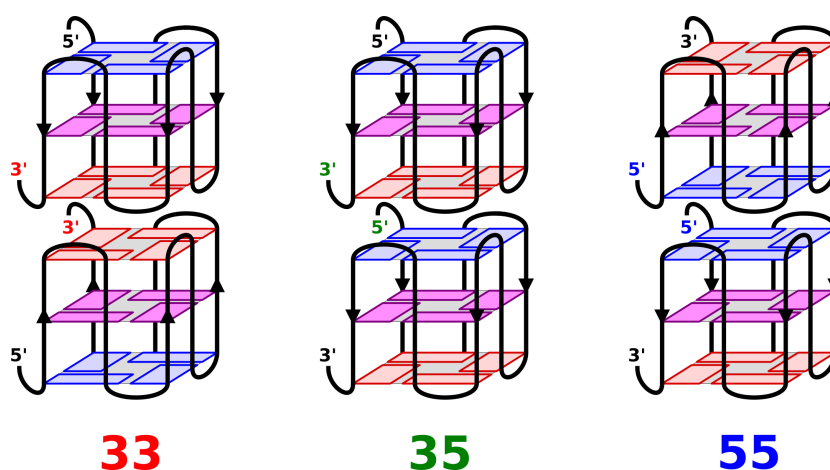
Kod PDB	Sekwencja	Źródło
2KM3	AGGGCTAGGGCTAGGGCTAGGG	[120]
148D	GGTTGGTGTGGTTGG	[121]
1C32, 1C34, 1C35, 1C38	GGTTGGTGTGGTTGG	[122]
1QDF, 1QDH	GGTTGGTGTGGTTGG	[123]
2N60	TTGTGTGGGTGGGTGGGT	[124]
2M6V	GGGTTGGGTTTTGGGTGGG	Nieopublikowane
1I34	GGTTTTGGCAGGGTTTTGGT	[125]
2KOW	TAGGGTAGGGTAGGGTAIGG	[126]
2MFU	TGGGTTTTGGGTTGGGTTTGGG	Nieopublikowane
2KF8	GGGTTAGGGTTAGGGTTAGGGT	[127]

Łatwo zauważyć, że w większości tych struktur pętle posiadają co najmniej dwie zasady azotowe (głównie tyminy). Oznacza to, że nie są to struktury, które mogą być brane pod uwagę jako struktury z możliwie najkrótszą sekwencją nici DNA. Co więcej z badań nad G-kwadrupleksami trójpłaszczyznowymi wynika, że występowanie pojedynczych tymin w pętlach jest wyjątkowo korzystne energetycznie dla ogólnej stabilności molekuly. Badania kalorymetryczne oraz spektroskopowe [128] pokazały, że krótkie sekwencje, takie jak  $G_3TG_3TG_3TG_3$ , tworzą G-kwadrupleksy posiadające trzy płaszczyzny oraz konformację równoległą z pętlami śmigłowymi. Co więcej posiadają także wysoką stabilność temperaturową. Natomiast sekwencje o innym składzie pętli lub o większej liczbie tymin w pętlach wykazywały jedynie marginalną stabilność lub brak przewagi jednej formy konformacyjnej. Bazując na takich informacjach, można założyć, że G-kwadrupleksy dwupłaszczyznowe posiadające pojedyncze tyminy w pętlach także powinny wykazywać znaczną stabilność względem innych sekwencji i tworzyć konformacje równoległe.



Niestety teoretyczne przewidywanie struktury G-kwadrupleksów na podstawie sekwencji nici DNA jest obecnie bardzo trudne. Konwencjonalne metody symulacji są w tym względzie nieużyteczne, ponieważ czasy formowania struktur G-kwadrupleksów ( $\gg \mu\text{s}$ ) są zdecydowanie poza ich zasięgiem. W związku z tymi trudnościami konieczne jest opracowanie nowej metody szacowania stabilności potencjalnych struktur G-kwadrupleksów.

## 1.10 Dimeryzacja jednoniciowych G-kwadrupleksów



RYSUNEK 1.11: Trzy możliwe sposoby dimeryzacji G-kwadrupleksów o topologii równoległej (mody dimeryzacji 3'-3', 3'-5' i 5'-5').

Od dawna uważano, że ze względu na swoją warstwową budowę, G-kwadrupleksy powinny wykazywać tendencję do agregacji poprzez oddziaływania powierzchniowe pomiędzy zewnętrznymi G-tetradami. W szczególności powinno to być możliwe dla G-kwadrupleksów o topologii równoległej, w których to zewnętrzne G-tetrazy są silnie eksponowane do rozpuszczalnika. Zgodnie z przewidywaniami, tworzenie struktur wyższych rzędów przez G-kwadrupleksy o konformacji równoległej zostało zaobserwowane w różnych badaniach kalorymetrycznych dla długich sekwencji o powtórzeniach GGGT [128] oraz TTAGGG [129]. Ponadto przy użyciu połączenia metod spektroskopii dichroizmu kołowego oraz spektrometrii mas, pokazano, że agregacja G-kwadrupleksów o konformacji równoległej silnie zależy od obecności zasad flankujących. Sugeruje to, że agregacja G-kwadrupleksów następuje przez ułożenie w stos kolejnych monomerów kontaktujących się przez zewnętrzne G-tetrazy. Poza tym wskazano, że dimery oraz trimery G-kwadrupleksów są najbardziej prawdopodobnymi formami zagregowanymi [130]. Powstawanie takich form zostało potwierdzone za pomocą spektroskopii NMR. Analiza równowagi termodynamicznej dimeryzacji G-kwadrupleksów o konformacji równoległej pokazała, że spośród trzech teoretycznie możliwych orientacji G-kwadrupleksów względem siebie (patrz Fig. 1.11), agregacja za pośrednictwem G-tetrad na 5'-końcach jest zdecydowanie najbardziej preferowana [131].

Preferencja do dimeryzacji G-kwadrupleksów poprzez 5'-końcowe G-tetrazy została następnie potwierdzona poprzez wyznaczenie wysokorozdzielczych struktur dimerów w technice NMR. Wśród tych struktur można wymienić modelowe sekwencje  $(GGA)_4$  oraz  $GIGT-(GGGT)_3$  [132, 133], jak również fragmenty sekwencji tworzące G-kwadrupleksy w genach *N. gonorrhoeae* takich jak pilE lub ludzkiej minisatelity CEB1 [134, 135]. Preferencja do agregacji poprzez G-tetrazy na 5'-końcach jest również odzwierciedlona w strukturach krystalograficznych różnych struktur równoległych G-kwadrupleksów. W tym przypadku sieć krystaliczna jest tworzona przez pojedyncze jednostki G-kwadrupleksów, które w większości układają się względem siebie poprzez G-tetrazy na 5'-końcach [61, 136, 137].

Sugeruje się, że dimeryzacja G-kwadrupleksów może dodatkowo stabilizować poszczególne jednostki G-kwadrupleksów tworzących dimer. Może mieć to istotne znaczenie dla efektywności upakowania sekwencji tworzących G-kwadrupleksy, w tym w obrębie telomerowego 3'-naddatka [138, 139]. Dimeryzacja G-kwadrupleksów może również odgrywać rolę w procesie skracania telomerów, który zachodzi stopniowo [140]. Jednakże przewidywanie, które konkretne struktury G-kwadrupleksów są podatne na agregację (dimeryzację) nie jest trywialne, ponieważ proces ten zależy od kilku czynników. Do tych czynników wlicza się ogólną topologię fałdowania, sekwencję pętli i zasad flankujących oraz stężenie i rodzaj kationów [129, 131]. Ponadto, równowagę dimeryzacji można modulować ligandami, takimi jak pochodne mezoporfiryny lub naftalenu. Sprzyjają one tworzeniu się dimerów zbudowanych z równoległych G-kwadrupleksów poprzez łączenie się z 3'-końcowymi G-tetradami, jednocześnie umożliwiając interakcję jednostek G-kwadrupleksów poprzez niezaangażowaną w inne oddziaływania 5'-końcową G-tetradę [136, 141]. Ponadto okazało się, że niektóre związki, takie jak kationowe porfiryny czy berberyny zwiększają stabilność dimerów G-kwadrupleksów poprzez interkalację pomiędzy G-tetradami występującymi na interfejsie dimeryzacji [142–147]. Ta ciekawa właściwość tych związków może mieć szczególne znaczenie przy projektowaniu ligandów ukierunkowanych na G-kwadrupleksy.

## 2 Założenia i cele pracy

Głównym celem niniejszej pracy jest zrozumienie molekularnego podłoża wybranych cech zachowania konformacyjnego różnych form kwasu deoksyrybonukleinowego. Po pierwsze zdecydowałem się na zbadanie mikroskopowych przyczyn różnorodności strukturalnej oraz stabilności termodynamicznej struktur typu G-kwadrupeksów, czyli niekanonicznych form DNA o potwierdzonym znaczeniu biologicznym. Następnie postanowiłem ustalić mechanizmy odpowiedzialne za zmiany konformacyjne zachodzące w klasycznej podwójnej helisie DNA pod wpływem obecności wybranych współrozpuszczalników.

W obu przypadkach zdecydowałem się na wykorzystanie współcześnie dostępnych metod symulacyjnych chemii obliczeniowej pozwalających na dogłębną analizę strukturalną oraz energetyczną badanego układu z rozdzielczością atomową. Metoda symulacyjna, którą wybrałem to w pełni atomowa dynamika molekularna.

W przeciwieństwie do klasycznej podwójnej helisy DNA, siły napędzające formowanie G-kwadrupeksów w środowisku wodnym nie zostały dobrze poznane. W związku z tym przedmiotem moich badań jest ustalenie głównych mechanizmów pojawiających się na różnym poziomie złożoności strukturalnej i odpowiedzialnych za stabilizację G-kwadrupeksów.

Pierwszym szczegółowym celem pracy jest zbadanie termodynamiki i kinetyki tworzenia wspólnego motywu strukturalnego wszystkich G-kwadrupeksów, jakim jest płaszczyzna guaninowa (G-tetrada). Aby osiągnąć ten cel, założyłem przede wszystkim wyznaczenie profilu energii swobodnej opisującego dysocjację pojedynczej G-tetrady na 3'-końcu w pełni uformowanym G-kwadrupeksie. Następnie, poprzez szczegółową analizę kolejnych etapów procesu dysocjacji G-tetrady, ustaleń towarzyszącą temu sekwencję zmian strukturalnych na poziomie molekularnym. Poprzez szczegółową dekompozycję wkładów entalpowych wnoszących do stabilności G-tetrady, planuję ustalić główne siły odpowiedzialne za stabilność G-tetrady. Równolegle oszacuję również towarzyszącą dysocjacji G-tetrady zmianę entropii konformacyjnej i jej główne źródła strukturalne oraz w ustaleń w jakim stopniu wnosi ona do całkowitej energii swobodnej dysocjacji G-tetrady.

Kolejny etap badań obejmuje zbadanie wpływu rejonów pętlowych na stabilność G-kwadrupeksów. W szczególności zamierzam przeprowadzić oszacowanie stabilności modelowych struktur G-kwadrupeksów w zależności od długości ich poszczególnych pętli. W tym celu opracuję metodę szacowania stabilności G-kwadrupeksów na podstawie czasu życia

reprezentatywnego zestawu modelowych struktur G-kwadrupleksów w symulacjach równowagowych. Struktury te wygeneruję poprzez opracowany przeze mnie protokół zwijania swobodnych nici DNA do struktury typu G-kwadrupleksu o arbitralnej topologii. Zakładam, że na podstawie oszacowanych stabilności dla zbioru reprezentatywnych sekwencji będę mógł wyciągnąć ogólne wnioski dotyczące wpływu długości i sekwencji rejonów pętlowych na stabilność struktur G-kwadrupleksów. Część uzyskanych przeze mnie wyników zamierzam skonfrontować, ze wynikami eksperymentalnymi uzyskanymi przez grupę badawczą prof. Z. Gdaniec z Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk.

Ostatnim elementem zaplanowanych badań nad G-kwadrupleksami jest zbadanie molekularnych determinantów procesu tworzenia struktur wyższych rzędów w następstwie dimeryzacji G-kwadrupleksów. W tym przypadku celem jest ustalenie preferencji dwóch G-kwadrupleksów o topologii równoległej do formowania dimerów o trzech możliwych sposobach ułożenia się względem siebie (modów dimeryzacji). Następnie zakładam przeprowadzenie szczegółowej analizy strukturalnej uzyskanego w symulacjach zestawu konformacji dimerów G-kwadrupleksów. Na podstawie tych analiz oraz dalszych dekompozycji entalpii dimeryzacji G-kwadrupleksów na poszczególne oddziaływania możliwe będzie dalej określenie sił sprzyjających oraz przeciwdziałających dimeryzacji. Ostatnim elementem tych badań jest ustalenie wpływu zwiększonej oraz zredukowanej siły jonowej na preferencję do poszczególnych modów dimeryzacji.

Druga część moich badań ma celu ustalenie molekularnych determinantów równowagi konformacyjnej podwójnej helisy DNA w obecności dwóch reprezentatywnych związków denaturujących – mocznika i betainy. Efektem działania obu substancji jest destabilizacja cząsteczki DNA, która przebiega w trochę inny sposób, zależny od sekwencji. W związku z tym po pierwsze założyłem wyznaczenie globalnych i lokalnych zmian konformacyjnych DNA powstających wskutek obecności w roztworze obu denaturantów. Następnie wyznaczę przestrzenny rozkład cząsteczek wody i cząsteczek denaturantów wokół DNA oraz towarzyszące temu zmiany solwatacyjne wokół cząsteczki DNA, takie jak zmiana liczby cząsteczek wody czy liczba tworzonych wiązań wodorowych. Na podstawie analizy poszczególnych wkładów energetycznych w układach planuję wyznaczyć oddziaływania odpowiedzialne za mechanizmy denaturacji DNA. Ta część badań będzie prowadzona w ramach współpracy z grupą badawczą prof. J. Stangreta z Katedry Chemii Fizycznej PG, w celu interpretacji danych uzyskanych za pomocą spektroskopii IR oraz kalorymetrii DSC i ITC.

# 3 Podstawy metod badawczych

## 3.1 Podstawy teoretyczne

Metody symulacji komputerowych są obecnie nieodłączną częścią badań prowadzonych w laboratorium, nierzadko generując lub potwierdzając wcześniej wysunięte hipotezy badawcze. Szczególnie popularną metodą chemii obliczeniowej jest dynamika molekularna, rozwijana obecnie intensywnie zarówno od strony teorii, jak i implementacji technicznej. Skutkiem tego są coraz większe możliwości predykcyjne tych metod umożliwiające np. przewidywanie struktur drugorzędowych białek, DNA oraz RNA a także projektowanie nowych ligandów o założonej aktywności biologicznej [148, 149]. Dodatkowo, ze względu na niskie koszty i możliwość badania układów molekularnych w warunkach zbliżonych do fizjologicznych, symulacje są niekiedy atrakcyjnym substytutem dla eksperymentów prowadzonych w laboratorium. Wyżej wymienione zalety skłoniły mnie do użycia dynamiki molekularnej jako narzędzia badawczego do osiągnięcia celów przedstawionych w poprzednim rozdziale.

### 3.1.1 Przybliżenie Borna-Oppenheimera

Podstawowym zadaniem dynamiki molekularnej jest wyznaczenie ewolucji czasowej układu atomów w przestrzeni fazowej poprzez całkowanie klasycznych równań ruchu, w wyniku czego otrzymywana jest odpowiednia reprezentacja parametrów makroskopowych badanych układów w stanie równowagi. Przy pomocy specjalnych technik możliwe jest także badanie procesów charakteryzujących się trudno dostępnymi regionami przestrzeni konformacyjnej. Pierwotnym założeniem dynamiki molekularnej jest zredukowanie złożoności rozwiązywania równań mechaniki kwantowej do całkowania Newtonowskich równań ruchu. Uzyskuje się to poprzez zastosowanie kilku przybliżeń, które są na ogół wystarczająco dobrze spełnione w warunkach zbliżonych do warunków standardowych. Zgodnie z postulatami nierelatywistycznej mechaniki kwantowej ewolucja układu kwantowego opisywana jest przez zależne od czasu równanie Schrödingera:

$$\hat{H}\psi = i\hbar \frac{\partial}{\partial t} \psi \quad (3.1)$$

W sytuacji, gdy Hamiltonian  $\hat{H}$  nie zależy w sposób jawny od czasu, rozwiązanie równania 3.1 sprowadza się do wyznaczenia jego stanów stacjonarnych, tj. rozwiązania problemu własnego dla Hamiltonianu  $\hat{H}$ :

$$\hat{H}\Psi_i = E_i\Psi_i \quad (3.2)$$

Wartości własne równania 3.2 oznaczają energię  $E_i$  stanów stacjonarnych, a wektory własne to amplitudy funkcji falowej w poszczególnych stanach stacjonarnych  $\Psi_i$ . Niestety równanie 3.2 może zostać rozwiązane analitycznie tylko dla najprostszych układów takich jak atom wodoropodobny. W przypadku układów wieloatomowych rozwiązanie problemu 3.2 można znacząco ułatwić przez wprowadzenie tzw. przybliżenia adiabatycznego. Zakłada ono, że w wyniku bardzo małej bezwładności elektronów (masa elektronu jest 1800 razy mniejsza od masy najprostszego jądra) struktura elektronowa atomów i cząsteczek dostosowuje się nieskończenie szybko do nowych położenia jąder. Przybliżenie adiabatyczne pozwala na rozdzielenie funkcji falowej z równania 3.2 na wkład elektronowy oraz jądrowy:

$$\Psi_i = \Psi_i(\mathbf{R}, \mathbf{r}) \approx \psi_i(\mathbf{r}; \mathbf{R})\phi_i(\mathbf{R}) \quad (3.3)$$

W powyższej formule elektronowa funkcja falowa  $\psi_i(\mathbf{r}; \mathbf{R})$  zależy od położenia elektronów  $\mathbf{r}$  oraz parametrycznie od położenia jąder  $\mathbf{R}$ , natomiast funkcja falowa jąder  $\phi_i(\mathbf{R})$  zależy wyłącznie od ich położenia  $\mathbf{R}$ . W przypadku ustalonej konfiguracji atomów funkcja falowa  $\psi_i(\mathbf{r}; \mathbf{R})$  jest rozwiązaniem tzw. elektronowego równania Schrödingera:

$$\hat{H}_{el}(\mathbf{R})\psi_i(\mathbf{r}; \mathbf{R}) = E_i^{el}(\mathbf{R})\psi_i(\mathbf{r}; \mathbf{R}) \quad (3.4)$$

W równaniu tym  $\hat{H}_{el}(\mathbf{R})$  jest tzw. Hamiltonianem elektronowym, który to zawiera operator energii kinetycznej elektronów oraz operatory wszystkich oddziaływań elektrostatycznych: elektron–jądro, elektron–elektron i jądro–jądro. W następnym kroku poprzez wstawienie równania 3.3 do równania 3.2 oraz zgodnie z podejściem Borna i Oppenheimera, wprowadzając małą korektę (tzw. poprawkę diagonalną), otrzymuje się równanie Schrödingera dla ruchu jąder atomowych:

$$[\hat{T}_n + E_i^{el}(\mathbf{R})]\phi_i(\mathbf{R}) = E_i\phi_i(\mathbf{R}) \quad (3.5)$$

$\hat{T}_n$  oznacza operator energii kinetycznej jąder, natomiast wyrażeniu  $E_i^{el}$  przypisana jest rola operatora energii potencjalnej generowanego przez elektronową funkcję falową z rozwiązania równania 3.4. Oznacza to, że wielkość  $E_i^{el}$ , nazywana także hiperpowierzchnią energii potencjalnej, reprezentuje potencjały, w których poruszają się jądra atomowe. Okazuje się, że dla tak skonstruowanego zagadnienia możliwe jest rozdzielenie ruchu rotacyjnego oraz oscylacyjnego układu jąder tworzących molekułę, natomiast ruch środka masy układu (translacyjny) daje się odseparować w sposób ścisły bez konieczności kwantowania. Podejście to pozwala uzyskać prosty opis energii  $N$ -jądrowego systemu poprzez funkcję  $3N - 6$  współrzędnych. Następnie korzystając z twierdzenia Hellmanna-Feynmana mówiącego o tym, że

ujemny gradient energii jest wartością oczekiwaną gradientu hamiltonianu, można połączyć wielkość klasyczną (siłę zachowawczą) z wielkością kwantową (wartością oczekiwaną operatora):

$$\mathbf{F}_k = -\frac{\partial}{\partial \mathbf{R}_k} E_i^{el}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N), \quad k = 1, 2, 3, \dots, N. \quad (3.6)$$

Następnie traktując jądra jako obiekty klasyczne (masy punktowe), można wyznaczyć ewolucję czasową jąder całkując klasyczne równania ruchu. Jedynym problemem pozostaje ściśle wyznaczenie  $E_i^{el}(\mathbf{R})$  dla układów o bardzo dużej liczbie atomów, co jest współcześnie nieosiągalne obliczeniowo. Dla stosunkowo małych układów problem ten można obejść poprzez odpowiednią aproksymację  $E_i^{el}(\mathbf{R})$ , zazwyczaj stosując jedną z przybliżonych metod kwantowomechanicznych (rachunek zaburzeń, metody oparte o zasadę wariacyjną lub teoria funkcjonału gęstości). Jednakże dla rzeczywistych solwatowanych układów biomolekularnych potrzebny jest wydajniejszy sposób na obliczenie hiperpowierzchni energii potencjalnej układu.

### 3.1.2 Empiryczne pola siłowe

Obliczenia tzw. „klasycznej” dynamiki molekularnej realizowane są poprzez zastąpienie  $E_i^{el}(\mathbf{R})$  przez model oddziaływań w postaci prostego potencjału  $V(\mathbf{R})$  (tzw. pól siłowych) kosztem nieodwracalnej utraty informacji o strukturze elektronowej. Pola siłowe, pomimo swojej prostej formy w postaci addytywnych wkładów reprezentujących poszczególne oddziaływania, pozwalają na adekwatne przybliżenie hiperpowierzchni energii potencjalnej. Największą zaletą korzystania z pól siłowych są stosunkowo małe koszty obliczeniowe. Ich wadą jest natomiast konieczność odpowiedniego dopasowania zestawów licznych parametrów opisujących poszczególne typy oddziaływań w układzie. Funkcyjna postać pól siłowych składa się z sumy tzw. wkładów wiążących modelujących wiązania kowalencyjne i lokalną geometrię cząsteczek oraz tzw. wkładów niewiązących opisujących łącznie siły elektrostatyczne oraz van der Waalsa. Konkretna postać wyrażenia na  $V(\mathbf{R})$  zależy od użytego pola siłowego, jednakże zazwyczaj przyjmuje ono następującą postać:

$$V(\mathbf{R}) = V_{bonded} + V_{nonbonded} \quad (3.7)$$

gdzie:

$V(\mathbf{R})$  - energia potencjalna,

$V_{bonded}$  - energia potencjalna członów wiążących,

$V_{nonbonded}$  - energia potencjalna członów niewiązących

Człon wiążący jest w głównej mierze odpowiedzialny za utrzymanie odpowiedniej geometrii wiązań oraz lokalnej struktury cząsteczek i jest on konstruowany następującą sumą wkładów:

$$\begin{aligned}
 V_{bonded} &= V_{bonds} + V_{angles} + V_{dihedr} + V_{impdihedr} \\
 &= \sum \frac{1}{2} k_{ij}^b (r_{ij} - r_{ij}^o)^2 \\
 &+ \sum \frac{1}{2} k_{ijk}^a (\Theta_{ijk} - \Theta_{ijk}^o)^2 \\
 &+ \sum \frac{k_{ijkl}^d}{2} [1 + \cos(n\Phi_{ijkl} - \Phi_{ijkl}^o)] \\
 &+ \sum \frac{k_{ijkl}^{id}}{2} (\xi_{ijkl} - \xi_{ijkl}^o)^2
 \end{aligned} \tag{3.8}$$

gdzie:

$V_{bonds}$  - wkład odpowiadający zmianom długości wiązań,

$V_{angles}$  - wkład odpowiadający zmianom wartości kątów płaskich,

$V_{dihedr}$  - wkład odpowiadający zmianom wartości kątów dwuściennych,

$V_{impdihedr}$  - wkład odpowiadający zmianom wartości niewłaściwych kątów dwuściennych

Pierwsze dwa człony we wzorze 3.8 ( $V_{bonds}$  i  $V_{angles}$ ) opisują zmiany długości wiązań oraz odkształcenia kątów płaskich i są odpowiedzialne za zdefiniowanie lokalnego ułożenia sąsiednich atomów względem siebie. Mają one postać prostych potencjałów harmonicznym, dla których wartości długości i kątów oscylują wokół wartości referencyjnych  $r_{ij}^o$  oraz  $\Theta_{ijk}^o$ . Kolejnym wkładem do części wiążącej pola siłowego jest człon odpowiedzialny za opis oddziaływań odpowiadających obrotom wokół pojedynczych wiązań. Potencjał dwuścienny  $V_{dihedr}$  jest modelowany przez funkcję trygonometryczną cosinus z periodycznością  $n$ , barierami potencjału  $k_{ijkl}$  oraz członem fazowym  $\Phi_{ijkl}^o$ . Natomiast w przypadku niewłaściwych kątów dwuściennych  $V_{impdihedr}$  odpowiadających m.in. za ograniczenie ruchów zginających i ruchów „poza płaszczyznę”, odchylenie od referencyjnej wartości kąta dwuściennego jest opisywane przez potencjał harmoniczny.

Z kolei oddziaływania niewiążące w symulowanym układzie są modelowane poprzez dwa człony. Chodzi tu o człony związane z oddziaływaniami typu Coulombowskiego oraz van der Waalsa (krótkozasięgowe odpychanie oraz przyciągające oddziaływania dyspersyjne):



$$\begin{aligned}
V_{nonbonded} &= V_{el} + V_{vdW} \\
&= \sum_{i=1}^N \sum_{j>i}^N \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \\
&\quad + \sum_{i=1}^N \sum_{j>i}^N \epsilon_{ij} \left[ \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^0}{r_{ij}} \right)^6 \right]
\end{aligned} \tag{3.9}$$

gdzie:

$V_{el}$  - energia oddziaływań elektrostatycznych,

$V_{vdW}$  - energia oddziaływań van der Waalsa

$q_i, q_j$  - ładunki atomu i-ego oraz atomu j-ego,

$\epsilon_0$  - przenikalność elektryczna próżni,

$\epsilon_r$  - stała dielektryczna,

$r_{ij}$  - odległość między atomem i-tym oraz j-tym

$\epsilon_{ij}$  - człon odpowiadający za głębokość studni potencjału,

$r_{ij}^0$  - położenie minimum potencjału,

Oddziaływanie elektrostatyczne przypadające na każde centrum oddziaływań wyznaczone jest wprost z prawa Coulomba poprzez sumowanie po wszystkich parach centrów oddziaływań w układzie. W praktyce podczas obliczeń brane są pod uwagę tylko te atomy, które znajdują się w różnych molekułach lub są oddzielone co najmniej 3 wiązaniami. Wynika to z tego, że część efektów elektrostatycznych jest już uwzględnionych w członach wiążących. Modelowanie oddziaływań van der Waalsa (vdW) odbywa się przy pomocy potencjału Lennarda-Jonesa lub innego o podobnej charakterystyce, z dwoma parametrami określającymi głębokość minimum potencjału  $\epsilon_{ij}$  oraz odległość, w której minimum jest osiągnięte  $r_{ij}^0$ . Opis oddziaływań vdW bazuje na dwóch różniących się znakiem członach, odpowiedzialnych kolejno za obserwowane eksperymentalnie przyciągające efekty dyspersyjne (wynikające głównie z oddziaływań pomiędzy fluktuującymi dipolami) oraz silne odpychanie występujące dla małych odległości międzyatomowych (wynikające z zakazu Pauliego).

Obecnie rozwijanych jest wiele rodzajów pól siłowych, a ich specjalne wersje służą modelowaniu konkretnych układów. Mimo tego w większości przypadków pola siłowe wykazują cechę uniwersalności, tj. możliwość przenoszenia zestawu parametrów uzyskanych dla modelowych cząsteczek do modelowania innych większych cząsteczek pokrewnych. Parametryzacja pola siłowego może odbywać się na kilka różnych sposobów, jednakże najczęściej robi się to w oparciu o jak najlepsze odwzorowanie eksperymentalnych wartości wielkości mierzalnych. W przypadku lokalnych parametrów geometrycznych ( $r_{ij}^0, \Theta_{ijk}^0$  i  $\zeta_{ijkl}^0$ ) ich wartości zazwyczaj ustala się na podstawie struktur małych modelowych cząsteczek uzyskanych

metodami krystalograficznymi lub NMR. A odpowiadające im stałe siłowe uzyskuje się poprzez przeprowadzenie analizy drgań normalnych i porównanie do eksperymentalnych częstości drgań charakterystycznych np. z spektroskopii w zakresie podczerwieni. Natomiast wysokości barier potencjałów dla obrotów wokół wiązań uzyskuje się w oparciu o zaawansowane metody *ab initio*. Tzw. ładunki cząstkowe (albo atomowe) dobierane są numerycznie, w taki sposób by jak najlepiej odtworzyć rzeczywisty rozkład potencjału elektrostatycznego generowanego przez cząsteczkę (np. w procedurze Merza-Kollmanna) [150]. Optymalizacja parametrów potencjału LJ przeprowadzana jest zazwyczaj przy uwzględnieniu określonych parametrów makroskopowych i termodynamicznych (gęstość, współczynnik dyfuzji itp.). Mimo że pola siłowe są prostą sumą poszczególnych par oddziaływań, to efekty ich działania obejmują także wielociałowe oddziaływania międzycząsteczkowe, w tym odpowiednią polaryzację cząsteczek w fazie skondensowanej.

Obecnie przy modelowaniu biomolekuł najczęściej używa się pól siłowych z rodziny OPLS [151], GROMOS [152] oraz najbardziej popularnych CHARMM [153] i AMBER [154]. Zazwyczaj umożliwiają one użycie kilku różnych modeli wody, które traktują ją jako cząsteczkę sztywną lub elastyczną. Każdy z modeli wody jest zoptymalizowany pod kątem pewnej konkretnej cechy fizycznej wody (gęstość, współczynnik dyfuzji itp.). Jednakże ze względów ekonomicznych najczęściej korzysta się z prostego trójpunktowego modelu TIP3P [155].

## 3.2 Metody obliczeniowe

### 3.2.1 Minimalizacja energii układu

W poprzednim podrozdziale przedstawiłem uzasadnienie tego, że odpowiednio dobrane pole siłowe z dobrym przybliżeniem oddaje właściwości hiperpowierzchni energii potencjalnej dla ruchu jąder. Wynika z tego, że dla racjonalnie zaprojektowanego układu symulacyjnego istnieją niskoenergetyczne konfiguracje układu. Ogólnie procedura wyznaczania lokalnych jak i globalnych niskoenergetycznych konfiguracji układu (punktów stacjonarnych funkcji wielu zmiennych) polega na analitycznym rozwiązaniu skomplikowanego układu  $3N$  równań nieliniowych:

$$\nabla_{\mathbf{R}}V(\mathbf{R}) = \mathbf{0} \quad (3.10)$$

Ze względu na złożoność funkcji energii pola siłowego  $V(\mathbf{R})$  rozwiązanie tego układu równań dla rzeczywistych systemów, gdzie  $N \gg 10$  nie jest możliwe analityczne, ale również często przy pomocy metod numerycznych. Jednakże przy założeniu, że dysponujemy

racjonalnym modelem początkowym układu, satysfakcjonującym rozwiązaniem powyższego problemu jest wyznaczenie jedynie lokalnych minimów funkcji, korzystając z jednej z szeregu metod iteracyjnych. Do standardowych metod zalicza się tzw. metodę sprzężonych gradientów (z ang. Conjugate Gradient). Jest ona rozwinięciem metody najmniejszego spadku (z ang. Steepest Descent) i w związku z tym także wymaga znajomości gradientu funkcji. Jednakże w jej przypadku uzyskuje się zdecydowanie szybszą zbieżność w okolicach minimum, ponieważ nowe kierunki poszukiwań wyznaczone są na podstawie dwóch poprzednio wyznaczonych gradientów. Jest to wystarczające postępowanie, ponieważ unikamy niestabilności numerycznych wynikających ze znacznych sił występujących w wysokoenergetycznych konfiguracjach układu. Tak uzyskane niskoenergetyczne konfiguracje układu są punktem początkowym do właściwych symulacji. Nie ma potrzeby lokalizowania globalnego minimum energii, ponieważ najbardziej prawdopodobne konformacje układu będą i tak generowane zgodnie z założoną temperaturą (rozkładem Boltzmanna), zazwyczaj fizjologiczną.

### 3.2.2 Dynamika molekularna

Przeprowadzone wyżej rozważania doprowadziły do wniosku, że w standardowych warunkach fizycznych atom z dobrym przybliżeniem zachowuje się jak cząstka klasyczna. W związku z tym do opisu jego trajektorii wystarczy zastosować dynamikę Newtona postaci:

$$\mathbf{F}_i(t) = m_i \frac{d^2 \mathbf{R}_i(t)}{dt^2} \quad i = 1, 2, 3, \dots, N \quad (3.11)$$

gdzie:

$\mathbf{F}_i(t)$  – wektor siły działającej na  $i$ -ty atom w chwili  $t$ ,

$m_i$  – masa  $i$ -ego atomu,

$\mathbf{R}_i(t)$  – wektor położenia  $i$ -ego atomu w chwili  $t$

Pewną kwestią pozostaje wyznaczenie sił działających na dany atom układu, ale generuje się je w prosty sposób z potencjalnego pola skalarnego na podstawie wybranego pola siłowego:

$$\mathbf{F}_i = -\nabla_{\mathbf{R}_i} V(\mathbf{R}) \quad i = 1, 2, 3, \dots, N \quad (3.12)$$

gdzie:

$\nabla_{\mathbf{R}_i}$  – gradient, wektor pochodnych cząstkowych  $(\frac{\partial}{\partial R_{i1}}, \frac{\partial}{\partial R_{i2}}, \dots, \frac{\partial}{\partial R_{in}})$ ,

$V(\mathbf{R})$  – skalarne pole potencjalne, zależne od położenia wszystkich atomów  $(\{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N\})$

Poprzez proste przyrównanie do siebie równania 3.11 oraz 3.12 otrzymuje się już praktyczne wyrażenie opisujące trajektorię układu w przestrzeni konfiguracyjnej:

$$\frac{d^2 \mathbf{R}_i(t)}{dt^2} = -\frac{1}{m_i} \nabla_{\mathbf{R}_i} V(\mathbf{R}) \quad i = 1, 2, 3, \dots, N \quad (3.13)$$

Ewolucję czasową układu można uzyskać przez scałkowanie tego układu  $3N$  równań różniczkowych drugiego rzędu. Wymagane jest do tego określenie początkowych położeń oraz prędkości każdego atomu w chwili ( $t = 0$ ), co jednocześnie determinuje początkową sumę wartości energii potencjalnej oraz energii kinetycznej układu. W praktyce układ równań 3.13 rozwiązuje się za pomocą metod numerycznych, zakładając dyskretyzację operatorów różniczkowania oraz przyjmując, że siły działające na atomy nie zmieniają się dla małego, ale skończonego kroku czasowego  $\delta t$ . Ogólny schemat algorytmu zakłada, że po czasie  $t + \delta t$  zgodnie z działającymi siłami układ przechodzi do nowej konfiguracji, dla której to następuje ponowne przeliczenie sił działających na atomy. Następnie dla tak nowo wyznaczonych sił następuje ponowna ewolucja układu o krok czasowy  $\delta t$ . Kolejne iteracje są kontynuowane do momentu uzyskania zadowalającej długości trajektorii.

### 3.2.3 Algorytmy całkowania równań ruchu

Numeryczne scałkowanie różniczkowego układu równań 3.13 oprócz dyskretyzacji operatorów różniczkowania wymaga również wprowadzenia pewnych rekurencyjnych zależności dla położeń oraz prędkości. W tym celu najczęściej stosuje się jedną z wersji algorytmu Verleta. Podstawą ich działania jest rozwinięcie funkcji położenia od czasu wokół punktu  $t$  w szereg Taylora, zaniedbanie wybranych wyrazów wyższych rzędów oraz uwzględnienie tego, że  $\frac{d^2 \mathbf{R}_i(t)}{dt^2} = \frac{\mathbf{F}_i(t)}{m_i}$ . Najprostsza wersja algorytmu Verleta daje możliwość wyznaczenia położenia  $i$ -tego atomu w chwili czasu  $t + \delta t$  oraz prędkości w chwili  $t$  poprzez następujące wyrażenia:

$$\mathbf{R}_i(t + \delta t) = 2\mathbf{R}_i(t) - \mathbf{R}_i(t - \delta t) + \frac{\mathbf{F}_i(t)}{m_i} \delta t^2 + O(\delta t^4) \quad (3.14)$$

$$\mathbf{v}_i(t) = \frac{\mathbf{R}_i(t + \delta t) - \mathbf{R}_i(t - \delta t)}{2\delta t} + O(\delta t^2) \quad (3.15)$$

Algorytm Verleta w postaci równań 3.14 i 3.15 przy wyznaczaniu nowych położeń atomów wymaga znajomości położeń wszystkich atomów w dwóch poprzednich krokach czasowych, jednakże nie utrudnia to implementacji technicznej. Z kolei informacja o nowo wyznaczonych prędkościach jest zawsze opóźniona o jeden interwał czasowy  $\delta t$  względem nowych położeń. Ma to istotne znaczenie w przypadku konieczności posiadania informacji o aktualnych wartościach energii kinetycznej lub temperatury. Inną niepożądaną cechą tego algorytmu jest obliczanie prędkości poprzez dzielenie dwóch bliskich sobie liczb. Skutkuje to tym, że wyliczane wartości silnie zależą od precyzji w reprezentacji liczb. Brak ścisłych wyników w obliczeniach w skrajnych przypadkach może prowadzić do niepożądanego dryfu energii układu. Sposobem na obejście trudności związanych z standardowym algorytmem

Verleta jest użycie modyfikacji algorytmu Verleta zwanej metodą „żabiego skoku” (z ang. leap-frog):

$$\mathbf{v}_i(t + \frac{\delta t}{2}) = \mathbf{v}_i(t - \frac{\delta t}{2}) + \frac{\mathbf{F}_i(t)}{m_i} \delta t + O(\delta t^3) \quad (3.16)$$

$$\mathbf{R}_i(t + \delta t) = \mathbf{R}_i(t) + \mathbf{v}_i(t + \frac{\delta t}{2}) \delta t + O(\delta t^3) \quad (3.17)$$

Jest to najpopularniejszy algorytm wykorzystywany do numerycznego całkowania równań ruchu, nie tylko w dynamice molekularnej. Równania 3.16 i 3.17 opisujące ewolucję położenia i prędkości są tym razem przesunięte względem siebie o pół kroku całkowania, jednakże łatwo wyeliminować tę niedogodność poprzez wyznaczenie nowej wartości  $v_i(t)$  na podstawie średniej arytmetycznej prędkości uzyskanych dla czasów  $t + \frac{\delta t}{2}$  i  $t - \frac{\delta t}{2}$ :

$$\mathbf{v}_i(t + \delta t) = \frac{\mathbf{v}_i(t - \frac{\delta t}{2}) + \mathbf{v}_i(t + \frac{\delta t}{2})}{2} \quad (3.18)$$

Wykazano również, że algorytmy Verleta nie zachowują całkowitej energii układu, lecz pseudo-hamiltonian układu, dzięki czemu całkowita energia oscyluje wokół wartości początkowej. Jest to głównie zasługą odwracalności algorytmu w czasie. Istnieją również bardziej złożone metody całkowania równań ruchu (np. algorytm Rungego-Kutty), jednak ich złożoność obliczeniowa z reguły dyskwalifikuje je z użycia w symulacjach dynamiki molekularnej. Co więcej wykazują one tendencję do eksponencjalnego rozbiegania się względem trajektorii rzeczywistej (tzw. brak symplektyczności algorytmu). Ogólnie algorytmy Verleta sprawdzają się w symulacjach dynamiki molekularnej, ponieważ posiadają wystarczającą stabilność numeryczną uzyskaną przy niskich kosztach obliczeniowych.

### 3.2.4 Krok czasowy

Dla powodzenia symulacji dynamiki molekularnej kluczowe znaczenie ma dobranie odpowiedniego kroku czasowego. Po pierwsze krok czasowy powinien być na tyle duży, aby uzyskana trajektoria objęła jak największy obszar przestrzeni fazowej. Może to jednak prowadzić do niestabilnościami algorytmów całkowania, gwałtownego wzrostu amplitudy drgań wiązań oraz braku zachowania energii układu, pędu, itd. Natomiast użycie zbyt krótkiego kroku czasowego może spowodować, że procesy badane w symulacji nie osiągną reprezentatywnego rozkładu konfiguracji. Jak wiadomo, dla standardowych układów molekularnych najszybszy ruch jest związany z oscylacjami długości wiązań kowalencyjnych z udziałem atomów wodoru. Częstotliwość tych drgań wymaga zastosowania kroku czasowego wynoszącego  $1 \text{ fs} = 10^{-15} \text{ s}$ , czyli około 10 razy więcej niż wynosi okres oscylacji tego typu wiązań (np. wiązanie C-H). Zwiększenie kroku czasowego do 2-4 fs możliwe jest poprzez zastosowanie więzów na długości wiązań X-H oraz na inne szybko oscylujące stopnie swobody. Co

więcej jest to czasami uzasadnione podejście, ponieważ w temperaturze pokojowej oscylacje lekkich atomów wodoru nie są adekwatnie modelowane jako klasyczne oscylacje harmoniczne, co może prowadzić do przeszacowania pojemności cieplnej układu.

### 3.2.5 Długości wiązań i więzy

W temperaturach typowanych dla warunków fizjologicznych częstości oscylacji długości wiązań i kątów płaskich są wysokie a ich amplitudy niewielkie. Poza tym ich wpływ na ogólną konformację oraz dynamikę makrocząsteczek, takich jak białko i DNA, jest niewielki. Pozwala to na nałożenie więzów na niektóre stopnie swobody co jest tożsamy z zamrożeniem ich w podstawowym stanie oscylacyjnym, niejako wyłączając je z tej części obliczeń. Jednym z algorytmów pozwalającym na ustalenie konkretnych długości wiązań jest algorytm SHAKE [156]. Jego postawą jest użycie techniki nieoznaczonych mnożników Lagrange'a. W efekcie równania ruchu ulegają nieznaczącej modyfikacji, a wyznaczenie nowych zmodyfikowanych położenia atomów nie jest trudne. W metodzie SHAKE procedura wyznaczania nowych położenia atomów w każdym kroku symulacji sprowadza się wprawdzie do wyznaczenia położenia atomów zgodnie ze standardowymi równaniami ruchu (np. 3.16 i 3.17), a następnie ich modyfikacji poprzez dodanie członów związanych z siłami wprowadzonych więzów. W przypadku prostego modelu cząsteczki wody istnieje analityczna wersja tego algorytmu pod nazwą SETTLE [157]. Istnieje również trochę szybszy, ale posiadający za to inne ograniczenia algorytm LINCS [158]. Oparty jest o rozważania geometryczne i w przypadku dynamiki molekularnej nie wymaga procedury iteracyjnej.

### 3.2.6 Temperatura i ciśnienie

Dla potrzeb dynamiki molekularnej temperaturę definiuje się jako miarę średniej energii kinetycznej układu. Jak się okazuje większe znaczenie, dla prawidłowości prowadzonych symulacji dynamiki molekularnej, ma utrzymywanie stałej temperatury układu niż niwelowanie błędów wynikających ze stosowanych przybliżeń obliczeniowych. Symulacje dynamiki molekularnej przeprowadza się zwykle po to, by wyznaczyć określone obserwowalne właściwości układu, jako średnie po wygenerowanym zespole statystycznym. Zaburzenia trajektorii, które powstają przez losowe błędy zaokrągleń numerycznych nie wpływają zwykle na ogólny obraz statystyczny systemu. Wynika z tego, że kluczowe jest zapewnienie stałości parametrów makroskopowych takich jak temperatura i ciśnienie systemu. Najprostszym sposobem na utrzymanie stałej temperatury jest cykliczne przeskalowywanie prędkości wszystkich atomów tak, aby rozkład ich wartości odpowiadał zakładanej temperaturze  $T$ . Jednakże w rzeczywistości stosuje się mniej radykalną metodę. Metoda słabego sprzężenia (z ang. weak-coupling method) zakłada „rozłożone w czasie” skalowanie wszystkich prędkości przy pomocy współczynnika:

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau_T} \left( \frac{T_0}{T(t)} - 1 \right)} \quad (3.19)$$

gdzie:

$\tau_T$  - stała czasowa określająca szybkość relaksacji do zadanej temperatury,

$\delta t$  - krok czasowy,

$T_0$  - temperatura do której dąży układ

Warto zaznaczyć, że narzucenie warunku na wartość temperatury na początku symulacji determinuje rozkład prawdopodobieństwa, z którego mogą zostać wyznaczone prędkości początkowe wszystkich atomów układu. Są one losowane z rozkładu Maxwella-Boltzmanna odpowiadającemu zadanej temperaturze (energii kinetycznej układu) przy warunku zerowania się ich sumy wektorowej.

Podobne podejście można zastosować do utrzymywania stałej wartości ciśnienia w układzie. W tym przypadku przeskalowywaniu podlegają jednak położenia atomów w układzie zamiast prędkości. Prowadzi to do zmian rozmiarów symulowanego systemu.

Istnieją również inne metody, które pozwalają na utrzymywanie stałej temperatury i ciśnienia w systemie. Metody rozszerzonego langrażjanu (z ang. extended-Lagrangian methods) polegają na dodaniu do układu dodatkowych stopni swobody, które symulują łącznie cieplną lub ciśnieniową. Jedną z tych metod jest łańcuchowy wariant termostatu Nosé-Hoovera. Jego użycie powoduje utrzymywanie temperatury wokół zadanej średniej. Odbywa się to dzięki temu, że nowe współrzędne pędowe występujące w zmodyfikowanych równaniach ruchu poddawane są sprzężeniu zwrotnemu. Każde odchylenie temperatury od średniej wartości jest niwelowane z zadaną stałą czasową. Wartość i kierunek tego sprzężenia uzależniony jest od różnic pomiędzy aktualną a zakładaną temperaturą.

### 3.2.7 Periodyczne warunki brzegowe

Ograniczenia sprzętowe zmuszają nas zwykle do rezygnacji z symulowania systemów o dużych objętościach. Powoduje to konieczność użycia pudeł symulacyjnych o stosunkowo niewielkich rozmiarach. Skończony wymiar pudła symulacyjnego wprowadza nienaturalne zachowania atomów na jego brzegach (tzw. efekty brzegowe). Łatwo pokazać, że dla większości symulowanych systemów znaczna część atomów znajduje się na brzegu pudła symulacyjnego. Graniczenie niektórych atomów z próżnią prowadzi do otrzymywania niefizycznego obrazu oddziaływań w systemie. Eliminacja tych przypadłości odbywa się najczęściej poprzez użycie periodycznych warunków brzegowych (z ang. Periodic Boundary Conditions). Ujmując najprościej idee PBC, symulacja przebiega w nieskończonej sieci okresowej, w której komórka elementarna reprezentowana jest przez pojedyncze pudło symulacyjne. Narzucenie periodycznych warunków na układ powoduje, że  $i$ -ta cząsteczka podlega oddziaływaniom

pochodzącym od wszystkich cząsteczek z pudła symulacyjnego oraz dodatkowo cząsteczek z pozostałych komórek sieci. Podejście takie wprowadza zatem ciągłość symulowanego ośrodka. Jednakże użycie zbyt małych wektorów określających wymiary systemu może prowadzić do pewnych artefaktów wynikających z tego, że symulowana molekula oddziałuje ze swoim własnym obrazem. Podczas symulacji zapamiętywane i używane potem do wszelkich analiz są jedynie współrzędne atomów z pojedynczego pudła symulacyjnego.

### 3.2.8 Oddziaływania niewiążące

Bardzo niesprzyjającym czynnikiem, mającym wpływ na szybkość obliczeń w dynamice molekularnej, jest konieczność ciągłego wyznaczania odległości międzyatomowych występujących w wyrażeniu na elektrostatyczną energię potencjalną. Procedura ta sprowadza się do kosztownego obliczenia pierwiastków kwadratowych z liczb rzeczywistych. Dodatkowo, konieczność liczenia podwójnych sum prowadzi do złożoności obliczeniowej rzędu  $O(N^2)$  ( $N$  – liczba atomów). Użycie periodycznych warunków brzegowych oraz konwencji najbliższego sąsiada nie redukuje niestety złożoności zagadnienia. Skrócenie czasu obliczeń możliwe jest dzięki zastosowaniu tzw. promienia odcięcia  $r_c$  (z ang. cut-off distance). Promień  $r_c$  jest to odległość, powyżej której zaniedbuje się wszelkie wkłady energetyczne pochodzące od centrów oddziaływań. Idea ta ma pewną zaletę polegającą na możliwości stworzenia listy sąsiadów, czyli listy atomów znajdujących się w promieniu  $r_c + \Delta r$  od centrum oddziaływania elektrostatycznego. Sprowadza to obliczenia do zagadnienia liniowego, którego złożoność wynosi już tylko  $O(N)$ . Niechcianą konsekwencją zastosowania promienia odcięcia są sztuczne nieciągłości energii oraz sił. Prowadzić mogą one do niestabilności oraz ciągłego dryfu energii całkowitej układu. Takim niepożądanym efektem przeciwdziała się poprzez użycie metody funkcji wyłączającej (z ang. switching function), polegającej na przemnożeniu w przedziale  $r_{switch} - r_c$  potencjału przez wielomian wygładzający tak, aby w punkcie  $r_c$  potencjał wynosił 0. Niestety wprowadzanie ograniczeń na zasięg oddziaływania elektrostatycznego często prowadzi do niefizycznych zachowań układu.

Wyjściem z tej niekorzystnej sytuacji jest użycie metody sum Ewalda. Podejście to, polega na zastąpieniu warunkowo i wolno zbiegającego szeregu Coulomba przez dwa szeregi zbiegające bezwzględnie oraz szybciej. Dodatkowo rezygnuje się z idei najbliższego sąsiada, co ostatecznie prowadzi do uzyskania złożoności obliczeniowej rzędu  $O(N^{\frac{3}{2}})$ . Modyfikacja tej metody jest algorytm sumowania Ewalda na siatce (z ang. Particle Mesh Ewald). Bazuje on na usprawnieniu obliczeń jednego z szeregów występującego w metodzie sum Ewalda. Ostatecznym wynikiem wprowadzenia algorytmu PME jest zredukowanie złożoności obliczeniowej do  $O(N \log N)$  a zatem znaczące skrócenie czasu obliczeń potrzebnego do pełnego opisu oddziaływań elektrostatycznych w układzie.



### 3.3 Elementy mechaniki statystycznej

#### 3.3.1 Analiza danych

W mechanice statystycznej wartości średnie zdefiniowane są jako:

$$\langle A \rangle_{zesp} = \int \int A(\mathbf{R}, \mathbf{p}) \rho(\mathbf{R}, \mathbf{p}) d\mathbf{R} d\mathbf{p} \quad (3.20)$$

gdzie  $A(\mathbf{R}, \mathbf{p})$  określa wartość obserwabli zależnej od mikrostanu układu, a  $\rho(\mathbf{R}, \mathbf{p})$  jest unormowaną gęstością prawdopodobieństwa wszystkich mikrostanów w przestrzeni fazowej, charakterystyczną dla danego zespołu. Wyznaczenie całek w równaniu 3.20 zazwyczaj jest bardzo trudne, a spowodowane jest to koniecznością znajomości  $\rho(\mathbf{R}, \mathbf{p})$  dla wszystkich możliwych mikrostanów zespołu. Można pokazać, że punkty trajektorii z symulacji dynamiki molekularnej generowane są z prawdopodobieństwem proporcjonalnym do rzeczywistego rozkładu równowagowego  $\rho(\mathbf{R}, \mathbf{p})$ . W takiej sytuacji bardziej przystępnym podejściem prowadzącym do oszacowania średniej  $\langle A \rangle_{zesp}$  jest policzenie średniej po czasie:

$$\langle A \rangle_{czas} = \frac{1}{\tau} \int_0^{\tau} A[\mathbf{R}(t), \mathbf{p}(t)] dt \quad (3.21)$$

pod warunkiem, że czas symulacji  $\tau$  jest wystarczająco długi, aby zadowalająco dobrze wypróbować dystrybucję mikrostanów układu. Hipoteza ergodyczna, jedna z podstaw mechaniki statystycznej, stwierdza:

$$\langle A \rangle_{zesp} = \lim_{\tau \rightarrow \infty} \langle A \rangle_{czas} \quad (3.22)$$

co usprawiedliwia zastosowane wyżej podejście. Oznacza to, że przeprowadzenie nieskończenie długiej symulacji powinno zapewnić przejście trajektorii przez wszystkie możliwe mikrostany. Idąc tą drogą, obserwabli pochodzące z eksperymentu, które z założenia są średnimi po zespole (równanie 3.21), mogą zostać porównane do średnich czasowych pochodzących z odpowiednio długich symulacji. Wyrażenie numeryczne, które jest używane do obliczenia  $\langle A \rangle_{czas}$  w symulacjach dynamiki molekularnej dane jest następująco:

$$\langle A \rangle_{czas} \approx \frac{1}{M} \sum_{i=0}^M A[\mathbf{R}(i\delta t), \mathbf{p}(i\delta t)] \quad (3.23)$$

gdzie  $M$  jest liczbą kroków czasowych.

#### 3.3.2 Energia swobodna

Fundamentalną koncepcją przydatną przy analizie procesów chemicznych i biologicznych w przypadku faz skondensowanych jest energia swobodna Gibbsa  $G$  (czasami nazywana też entalpią swobodną). Zmiany energii swobodnej Gibbsa decydują o spontaniczności procesów. Proces przebiega spontanicznie od stanu A do stanu B, tylko wówczas, gdy różnica

$\Delta G(A \rightarrow B)$  pomiędzy tymi stanami jest ujemna. Przewidywanie kierunków spontanicznych procesów na podstawie symulacji wymaga zatem wyznaczenia (oszacowania) różnic energii swobodnej w układzie. W niektórych przypadkach, oprócz różnic energii swobodnej pomiędzy (skrajnymi) stanami, interesuje nas pełny profil energii swobodnej wzdłuż zdefiniowanego stopnia swobody (tzw. współrzędna reakcji). Znajomość pełnego profilu energii swobodnej  $G(\xi)$  zapewnia zarówno kompleksowy opis termodynamiki procesu jak i możliwość badania jego kinetyki (np. poprzez tzw. dynamikę brownowską). Równowaga termodynamiczna pomiędzy rozważanymi stanami jest scharakteryzowana albo przez różnice energii swobodnej pomiędzy nimi, albo przez stosunek stałych szybkości przejścia pomiędzy nimi. Profil energii swobodnej jest zdefiniowany przez:

$$G(\xi) = G(\xi^*) - k_B T \ln \left[ \frac{\langle \rho(\xi) \rangle}{\langle \rho(\xi^*) \rangle} \right] \quad (3.24)$$

gdzie  $k_B$  jest stałą Boltzmanna,  $\xi^*$  i  $G(\xi^*)$  są dowolnymi stałymi, natomiast  $\langle \rho(\xi) \rangle$  jest funkcją daną wzorem:

$$\langle \rho(\xi) \rangle = \frac{\int \delta(\xi'(\mathbf{R}) - \xi) e^{-H(\mathbf{R}, \mathbf{p})/k_B T} d\mathbf{R}, d\mathbf{p}}{\int e^{-H(\mathbf{R}, \mathbf{p})/k_B T} d\mathbf{R}, d\mathbf{p}} \quad (3.25)$$

w której  $\xi'(\mathbf{R})$  jest funkcją definiującą współrzędną reakcji, która zależy od jednej lub większej liczby stopni swobody.  $\xi'(\mathbf{R})$  może być odległością, kątem lub bardziej skomplikowaną funkcją współrzędnych kartezjańskich.

Okazuje się, że profil energii swobodnej  $G(\xi)$  jest równoważny innej ważnej koncepcji, a mianowicie potencjałowi średniej siły (z ang. Potential of Mean Force). Jest to potencjał będący rezultatem całkowania średniej siły działającej na cząstkę (cząstki) wzdłuż współrzędnej reakcji  $\langle F(\xi) \rangle_\xi$ . Łatwo pokazać, że  $\partial G(\xi)/\partial \xi = -\langle F(\xi) \rangle_\xi$  [159]. W praktyce możliwe jest stosowanie tych dwóch wyrażeń zamiennie.

Wiele symulacji dynamiki molekularnej układów chemicznych lub biologicznych jest poświęconych wyznaczeniu profilu energii swobodnej wzdłuż wybranego stopnia swobody. Niestety, niewykonalne jest uzyskanie wiarygodnego oszacowania PMF w ramach skończonego czasu konwencjonalnej symulacji. Dzieje się tak, ponieważ zazwyczaj bardzo trudno jest pokonać bariery energii swobodnej, które rozdzielają stany termodynamiczne. W związku z tym pojawiły się pewne metody, które omijają ten problem. W tej pracy użyłem głównie metod znanych jako *Umbrella Sampling* (US, tzw. próbkowanie parasolkowe) oraz *Adaptive Biasing Force* (ABF, tzw. adaptacyjna siła wymuszająca).

### 3.3.3 Umbrella Sampling

*Umbrella Sampling* jest jednym z najbardziej popularnych narzędzi, służących do określania funkcji gęstości prawdopodobieństwa  $\rho(\xi)$  i związanego z nią profilu energii swobodnej  $G(\xi)$ . W metodzie tej symulacja prowadzona jest w obecności sztucznie dodanego potencjału

$U_{bias}(\xi)$  [160], a więc trajektorie generowane są przez potencjał  $V(\mathbf{R}) + U_{bias}(\xi)$ . Celem przyłożenia do układu dodatkowego potencjału jest zwiększenie próbkowania w istotnym zakresie współrzędnej  $\xi$ . W praktyce, aby uzyskać zadowalające próbkowanie w pełnym zakresie zmienności  $\xi$ , należy zastosować serię symulacji z różnymi potencjałami  $U_{bias}^i(\xi)$ , mającymi za zadanie utrzymywanie układu w różnych obszarach  $\xi$ . Liczba tych tzw. okien musi być na tyle duża, aby zapewnić dobre próbkowanie całego zakresu współrzędnej reakcji. Najprostszym, a jednocześnie najczęściej wybieranym potencjałem jest potencjał harmoniczny  $U_{bias}^i(\xi) = k_i(\xi - \xi_i)^2/2$ . Stałe siłowe  $k_i$  dla każdego potencjału nadawane są indywidualnie, natomiast położenia ich minimów  $\xi_i$  są tak dobierane, aby uzyskać jak najbardziej równomierne próbkowanie badanego zakresu współrzędnej reakcji. W celu wyznaczenia właściwego profilu energii swobodnej, pozbawionego wpływu dodanych potencjałów rezultaty każdego okna symulacyjnego są korygowane następująco:

$$G_i(\xi) = -k_B T \ln \rho_i^{bias}(\xi) - U_{bias}^i(\xi) + F_i \quad (3.26)$$

gdzie  $\rho_i^{bias}(\xi)$  jest zaburzona gęstością prawdopodobieństwa uzyskaną w symulacji poprzez dodanie potencjału  $U_{bias}^i(\xi)$ , a stała  $F_i$  jest zdefiniowana przez:

$$e^{-F_i/k_B T} = \langle e^{-U_{bias}^i(\xi)/k_B T} \rangle \quad (3.27)$$

co odpowiada dodatkowej energii swobodnej związanej z wprowadzonym potencjałem zaburzającym.

Istnieje wiele metod służących do wyznaczenia właściwej gęstości prawdopodobieństwa  $\rho(\xi)$  na podstawie danych pochodzących z symulacji typu *Umbrella Sampling*. Obecnie najczęściej stosowaną metodą jest tzw. metoda ważonych histogramów (z ang. Weighted Histogram Analysis Method) [161]. Podstawą idei WHAM jest wyznaczenie niezaburzonej funkcji gęstości  $\rho(\xi)$  na podstawie ważonej sumy niezaburzonych funkcji gęstości pochodzących z poszczególnych okien. Wagi dobrane są tak, aby minimalizować błąd statystyczny. Na podstawie histogramów  $\rho_i^{bias}(\xi)$  zastosowanie procedury WHAM daje taką  $\rho(\xi)$ , której odpowiada najmniejsza niepewności statystyczna:

$$\rho(\xi) = \frac{\sum_{i=1}^{N_u} n_i \rho_i^{bias}(\xi)}{\sum_{i=1}^{N_u} n_i e^{-[U_{bias}^i(\xi) - F_i]/k_B T}} \quad (3.28)$$

gdzie  $e^{-F_i/k_B T} = \int e^{-U_{bias}^i(\xi)/k_B T} \rho(\xi) d\xi$ , a sumowanie przebiega po wszystkich oknach symulacyjnych ( $N_u$ ). Równanie 3.28 rozwiązywane jest poprzez procedurę iteracyjną przy początkowym założeniu wartości  $F_i$ . Chociaż *Umbrella Sampling* jest metodą szeroko stosowaną to posiada pewne niedogodności. Po pierwsze, nawet użycie  $U_{bias}^i(\xi) = -G(\xi)$  nie gwarantuje jednolitego próbkowania, wyniki mogą być nadal ubogie statystycznie w takich regionach

przestrzeni konfiguracyjnej, które szczególnie wpływają na zmianę energii swobodnej. Ponadto, dobranie odpowiedniego  $U_{bias}^i(\xi)$  jest trudne, szczególnie dla nowych problemów. W takich wypadkach stosuje się zwykle metodę prób i błędów, obserwując czy na początku symulacji uzyskuje się założone próbkowanie, ewentualnie przerywając symulację i zmieniając parametry potencjału  $U_{bias}^i(\xi)$ .

### 3.3.4 Adaptive Biasing Force

W przeciwieństwie do metody US, *Adaptive Biasing Force* w celu wyznaczenia PMF nie potrzebuje informacji o gęstości prawdopodobieństwa  $\rho(\xi)$ . Zamiast tego metoda ABF wyznacza średnią siłę działającą wzdłuż współrzędnej reakcji, która po scałkowaniu wzdłuż  $\xi$  daje PMF. W trakcie trwania symulacji zliczane są chwilowe wartości siły działające w małych przedziałach  $d\xi$ . Średnie z każdego przedziału mogą zostać użyte do policzenia pochodnej  $d\tilde{G}(\xi)/d\xi$ , gdzie  $\tilde{G}(\xi)$  jest bieżącym oszacowaniem funkcji energii swobodnej. W celu uzyskania wiarygodnej średniej  $\langle F(\xi) \rangle_{\xi}$  dla całego zakresu  $\xi$ , cała ścieżka reakcji powinna zostać dobrze i relatywnie równomiernie wypróbkowana. W podejściu ABF, uzyskanie równomiernego próbkowania realizowane jest przez zastosowanie dodatkowej siły zaburzającej. Pomaga ona w przezwyciężeniu barier energii swobodnej, np. poprzez wypychanie systemu z minimum potencjału oraz przepychanie przez bariery. Siła ta ( $F^{ABF}$ ) jest zdefiniowana jako pochodna po stale aktualizowanym PMF:

$$F^{ABF} = -\langle F(\xi) \rangle_{\xi} \nabla_x \xi = \frac{d\tilde{G}(\xi)}{d\xi} \nabla_x \xi \quad (3.29)$$

gdzie  $\langle F(\xi) \rangle_{\xi}$  oznacza bieżącą średnią siły  $F(\xi)$  działającej wzdłuż współrzędnej reakcji  $\xi$ . Dodawanie  $F^{ABF}$  podczas trwania symulacji prowadzi do wygładzania powierzchni energii potencjalnej, a w konsekwencji do bardziej wydajnego próbkowania.

Dla dalszego zwiększenia efektywności obliczeń, współrzędna reakcji może zostać podzielona na okna (mniejsze zakresy współrzędnej reakcji  $[\xi_i, \xi_{i+1}]$ ) dla których generowana jest oddzielna trajektoria oraz oddzielny zestaw  $\langle F(\xi) \rangle_{\xi}$ . W przeciwieństwie do poprzedniej metody nie ma konieczności łączenia wyników z poszczególnych okien symulacyjnych.

## 4 Procedury symulacyjne

Poniższy rozdział zawiera szczegółowe informacje dotyczące użytych przeze mnie metod budowy układów oraz protokołów symulacyjnych, potrzebnych do odtworzenia uzyskanych przeze mnie wyników, dotyczących każdego z prowadzonych projektów.

### 4.1 Stabilność G-tetrazy w uformowanym G-kwadrupleksie

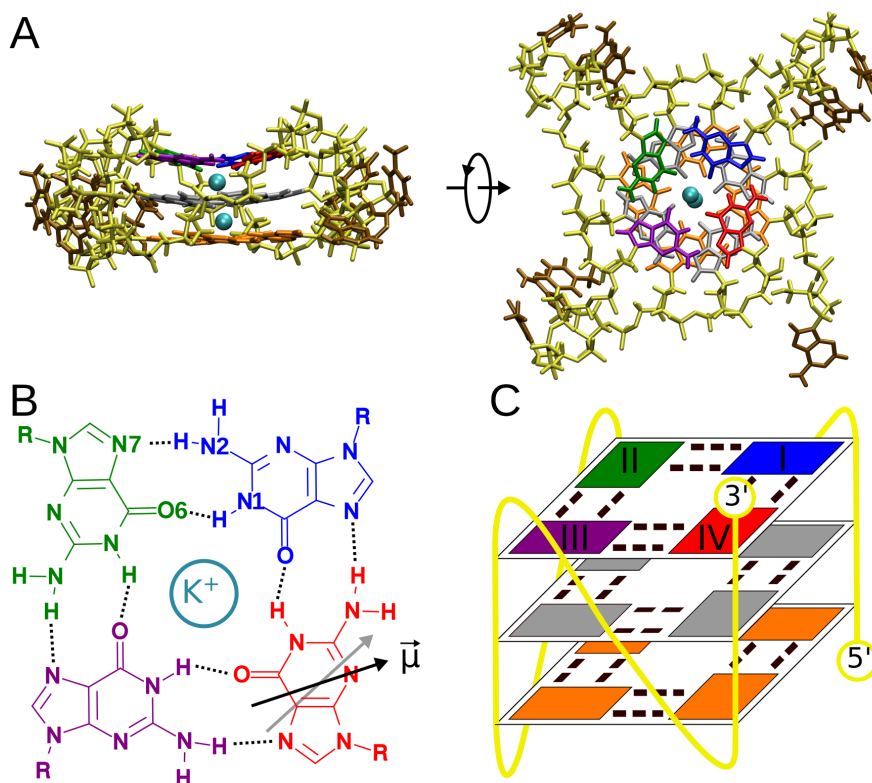
#### 4.1.1 Konstrukcja modelu i szczegóły symulacji

Początkowe współrzędne atomów G-kwadrupleksu o konformacji równoległej wraz z 3 jonami potasu w centralnym kanale zostały wygenerowane na podstawie struktury krystalograficznej o rozdzielczości 2,1 Å (kod PDB 1KF1) (patrz Rys. 4.1A,C) [61]. Układ składający się z pojedynczej struktury jednoniciowego DNA o ludzkiej sekwencji telomerowej d[AGGG(TTAGGG)<sub>3</sub>] w dwunastościennym pudle symulacyjnym o rozmiarach 8×8×5,6 nm uzupełniłem przez 11469 cząsteczki wody typu TIP3 [162] przy zastosowaniu fizjologicznego stężenia jonów (150 mM KCl). Do symulacji DNA oraz jonów użyłem pola siłowego CHARMM27 [163].

Symulacje były wykonane przy pomocy pakietu Gromacs 4.5.5 [164] w zespole statystycznym NPT o temperaturze 300 K i ciśnieniu 1 bara. Periodyczne warunki brzegowe zostały zastosowane dla wszystkich trzech wymiarów, a oddziaływania elektrostatyczne były liczone przy użyciu metody Particle Mesh Ewald (PME) [165] z promieniem odcięcia równym 1 nm oraz szerokością siatki Fouriera wynoszącą 0,1 nm. Dla oddziaływań typu Lennarda-Jonesa użyłem promienia odcięcia równego 1 nm. Długości wiązań w DNA były zamrożone poprzez użycie algorytmu P-LINCS [158], a w przypadku cząsteczek wody dzięki algorytmowi SETTLE [157]. Równania ruchu były rozwiązywane za pomocą algorytmu „żabiego skoku” (z ang. leap-frog) z krokiem czasowym równym 2 fs. Przed przystąpieniem do właściwych symulacji typu energii swobodnej, symulowałem analogiczny system przez 4 μs w celu scharakteryzowania właściwości równowagowych G-kwadrupleksów w stanie natywnym. Wizualizacje przygotowałem za pomocą programu VMD [166], wykresy za pomocą programu Gnuplot [167], natomiast schematy za pomocą programu Inkscape [168].

#### 4.1.2 Sterowana dynamika molekularna

Do przygotowania początkowego zestawu konfiguracji dla symulacji typu energii swobodnej cała G-tetradą położona na 3'-końcu (guaniny oznaczone przez I, II, III, IV na Rys. 4.1C)



RYSUNEK 4.1: A) Początkowa konformacja symulowanego równoległego G-kwadrupeksu o ludzkiej sekwencji telomerowej, zaczerpnięta ze struktury rentgenowskiej [61]: reszty guaniny tworzące 3'-końcową G-tetradę są zaznaczone na niebiesko (I), zielono (II), fioletowo (III) i czerwono (IV); dwie pozostałe G-tetrazy są zaznaczone na szaro (środkowa) i pomarańczowo (5'-końcowa), tyminy i adeniny zaznaczono na brązowo, a szkielet fosfocukrowy na żółto; dwa jony potasu skoordynowane w kanale centralnym G-kwadrupeksu są przedstawione jako sfery van der Waalsa. W całym rozdziale zastosowano ten sam schemat kolorów. B) Struktura chemiczna pojedynczej G-tetrazy, w której cztery guaniny są ze sobą połączone poprzez cykliczne wiązania wodorowe typu Hoogsteena. Czarno-szare strzałki przedstawiają orientację momentów dipolowych guaniny IV i guaniny zlokalizowanych bezpośrednio pod nią. C) Schemat przedstawiający strukturę równoległego G-kwadrupeksu oraz położenie badanej 3'-końcowej G-tetrazy.

została poddana wymuszonej dysocjacji poprzez zastosowanie zewnętrznego potencjału. Potencjał przyłożyłem do współrzędnej reakcji zdefiniowanej jako całkowita liczba natywnych wiązań wodorowych  $N_{hb}$  tworzonych w G-tetradzie (Rys. 4.1B). Stałą siłową potencjału działającego przez cały czas symulacji (100 ns) ustawiłem na 2092 kcal/(mol·nm<sup>2</sup>). Całą procedurę przeprowadziłem przy użyciu wtyczki PLUMED 1.3 [169] współpracującej z pakietem obliczeniowym Gromacs. Całkowitą liczbę natywnych wiązań wodorowych  $N_{hb}$  zdefiniowałem jako sumę funkcji aktywacji (szczegóły opisane w Załączniku A i na Rys. A.1) dla wszystkich par atomów (akceptor-donor) tworzących natywne wiązania wodorowe w G-tetradzie. Dla w pełni utworzonej płaszczyzny powstaje  $N_{hb} = 8$  wiązań wodorowych pomiędzy atomami N1–O6 i N2–N7 (Fig. 4.1B), natomiast  $N_{hb} = 0$  wiązań wodorowych oznacza kompletną dysocjację płaszczyzny.

### 4.1.3 Symulacje typu energii swobodnej dla dysocjacji G-tetrazy

W celu zbadania stabilności termodynamicznej G-tetrazy na 3'-końcu użyłem metody *Umbrella Sampling*, która doskonale sprawdza się do wyznaczenia profili energii swobodnej wzdłuż kolektywnych współrzędnych takich jak całkowita liczba natywnych wiązań wodorowych. Cały interesujący mnie zakres wybranej współrzędnej kolektywnej (0–8 wiązań wodorowych) podzieliłem na 33 równoodległe okna symulacyjne z konfiguracjami początkowymi wygenerowanymi z trajektorii dynamiki sterowanej. Każde takie okno było symulowane przez czas 1000 ns z potencjałem harmonicznym o stałej siłowej równej  $1255,2 \text{ kcal}/(\text{mol} \cdot \text{nm}^2)$ , przyłożonym do wybranej współrzędnej kolektywnej. Profil energii swobodnej został wyznaczony na podstawie ostatnich 900 ns symulacji przy użyciu metody analizy ważonych histogramów (z ang. Weighted Histogram Analysis Method) [161]. Niepewności oszacowałem przy użyciu statystycznej metody analizy błędów „bootstrap” z uwzględnieniem korelacji pomiędzy analizowanymi seriami czasowymi danych.

### 4.1.4 Wymuszona dysocjacja pojedynczej guaniny

W celu wyznaczenia profilu energii swobodnej dla procesu dysocjacji pojedynczej guaniny z 3'-końcowej G-tetrazy równoległego G-kwadrupeksu przeprowadziłem symulację z użyciem metody adaptacyjnej siły wymuszającej (z ang. Adaptive Biasing Force) [170, 171] zaimplementowanej w programie NAMD 2.8 [172]. Budowa systemu oraz zestaw parametrów pola siłowego były takie same jak w symulacjach dysocjacji 3'-końcowej G-tetrazy. Symulacje dynamiki molekularnej przeprowadziłem w zespole statystycznym NPT wykorzystując dynamikę Langevina oraz zgodny z nią algorytm Noséa-Hoovera Langevina utrzymujący stałe ciśnienie [173]. Zastosowałem również periodyczne warunki brzegowe, natomiast oddziaływania elektrostatyczne były wyznaczone przy zastosowaniu metody PME [165] z promieniem odcięcia równym 1,2 nm oraz gęstością siatki Fouriera wynoszącą 0,1 nm. Dla oddziaływań typu Lennarda-Jonesa użyłem promienia odcięcia równego 1,2 nm. Na wiązania w cząsteczce DNA obejmujące również atomy wodoru nałożyłem więzy przy użyciu algorytmu SHAKE [174], natomiast więzy na geometrię cząsteczki wody nałożyłem przy użyciu algorytmu SETTLE [157]. Równania ruchu zostały rozwiązane przy użyciu prędkościowej wersji algorytmu Verleta z krokiem czasowym wynoszącym 2 fs.

Aby przygotować początkowe konfiguracje dla symulacji typu energii swobodnej przeprowadziłem na badanym układzie sterowaną dynamikę molekularną o długości 100 ns z dodatkowym potencjałem harmonicznym (ze stałą siłową  $2092 \text{ kcal}/(\text{mol} \cdot \text{nm}^2)$ ) wymuszającym dysocjację pojedynczej guaniny (IV na Rys. 4.1) od 3'-końcowej G-tetrazy w strukturze równoległego G-kwadrupeksu. Potencjał został przyłożony do współrzędnej reakcji zdefiniowanej jako odległość między środkiem masy wszystkich ciężkich atomów 3'-końcowej

guaniny a środkiem masy wszystkich ciężkich atomów pozostałych reszt guaninowych. Następnie z symulacji sterowanej dynamiki molekularnej wybrałem osiem konfiguracji, z których utworzyłem osiem zachodzących na siebie okien symulacyjnych typu ABF każde o szerokości 0,3 nm. Dzięki temu współrzędna reakcji obejmowała zakres od 0,6 nm (dla w pełni utworzonej G-tetrady) do 1,8 nm (dla zdysocjowanej guaniny IV). W każdym z tych okien system był symulowany przez 300 ns. Procedura zakładała zapamiętywanie chwilowych wartości siły działającej wzdłuż współrzędnej reakcji z dokładnością do 0,05 Å. Profil energii swobodnej wyznaczyłem na podstawie ostatnich 280 ns wszystkich okien symulacyjnych. Wartości błędów na profilach energii swobodnej oszacowałem jako błędy standardowe skorygowane o autokorelację szeregów czasowych.

Aby zdekomponować energię swobodną na wkład entalpowy i entropowy wyznaczyłem profile energii swobodnej w trzech różnych temperaturach: 295, 300 i 305 K (patrz Rys. A.10) z wykorzystaniem tego samego protokołu symulacyjnego. Z założenia o tym, że ciepło właściwe układu jest stałe w każdym punkcie współrzędnej reakcji ( $r$ ) wynika, że  $\Delta H$  i  $\Delta S$  nie zależą pośrednio od temperatury ( $\frac{\partial \Delta H(r,T)}{\partial T} = 0$  i  $\frac{\partial \Delta S(r,T)}{\partial T} = 0$ ). Oznacza to, że wkład entropowy jest prostą pochodną energii swobodnej względem temperatury przy założeniu stałego ciśnienia  $\Delta S(r) = \left( \frac{d\Delta G(r,T)}{dT} \right)_{N,P}$ . W każdym punkcie współrzędnej reakcji pochodna ta była aproksymowana przez nachylenie funkcji liniowej dopasowanej do wartości energii swobodnej w trzech różnych temperaturach.

## 4.2 Rola pętli w stabilności G-kwadrupleksów

### 4.2.1 Konstrukcja modeli i szczegóły symulacji

Początkowe konfiguracje modelowych dwupłaszczyznowych G-kwadrupleksów zostały wygenerowane za pomocą zaproponowanej przeze mnie procedury zwijania struktur G-kwadrupleksów o arbitralnej topologii i sekwencji (szczegółowy opis w podrozdziale 4.2.2). Struktury wygenerowałem na bazie rdzeni guaninowych czterech reprezentatywnych topologii występujących w wyznaczonych eksperymentalnie strukturach G-kwadrupleksów: równoległej (kod PDB 1KF1) [61], antyrównoległej (143D) [69], hybrydowej 3+1 (2JSM) [175] oraz krzeselkowej (2KM3) [176]. Dla dwóch z tych struktur (143D i 2JSM) uzyskałem po dwie różne wersje rdzeni guaninowych (143D\_2, 2JSM\_2), ponieważ w ich trójpłaszczyznowych strukturach znajduje się po jednej G-tetradzie różniącej się od dwóch pozostałych zestawem kątów glikozydowych guanin (patrz Tabela 4.1). Wybór referencyjnych topologii wynika z posiadania przez nie zróżnicowanych konfiguracji pętli. Taki wybór pozwoli na dobre scharakteryzowanie ogólnej zależności pomiędzy stabilnością struktury a rodzajem pętli i jej położeniem w sekwencji. Następnie dla każdej z sześciu referencyjnych topologii utworzyłem serię modelowych struktur o sekwencjach 5'-XGGX<sub>n1</sub>GGX<sub>n2</sub>GGX<sub>n3</sub>GG-3', gdzie X to dowolna zasada, a  $n_1$ ,  $n_2$  i  $n_3$  to liczba zasad z przedziału 1–4.



TABELA 4.1: Zestawienie właściwości strukturalnych referencyjnych konformacji pochodzących z wyznaczonych eksperymentalnie struktur G-kwadrupleksów.

Topologia	Kod PDB	Pętla od 5'-końca	Pętla środkowa	Pętla od 3'-końca	Kąty glikozydowe w G-tetradach
Równoległa	1KF1	śmigłowa	śmigłowa	śmigłowa	<i>anti-anti-anti-anti</i> <i>anti-anti-anti-anti</i> <i>anti-anti-anti-anti</i>
Antyrównoległa	143D	boczna	diagonalna	boczna	<i>anti-syn-syn-anti</i> <i>syn-anti-anti-syn</i> <i>anti-syn-syn-anti</i>
Hybrydowa	2JSM	boczna	boczna	śmigłowa	<i>syn-anti-syn-syn</i> <i>anti-syn-anti-anti</i> <i>anti-syn-anti-anti</i>
Krzesełkowa	2KM3	boczna	boczna	boczna	<i>syn-anti-syn-anti</i> <i>anti-syn-anti-syn</i>

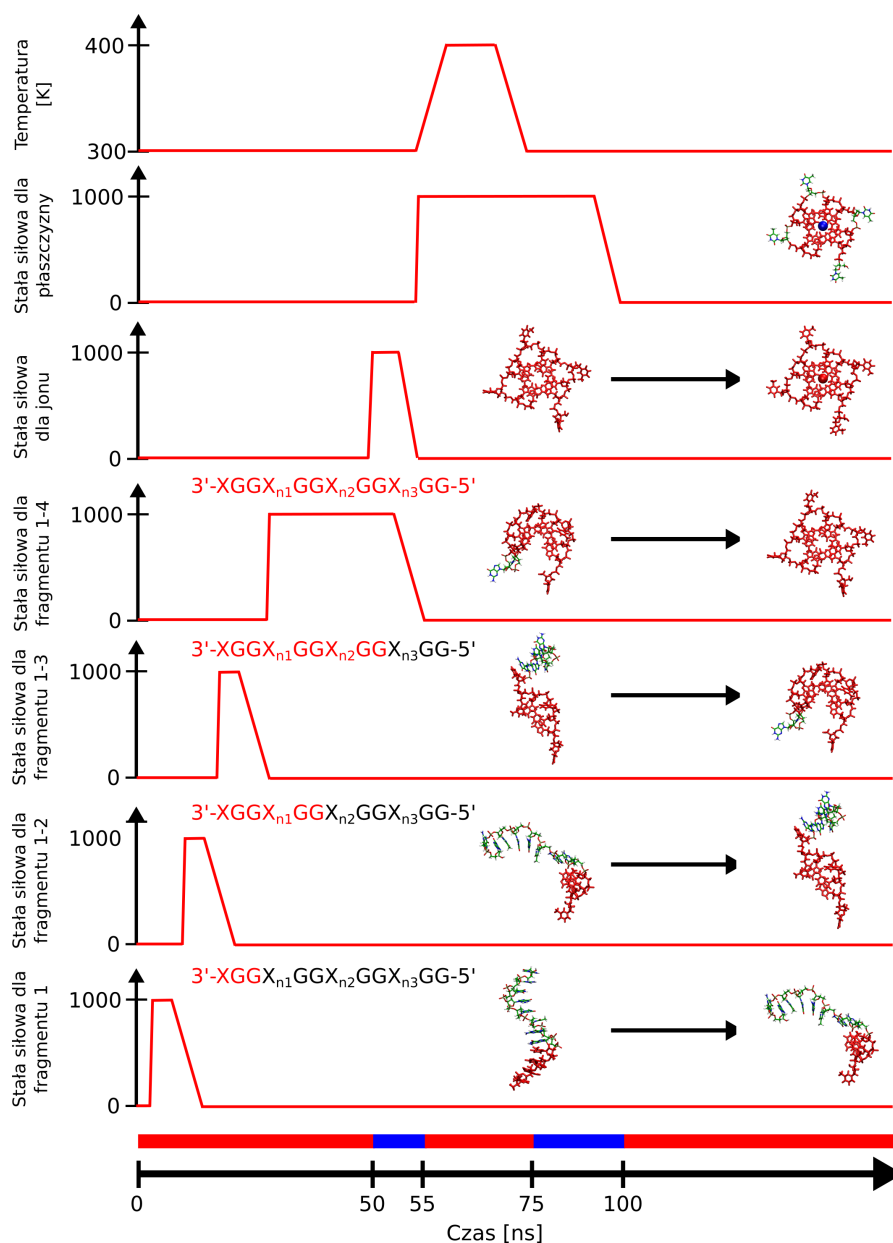
Dodatkowo dla celów weryfikacji zaproponowanej procedury zwijania G-kwadrupleksów zwinąłem *de novo* 3 wyznaczone eksperymentalnie struktury G-kwadrupleksów trój- i dwupłaszczyznowych (1KF1, 143D, 2JSM) oraz jedną dwupłaszczyznową (148D) [121].

Każda z wygenerowanych modelowych struktur G-kwadrupleksów została umieszczona w dwunastościennym pudle symulacyjnym z krawędziami w odległości 1,2 nm do najbliższego atomu DNA. Następnie pudło symulacyjne uzupełniłem odpowiednią liczbą cząsteczek wody typu TIP3 [162] oraz zneutralizowałem układ stosując fizjologiczne stężenie jonów (150 mM KCl). Do symulacji DNA oraz jonów użyłem pola siłowego CHARMM36 [163].

Wszystkie symulacje były wykonane przy pomocy pakietu Gromacs 5.0.4 [164] w zespole statystycznym NPT o temperaturze 300 K i ciśnieniu 1 bara. Periodyczne warunki brzegowe zostały zastosowane dla wszystkich trzech wymiarów, a oddziaływania elektrostatyczne były liczone przy użyciu metody Particle Mesh Ewald (PME) [165] z promieniem odcięcia równym 1 nm oraz szerokością siatki Fouriera wynoszącą 0,1 nm. Dla oddziaływań typu Lennarda-Jonesa użyłem promienia odcięcia równego 1 nm. Długości wiązań w DNA były zamrożone poprzez użycie algorytmu P-LINCS [158], a w przypadku cząsteczek wody za pomocą algorytmu SETTLE [157]. Równania ruchu były rozwiązywane za pomocą algorytmu „żabiego skoku” (z ang. leap-frog) z krokiem czasowym równym 2 fs.

#### 4.2.2 Procedura zwijania dwupłaszczyznowych G-kwadrupleksów

W celu wyznaczenia wpływu różnych pętli oraz ich długości na stabilność modelowych struktur G-kwadrupleksów opracowałem nową procedurę zwijania tych struktur. Używane



RYSUNEK 4.2: Schemat przedstawia kolejne etapy procesu zwijania na przykładzie G-kwadrupleksów dwupłaszczyznowych. Na rysunkach strukturalnych kolorem czerwonym zaznaczone są fragmenty nici DNA, które w danym momencie uczestniczą w procesie zwijania. Zwijanie struktur uzyskane jest za pomocą dodatkowych potencjałów harmonicznych działających na kolejne, coraz dłuższe fragmenty nici DNA.

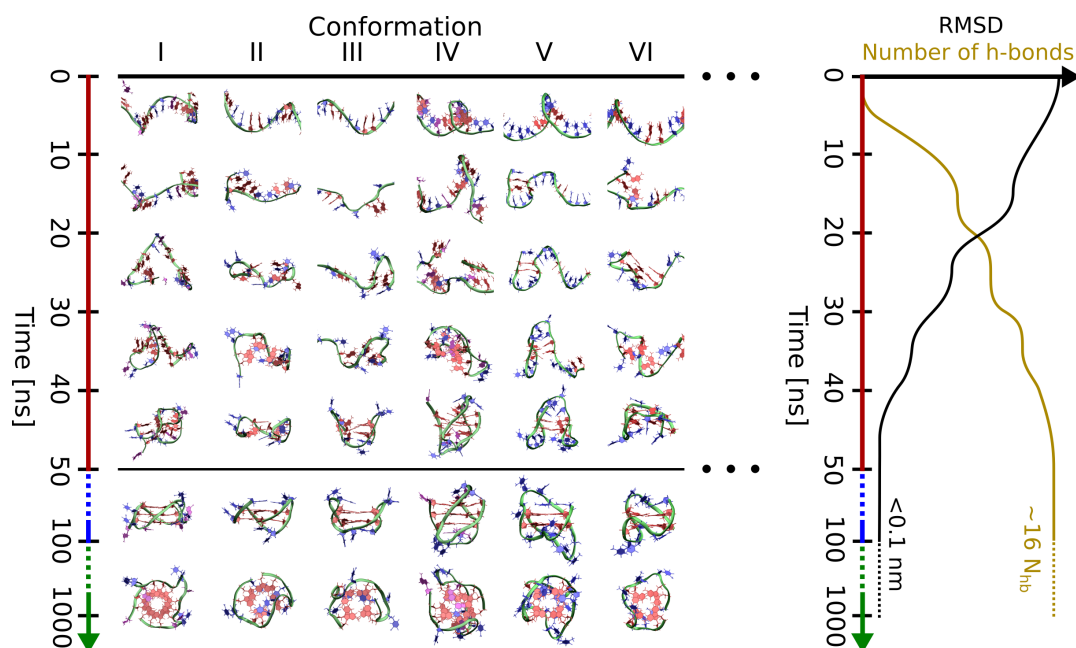
do tej pory metody zwijania G-kwadrupleksów polegały na przeprowadzeniu wielu kosztownych równoległych symulacji z temperaturową wymianą replik [177] lub użyciu skomplikowanej kombinacji kilku kolektywnych współrzędnych reakcji w symulacjach typu energii swobodnej [178].

Opracowana przeze mnie metoda zwijania struktur G-kwadrupleksów o arbitralnej topologii i sekwencji opiera się na sukcesywnym zwijaniu kolejnych fragmentów swobodnej nici DNA do referencyjnego modelu strukturalnego danej topologii. Jak pokazano wcześniej

każda z 26 teoretycznych topologii G-kwadrupleksów niezależnie od sekwencji cechuje się takim samym zestawem G-tetrad w rdzeniu guaninowym [60], który może służyć jako szablon do generowania modelowych G-kwadrupleksów. Wynik z tego, że aby zwinąć nić DNA do struktury G-kwadrupleksu o arbitralnej topologii należy zwinąć przede wszystkim poszczególne fragmenty nici z guaninami (GG) związane z G-tetradami, w takiej kolejności od 5'-końca w jakiej występują w strukturze referencyjnego rdzenia guaninowego. Natomiast adekwatne typy pętli występujące wybranej topologii formują się automatycznie podczas zwijania.

Ogólnie proces zwijania pełnowartościowej struktury G-kwadrupleksu jest podzielony na kilka etapów. Dla struktur dwupłaszczyznowych początkowy etap obejmuje zwijanie pierwszego fragmentu GG sekwencji od 5'-końca (5'-XGG) w celu utworzenia pierwszej części rdzenia guaninowego o odpowiednich kątach glikozydowych guanin (fragment 1 na Rys. 4.2). Później następują kolejne etapy formowania rdzenia guaninowego i przy okazji rejonów pętlowych. Obejmują one coraz to dłuższe fragmenty sekwencji nici DNA, odpowiednio fragment 1-2 o długości 5'-XGGX<sub>n1</sub>GG, fragment 1-3 o długości 5'-XGGX<sub>n1</sub>GGX<sub>n2</sub>GG i w końcu cała nić DNA 5'-XGGX<sub>n1</sub>GGX<sub>n2</sub>GGX<sub>n3</sub>GG-3'. Zwijanie rdzenia guaninowego jest realizowane za pomocą sterowanej dynamiki molekularnej poprzez zastosowanie zewnętrznego potencjału zmniejszającego wartość średniokwadratowego odchylenia położenia atomów (z ang. Root Mean Square Deviation) wyznaczonego pomiędzy położeniem ciężkich atomów guanin ze zwijanego fragmentu nici DNA a położeniem ciężkich atomów guanin z referencyjnego rdzenia guaninowego (patrz Rys. 4.3). Całkowity czas zwijania rdzenia guaninowego i relaksacji struktury wynosi 100 ns. Zwijaniu rdzenia guaninowego towarzyszy również formowanie natywnych wiązań wodorowych takich jakie występują w referencyjnym modelu rdzenia guaninowego (Rys. 4.3). W przypadku G-kwadrupleksów dwupłaszczyznowych całkowita liczba wiązań wodorowych typu Hoogsteena wynosi 16.

Kolejny etap generowania pełnowartościowej struktury G-kwadrupleksu polega na umieszczeniu jonu K<sup>+</sup> w środku rdzenia guaninowego. Za pomocą dodatkowego potencjału zmniejszana jest odległość pomiędzy jonom a środkiem masy rdzenia guaninowego. Ostatni etap odpowiada za właściwą relaksację rejonów pętlowych oraz rdzenia guaninowego G-kwadrupleksu. Jest to konieczne, ponieważ zwijanie nie jest procesem równowagowym, w którym mogą powstać naprężenia mechaniczne w strukturze. Relaksacja polega na nałożeniu więzów, najpierw na całą strukturę G-kwadrupleksu, a następnie stopniowym ograniczeniu więzów jedynie do rdzenia guaninowego i jednoczesnym podgrzaniu układu do 400 K w czasie 25 ns. Dzięki takiemu zabiegowi pętla w G-kwadrupleksie mają możliwość przyjęcia korzystnych energetycznie konformacji oraz relaksacji termicznej. Na koniec następuje stopniowe wyłączenie więzów i ochłodzenie systemu do 300 K w czasie 25 ns, po czym mogą nastąpić właściwe symulacje równowagowe.



RYSUNEK 4.3: Schemat przedstawia ogólną ideę procesu zwijania modelowych G-kwadrupleksów dwupłaszczyznowych. Po czasie 100 ns następuje właściwa symulacja równowagowa mająca na celu oszacowanie stabilności struktur. Wykres przedstawia wartość RMSD wyliczoną dla fałdowanej nici DNA względem modelowej struktury o danej topologii. Na tym samym wykresie przedstawiono również liczbę natywnych wiązań wodorowych tworzonych w formowanym rdzeniu guaninowym.

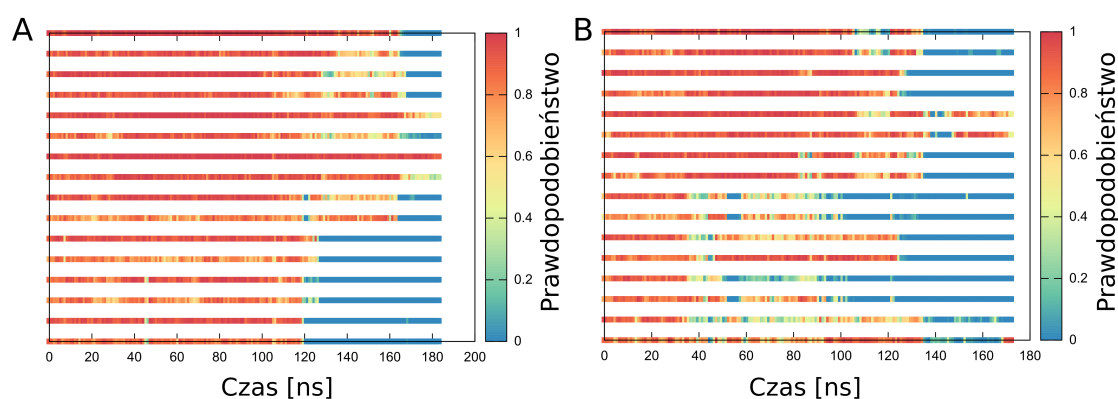
#### 4.2.3 Szacowanie stabilności G-kwadrupleksów

Idea szacowania stabilności termodynamicznej G-kwadrupleksów bazuje na pomiarze średniego oczekiwanego czasu życia wygenerowanych struktur G-kwadrupleksów. Ze względu na brak modelu matematycznego opisującego prawdopodobieństwo czasu życia struktur G-kwadrupleksów zakładam, że wyznaczony przeze mnie w symulacjach równowagowych czas życia (czas potrzebny do rozwinięcia) struktur G-kwadrupleksów jest miarą ich stabilności. Oczekiwany czas życia (do pierwszego rozwinięcia struktury) w przybliżeniu będzie zależał od bariery kinetycznej towarzyszącej temu procesowi. Oznacza to, że w przypadku pomiaru krótkiego czasu życia wygenerowanej struktury, a co za tym idzie niskiej bariery kinetycznej, szacowana stabilność wygenerowanej struktury jest mała. W przypadku pomiaru długich czasów życia badanej struktury, można stwierdzić, że bariera towarzysząca rozfałdowaniu jest duża i sprzyja potencjalnej stabilności struktury.

W praktyce, aby wiarygodnie oszacować oczekiwany czas życia, każdy wariant wygenerowanego modelowego G-kwadrupleksu był równolegle symulowany w 16 wersjach różniących się warunkami początkowymi (prędkościami początkowymi). Średni oczekiwany czas życia był wyznaczany, przy uwzględnieniu informacji o czasach życia w poszczególnych symulacjach oraz liczbie symulacji, w których nastąpiło rozfałdowanie struktury, na podstawie wzoru 4.1.

$$t = \frac{t_1 + t_2 + t_3 + \dots + t_n}{N_{end}} \quad (4.1)$$

gdzie  $t_1, t_2, t_3, \dots, t_n$  to zaobserwowane czasy życia w poszczególnych symulacjach, a  $N_{end}$  to liczba symulacji w których nastąpiło rozfałdowanie struktur G-kwadrupleksów. W celu efektywnego gospodarowania zasobami obliczeniowymi każda symulacja była zatrzymywana w momencie rozfałdowania struktury G-kwadrupleksu lub po osiągnięciu  $1 \mu\text{s}$  lub w niektórych przypadkach  $2 \mu\text{s}$  czasu symulacji. Strukturę, w której rozpadowi uległa jakakolwiek płaszczyzna guaninowa, czyli zerwane zostały natywne wiązania wodorowe, uznałem za rozfałdowaną.



RYSUNEK 4.4: Dwa typowe przebiegi symulacji równowagowych modelowych G-kwadrupleksów z wyszczególnieniem prawdopodobieństwa formowania każdego z 16 natywnych wiązań wodorowych w płaszczyznach guaninowych na przykładzie sekwencji  $\text{AGGT}_2\text{AGGT}_2\text{AGGT}_2\text{AGGT}_2\text{A}$  oraz konformacji równoległej. Prawdopodobieństwo formowania wiązania przedstawiono za pomocą skali kolorów, kolor czerwony oznacza 100% formowanie wiązania wodorowego, natomiast kolor niebieski oznacza brak wiązania wodorowego.

Rys. 4.4 przedstawia dwa typowe przebiegi symulacji równowagowych modelowych G-kwadrupleksów z wyszczególnieniem prawdopodobieństwa występowania każdego z 16 wiązań wodorowych w płaszczyznach guaninowych. W pierwszym przypadku (Rys. 4.4A) po około 120 ns symulacji równowagowej rozpadowi ulega 6 kolejnych wiązań wodorowych, a następnie po kolejnych 40 ns rozpadowi ulega reszta natywnych wiązań wodorowych. Taka kaskada zrywanych wiązań wodorowych jest charakterystyczna dla struktur nie wykazujących stabilności. W żadnej z 16 wersji symulacji (inny przykład na Rys. 4.4B) nie nastąpiło ponowne spontaniczne uformowanie jakiegokolwiek fragmentu początkowej struktury modelowego G-kwadrupleksu. Obserwacje wynikające z tych symulacji pozwoliły mi założyć, że czas rzędu  $1\text{-}2 \mu\text{s}$  jest wystarczający, aby uznać daną strukturę za potencjalnie stabilną i wartą dalszych analiz.

## 4.3 Dimeryzacja G-kwadrupleksów o konformacji równoległej

### 4.3.1 Konstrukcja modeli i szczegóły symulacji

Dimeryzacja G-kwadrupleksów o konformacji równoległej może zachodzić na 3 różne sposoby (patrz Rys. 1.11). W związku z tym przygotowałem 3 różne układy symulacyjne, każdy zawierający po dwie struktury równoległych G-kwadrupleksów o ludzkiej sekwencji telomerowej d[AG<sub>3</sub>(TTAG<sub>3</sub>)<sub>3</sub>] wstępnie odpowiednio ułożone względem siebie (w odpowiednim modzie dimeryzacji). Początkowe struktury G-kwadrupleksów, pierwotnie zawierające dwa jony K<sup>+</sup> koordynujące w centralnym kanale G-kwadrupleksu, pobrałem z bazy Protein Data Bank (kod PDB: 1KF1) [61]. Następnie utworzyłem dwunastościenne pudła symulacyjne o długości wektorów komórki elementarnej wynoszącej 10,1 nm. Każde pudło symulacyjne uzupełniłem 22826 cząsteczkami wody typu TIP3P [155] oraz ustawiłem fizjologiczne stężenie jonów (150 mM soli KCl). Dla każdego z tych 3 układów przygotowałem również po dwa analogiczne wersje układów różniące się jedynie pod względem stężenia soli. Pierwszy układ zawierał podwójne stężenie soli (300 mM KCl), natomiast drugi maksymalnie obniżone stężenie soli (60 mM), jednakże nadal zapewniające całkowitą neutralizację układu. Najniższe stężenie odpowiadało jedynie 42 przeciwjonom K<sup>+</sup>, włączając w to 2 jony w centralnym kanale. Do opisu cząsteczek DNA i jonów wybrałem pole siłowe CHARMM36 [153], ponieważ jak wcześniej pokazano odpowiednio odtwarza struktury oraz stabilność równoległych G-kwadrupleksów w roztworach wodnych [179–183]. W celu potwierdzenia niezależności uzyskanych wniosków od użytego pola siłowego zdecydowałem się również na użycie pola siłowego AMBER w wersji parmbsc1 [184], jednakże już tylko dla wersji z fizjologicznym stężeniem soli KCl. Przygotowałem również dodatkową symulację zawierającą ligand potencjalnie pośredniczący w dimeryzacji G-kwadrupleksów. Do tego celu wybrałem związek 3,4-TMPyPz. Główną parametryzację tego związku wykonałem za pomocą programu CGenFF [185]. Ładunki cząstkowe wyznaczyłem w programie Gaussian [186] za pomocą obliczeń metodą Hartree-Focka stosując bazę 6-31G\* oraz poprzez procedurę dopasowania potencjału elektrostatycznego Merza-Kollmana. Wartości numeryczne ładunków cząstkowych przedstawiłem w Tabeli B.1 w Załączniku B.1.

Symulacje dynamiki molekularnej prowadziłem za pomocą programu Gromacs w wersji 5.0.4 [164]. Symulacje przeprowadziłem w zespole statystycznym NPT z termostatem w wersji „v-rescale” [187] i temperaturą ustawioną na 300 K oraz ciśnieniem ustawionym na 1 bar kontrolowanym przez barostat Parrinello-Rahmana [188]. Zastosowałem okresowe warunki brzegowe we wszystkich kierunkach, natomiast oddziaływania elektrostatyczne były obliczane przy wykorzystaniu sumowania Ewalda (PME) [165] z promieniem odcięcia 1,2 nm oraz odległościami na siatce Fouriera 0,12 nm. Poza tym użyłem promień odcięcia równy 1,2 nm dla oddziaływań Lennarda-Jonesa. Na wszystkie długości wiązań w cząsteczce

DNA nałożyłem więzy metodą P-LINCS [158], natomiast dla cząsteczek wody użyłem metody SETTLE [157]. Całkowanie równań ruchu odbyło się za pomocą algorytmu Verleta w wersji „żabiego skoku” (z ang. leap-frog) z krokiem czasowym wynoszącym 2 fs.

### 4.3.2 Symulacje typu energii swobodnej

Głównym celem badań było ustalenie względnej stabilności trzech różnych sposobów agregacji dimerów (modów dimeryzacji) złożonych z dwóch równoległych G-kwadrupleksów. Odpowiadające temu profile energii swobodnej wyznaczyłem za pomocą metody *Umbrella Sampling* połączonej z wymianą replik (z ang. Replica Exchange Umbrella Sampling). Obliczenia przeprowadziłem przy użyciu wtyczki PLUMED 2.0 [169] połączonej z programem Gromacs. Początkowe konfiguracje układów na potrzeby dalszych symulacji REUS wygenerowałem w kilku kolejnych krokach. Najpierw przeprowadziłem krótkie symulacje równowagowe (<250 ns) dwóch niepołączonych monomerów G-kwadrupleksów. Jako współrzędną reakcji wybrałem odległość separacji pomiędzy środkami mas rdzeni guaninowych każdego z nich, której początkowa wartość była w zakresie 1,9-2,0 nm. W celu ułatwienia spontanicznej agregacji G-kwadrupleksów oba monomery początkowo zorientowałem równolegle względem siebie. Jak się później okazało jedynie w przypadku modów 3'-5' i 5'-5' zaobserwowałem spontaniczne wiązanie poprzez bezpośrednie oddziaływania płaszczyzn guaninowych. Nastąpiło to odpowiednio w 180 i 210 ns symulacji. Aby uzyskać podobne stany dimeryzacji również dla modu 3'-3', który nie wykazywał spontanicznego formowania, przeprowadziłem dodatkową sterowaną symulację dynamiki molekularnej. Przy użyciu dodatkowego potencjału harmonicznego, którego minimum poruszało się wzdłuż zastosowanej współrzędnej reakcji, dwa monomery G-kwadrupleksów były sukcesywnie zbliżane do siebie w czasie 100 ns symulacji. Stałą siłową tego potencjału ustawiłem na 286,1 kcal/(mol·nm<sup>2</sup>). Następnie, aby uzyskać stany z pozostałego zakresu współrzędnej reakcji tj. od stanu związanego (1,0 nm) do stanu całkowitej dysocjacji (3,5 nm), dla każdego modu dimeryzacji G-kwadrupleksy zostały poddane działaniu tego samego potencjału co poprzednio, jednakże poruszającego się przeciwnie.

W celu zapewnienia lepszej zbieżności profili energii swobodnej, w szczególności dla zakresu od 1,0 do 2,0 nm, początkowe konfiguracje układów wygenerowałem naprzemiennie z trajektorii spontanicznej asocjacji oraz wymuszonej dysocjacji. Natomiast konfiguracje układów dla zakresu 2,0–3,5 nm wygenerowałem jedynie z trajektorii wymuszonej dysocjacji. W wyniku tego uzyskałem 25 równomiernie rozłożonych okien REUS oddalonych od siebie o 0,1 nm. Każde z tych okien symulowałem przez 0,5  $\mu$ s, wykorzystując potencjał harmoniczny o stałej siłowej równej 286,1 kcal/(mol·nm<sup>2</sup>) ograniczający układ w wybranym obszarze współrzędnej reakcji. Próba wymian konfiguracji pomiędzy sąsiednimi oknami następowała

co 2 ps, natomiast współczynnik akceptacji wyniósł  $\sim 21\%$ . Warto zaznaczyć, że nie zastosowałem żadnych dodatkowych więzów ograniczających względną orientację monomerów G-kwadrupeksów. Profile energii swobodnej wyznaczyłem na podstawie ostatnich 450 ns uzyskanych trajektorii stosując standardową metodę analizy ważonych histogramów (WHAM 2.0.9) [189]. Niepewności oszacowałem za pomocą metody „bootstrap” uwzględniającej korelację w analizowanych szeregach czasowych. Standardowe energie swobodne dimeryzacji oszacowałem na podstawie profili energii swobodnej przy użyciu prostego wyrażenia:

$$\Delta G_{dim}^{\circ} = -k_B T \ln \left( \frac{1}{V_0} \int_0^R dr \int_0^{\pi} d\theta \int_0^{2\pi} d\phi r^2 \sin \theta \exp \left( -\frac{G(r)}{k_B T} \right) \frac{1}{4\pi r^2} \right) \quad (4.2)$$

gdzie  $G(r)$  to profil energii swobodnej z symulacji REUS,  $(4\pi r^2)^{-1}$  to człon korekty radialnej,  $V_0$  to objętość standardowa ( $1661 \text{ \AA}^3$ ) związana z standardowym stężeniem 1 M,  $R$  jest górną granicą definiującą stan związany,  $k_B$  jest stałą Boltzmann, a  $T$  temperaturą [190, 191].

## 4.4 Wpływ osmolitów na DNA

### 4.4.1 Konstrukcja modeli i szczegóły symulacji

Początkowe współrzędne dwuniciowej helisy B-DNA o długości 16 par zasad i sekwencji 5'-AGTCTAACTTGCATCT-3' wygenerowałem przy użyciu programu X3DNA [192]. We wszystkich systemach modelową cząsteczkę DNA otoczyłem cząsteczkami wody TIP3P [193] i odpowiednią liczbą cząsteczek osmolitu w dwunastościennym pudle symulacyjnym o wymiarach  $6,62 \text{ nm} \times 6,62 \text{ nm} \times 6,62 \text{ nm}$ . Symulacje obejmowały 14 różnych roztworów: roztwór bez denaturantów, po 6 roztworów zawierających mocznik lub TMG o stężeniach od 1,0 do 3,5 M co 0,5 M oraz dodatkowy roztwór z mocznikiem o stężeniu 5,0 M (szczegółowe składy poszczególnych układów w Załączniku C.1). Następnie dodałem 30 jonów sodu w celu zneutralizowania ładunku szkieletu DNA. Pole siłowe CHARMM27 [194] zostało użyte do DNA i jonów. Parametry pola siłowego dla mocznika zostały wygenerowane w CHARMM General Force Field (CGenFF) [185], a dla betainy zastosowałem wcześniej opracowany zestaw parametrów. Zestaw został wygenerowany przez analogię z parametrów pola siłowego CHARMM dla grup fosfolipidowych (grupa trimetyloamoniowa z fosfatydylocholiny; grupa karboksylowa z fosfatydyloseryny) [195].

Wszystkie symulacje dynamiki molekularnej przeprowadzono przy użyciu pakietu Gromacs 4.6.5 [164] w zespole statystycznym NPT przy temperaturze 300 K i ciśnieniu równym 1 bar. Zastosowałem periodyczne warunki brzegowe we wszystkich kierunkach, a oddziaływania elektrostatyczne były obliczane za pomocą metody PME z promieniem odcięcia równym 1 nm i gęstością siatki Fouriera równą 0,1 nm. Dla oddziaływań Lennarda-Jonesa zastosowałem promień odcięcia równy 1 nm. Długości wiązań w cząsteczce DNA i osmolitów ograniczyłem za pomocą algorytmu P-LINCS [158], natomiast dla wody zastosowałem



algorytm SETTLE [157]. Równania ruchu były całkowane przy użyciu algorytmu „żabiego skoku” z krokiem czasowym 2 fs. Całkowity czas symulacji dla każdego systemu wynosił 1  $\mu$ s.

Zawartość guanin i cytozyn w modelowej cząsteczce DNA (bez uwzględnienia końcowych zasad 3' i 5') odpowiada średniej zawartości guanin i cytozyn w typowej sekwencji DNA grasicy cielej (ctDNA; ~40% GC, ~60% AT), która została użyta w eksperymentach. 3'- i 5'-końców zostały pominięte w analizach z powodu nierealistycznego wpływu solwatacji na rozplatanie końców podwójnej helisy.

#### 4.4.2 Wymuszona denaturacja DNA

W celu symulowania DNA także w stanie zdenaturowanym wpierw dla wszystkich układów przeprowadziłem wymuszoną denaturację helisy DNA. Dwie nici DNA zmuszono do dysocjacji w 500 ns symulacji poprzez zastosowanie dodatkowego potencjału harmonicznego (o stałej siłowej 2092 kcal/(mol·nm<sup>2</sup>)) do współrzędnej kolektywnej zdefiniowanej jako odległość minimalna (MinDist) między wszystkimi grupami fosforanowymi z dwóch osobnych nici DNA (patrz szczegóły w Załączniku C.1).

Stan w pełni zdysocjowany został osiągnięty, gdy wartość odległości minimalnej przekroczyła 2 nm. Ostatecznie układy poddano symulacjom równowagowym dynamiki molekularnej przez 1  $\mu$ s utrzymując MinDist 2 nm za pomocą dodatkowego potencjału harmonicznego (o stałej siłowej 2092 kcal/(mol·nm<sup>2</sup>)). Cała procedura została przeprowadzona przy użyciu wtyczki PLUMED 2.0 [169] w połączeniu z programem Gromacs 4.6.5.

# 5 Wyniki oraz ich dyskusja

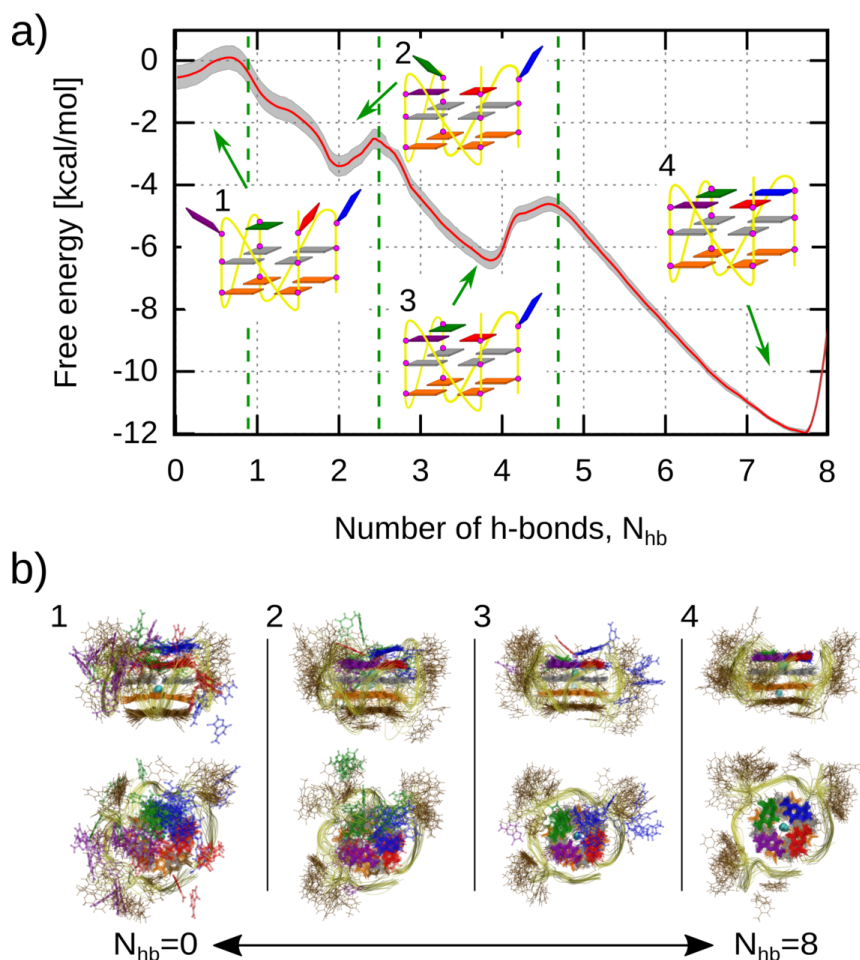
## 5.1 Stabilność G-tetrad w uformowanym G-kwadrupleksie

Ostatnie badania wykorzystujące symulacje typu dynamiki molekularnej pozwoliły na dokładne wyznaczenie dynamiki konformacyjnej różnych topologii G-kwadrupleksów oraz potwierdziły ich stabilność w  $\mu\text{s}$  skali czasowej [196–198]. Jednakże równowaga sił między- i wewnątrzcząsteczkowych odpowiedzialnych za tę stabilność nie została jednak dotąd dobrze opisana. Obliczenia chemii kwantowej i dynamiki molekularnej z wykorzystaniem modelowych układów dały ważny wgląd w to, jak względna stabilność różnych zestawów G-tetrad zależy od takich parametrów, jak względne ułożenie nici DNA, konformacje kąta glikozydowego oraz odległość i kąt skręcenia pomiędzy G-tetradami [199–201]. Badania te przewidywały preferencje energetyczne dla niektórych układów G-tetrad zasadniczo zgodnie z rozkładem konformacji znalezionych w eksperymentalnie wyznaczonych strukturach G-kwadrupleksów. Dopiero niedawno możliwe stało się bezpośrednie oszacowanie poszczególnych wkładów przyczyniających się do stabilizacji G-kwadrupleksów za pomocą eksperymentów operujących na pojedynczych cząsteczkach. W szczególności Ghimire et al. zastosowali wspomaganą chemicznie metodę szczypiec laserowych (z ang. laser-tweezer), aby pokazać, że oddziaływania w pętłach G-kwadrupleksów mogą być co najmniej tak samo ważne dla stabilności G-kwadrupleksów jak asocjacja guanin z wytworzeniem G-tetrad [202]. Pomimo powyższych wysiłków, pełny i kompleksowy obraz stabilizacji G-kwadrupleksów bazujący na informacjach strukturalnych i zapewniający na poziomie molekularnym zrozumienie sił napędowych występujących przy zwijaniu G-kwadrupleksów pozostaje nieznany.

Aby zatem szczegółowo zrozumieć podłoże stabilności G-kwadrupleksów w pierwszej kolejności wyznaczyłem profil energii swobodnej dla dysocjacji pojedynczej G-tetrady w kontekście w pełni uformowanej struktury G-kwadrupleksu. W tym celu wykorzystałem dynamikę molekularną z jawnym modelem rozpuszczalnika w połączeniu z metodą *Umbrella Sampling*, uzyskując całkowity czas symulacji wynoszący  $35 \mu\text{s}$ . Jako modelowy system wybrałem 3'-końcową G-tetradę w telomerowym G-kwadrupleksie o topologii równoległej, który wydaje się być najbardziej stabilną konformacją w warunkach stłoczenia molekularnego, a zatem może być dominującą formą G-kwadrupleksów w genomie DNA [79, 203]. Na podstawie profilu energii swobodnej oszacowałem termodynamikę i kinetykę dysocjacji i asocjacji G-tetrady oraz przewidziałem, że oba procesy przebiegają w sposób sekwencyjny i zależny od wiązania kationu w kanale G-kwadrupleksu. Ustaliłem również, że tworzenie G-

tetrad jest napędzane entalpowo, przy czym dominujący wkład pochodzi od oddziaływań z udziałem szkieletu fosfocukrowego DNA, a nie – jakby się mogło wydawać – zasad guaninowych z rozważanej G-tetrazy. Poza tym ogólny spadek entalpii towarzyszący formowaniu G-tetrazy jest w dużej mierze kompensowany przez spadek entropii wywołany związaniem się guanin oraz optymalizacją geometrii szkieletu.

### 5.1.1 Profil energii swobodnej dysocjacji G-tetrazy



RYSUNEK 5.1: A) Profil energii swobodnej dla formowania/dysocjacji pojedynczej tetrazy guaninowej z równoległej struktury G-kwadrupeksu wzdłuż współrzędnej zdefiniowanej jako całkowita liczba natywnych wiązań wodorowych wewnątrz G-tetrazy. Schematy G-kwadrupeksu reprezentują 4 typowe struktury pojawiające się wzdłuż ścieżki formowania/dysocjacji G-tetrazy. B) Nałożenie 20 reprezentatywnych struktur G-kwadrupeksów nawiązujących do każdego z 1–4 stanów. Struktury pokazują stopień uformowania wybranej G-tetrazy oraz jej dynamikę. Położenie i zakres poszczególnych stanów określiłem na podstawie profilu energii swobodnej z punktu A. Struktury do nałożenia wybrałem co  $0.2 \mu s$  z  $4 \mu s$  fragmentów symulacji odpowiadających centrum każdego stanu.

W celu zbadania stabilności konformacyjnej wspólnego motywu strukturalnego wszystkich G-kwadrupleksów wyznaczyłem z pomocą techniki US profil energii swobodnej dla dysocjacji G-tetrazy na 3'-końcu struktury G-kwadrupleksu o topologii równoległej. Jako współrzędną kolektywną opisującą badane przejście konformacyjne użyłem całkowitą liczbę natywnych wiązań wodorowych (h-bonds), które tworzone są w G-tetradzie ( $N_{hb}$ ).

Profil energii swobodnej z Rys. 5.1A pokazuje, że G-tetrada na 3'-końcu w stanie związanym, zdefiniowanym poprzez 8 natywnych wiązań wodorowych (stan 4 na Rys. 5.1), jest o  $\sim 12$  kcal/mol bardziej stabilna niż w stanie pełnej dysocjacji, tj. przy braku natywnych wiązań wodorowych (stan 1 na Rys. 5.1). Rys. 5.1B i Rys. A.2 pokazują również, że w stanie zdysocjowanym guaniny standardowo tworzące G-tetradę są silnie eksponowane do roztworu, jednakże czasami są także zaangażowane w lokalne oddziaływania z tyminami z pętli (oznaczone kolorem brązowym) oraz z guaninami z centralnej G-tetrazy (oznaczonej kolorem szarym). W tym miejscu warto także zaznaczyć, że nawet przy kompletnej dysocjacji G-tetrazy na 3'-końcu, reszta struktury G-kwadrupleksu pozostaje nadal stabilna w skali mikrosekund, co zgadza się z danymi eksperymentalnymi wskazującymi na możliwość formowania stabilnych G-kwadrupleksów zawierających tylko dwie G-tetrazy [204].

Rys. 5.1A pokazuje także, że podczas przejścia od globalnego minimum energii swobodnej w punkcie  $N_{hb}=8$  wiązań wodorowych (stan 4) do stanu pełnej dysocjacji w punkcie  $N_{hb}=0$  (stan 1) system przechodzi przez dwa lokalne minima w pobliżu punktów odpowiadających czterem (stan 3) oraz dwóm (stan 2) wiązaniom wodorowych. Jak można było oczekiwać, koszt dysocjacji pierwszej guaniny (6 kcal/mol), związany dokładnie z zerwaniem 4 wiązań wodorowych, jest w przybliżeniu dwa razy większy niż dla dwóch kolejnych etapów (3 kcal/mol), dla których zrywane są już tylko 2 wiązania wodorowe. W dalszych etapach analiz poprzez scałkowanie gęstości prawdopodobieństwa  $\rho(N_{hb}) \propto \exp[-\beta G(N_{hb})]$ , gdzie  $G(N_{hb})$  jest profilem energii swobodnej, obliczyłem, że równowagowe prawdopodobieństwo spontanicznej dysocjacji pojedynczej guaniny ze stanu w pełni uformowanej G-tetrazy jest w przybliżeniu równe 0,01%. Natomiast prawdopodobieństwo dysocjacji kolejnej guaniny jest zdecydowanie wyższe i wynosi już 0,6%.

Przebieg uzyskanego profilu energii swobodnej dodatkowo sugeruje, że dysocjacja tak samo jak formowanie G-tetrazy na 3'-końcu następuje w sposób sekwencyjny, gdzie kolejne guaniny dysocjują/asocjują jedna po drugiej. Wyznaczony przeze mnie średni czas pierwszego przejścia (z ang. Mean First Passage Time) dla dysocjacji pierwszej guaniny wynosi około 250 ns (szczegóły obliczeń w Załączniku A) podczas gdy stała szybkości równa się  $0,004 \text{ ns}^{-1}$ . Obie wartości wskazują, że jest to najwolniejszy etap dysocjacji, przez co determinuje on stabilność kinetyczną struktury. Ogólnie guanina I (niebieska na Rys. 5.1) wykazuje największą skłonność do dysocjacji ze stanu w pełni uformowanej G-tetrazy i prawdopodobnie jest to związane z jej położeniem najbliższym 5'-końca łańcucha DNA. Proces odejścia pojedynczej guaniny jest także w nieznacznym stopniu realizowany przez guaninę IV (na

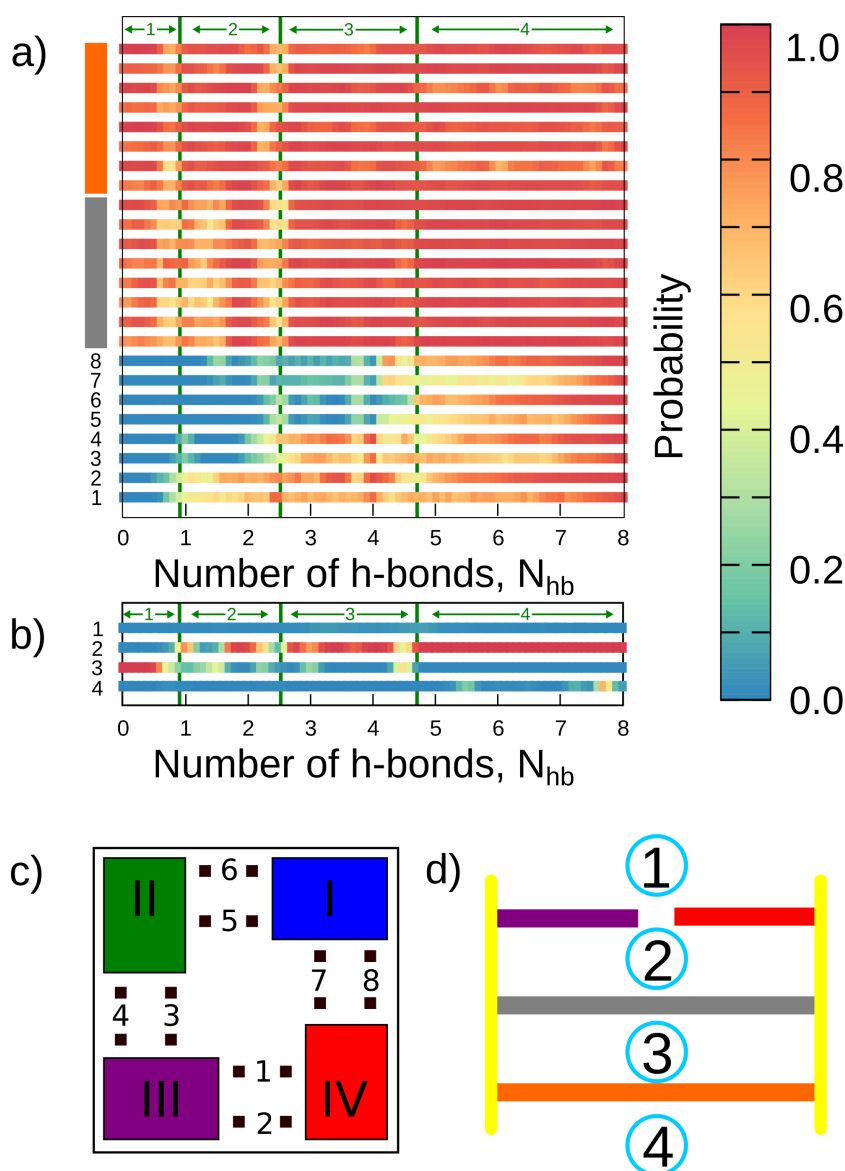
3'-końcu, czerwona na Rys. 5.1). Te obserwacje są również zgodne z symulacjami równowagowymi o długości 4  $\mu$ s, w których zaobserwowałem dysocjację pojedynczej guaniny I po czasie 750 ns, co można bezpośrednio zobaczyć na wykresie przedstawiającym ewolucję wiązań wodorowych w czasie (patrz Rys. A.3). Mimo, że profil energii swobodnej sugeruje, że dysocjacja drugiej guaniny (guanina II lub w mniejszym stopniu guanina IV, Rys. 5.1) jest o jeden rząd wielkości szybsza, z średnim czasem pierwszego przejścia równym  $\sim 40$  ns ( $k_{\text{off}} = 0,025 \text{ ns}^{-1}$ ), to jednak nie obserwowałem tego typu dysocjacji w symulacji równowagowej. Szczegółowa analiza wykazała, że ta niespodziewana stabilność jest rezultatem tworzenia się trypletów guaninowych, w których każda guanina bierze udział w tworzeniu trzech nienatywnych wiązań wodorowych z dwoma innymi guaninami (Rys. A.4). Ta obserwacja sugeruje, że stan pośredni, z płaszczyzną guaninową tworzoną przez jedynie trzy guaniny, może być w rzeczywistości bardziej stabilny niż wynika to z profilu energii swobodnej. Natomiast uzyskany przeze mnie stan pośredni G-trypletu generowany jest (stan 3 na Rys. 5.1) poprzez asocjację dwóch skrajnych guanin (II oraz IV) i tworzących wraz z centralną tylko 2 wiązania wodorowe (guanina III). W konsekwencji dysocjacja drugiej guaniny może być w rzeczywistości znacząco wolniejsza niż wynikałoby to z profilu energii swobodnej.

### 5.1.2 Mechanizm dysocjacji G-tetrady

W celu szczegółowego zbadania sekwencji zdarzeń występujących podczas stopniowej dysocjacji G-tetrady na 3'-końcu, policzyłem jak zmieniają się prawdopodobieństwa formowania natywnych wiązań wodorowych pomiędzy guaninami w zależności od  $N_{hb}$ . W związku z tym oryginalne dane z symulacji typu US zostały właściwie przeważone, aby uzyskać odpowiadające im prawdopodobieństwa równowagowe. Użyte przeze mnie wagi dla każdego  $i$ -tego okna symulacyjnego miały postać  $\exp[(V_i(N_{hb}) - F_i) / k_B T]$ , gdzie  $V_i$  jest nałożonym potencjałem ograniczającym, a  $F_i$  stałą o znaczeniu energii swobodnej wynikającą z przyłożenia tego potencjału.

Prawdopodobieństwa tworzenia wiązań wodorowych z Rys. 5.2A potwierdzają sekwencyjną naturę procesu dysocjacji, w którym to przy odejściu pierwszej guaniny następuje jednoczesne zrywanie 4 wiązań wodorowych oraz średnio 2 wiązań wodorowych dla pozostałych płaszczyzn guaninowych. Jak już wspomniałem największą skłonność do dysocjacji z w pełni uformowanej G-tetrady wykazuje guanina I (niebieska na Rys. 5.2C). Ustaliłem również, że odejście guaniny I następuje z prawdopodobieństwem 0,8, później następuje dysocjacja guaniny II (zielona) i na końcu zaś dochodzi do zrywania kontaktu pomiędzy guaniną III (fioletowa) i guaniną IV (czerwona).

Rys. 5.2A pokazuje także, że w odpowiedzi na dysocjację pierwszej oraz drugiej guaniny z G-tetrady na 3'-końcu prawdopodobieństwa formowania pozostałych wiązań wodorowych w jedynie częściowo utworzonej G-tetradzie stopniowo zmniejszają się do  $\sim 0,5$ . Ten znaczny spadek prawdopodobieństwa wskazuje na to, że wiązania wodorowe mają kooperatywną



RYSUNEK 5.2: A) Prawdopodobieństwa formowania 24 natywnych wiązań wodorowych w trzech G-tetradach w funkcji współrzędnej  $N_{hb}$ . Wiązania wodorowe zostały zdefiniowane przy użyciu standardowego kryterium geometrycznego, tzn. wiązanie wodorowe istnieje, gdy odległość D–A jest mniejsza niż  $3.5 \text{ \AA}$ , a kąt A–D–H jest mniejszy niż  $40.0^\circ$ , gdzie D oznacza donora, natomiast A akceptora w wiązaniu wodorowym. B) Prawdopodobieństwa obecności jonów w 4 możliwych miejscach obsadzenia w kanale w funkcji współrzędnej  $N_{hb}$ . C) Schemat rozważanej G-tetrady z ponumerowanymi wiązaniami wodorowymi (1–8) oraz D) schemat miejsc obsadzenia (1–4) przez jony miejsc wiążących pomiędzy i na zewnątrz G-tetrad.

naturę. W konsekwencji, płaszczyzna utworzona przez mniej niż 4 guaniny jest charakteryzowana zmniejszoną stabilnością strukturalną. Dwie pozostałe G-tetrady z G-kwadrupeksu (szara oraz pomarańczowa na Rys. 5.2A) na ogół pozostają nienaruszone podczas procesu dysocjacji terminalnej G-tetrady, jednakże można zaobserwować niewielkie zmniejszenie prawdopodobieństwa wiązań wodorowych, zwłaszcza podczas zdarzeń dysocjacji guanin.

Istnieją dwa typy chemiczne wiązań wodorowych stabilizujących G-tetradę: i) wiązania

wodorowe pomiędzy grupą aminową N2 i akceptorem N7 („zewewnętrzne wiązania wodorowe”, oznaczone przez cyfry 2, 4, 6 i 8 na Rys. 5.2C), oraz ii) wiązania wodorowe pomiędzy grupą aminową N1 oraz akceptorem O6 („wewnętrzne wiązania wodorowe”, oznaczone przez cyfry 1, 3, 5 i 7). Rys. 5.2A wskazuje na to, że te dwa typy wiązań wodorowych nie są równoważne pod względem stabilności podczas procesu dysocjacji. Mówiąc precyzyjniej, tylko w przypadku w pełni uformowanej G-tetrady zewnętrzne i wewnętrzne wiązania wodorowe tworzą się z takim samym prawdopodobieństwem. Dla stanów pośrednich to jednak zewnętrzne wiązania wodorowe są znacznie bardziej stabilne. Ta różnica w stabilności wydaje się mieć związek z występowaniem pewnych więzów geometrycznymi wynikających z tworzenia niekanonicznych wiązań wodorowych typu Hoogstena. W konsekwencji zewnętrzne wiązania wodorowe są krótsze (średnia odległość donor-akceptor wynosi 2,98 Å) i bardziej współliniowe (średni kąt akceptor–donor–wodór wynosi 15°) niż w przypadku wewnętrznych wiązań wodorowych (wartości odpowiednio 3,10 Å i 24°). Angażowanie atomów O6 w koordynowanie jonów  $K^+$  może także lokalnie zmniejszać tendencję do tworzenia wewnętrznych wiązań wodorowych.

Wielokrotnie pokazano, że kationy jednowartościowe, takie jak  $K^+$ ,  $Na^+$  czy  $NH_4^+$ , są niezbędne do formowania G-kwadrupleksów. Jest to skutkiem tego, że jony wykazują preferencję do bycia koordynowanymi w przestrzeniach pomiędzy G-tetradami przez osiem grup karbonylowych O6 wystawionych do centralnego kanału (patrz Rys. 4.1), tym samym zapewniając im dodatkową stabilizację [86, 88, 205]. W celu zbadania zmian obsadzenia miejsc wiążących jony podczas przejścia ze stanu zdysocjowanego do stanu z w pełni uformowaną G-tetradą, wyznaczyłem prawdopodobieństwa poszczególnych obsadzeń przez kation  $K^+$  w funkcji współrzędnej  $N_{hb}$ . Bezwzględne prawdopodobieństwa wyznaczyłem na podstawie oryginalnych danych z symulacji *Umbrella Sampling* przy użyciu tej samej procedury przeważenia danych jaką zastosowałem wcześniej dla wiązań wodorowych. Dane miejsce wiązania jest uznane za obsadzone, jeśli jakikolwiek jon  $K^+$  znajduje się w odległości 0,15 nm od geometrycznego środka tego miejsca. Z uwagi na fakt iż w moich symulacjach równowagowych zaobserwowałem, że jony  $K^+$  wykazują wyraźną tendencję do wiązania się z czterema grupami karbonyłowymi O6 również z zewnętrznych G-tetrad (pozycje 1 i 4 na Rys. 5.2D), również te dwa dodatkowe miejsca zostały uwzględnione w analizie.

Z otrzymanych prawdopodobieństw uwidocznionych na Rys. 5.2B można zauważyć, że gdy G-tetradą z 3'-końca jest w stanie całkowicie zdysocjowanym (stan 1) miejsce wiązania tworzone przez 2 pozostałe G-tetrady (pozycja 3) jest obsadzone z prawdopodobieństwem bliskim 1. Co ciekawe, po asocjacji dwóch pierwszych guanin z wybranej G-tetradą, tzn. po osiągnięciu stanu 2 z Rys. 5.1, prawdopodobieństwo obsadzenia pozycji 2 pomiędzy płaszczyzną guaninową na 3'-końcu oraz środkową płaszczyzną znacząco wzrasta do 0,7 kosztem pozycji 3. Wynik ten wskazuje na to, że obecność sześciu grup karbonylowych O6 w miejscu

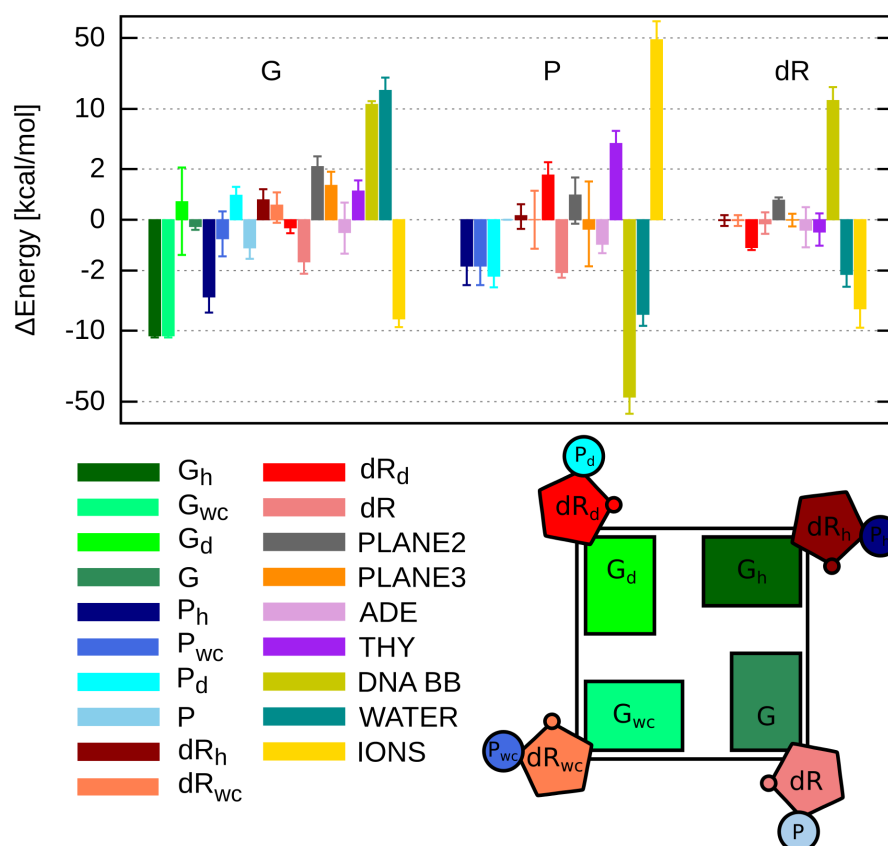
wiązania 2 jest najwyraźniej wystarczająca, aby związać jon potasu z pozycji 3. Przyłączenie dwóch kolejnych guanin skutkuje dalszym wzrostem obsadzenia pozycji 2 z prawdopodobieństwem zbliżającym się do jedności dla kompletnie uformowanej G-tetrady w stanie 4. Jednocześnie prawdopodobieństwo obsadzenia pozycji 3 spada do niemal zera. Wyraźna preferencja do częstszego obsadzania pozycji 2 niż pozycji 3 została także zaobserwowana w moich symulacjach równowagowych, w których koordynowanie pojedynczego jonu  $K^+$  w centralnym kanale jest wystarczające do utrzymania stabilności G-kwadrupleksu o strukturze równoległej przez czas kilku mikrosekund (Rys. A.5). Aby poddać obserwowane efekty dalszej analizie, przy użyciu dodatkowej symulacji metodą metadynamiki (patrz Załącznik A), wyznaczyłem profil energii swobodnej dla przejścia pojedynczego jonu  $K^+$  pomiędzy miejscami wiązania 2 i 3. Otrzymana funkcja energii swobodnej (Rys. A.6) potwierdza, że oba miejsca nie są równoważne pod względem ich powinowactwa do kationów  $K^+$ , przy czym wiązanie w miejscu 2 jest o  $\sim 2$  kcal/mol bardziej korzystne niż do miejsca 3. Co ciekawe, dominacja natywnej struktury G-kwadrupleksu (stan 4) z pojedynczym jonem związanym w centralnym kanale, wskazuje na to, że ze względu na odpychanie elektrostatyczne koordynacja  $K^+$  w jednym miejscu znacznie zmniejsza powinowactwo wiązania jonów w drugim. Należy jednak pamiętać, że ten efekt wykluczania jonów jest prawdopodobnie przeszacowany w moim modelu, ponieważ jak pokazano, większość addytywnych pól sił ma tendencję do zawężania odpychania kation–kation [198].

Fakt stabilności G-tetrady w czasie mikrosekund mimo braku obecności koordynującego jednowartościowego jonu stawia pytanie odnośnie rzeczywistych czynników determinujących stabilność G-kwadrupleksów. Należy jednak zaznaczyć, że w moich symulacjach obecność co najmniej jednego jonu  $K^+$  wewnątrz kanału jest konieczna dla stabilizacji G-kwadrupleksu. Natomiast gdy całkowite prawdopodobieństwo obsadzenia spada poniżej 1 ( $N_{hb}$  w zakresie 0,7–1,2 lub 2,2–2,6) stabilność całej struktury zmniejsza się, co również znajduje odzwierciedlenie w prawdopodobieństwach wewnątrzcząsteczkowych wiązań wodorowych w G-tetradach (Rys. 5.2A).

Rys. 5.2B pokazuje także, że pozycja 4, na zewnątrz 5'-końcowej G-tetrady, jest także dość często obsadzana przez jony  $K^+$ . Jest to szczególnie widoczne w stanie natywnym ( $N_{hb}$  w zakresie 7–8), w którym pozycja 3 jest zazwyczaj nieobsadzona, dzięki czemu nie wpływa niekorzystnie na energię wiązania w pozycji 4. Co więcej koordynowanie jonów  $K^+$  w pozycji 4 wydaje się być wspomagane przez adeniny, które zasadniczo posiadają pewną preferencję do wiązania się do 5'-końcowej G-tetrady (patrz Rys. 5.1B oraz Rys. A.2). W ten sposób tworzą one korzystniejsze miejsce wiązania oraz zapewniają dodatkową koordynację jonu (typowe struktury są pokazane na Rys. A.7). Natomiast, gdy asocjacja dodatkowych zasad nie jest możliwa przez zawadę przestrzenną, tak jak ma to miejsce przy 3'-końcowej G-tetradzie, nie obserwuje się znaczącego obsadzenia tego miejsca jonami (Rys. A.5).



## 5.1.3 Energetyczne determinanty stabilności G-tetrad



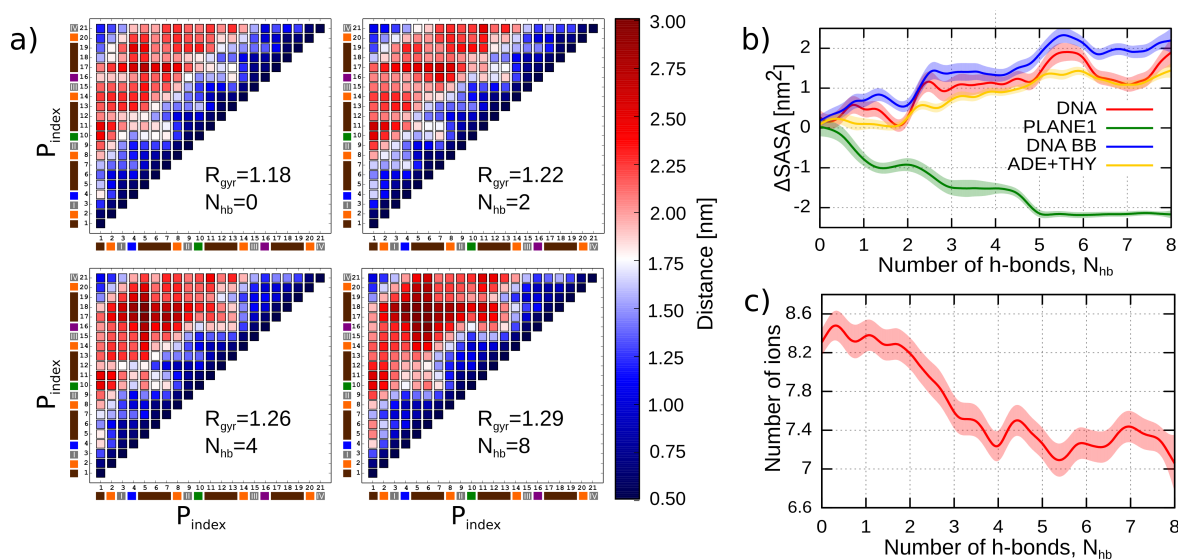
RYSUNEK 5.3: Poszczególne wkłady wnoszące do entalpii tworzenia G-tetrad wynikające z oddziaływania między nukleotydami guaninowymi I-IV, podzielonymi na guaninę (G), grupę fosforanową (P), dezoksyrybozę (dR) i pozostałymi elementami strukturalnymi układu. Tożsame wkłady energii oddziaływań uśredniono, aby uzyskać wkład entalpowy odpowiadający asocjacji pojedynczego nukleotydu guaninowego do G-tetrad. Dolne indeksy „wc” i „h” oznaczają oddziaływanie z nukleotydem guaninowym, który kontaktuje się z analizowanym nukleotydem odpowiednio jego krawędzią Watsona-Cricka lub Hoogsteena. Indeks dolny „d” oznacza oddziaływanie z przeciwnym nukleotydem guaninowym, a brak indeksu dolnego oznacza oddziaływanie wewnątrzcząsteczkowe. Resztę struktury G-kwadrupleksu podzielono na: środkową G-tetradę (PLANE2), 5'-kończącą G-tetradę (PLANE3), reszty adeninowe (ADE), tyminowe (THY) oraz szkielet fosfocukrowy (DNA BB).

W celu zidentyfikowania głównych determinantów strukturalnych odpowiedzialnych za stabilność G-kwadrupleksu przeprowadziłem dekompozycję entalpii formowania G-tetrad na 3'-końcu. Uzyskałem to poprzez wyznaczenie różnic w entalpii pomiędzy stanem 4 oraz 1 z Rys. 5.1A dla wkładów pochodzących od poszczególnych elementów strukturalnych G-kwadrupleksu. Wkłady wyznaczyłem jako różnice energii oddziaływania elektrostatycznego oraz van der Waalsa uśrednione po całym zespole równowagowym (zaniedbując znikomą pracę objętościową), co przedstawiłem na Rys. A.8 i Rys. A.9.

Jak można zobaczyć na Rys. A.8 struktura G-kwadrupeksu o topologii równoległej wykazuje czterokrotną (pseudo) symetrię obrotową (w której oś symetrii przechodzi przez centralny kanał), co wyraźnie znajduje odzwierciedlenie w wartościach energii poszczególnych wkładów do entalpii. Większość równoważnych, ze względu na symetrię, typów oddziaływań ma bardzo podobne wartości energii dla wszystkich czterech nukleotydów guanin (I–IV). Szczególnie jest to widoczne dla oddziaływania pomiędzy dowolną guaniną z rozważanej G-tetrady a jej sąsiednią guaniną lub pomiędzy grupą fosforanową oraz pozostałymi elementami łańcucha DNA. W celu scharakteryzowania generalnych sił stabilizujących działających na pojedynczy nukleotyd uśredniłem wkłady energetyczne z Rys. A.8 zgodnie z zaobserwowaną czterokrotną symetrią. Wynik przedstawiłem na Rys. 5.3.

Po pierwsze, analiza entalpii formowania G-tetrady pokazuje, że wkład entalpowy do energii swobodnej jest wynikiem subtelnej kompensacji dużych wkładów o przeciwnych znakach. Tak jak można się było spodziewać, jedna z głównych sił napędzających formowanie G-tetrady (spadek entalpii o 11 kcal/mol) pochodzi od wiązania pojedynczej guaniny do innej sąsiedniej guaniny krawędzią Hoogstena lub Watsona-Cricka, w wyniku tworzenia dwóch wiązań wodorowych. Ze względu na symetrię układu zysk energetyczny jest taki sam w obu przypadkach. Z drugiej strony ten zysk energetyczny jest jednocześnie kompletnie niwelowany przez utratę korzystnych oddziaływań guanin z wodą (15 kcal/mol). Układ zyskuje dodatkową stabilizację ( $-7,5$  kcal/mol na guaninę) dzięki koordynacji jonu potasu pomiędzy G-tetradami co silnie sprzyja formowaniu płaszczyzny. Dość niespodziewanie okazuje się, że oddziaływanie pomiędzy guaniną z rozważanej G-tetrady a płaszczyznami guaninowymi leżącymi poniżej jest tylko nieznacznie niekorzystne, z energiami oddziaływania wynoszącymi odpowiednio  $\sim 2$  oraz 1 kcal/mol. Efekt ten jest związany z faktem posiadania przez guaniny znacznego momentu dipolowego równego 6,71 D (Rys. 4.1B). W G-kwadrupeksach o topologii równoległej guaniny są ułożone prawie idealnie jedna nad drugą, co powoduje niekorzystne współliniowe ułożenie momentów dipolowych. Te niekorzystne oddziaływania elektrostatyczne są tylko częściowo kompensowane przez oddziaływania typu  $\pi - \pi$  (uwzględniane w polu siłowym w członie VdW). Zatem dodatkowa stabilizacja (7,5 kcal/mol), zapewniona przez koordynację jonu  $K^+$  między płaszczyznami (pozycja 2 na Rys. 5.2D), może w rzeczywistości być konieczna dla promowania tworzenia G-tetrady, zgodnie ze wszystkimi znanymi faktami eksperymentalnymi dotyczącymi formowania G-kwadrupeksów [86, 206, 207]. Ta indukowana jonami stabilizacja może być nawet silniejsza, biorąc pod uwagę możliwe zaniżenie energetyki przyciągania kationu przez guaniny [198]. Co ważne, subtelna równowaga w oddziaływaniach między resztami guaninowymi a ich bezpośrednim otoczeniem skutkuje ogólną stabilizacją energetyczną G-tetrady ( $\sim 2$  kcal/mol). Co więcej powinna być ona w dużej mierze niezależna od topologii G-kwadrupeksu oraz szczegółów ścieżki dysocjacji. Można stąd oczekiwać, że będzie miała takie samo kluczowe

znaczenie dla stabilności innych G-tetrad. Jednakże stabilizacja wynikająca z lokalnych oddziaływań guanin z innymi bliskimi elementami układu jest w znacznej mierze jedynie drugorzędna. Jak już wcześniej zasugerowano [202], dalekozasięgowe oddziaływania elektrostatyczne, obejmujące zwłaszcza pętle, może być co najmniej tak samo ważny dla stabilności G-kwadrupleksu. Nawiązując do tego, odkryłem, że tworzenie G-tetrazy na 3'-końcu w równoległym G-kwadrupleksie jest dodatkowo wspomagane przez korzystne ułożenie momentów dipolowych guanin względem grup fosforanowych od strony krawędzi Hoogsteena i w mniejszym stopniu grup fosforanowych związanych z krawędzią Watsona-Cricka Rys. 5.3.



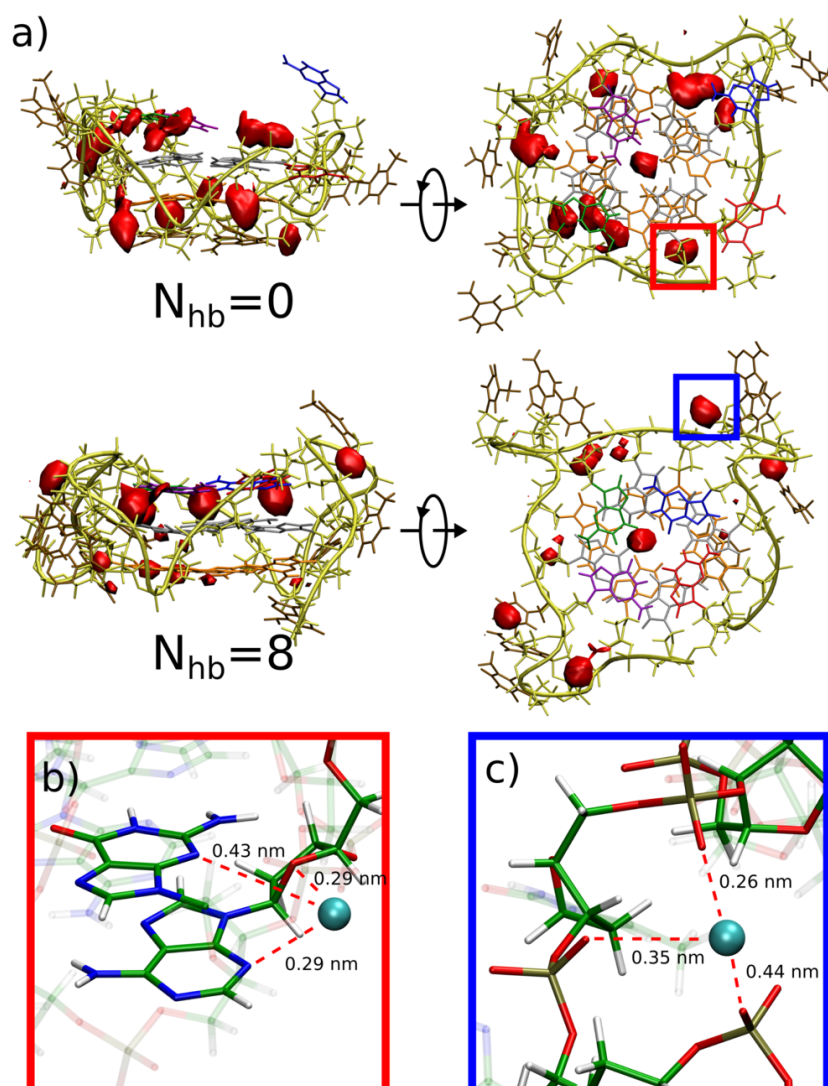
RYSUNEK 5.4: A) Mapa odległości dla wszystkich 21 fosforanów z łańcucha DNA. Promień żyracji ( $R_{gyr}$ ) policzyłem dla wszystkich ciężkich atomów z G-kwadrupleksu). B) Zmiany powierzchni dostępnej dla rozpuszczalnika rozłożone na wkłady pochodzące od różnych fragmentów struktury DNA. DNA – 715 atomów, PLANE1 – 60 atomów, DNA BB – 515 atomów, ADE+THY – 140 atomów. C) Zmiany liczby jonów w pobliżu DNA.

Duże znacznie członów związanych z oddziaływaniami dalekozasięgowymi dla tworzenia natywnej formy G-kwadrupleksu jest również odzwierciedlone w wewnętrznych oddziaływaniach łańcucha DNA. Z powodu ogólnego występowania odpychania elektrostatycznego można oczekiwać, że oddziaływania między grupami fosforanowymi I-IV zdestabilizują G-tetradę w stosunku do jej stanu zdysocjowanego. Nieoczekiwanie moje dane pokazują jednak, że omawiane oddziaływania fosforanów wpływają korzystnie na stabilność G-tetrazy. Związane z tym średnie zmiany energii wynoszą  $-1,5$  kcal/mol dla oddziaływań z fosforanem z krawędzi Hoogsteena,  $-1,5$  kcal/mol dla oddziaływań z fosforanem z krawędzi Watsona-Cricka oraz  $-2,5$  kcal/mol dla oddziaływań z przeciwległym fosforanem. Taki wynik wskazuje na to, że grupy fosforanowe I-IV muszą (średnio) oddalać się od siebie w trakcie tworzenia G-tetrazy. Aby dokładniej zbadać towarzyszące temu zmiany strukturalne, wyznaczyłem wartości odległości między parami wszystkich grup fosforanowych oraz

ich promień bezwładności w funkcji liczby wiązań wodorowych  $N_{hb}$ . Wyniki przedstawione na Rys. 5.4 pokazują, że nie tylko fosforany I–IV oddalają się od siebie i od innych fosforanów (średnio o 5,6%), ale także zwiększa się odległość separacji pomiędzy większością pozostałych grup fosforanowych (o 9,6%), w konsekwencji osiągając maksymalną wartość dla w pełni utworzonej G-tetrazy. To napędzane silnym odpychaniem względne przesunięcie grup fosforanowych powoduje duże zmiany energii oddziaływania między fosforanami I–IV a resztą szkieletu DNA (Rys. 5.3) oraz w obrębie samego szkieletu DNA (Rys. A.8). W związku z tym można zaproponować, że w G-kwadrupleksie o topologii równoległej fosforany kinetycznie stabilizują strukturę przez energetyczną optymalizację konformacji, która zapobiega lokalnym zaburzeniom struktury wywołanym np. przez dysocjację G-tetrazy.

Formowanie G-tetrazy i towarzyszące temu zmiany strukturalne są także związane ze zmianą pola powierzchni dostępnej dla rozpuszczalnika (z ang. solvent accessible surface area; SASA), czego analizy przestawiłem na Rys. 5.4B. Największe zmiany powierzchni są związane z wkładem pochodzącym od łańcucha DNA, powodując wzrost SASA o  $\sim 2 \text{ nm}^2$  i to pomimo jednoczesnego zmniejszenia się SASA dla guanin z rozważanej G-tetrazy (Rys. 5.4B). Jak można zobaczyć na Rys. 5.3 i Rys. A.8, zaobserwowany wzrost powierzchni łańcucha DNA wspomaga solwatację grup fosforanowych (o  $\sim 7 \text{ kcal/mol}$ ) i prowadzi do dodatkowej stabilizacji energetycznej G-tetrazy.

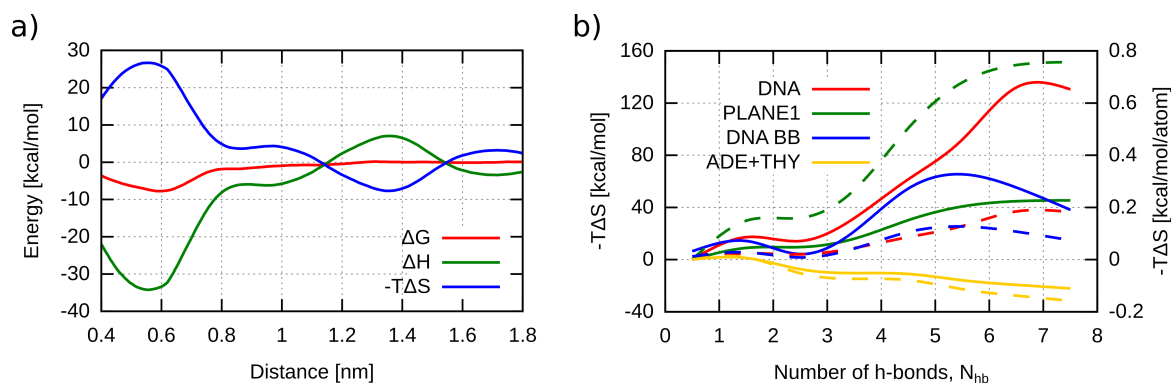
Rys. 5.3 i Rys. A.8 dodatkowo pokazują, że podczas formowania G-tetrazy wkłady energetyczne pochodzące od oddziaływań typu fosforan-fosforan i fosforan-woda są sumarycznie kompensowane przez utratę korzystnych oddziaływań elektrostatycznych między grupami fosforanowymi i jonami  $\text{K}^+$  obecnymi w rozpuszczalniku. Jest to dość nieoczekiwany efekt, szczególnie biorąc pod uwagę wydawałoby się zwiększoną ekspozycję fosforanów do rozpuszczalnika po utworzeniu G-tetrazy. Aby zrozumieć mechanizm odpowiedzialny za ten wynik, po pierwsze wyznaczyłem zmianę liczby jonów  $\text{K}^+$  w bezpośrednim kontakcie z G-kwadrupleksem w funkcji  $N_{hb}$ . Rys. 5.4C pokazuje, że tworzenie G-tetrazy jest rzeczywiście związane z utratą  $\sim 1 \text{ K}^+$  jonu bezpośrednio związanego z G-kwadrupleksem. Prawdopodobieństwo znalezienia jonu  $\text{K}^+$  w sąsiedztwie powierzchni G-kwadrupleksu (Rys. 5.5A) pokazuje, że przejściu od stanu całkowicie zdysocjowanego ( $N_{hb} = 0$ ) do natywnego ( $N_{hb} = 8$ ) towarzyszą wyraźne zmiany w przestrzennym rozkładzie jonów. W stanie natywnym jony wiążą się głównie do centralnego kanału G-kwadrupleksu oraz we wnękach utworzonych przez grupy fosforanowe pętli TTA (Rys. 5.5C), co również zaobserwowano w poprzednich symulacjach dynamiki molekularnej [196, 198, 208]. W przeciwieństwie do stanu zdysocjowanego w stanie natywnym istnieją również cztery dobrze zdefiniowane miejsca wiązania jonów wokół 5'-końcowej G-tetrazy. Tworzone są one przez atomy N7 reszt guaninowych oraz asocjujących adenin oraz atom O4' dezoksyrybozy (Rys. 5.5B). Te nowo utworzone miejsca wiązania są w dużej mierze odpowiedzialne za wiązanie średnio jednego dodatkowego jonu  $\text{K}^+$  z powierzchnią G-kwadrupleksu.



RYSUNEK 5.5: A) Przestrzenny rozkład jonów potasu wokół G-kwadrupleksu dla zdysocjowanej (górny panel) i w pełni uformowanej (dolny panel) 3'-końcowej G-tetrazy, przedstawiony jako izopowierzchnie gęstości prawdopodobieństwa o wartości 20 razy większej niż obserwowane w roztworze. Dwa dobrze określone miejsca wiązania jonu  $K^+$  oznaczone czerwonymi i niebieskimi ramkami zostały przedstawione w szczegółach na panelach poniżej. B) Miejsce wiązania jonu potasu w pobliżu 5'-końcowej G-tetrazy. C) Miejsce wiązania potasu utworzone przez grupy fosforanowe pętli TTA.

#### 5.1.4 Entropia formowania G-tetrazy

Przedstawiona wyżej dekompozycja oddziaływań połowicznie składających się na energie swobodną wskazała, że tworzenie G-tetrazy jest entalpowo korzystne. Jednakże można spodziewać się co najmniej częściowej kompensacji tego zysku energetycznego w wyniku spadku entropii wynikającego z utraty swobody konformacyjnej po asocjacji guanin. Niestety



RYSUNEK 5.6: A) Wkład entalpowy ( $\Delta H$ ) i entropowy ( $-T\Delta S$ ) do profilu energii swobodnej ( $\Delta G$ ) dla dysocjacji/asocjacji pojedynczej guaniny (IV na Rys. 4.1) z 3'-końcowej G-tetrazy równoległego G-kwadrupeksu. B) Zmiany entropii konformacyjnej dla dysocjacji/asocjacji G-tetrad obliczone w ramach przybliżenia quasi-harmonicznego dla całego G-kwadrupeksu DNA i jego poszczególnych składników, oznaczone zgodnie z Rys. 5.3 (linie ciągłe). Należy podkreślić, że ze względu na występowanie korelacje poszczególne wkłady nie są addytywne (suma stanowi górne ograniczenie rzeczywistych zmian entropii). Dodatkowo profile entropii przypadającej na atom są pokazane jako linie przerywane. Analiza zbieżności wartości entropii pokazana jest na Rys. A.11.

precyzyjne wyznaczenie zmian entropii jest dość trudne, ponieważ w ogólności wymagałoby precyzyjnego wyznaczenia tożsamy profili energii swobodnej w kilku różnych temperaturach. Z tego powodu w pierwszej kolejności skupiłem się na analizie procesu asocjacji jedynie pojedynczej guaniny. Takie podejście pozwoliło mi na obniżenie kosztów obliczeniowych o rząd wielkości zaś ze względu na symetrię problemu, umożliwiło mi wyciągnięcie jakościowych wniosków dla całej G-tetrazy. W szczególności, aby oszacować kompensację entropia/entalpia w trakcie tworzenia pełnej G-tetrazy, zastosowałem protokół symulacyjny *Adaptive Biasing Force* [170]. W rezultacie otrzymałem profile energii swobodnej dla dysocjacji pojedynczej reszty guaninowej (IV na Rys.4.1C) z 3'-końcowej G-tetrazy w trzech różnych temperaturach (295, 300 i 305 K). Uzyskana zależność profili energii swobodnej od temperatury (Rys. A.10) pozwoliła mi na rozdzielenie energii swobodnej asocjacji na wkłady entalpowe ( $\Delta H$ ) i entropowe ( $-T\Delta S$ ). Przedstawiłem je wraz z profilem energii swobodnej w 300 K na Rys. 5.6.

Warto zauważyć, że wzrost energii swobodnej towarzyszący asocjacji guaniny IV z wstępnie utworzonym G-tripleksem ( $-7$  kcal/mol) dobrze zgadza się ze zmianą wartości energii swobodnej po przejściu ze stanu 4 do stanu 3 na Rys. 5.1A (6 kcal/mol). Uzyskałem bardzo porównywalne wyniki mimo tego, że poprzednie przejście angażowało inną guaninę (głównie I), a zastosowane współrzędne reakcji znacznie się różniły (liczba natywnych wiązań wodorowych vs. odległość separacji centrów mas). Ta spójność potwierdza, że uzyskane przeze mnie profile mają odpowiednią zbieżność i że energetyka asocjacji odzwierciedla symetryczną naturę struktur G-tetrad. Co więcej, Rys. 5.6A pokazuje, że całkowita energia swobodna asocjacji jest rzeczywiście wynikiem kompensacji między korzystnym wkładem entalpowym

( $\Delta H = -35$  kcal/mol) i niekorzystną zmianą entropią ( $-T\Delta S = +29$  kcal/mol). Podobny efekt kompensacji energii/entropii został zaobserwowany wcześniej dla zwijania G-kwadrupleksów na podstawie pomiarów kalorymetrycznych [209, 210].

Jak można zobaczyć na Rys. 5.6A, zmiany entalpii i entropii występujące podczas procesu dysocjacji guaniny są stosunkowo duże w porównaniu do stabilizacji energetycznej całej G-tetrazy wynikającej z oddziaływania między resztami guanin i ich bezpośrednim otoczeniem (wcześniej oszacowane na  $\sim 2$  kcal/mol). Ta rozbieżność w skalach energii wskazuje na to, że dysocjacja reszty guaninowej z G-tetrazy wymaga rozległego przearanżowania konformacji G-kwadrupleksu obejmującego strukturę szkieletu fosfocukrowego DNA. Ten wniosek jest zgodny z opisanym wyżej procesem optymalizacji energetycznej geometrii szkieletu DNA poprzez maksymalizację odległości pomiędzy fosforanami. Wygląda na to, że lokalna optymalizacja geometrii szkieletu w stanie natywnym ( $N_{hb} = 8$ ) przyczynia się znacząco do zmniejszenia entropii konformacyjnej G-kwadrupleksu po asocjacji guaniny.

W celu dalszej analizy zmian entropii konformacyjnej podczas formowania G-tetrazy zastosowałem analizę quasi-harmoniczną [211]. W celu wyznaczenia i diagonalizacji macierzy kowariancji opisującej fluktuacje przestrzenne atomów w funkcji  $N_{hb}$ , ponownie wykorzystałem przeważone dane z symulacji US dla całego G-kwadrupleksu jak również dla jego głównych elementów strukturalnych. Uzyskane profile entropii (Rys. 5.6B) potwierdzają, że powstawaniu G-tetrazy towarzyszy duży spadek entropii ( $-T\Delta S = +130$  kcal/mol), który ze względu na symetrię układu jest około cztery razy większy niż oszacowany wcześniej dla pojedynczej guaniny na podstawie temperaturowej zależności energii swobodnej. Chociaż poszczególne wkłady entropowe uzyskane metodą quasi-harmoniczną nie są ściśle addytywne z powodu nieokreślonej korelacji pomiędzy nimi, zmiany w entropii szkieletu DNA i reszt guaninowych ( $-T\Delta S = +40$  kcal/mol) wskazują, że zmniejszone fluktuacje konformacyjne tych elementów przyczyniają się mniej więcej w równym stopniu do ogólnego spadku entropii. Z drugiej strony entropia konformacyjna reszt tyminowych i adeninowych z pętli wzrasta po utworzeniu G-tetrazy ( $-T\Delta S = -25$  kcal/mol), ponieważ w stanie natywnym są maksymalnie eksponowane do środowiska wodnego i próbują wiele możliwych konformacji.

Aby lepiej zobrazować zamiany strukturalne towarzyszące wypadkowemu spadkowi entropii, na Rys. A.12 przedstawiłem porównanie rozkładów równowagowych najważniejszych kątów torsyjnych określających geometrię nukleotydów tworzących rozważaną G-tetradę w stanie natywnym i zdysocjowanym. Tak jak można było oczekiwać, te wszystkie rotacyjne stopnie swobody są bardziej ograniczone w stanie natywnym, co znajduje odzwierciedlenie w obserwowanym spadku entropii konformacyjnej. Jednocześnie nie widać jednak wyraźnych przejść konformacyjnych podczas formowania G-tetrazy. Największa zmiana konformacyjna występuje dla kąta glikozydowego  $\chi_n$ , co wskazuje na to, że tworzenie G-

tetrad wymaga przyjęcia dobrze określonej orientacji guaniny względem pierścienia 2-dezoksyrybozowego i to dokładnie w konformacji *anti* ( $\sim 120^\circ$ , zgodnie z oczekiwaniami dla G-kwadrupeksów o topologii równoległej). Pozostałe kąty dla których obserwuje się wyraźne zwięźenie rozkładu to:  $\beta_n$ ,  $\alpha_n$  i  $\delta_n$ . Odpowiadają one za lokalną geometrię szkieletu DNA, w szczególności ułożenie guanozyny względem jej 5'-końcowej grupy fosforanowej. Można również zaobserwować, że dla stanu natywnego obserwuje się najlepiej zdefiniowaną konformację pierścienia deoksyrybozy, gdzie konformacje są ograniczone do C2'-endo (76%) i C1'-egzo (18%). Rozkład konformacji pierścienia deoksyrybozy jest za to znacznie bogatszy w stanie zdysocjowanym G-tetrady (48% C2'-endo, 38% C1'-egzo, 6% C3'-egzo i 5% O4'-endo).

### 5.1.5 Podsumowanie

W tym rozdziale przedstawiłem wyniki badań dotyczących molekularnych podstaw powstania i stabilności tetrady guaninowej (G-tetrady), która jest podstawowym elementem budulcowym każdego G-kwadrupeksu. W tym celu scharakteryzowałem termodynamikę i kinetykę dysocjacji 3'-końcowej G-tetrady w G-kwadrupeksie o topologii równoległej, stosując obliczenia typu energii swobodnej oparte na dynamice molekularnej. Wyznaczone przeze mnie profile energii swobodnej ujawniły, że dysocjacja i ponowna asocjacja G-tetrady zachodzi przez sekwencyjny i kooperacyjny mechanizm, który zależy od wiązania jednowartościowego jonu w centralnym kanale G-kwadrupeksu.

Na podstawie dekompozycji energii swobodnej dysocjacji na poszczególne wkłady entalpowe wywnioskowałem ponadto, że oddziaływania między resztami guaniny a resztą układu zapewniają tylko mało istotny wkład do ogólnej stabilności G-tetrady. Wynika to stąd, że stabilizujące natywne wiązania wodorowe Hoogsteena oraz oddziaływania wiążące jony są prawie całkowicie skompensowane przez niekorzystną desolvatację oraz szczególnie niekorzystne równoległe ułożenie momentów dipolowych guanin. Zgodnie z moimi symulacjami, G-tetradą jest stabilizowana przez oddziaływania obejmujące szkielet fosfocukrowy oraz pętle TTA, co zgadza się z wcześniejszymi danymi eksperymentalnymi [202, 212, 213]. Moje dane szczegółowo wskazują na to, że ta dodatkowa siła napędowa wynika z optymalizacji geometrii szkieletu DNA poprzez maksymalizację odległości między fosforanami. Jakakolwiek perturbacja wydaje się zakłócać lokalnie optymalny rozkład fosforanów w stanie natywnym. Ten wkład może również zwiększyć stabilność kinetyczną G-kwadrupeksów DNA, co tłumaczyłoby konieczność użycia dużych sił w celu rozfałdowania struktury G-kwadrupeksu w technice AFM [202]. Szczegółowa analiza przestrzennego rozkładu zmian entalpii i entropii dysocjacji wykazała ponadto, że korzyść energetyczna wynikające z tworzenia G-tetrady jest w dużej mierze kompensowana przez spadek entropii towarzyszący łączeniu guanin i optymalizacji struktury szkieletu DNA.



## 5.2 Rola pętli w stabilności G-kwadrupleksów

W celu ustalenia wpływu poszczególnych rodzajów pętli na stabilność G-kwadrupleksów należałoby przebadać wszystkie możliwe kombinacje sekwencji DNA potencjalnie tworzących G-kwadrupleksy oraz ustalić czy dana sekwencja ma możliwość formowania stabilnej struktury G-kwadrupleksu. Następnie konieczne byłoby ustalenie najbardziej prawdopodobnej konformacji i struktury przyjmowanej przez daną sekwencję. Jest to trudne i żmudne przedsięwzięcie eksperymentalne. Niestety również użycie symulacji komputerowych do tego celu jest niepraktyczne, ponieważ osiągnięcie czasów symulacji pozwalających na uzyskanie spontanicznego formowania G-kwadrupleksu ze stanu kompletnie rozciągniętej nici DNA jest obecnie nierealne.

W związku z tymi trudnościami postanowiłem zmodyfikować standardowe podejście badawcze do tego rodzaju problemów. Kosztowne metody symulacyjne bazujące na wyznaczeniu stabilności w symulacjach typu energii swobodnej zostały zastąpione podejściem obserwacyjnym sprawdzającym jedynie czy wygenerowana struktura G-kwadrupleksu ulega rozpadowi lub pozostaje bliska swojej formy początkowej (zwinionej). W zaproponowanym podejściu za naturalną miarę stabilności danej struktury wybrałem średni oczekiwany czas życia struktury zwinionej. Konkretnie za rozfałdowanie struktury uznawałem moment zerwania wszystkich wiązań wodorowych (dysocjacji guanin) w którejkolwiek płaszczyźnie guaninowej.

Dla kilku reprezentatywnych sekwencji, mogących potencjalnie formować dwupłaszczyznowe G-kwadrupleksy, wygenerowałem modelowe struktury G-kwadrupleksów o różnych reprezentatywnych konformacjach, stosując w tym celu specjalnie opracowany protokół zwijania G-kwadrupleksów (szczegóły w podrozdziale 4.2.2). Procedura generowania modelowych struktur G-kwadrupleksów jest realizowana na podstawie wiedzy o znanych i dobrze przebadanych strukturach eksperymentalnych oraz dzięki temu, że cechą charakterystyczną G-kwadrupleksów o zadanej konformacji jest występowanie w nich identycznych rdzeni guaninowych. Następnie poprzez przeprowadzenie odpowiednio długich symulacji konwencjonalnej dynamiki molekularnej badałem średnie oczekiwane czasy życia poszczególnych form modelowych G-kwadrupleksów.

### 5.2.1 Weryfikacja procedury zwijania G-kwadrupleksów

W pierwszej kolejności przeprowadziłem weryfikację zaproponowanego przeze mnie protokołu zwijania i generowania struktur G-kwadrupleksów o dowolnej topologii i sekwencji. W tym celu zwiniałem *de novo* 4 wyznaczone eksperymentalnie struktury G-kwadrupleksów oraz przeprowadziłem symulacje równowagowe wygenerowanych struktur. Symulacje dotyczyły trzech trójpłaszczyznowych G-kwadrupleksów o sekwencji telomerowej (1KF1, 143D i 2JSM) oraz jednej struktury z dwiema G-tetradami (148D). Każda z tych struktur cechuje się

inną topologią oraz innym zestawem pętli (Tabela 4.1). Stanowią one reprezentatywny zbiór testowy dla ustalenia użyteczności mojej procedury.

TABELA 5.1: Czasy życia modelowych struktur G-kwadrupleksów w poszczególnych symulacjach równowagowych dla sekwencji telomerowej AGGGT<sub>2</sub>AGGGT<sub>2</sub>AGGGT<sub>2</sub>AGGG oraz GGT<sub>2</sub>GGTGTGGT<sub>2</sub>GG.

Numer symulacji	równ. 1KF1	antyrówn. 143D	hybry. 2JSM	antyrówn. 148D
1	1000	1000	599	1000
2	1000	1000	1000	1000
3	295	1000	1000	201
4	1000	1000	1000	47
5	85	1000	1000	1000
6	1000	1000	1000	310
7	166	1000	1000	1000
8	1000	1000	88	1000
9	1000	61	1000	106
10	1000	1000	891	1000
11	1000	1000	1000	1000
12	1000	1000	1000	1000
13	1000	1000	1000	128
14	1000	1000	1000	215
15	1000	1000	21	1000
16	369	1000	1000	1000
Średni oczekiwany czas życia [ns]	3229	15061	3400	1835

Tabela 5.1 przedstawia zaobserwowane czasy życia modelowych G-kwadrupleksów w poszczególnych symulacjach równowagowych oraz wynikające z nich średnie czasy życia. Ogólnie większość wersji modelowych struktur pozostaje stabilna, nie wykazując oznak rozfałdowania w postaci zrywanych wiązań wodorowych w G-tetradach. Co więcej struktury te nie zmieniają się znacząco w czasie w stosunku do swoich odpowiedników wyznaczonych eksperymentalnie ( $RMSD < 0,3$  nm). Tylko w nielicznych symulacjach doszło do zerwania wiązań wodorowych i rozfałdowania G-kwadrupleksów. Oznacza to, że struktury G-kwadrupleksów zwinięte za pomocą mojej procedury mają odpowiednio zrelaksowaną strukturę oraz znajdują się w wystarczająco głębokim minimum energetycznym. Co ważne, średni oczekiwany czas życia dla struktury równoległej (1KF1) jest trochę krótszy od dwóch pozostałych struktur telomerowych. Oznacza to, że w symulacjach ta struktura wykazuje mniejszą stabilność od struktur antyrównoległej (143D) oraz hybrydowej (2JSM), co jest zgodne z danymi eksperymentalnymi [214]. W przypadku badanego G-kwadrupleksu dwupłaszczyznowego (148D) średni oczekiwany czas życia jest stosunkowo niewielki, wskazując na niską stabilność struktury. Jest to wynik zgodny z eksperymentalnie wyznaczoną stabilnością wynoszącą  $\Delta G = -1,9-3,2$  kcal/mol [215].

Przedstawione wyniki potwierdzają efektywność zaproponowanej metody zwijania G-kwadrupleksów oraz szacowania stabilności na podstawie średniego oczekiwanego czasu życia.

## 5.2.2 Stabilność modelowych G-kwadrupleksów

Właściwa część badań miała na celu ustalenie wpływu długości petli, ich rodzajów oraz położenia na stabilność modelowych G-kwadrupleksów. Celem było także ustalenie najkrótszych sekwencji potencjalnie tworzących G-kwadrupleksy. Dla wcześniej wybranych sześciu referencyjnych topologii przeprowadziłem symulacje równowagowe na kilku wybranych sekwencjach o różnej długości oraz zawierających różne zasady azotowe pomiędzy fragmentami GG. Tabela 5.2 przedstawia średnie czasy życia modelowych struktur G-kwadrupleksów uzyskane dla 14 wybranych sekwencji tworzących jedną z sześciu wybranych konformacji.

W pierwszej kolejności sprawdziłem możliwość formowania dwupłaszczyznowych G-kwadrupleksów dla najkrótszej możliwej sekwencji, czyli TGGTGGTGGTGG. Okazało się, że sekwencja ta ma możliwość tworzenia stabilnych G-kwadrupleksów wyłącznie o konformacji równoległej. Jak już wcześniej wskazano, pętle śmigłowe występujące w konformacjach równoległych rzeczywiście są preferowane dla petli o krótkich sekwencjach. Natomiast w przypadku pozostałych konformacji struktury modelowych G-kwadrupleksów podlegają silnym naprężeniom mechanicznym i ulegają natychmiastowemu rozpadowi w czasie poniżej 1 ns. W związku z tą obserwacją, dla pozostałych sekwencji z pojedynczą zasadą symulacje prowadzone były tylko dla topologii równoległej.

TABELA 5.2: Średnie czasy życia modelowych struktur G-kwadrupleksów uzyskane dla wybranych sekwencji tworzących jedną z 6 referencyjnych konformacji. Średnie czasy życia są obliczone na podstawie 16 niezależnych symulacji równowagowych i podane są w nanosekundach. Znak – oznacza brak danych. Dla przejrzystości czasy symulacji zaokrągliłem do jedności.

Sekwencja	równ. (1KF1)	antyrówn. I (143D)	antyrówn. II (143D_2)	hyb. I (2JSM)	hyb. II (2JSM_2)	krzeselkowa (2KM3)
TGGTGGTGGTGG	1251	1	1	2	2	1
AGGAGGAGGAGG	1622	-	-	-	-	-
CGGCGGCGGCGG	368	-	-	-	-	-
GGGGGGGGGGGG	991	-	-	-	-	-
abGGabGGabGGabGG	684	-	-	-	-	-
TGGT <sub>2</sub> GGT <sub>2</sub> GGT <sub>2</sub> GG	180	5	4	7	73	42
TGGT <sub>3</sub> GGT <sub>3</sub> GGT <sub>3</sub> GG	447	79	1229	222	441	1219
TGGT <sub>4</sub> GGT <sub>4</sub> GGT <sub>4</sub> GG	334	364	16000	561	2078	7756
TGGTGGT <sub>3</sub> GGTGG	572	2	1	20	44	748
TGGTGGT <sub>3</sub> GGT <sub>2</sub> GG	526	4	3	145	73	529
TGGT <sub>2</sub> GGT <sub>3</sub> GGTGG	198	16	9	10	129	511
TGGT <sub>2</sub> GGT <sub>3</sub> GGT <sub>2</sub> GG	272	127	414	145	201	899
TGGT <sub>3</sub> GGT <sub>3</sub> GGT <sub>2</sub> GG	312	284	756	225	235	808
AGGT <sub>2</sub> AGGT <sub>2</sub> AGGT <sub>2</sub> AGG	299	839	1017	693	637	1062

W następnym kroku swoich badań sprawdziłem, czy możliwe jest zwiększenie stabilności struktur o konformacji równoległej powstałych na bazie sekwencji TGGTGGTGGTGG poprzez mutację wszystkich tymin na inne zasady azotowe. Tyminy zostały zastąpione przez: adeniny, cytozyny, guaniny lub uszkodzenia bez zasady azotowej (z ang. abasic sites; ab).

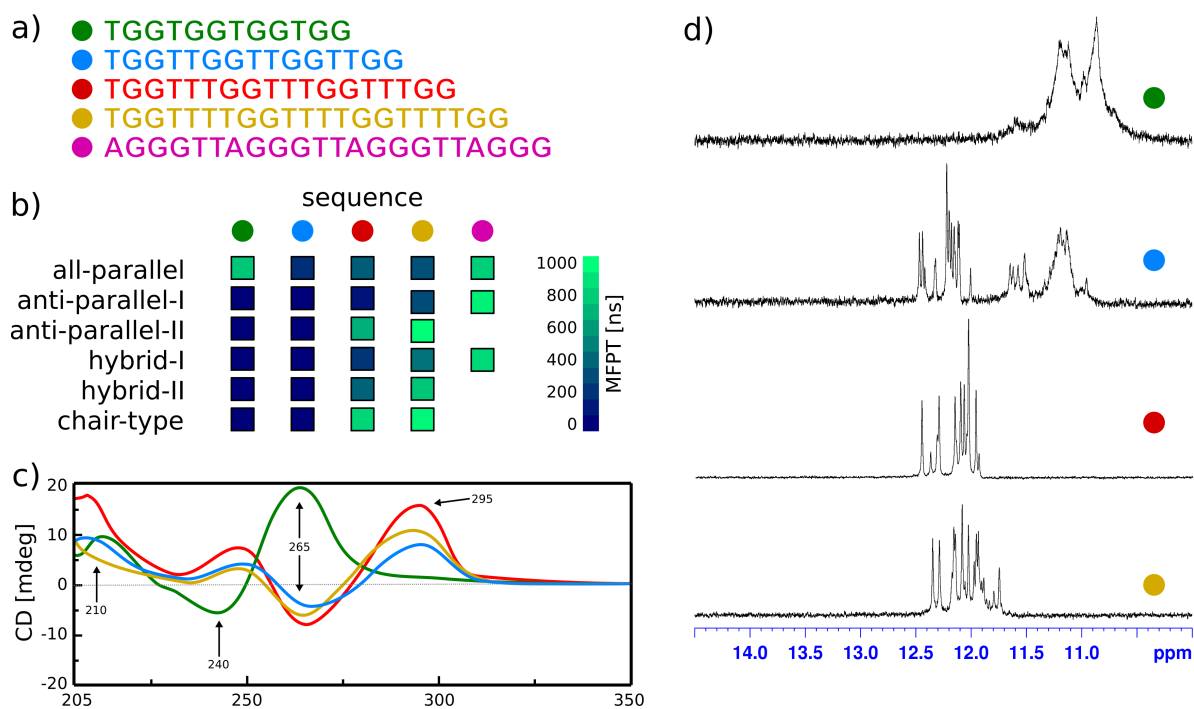
Ogólnie stabilność struktur o konformacji równoległej nie uległa znacznej zmianie po mutacji sekwencji. Jednakże użycie sekwencji CGGCGGCGGCGG wyraźnie obniżyło średnie czasy życia tych struktur. Może to wynikać z dodatkowych oddziaływań cytozyn z pozostałymi zasadami.

Ciekawie przedstawiają się także wyniki dalszego wydłużania sekwencji o kolejne tyminy w pętłach. Użycie sekwencji TGGT<sub>2</sub>GGT<sub>2</sub>GGT<sub>2</sub>GG nie poprawia średnich czasów życia konformacji innych niż równoległa. Oznacza to, że tylko dwie zasady azotowe obecne przynajmniej w jednym z trzech rejonów pętlowych są nadal nie wystarczające do utworzenia stabilnej struktury innej niż równoległa. Natomiast średni czas życia dla konformacji równoległej oraz wynikająca z tego stabilność struktury została znacznie obniżona. Co więcej, dalsze symetryczne wydłużanie sekwencji pozostawia średnie czasy życia struktur o konformacji równoległej na podobnym poziomie i sugeruje małą stabilność tych struktur. Natomiast wydłużanie sekwencji pętli do 3 i 4 tymin wyraźnie sprzyja powstawaniu konformacji hybrydowej II (2JSM\_2), konformacji antyrównoległej II (143D\_2) oraz konformacji krzeselkowej (2KM3). Wskazuje to na preferencję pętli bocznych i diagonalnych do dłuższych sekwencji.

Pojawia się pytanie dlaczego przy zmianie długości sekwencji z TGGT<sub>2</sub>GGT<sub>2</sub>GGT<sub>2</sub>GG na TGGT<sub>3</sub>GGT<sub>3</sub>GGT<sub>3</sub>GG następuje wyraźna stabilizacja struktur o topologii antyrównoległej i hybrydowej. W celu odpowiedzenia na to pytanie, przygotowałem modelowe struktury G-kwadrupleksów o niesymetrycznej długości każdej pętli. Wiadomo, że struktury antyrównoległe, czyli posiadające środkową pętlę diagonalną, wymagają sekwencji o długości trzech zasad w tej pętli. Pętłe diagonalne o długości dwóch zasad wprowadzają zbyt silne naprężenia mechaniczne w takiej strukturze G-kwadrupleksu. Okazało się jednak, że wydłużenie jedynie pętli diagonalnej o jedną tyminę dla różnych wersji sekwencji (TGGTGGT<sub>3</sub>GGTGG, TGGTGGT<sub>3</sub>GGT<sub>2</sub>GG, TGGT<sub>2</sub>GGT<sub>3</sub>GGTGG, TGGT<sub>2</sub>GGT<sub>3</sub>GGT<sub>2</sub>GG) nie gwarantuje uzyskania dużej stabilności konformacji antyrównoległej, czyli takiej, jak ma to miejsce w sekwencji z trzema tyminami we wszystkich pętłach. Dodatkowo zauważyłem, że struktury w konformacji antyrównoległej i hybrydowej posiadają szczególnie szeroką wersję bruzdy związanej z pierwszą pętlą od strony 5'-końca. Wywnioskowałem z tego, że taka bruzda wymaga jednak pętli o sekwencji dłuższej niż dwie zasady. Czasy życia uzyskane dla sekwencji TGGT<sub>3</sub>GGT<sub>3</sub>GGT<sub>2</sub>GG, rzeczywiście potwierdzają wysuniętą hipotezę.

Dość niespodziewany efekt uzyskałem dla sekwencji AGGT<sub>2</sub>AGGT<sub>2</sub>AGGT<sub>2</sub>AGG. Posiada ona taką samą długość co sekwencja TGGT<sub>3</sub>GGT<sub>3</sub>GGT<sub>3</sub>GG, ale inny zestaw zasad azotowych w pętłach, charakterystyczny dla sekwencji telomerowych. Użycie takiej sekwencji telomerowej spowodowało zdecydowane wzmocnienie stabilności wszystkich topologii, oprócz równoległej. Oznacza to, że sekwencja ta cechuje się możliwością formowania struktur G-kwadrupleksów o większej różnorodności. Może wynikać to z zaangażowania tymin oraz adenin z pętli w różnego typu oddziaływania pomiędzy sobą.

Moje przewidywania dotyczące potencjalnej możliwości formowania stabilnych dwupłaszczyznowych G-kwadrupleksów dla pewnych symetrycznych sekwencji z pętlami tyminowymi zostały również częściowo potwierdzone przez współpracującą ze mną grupę badawczą z Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk w Poznaniu. Po pierwsze dodatnie pasmo w 265 nm oraz ujemne pasmo w 240 nm na uzyskanych widmach CD wskazują na formowanie G-kwadrupleksów o topologii równoległej (zielona linia na Rys. 5.7C). Pozostałe trzy sekwencje wykazują preferencję do formowania struktur o topologii antyrównoległej, jednakże intensywność odpowiednich pas na widmie CD wskazuje na to, że mogą przyjmować kilka różnych form konformacyjnych. Podobne wnioski można wyciągnąć na podstawie jednowymiarowych widm NMR z Rys. 5.7D, z tą różnicą że jedynie sekwencja z trzema tyminami w pętlach generuje spójne widmo sugerujące występowanie jednej formy G-kwadrupleksu o topologii antyrównoległej. Ma to swoje odzwierciedlenie w wynikach uzyskanych przeze mnie i przedstawionych w formie graficznej na Rys. 5.7B.



RYSUNEK 5.7: A) Sekwencje tyminowe (oprócz telomerowej) zweryfikowane w eksperymentach CD oraz jednowymiarowego NMR. B) Wyniki średnich czasów życia poszczególnych modelowych struktur G-kwadrupleksów. C) Widmo CD dla wybranych sekwencji. D) Jednowymiarowe widmo NMR dla struktur tworzonych przez wybrane sekwencje. Dane przedstawione za zgodą prof. dr hab. Zofii Gdaniec.

### 5.2.3 Podsumowanie

Różnorodność teoretycznych topologii i struktur przyjmowanych przez G-kwadrupleksy w połączeniu z również znaczną liczbą sekwencji potencjalnie tworzących G-kwadrupleksy,

sprawia, że przewidywanie stabilności oraz struktury G-kwadrupeksów na podstawie sekwencji nici DNA jest niemożliwe. Tak wyznaczone cele wymagają opracowania nowych metod symulacyjnych.

Na potrzeby tych badań opracowałem metodę szacowania stabilności modelowych struktur G-kwadrupeksów oraz nową procedurę zwijania G-kwadrupeksów o dowolnej sekwencji i arbitralnej topologii. Obie metody z powodzeniem przetestowałem na wcześniej wyznaczonych strukturach G-kwadrupeksów.

Następnie na podstawie zmierzonego średniego czasu życia, oszacowałem stabilności zestawu modelowych struktur dwupłaszczyznowych G-kwadrupeksów. Zestaw struktur obejmował 6 reprezentatywnych topologii G-kwadrupeksów oraz 14 sekwencji różniących się obszarami pętlowymi. Na podstawie średnich czasów życia poszczególnych wariantów G-kwadrupeksów ustaliłem, że najkrótsza sekwencja TGGTGGTGGTGG preferuje fałdowanie do G-kwadrupeksów o topologii równoległej z pętlami śmigłowymi. Ustaliłem również, że zwinięcie stabilnego G-kwadrupeksu o topologii antyrównoległej z środkową pętlą diagonalną jest możliwe tylko dla sekwencji posiadających w środkowym rejonie pętlowym przynajmniej 3 zasady azotowe. Co więcej podobną preferencję wykazuje pierwszy rejon pętlowy od strony 5'-końca. Ustaliłem również, że sekwencje z 3 i 4 zasadami w pętlach nie formują stabilnych G-kwadrupeksów o konformacji równoległej, ale wykazują stabilność dla G-kwadrupeksów o wielu innych konformacjach. Przewidywania te zostały również częściowo potwierdzone przez współpracującą ze mną grupę badawczą z Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk w Poznaniu.

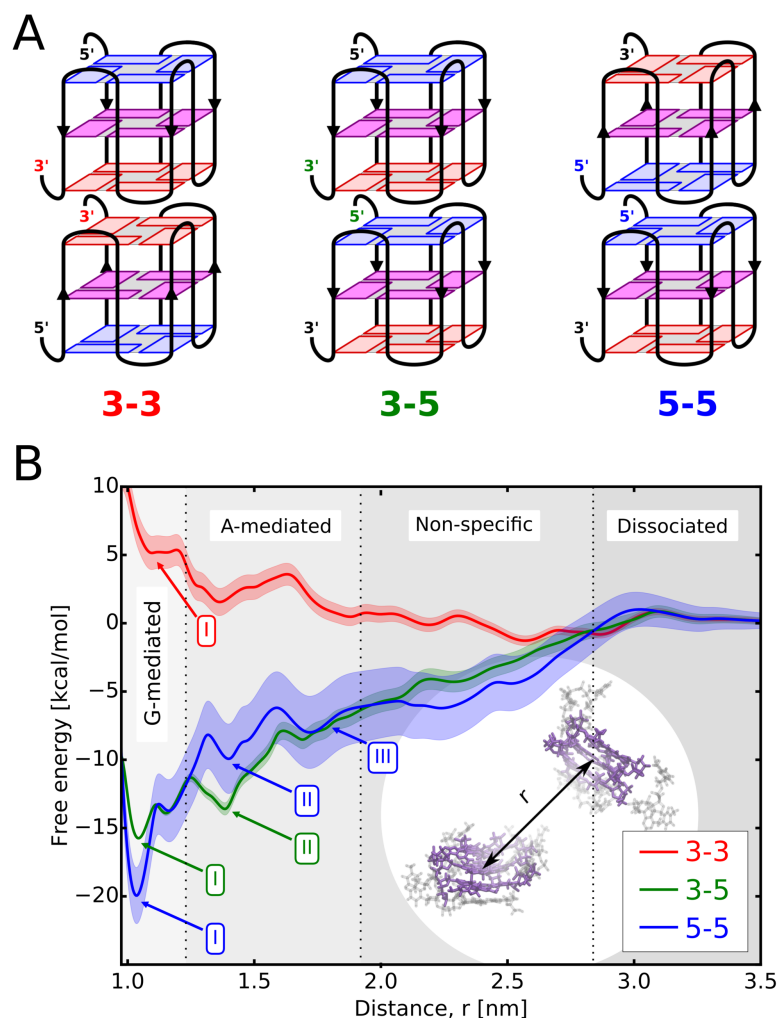
### 5.3 Dimeryzacja G-kwadrupleksów o konformacji równoległej

Wcześniejsze badania przeprowadzone z pomocą symulacji dynamiki molekularnej wykazały, że preformowane układy dimerów i trimerów zbudowane z równoległych G-kwadrupleksów i zasocjowane za pośrednictwem 3' oraz 5'-końcowych G-tetrad są stabilne w skali czasowej 15 ns [216]. Ponadto obliczenia metodą MM-PBSA sugerują zwiększoną stabilność dimerów równoległych G-kwadrupleksów w stosunku do ich stanów monomerycznych [217]. Dzięki obliczeniom chemii kwantowej oraz dynamiki molekularnej w połączeniu z przeszukiwaniem baz strukturalnych zostały również wyznaczone bezwzględne energie tworzenia prostych modeli izolowanych rdzeni guaninowych i różnych potencjalnie możliwych interfejsów agregacji [218]. Jednakże, pomimo tych badań, nie zostały dotąd jednoznacznie scharakteryzowane siły napędowe odpowiedzialne za dimeryzację, a zwłaszcza za silną preferencję do formowania dimeru typu 5'-5' (mod 5'-5'), co zdecydowanie utrudnia przewidywanie i kontrolowanie procesu dimeryzacji [219].

W celu lepszego zrozumienia tych sił napędowych, w niniejszym rozdziale przedstawiłem wyniki badań stabilności dimerów G-kwadrupleksowych. Przede wszystkim wyznaczyłem profile energii swobodnej dla wszystkich trzech możliwych sposobów agregacji G-kwadrupleksów o topologii równoległej (3'-3', 3'-5' i 5'-5'). W swoich badaniach korzystałem z pełnoatomowych symulacji metodą dynamiki molekularnej wraz jawnym modelem rozpuszczalnika. Łączny czas uzyskanych przeze mnie symulacji to  $\sim 100 \mu\text{s}$ . Co najważniejsze, moje obliczenia prawidłowo odtworzyły termodynamiczną preferencję do modu dimeryzacji, w którym agregacja zachodzi poprzez bezpośrednią agregację 5'-końcowych G-tetrad. Wskazałem również na to, że główny mechanizm napędzający dimeryzację jest wynikiem równowagi pomiędzy elektrostatycznym odpychaniem szkieletów fosfocukrowych DNA a korzystnym wiązaniem przeciwjonów w pobliżu interfejsu dimeru. Ponadto pokazałem, że uzyskane informacje na temat czynników rządzących dimeryzacją można wykorzystać do przewidywania, w jaki sposób zmiany stężenia soli lub obecność ligandów wiążących się do G-kwadrupleksów może przesuwac równowagę pomiędzy modami dimeryzacji.

#### 5.3.1 Preferencje G-kwadrupleksów do dimeryzacji

W celu zbadania względnej stabilności dimerów tworzonych przez telomerowe G-kwadrupleksy o topologii równoległej, wyznaczyłem profile energii swobodnej dla trzech głównych modów dimeryzacji zdefiniowanych przez sposób agregacji G-tetrad na 5'-końcu i 3'-końcu: 3'-3', 3'-5', 5'-5' (zobacz Rys. 5.8A). Jako kolektywną współrzędną reakcji adekwatnie opisującą proces dimeryzacji użyłem odległość pomiędzy środkami mas rdzeni guaninowych dwóch asocjujących G-kwadrupleksów (patrz wstawka na Rys. 5.8B). Mapy prawdopodobieństwa orientacji struktur G-kwadrupleksów względem siebie (Rys. B.1 w Załączniku B.1) potwierdzają prawidłową agregację oraz pokazują, że dla większych odległości G-kwadrupleksy



RYSUNEK 5.8: A) Schemat przedstawiający trzy możliwe sposoby dimeryzacji (mody) dwóch G-kwadrupleksów o konformacji równoległej. Dimery powstają poprzez bezpośrednie oddziaływanie G-tetrad od strony końców odpowiednio: 3'-3', 3'-5' oraz 5'-5'. G-tetrad na 5'-końcu jest oznaczona kolorem niebieskim, środkowa kolorem różowym, natomiast G-tetrad na 3'-końcu kolorem czerwonym. B) Profile energii swobodnej dla formowania powyższych dimerów jako funkcja odległości między agregującymi monomerami. Profile wyznaczylem przy użyciu pola siłowego CHARMM36; analogiczne wyniki uzyskane z użyciem Amber-parmbc1 przedstawiłem w Załączniku B.1 na Rys. B.3. Obszary o różnym stopniu szarości reprezentują kolejne etapy procesu dimeryzacji. Ten sam schemat kolorów dla G-tetrad oraz modów dimeryzacji zastosowałem na pozostałych rysunkach.

przyjmują bardzo szeroki zakres różnych względnych orientacji względem siebie. Prawdopodobnie jest to również jeden z czynników korzystnie wpływających na szybką zbieżność profili energii swobodnej dimeryzacji (patrz Rys. B.2).

Profile energii swobodnej ( $G(r)$ ), które przedstawiłem na Rys. 5.8B przedstawiają nieoczekiwane duże różnice w stabilności trzech rozważanych struktur dimerów, przy czym mod 5'-5' jest o około 4 kcal/mol stabilniejszy od modu 3'-5' oraz o około 24 kcal/mol stabilniejszy od modu 3'-3'. Scałkowaniu odpowiednich rozkładów prawdopodobieństwa,  $\rho(r) \propto$

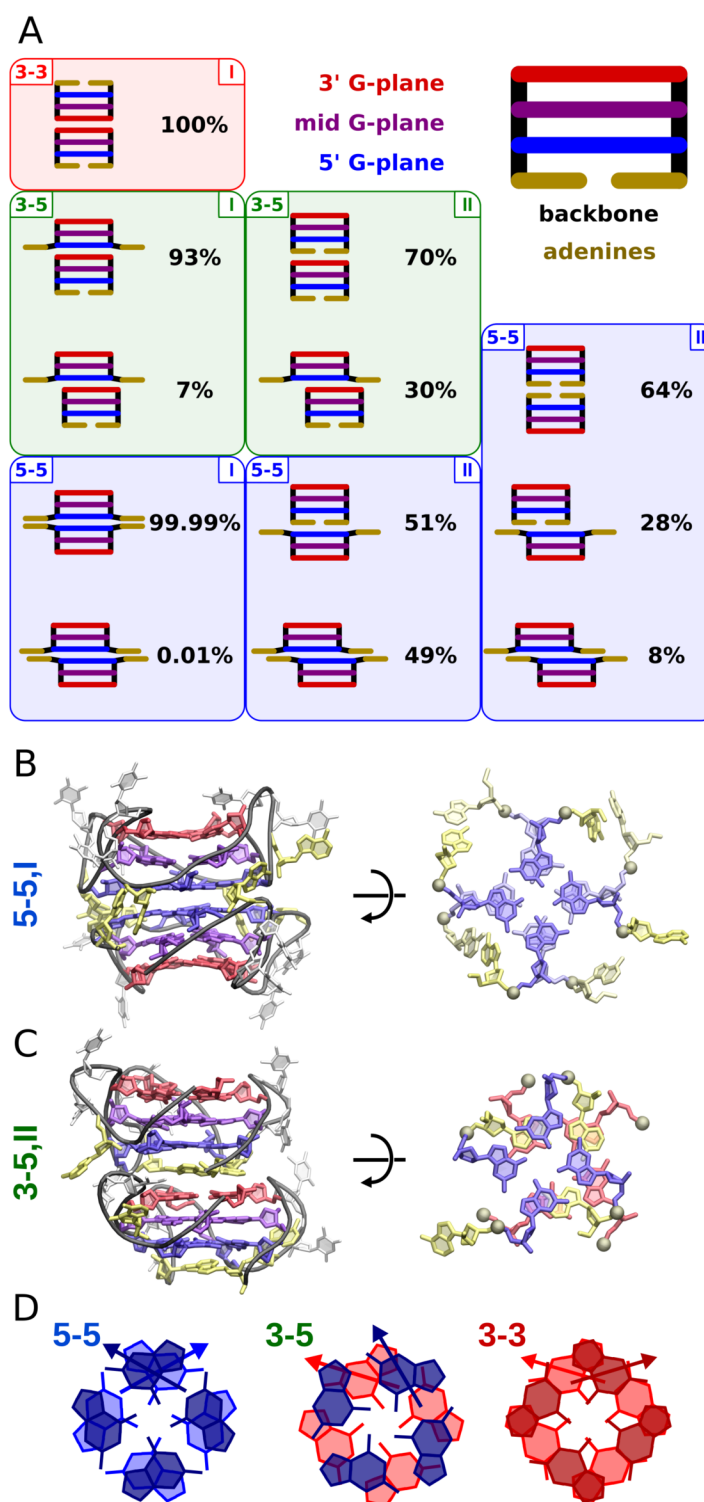


TABELA 5.3: Standardowe energie swobodne wiązania,  $\Delta G_{\text{dim}}^{\circ}$  dla trzech rozważanych modów dimeryzacji obliczone na podstawie profili na Rys. 5.8B, oddzielnie dla stanu G-mediated oraz A-mediated wraz z odpowiadającymi populacjami równowagowymi.

Mod dimeryzacji	$\Delta G_{\text{dim}}^{\circ}$ [kcal/mol]		Populacja [%]	
	G-med.	A-med.	G-med.	A-med.
3-3	$6,4 \pm 1,1$	–	$\sim 0$	–
3-5	$-13,5 \pm 0,1$	$-11,7 \pm 0,3$	0,1	$\sim 0$
5-5	$-17,5 \pm 0,1$	$-8,6 \pm 1,8$	99,9	$\sim 0$

$\exp(-G(r)/kT)$ , wskazuje, że orientacja 5'-5' występuje z 99,9% prawdopodobieństwem, 3'-5' z 0,1% prawdopodobieństwem, podczas gdy mod 3'-3' praktycznie nie występuje (patrz Tabela 5.3). Co ważne, wyraźna preferencja dla modu 5'-5' występująca zarówno w polu siłowym CHARMM36 jak i parmbc1 (Rys. B.3) jest w pełni zgodna z dostępnymi danymi eksperymentalnymi. Stosując spektroskopię NMR, stwierdzono wcześniej, że równoległe G-kwadrupeksy DNA składające się z powtarzających się cyklicznie fragmentów GGGC, GGGT i GGA (kody PDB: 2LE6 i 1MYQ) [131–133], występujące w ludzkich minisatelitach CEB1 (2MB4) oraz bakteryjne G-kwadrupeksy z genu *pile* (2LXV) [134, 135] wykazują silną tendencję do tworzenia dimerów za pośrednictwem interfejsu 5'-5'. Takie same preferencje zaobserwowano dla G-kwadrupeksów dla ludzkiej sekwencji telomerowej w warunkach stłoczenia molekularnego [203] i dla G-kwadrupeksów TERRA [220, 221]. W badaniach krystalograficznych również stwierdzono, że telomerowe G-kwadrupeksy o śmigłowych typach pętli krystalizują preferencyjnie w modzie 5'-5' [61].

Przebieg wyznaczonych profili oraz obecność kilku wyraźnych minimów energii sugeruje występowanie dużej różnorodności struktur dimerycznych, nawet w obrębie jednego modu dimeryzacji. Średnia odległość pomiędzy sąsiednimi G-tetradami w rdzeniu guaninowym wynosi około 0,35 nm, co oznacza, że najgłębsze minimum przy odległości 1,1 nm musi odpowiadać bezpośredniemu związaniu między końcowymi G-tetradami (stany oznaczone jako G-mediated na Rys. 5.8B). Z kolei położenie lokalnych minimów przy 1,4 i 1,7 nm sugeruje, że zasady adeninowe mogą również w pośredniczyć w tworzeniu dimerów (stany A-mediated na Rys. 5.8B). Może odpowiadać za to podobny mechanizm jaki ma miejsce w monomerycznych G-kwadrupeksach telomerowych, gdzie adenin z rejonów pętlowych mają tendencję do wiązania się z 5'-kończącą G-tetradą [179, 222]. Tworzenie mniej stabilnych dimerów typu A-mediated skutkuje obecnością dwóch dodatkowych minimów dla modu 5'-5' (minima II i III) i jednego dla modu 3'-5' (II) (patrz także szczegółowa analiza strukturalna poniżej). Dla odległości w zakresie od 1,9 do 2,9 nm zaobserwowałem różne słabo zdefiniowane i niespecyficzne oddziaływania między dwoma monomerami, a powyżej 2,9 nm G-kwadrupeksy nie są już w kontakcie (co odpowiada plateau na profilu energii swobodnej).



RYSUNEK 5.9: A) Schematyczne przedstawienie względnych ułożeń struktur G-kwadrupleksów w trzech różnych modach dimeryzacji we wszystkich oznaczonych minimach z Rys. 5.8B. Frakcje struktur w poszczególnych minimach uzyskano przez odpowiednie przeważenie zliczeń wynikających z symulacji *Umbrella Sampling*. B) Widok z boku i widok z góry na interfejs dimeru w stanie I w modzie 5'-5'. C) Widok z boku i widok z góry na interfejs dimeru w stanie II w modzie 3'-5'. Struktury reprezentatywne pozostałych stanów zdefiniowanych na Rys. 5.8B przedstawiłem na Rys. B.5. D) Względne ułożenie guanin i ich momentów dipolowych dla 3 rozważanych interfejsów. Ułożenia wyznaczyłem na podstawie przestrzennych map gęstości położenia atomów z guanin (Rys. B.6).

### 5.3.2 Szczegóły struktur dimerów

W celu potwierdzenia charakterystyki minimów występujących w profilach energii swobodnej i potrzeby zbadania tego jak zróżnicowane strukturalnie są różne mody dimeryzacji, przeprowadziłem klastrowanie pełnego wygenerowanego w symulacji zespołu struktur dimerów z sześciu minimów wskazanych na Rys. 5.8B. Następnie przeprowadziłem analizę geometryczną struktur z najliczniejszych uzyskanych klastrów. Klastrowanie przeprowadzone zostało przy użyciu hybrydowego algorytmu (łączącego metody k-centers i k-medoids) na podstawie kryterium odległości RMSD (z ang. Root Mean Square Deviation) obliczonym dla wszystkich ciężkich atomów dwóch rdzeni guaninowych z promieniem odcięcia wynoszącym 0,3 nm [223].

Rys. 5.9A przedstawia schematy reprezentatywnych struktur dimerycznych w poszczególnych minimach wraz z ich procentowymi udziałami w danym stanie. Zgodnie z oczekiwaniami najniższe położone minima energii swobodnej (oznaczone jako I na Rys. 5.9B) odpowiadają dobrze zdefiniowanym strukturalnym dimerom z pełnym nakładaniem się na siebie dwóch G-tetrad tworzących interfejs. Powierzchnia kontaktu zasocjowanych G-tetrad wynosi około  $2,2 \text{ nm}^2$  (Rys. B.4), co przyczynia się do stabilizacji całej struktury. W dominującym stanie modu 5'-5' (5-5, na Rys. 5.9B) względne ułożenie G-tetrad umożliwia maksymalne oddziaływanie pomiędzy pierścieniami pirymidowymi (tj. 6-członowymi) guanin jednej z G-tetrad a zachodzącymi na nie pierścieniami imidazolowymi (5-członowymi) drugiej G-tetrad (Rys. 5.9D oraz Rys. B.6). W takim ułożeniu momenty dipolowe guanin na interfejsie są zorientowane w przybliżeniu prostopadle względem siebie, prowadząc do najkorzystniejszego oddziaływania elektrostatycznego między G-tetradami wśród trzech badanych modów dimeryzacji (Rys. 5.9D). Warto zauważyć, że to samo ułożenie guanin na interfejsie zostało już wcześniej zaobserwowane doświadczalnie dla dimerów typu 5'-5' tworzonych przez równoległe G-kwadrupleksy, zarówno w roztworze wodnym [133], jak i w strukturze krystalicznej [61] (patrz Rys. B.7 dla porównania struktur), co dodatkowo potwierdza moje przewidywania. Na podstawie dalszych analiz można stwierdzić, że dla innych struktur dimerów w stanie G-mediated guaniny na interfejsie prawie wcale nie nakładają się na siebie (3'-5', I), albo kontaktują się tylko poprzez ich pierścienie imidazolowe (3'-3', I), co sugeruje mniej korzystne oddziaływania typu warstwowego (Rys. 5.9D).

Klastrowanie struktur pokazuje również, że w dominującym stanie G-mediated modu 5'-5' dwie struktury G-kwadrupleksów są skręcone względem siebie tak, że ich łańcuchy fosfocukrowe o 4-krotnej symetrii obrotowej tworzą regularny układ z naprzemiennym ułożeniem łańcuchów (Rys. 5.9B). Oznacza to, że są one przesunięte w fazie względem siebie o połowę okresu, tj.  $\sim 45^\circ$  (Rys. B.8), co jednocześnie zapewnia, że wszystkie grupy fosforanowe z jednego monomeru są w maksymalnej możliwej odległości od ich najbliższego sąsiada

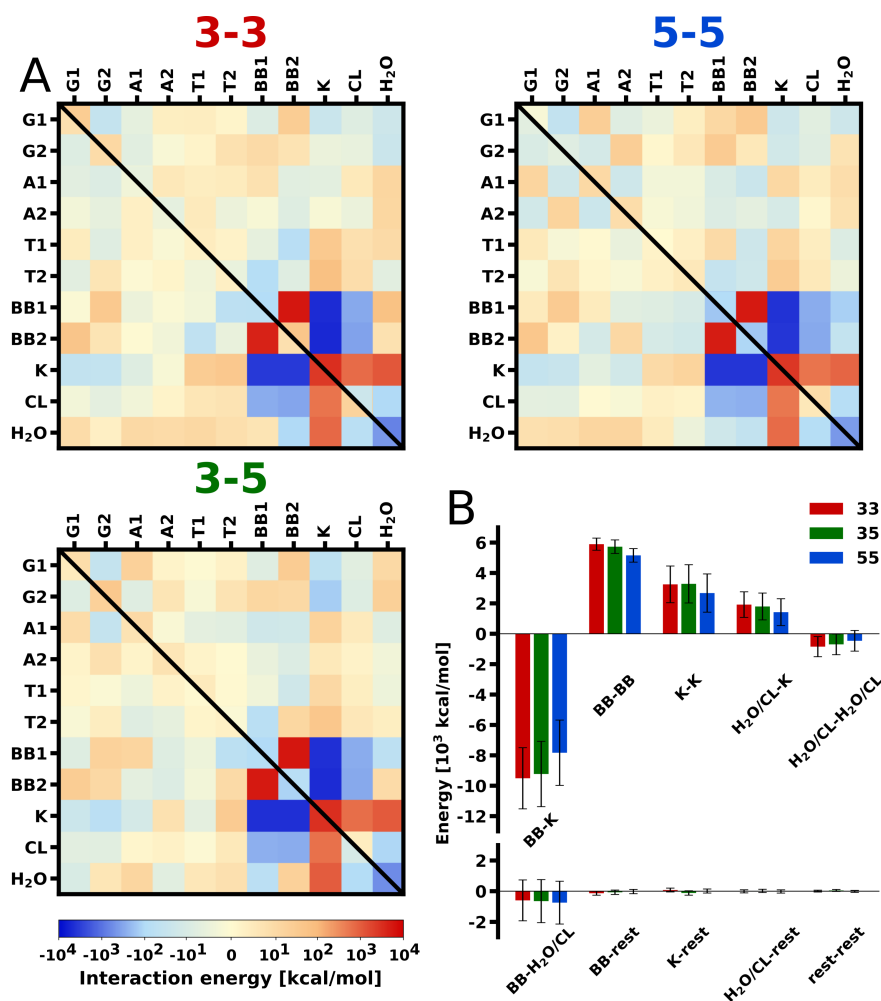
w drugim monomerze Rys. B.9. To charakterystyczne ułożenie monomerów względem siebie w stanie G-mediated przyczynia się prawdopodobnie do minimalizacji odpychania elektrostatycznego między silnie ujemnie naładowanymi łańcuchami DNA (patrz także dalej). Podobną cechę zaobserwowano już wcześniej w strukturze krystalicznej telomerowych G-kwadrupeksów o topologii równoległej [61]. Efekt ten najpewniej sprzyja stabilności sieci krystalicznej. W stanach G-mediated modów 3'-5' oraz 3'-3', względne ułożenie łańcuchów fosfocukrowych są inne i znacznie mniej regularne (Rys. B.8). Powoduje to, że najbliższe grupy fosforanowe z dwóch osobnych nici DNA są bliżej siebie w porównaniu do modu 5'-5' (Rys. B.9). Jest to szczególnie widoczne w przypadku modu 3'-3', gdzie średnio około 3 pary sąsiednich fosforanów są oddalone od siebie o mniej niż 0,6 nm. Jednocześnie nie zaobserwowałem żadnych znaczących zmian w obszarach pętli oraz zmian konformacji pierścienia deoksyrybozy indukowanych dimeryzacją w poszczególnych monomerach G-kwadrupeksach, nie mogą one więc tłumaczyć powyższych różnic (patrz Rys. B.10 i Rys. B.11).

Co ciekawe, struktury dimerów w stanie G-mediated modu 5'-5' przejawiają również dodatkową stabilizację wywołaną przez obecność zewnętrznego pierścienia utworzonego przez adeniny wiążące się w pobliżu interfejsu dimeryzacji (prawy panel na Rys. 5.9B). Analiza powierzchni kontaktu adenin (Rys. B.4) rzeczywiście potwierdza powstawanie od 2 do 3 kontaktów typu adenina-adenina pochodzących z dwóch różnych nici DNA. Ta obserwacja dobrze zgadza się z danymi rentgenowskimi i NMR również wskazującymi na odpowiednio 2 i 3 wiązania adenina-adenina pomiędzy dimeryzującymi G-kwadrupeksami [61, 132].

Pośród wszystkich struktur zidentyfikowanych przez klastrowanie w stanie G-mediated tylko bardzo mały odsetek – odpowiednio 0,01% i 7% w modzie 5'-5' i 3'-5' – wykazuje częściowe nakładanie się na siebie dwóch G-tetrad (ok. 1,68–1,84 nm<sup>2</sup>), które dodatkowo są przesunięte o około 0,41–0,49 nm względem siebie (Rys. B.12). Natomiast pozostałe minima na powierzchni energii swobodnej (odpowiednio II i III na Rys. 5.8B) są znacznie bardziej zróżnicowane strukturalnie. Odpowiadają one głównie takim strukturom, w których jedna (II, Rys. 5.9C) lub dwie (III) warstwy adeninowe z ok. 2–3 adeninami na warstwę pośredniczą w dimeryzacji G-kwadrupeksów (stany A-mediated). Jednakże w obu typach minimum znaczną część (30–49%) struktur posiada powierzchnię kontaktu równą jedynie 0,89–1,84 nm<sup>2</sup> (zobacz Rys. B.4 i Rys. B.12), co sugeruje, że G-tetrazy z interfejsu dimeryzacji zachodzą na siebie tylko częściowo.

### 5.3.3 Siły napędowe dimeryzacji

W celu zidentyfikowania głównych sił napędowych i energetycznych determinantów stabilności dimerów G-kwadrupeksowych, wyznaczyłem całkowitą entalpię dimeryzacji ( $\Delta H$ ) oraz zdekomponowałem ją dla każdego z trzech modów dimeryzacji (zarówno w stanach G-mediated jak i A-mediated) na wkłady pochodzące od poszczególnych elementów strukturalnych układu. Wkłady przedstawione jako macierze oddziaływań na Rys. 5.10A są sumą



RYSUNEK 5.10: A) Poszczególne wkłady do entalpii dimeryzacji ( $\Delta H$ ) wyznaczone jako zmiany energii oddziaływań pomiędzy poszczególnymi elementami strukturalnymi układu: guaniny (G), adeniny (A), tyminy (T), łańcuch fosfocukrowy DNA (BB), kationy potasu (K), aniony chlorkowe (CL) i woda ( $H_2O$ ) w stanie G-mediated (górny trójkąt macierzy) i w stanie A-mediated (dolny trójkąt macierzy); poszczególne monomery G-kwadrupleksów oznaczono jako 1 i 2. B) Zsumowane wkłady do  $\Delta H$  w stanach G-mediated, bez rozróżnienia na poszczególne monomery.

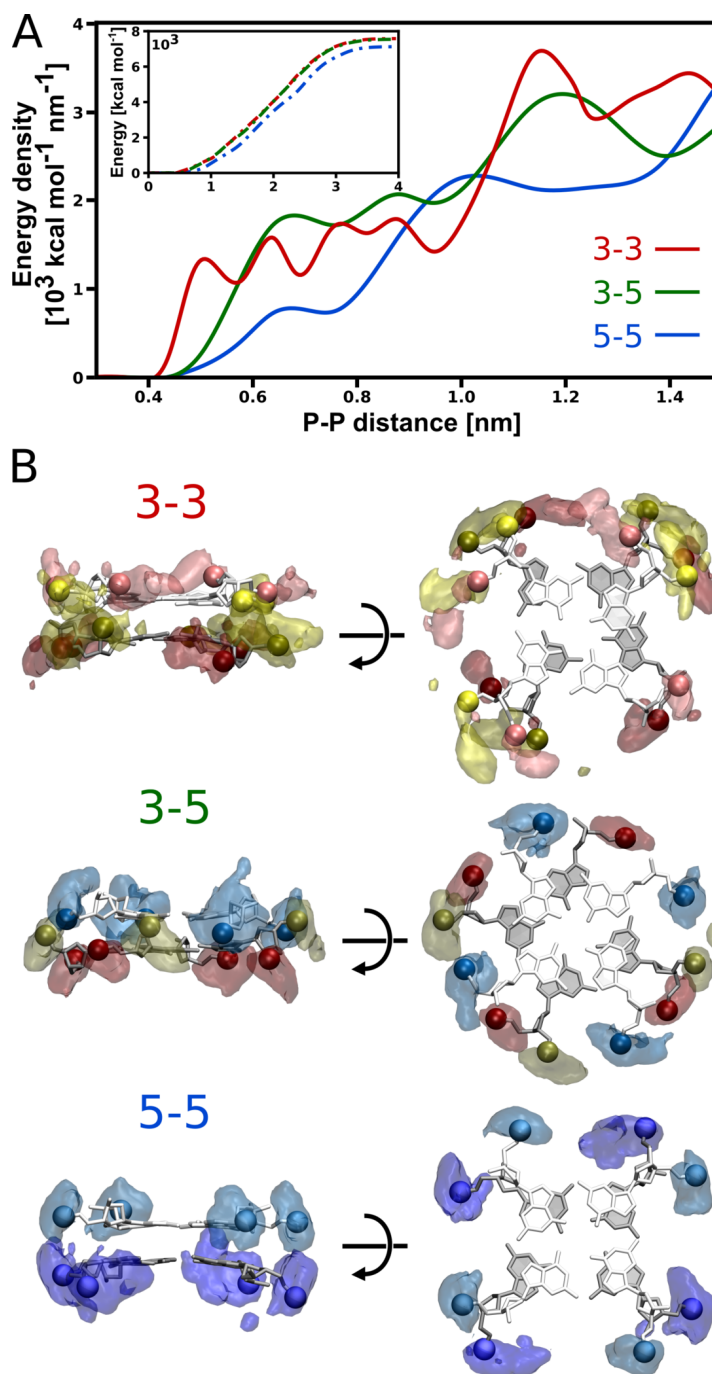
energii elektrostatycznych oraz van der Waalsa i wyznaczyłem je jako średnie różnice entalpii między pomiędzy stanem dimerycznym i monomerycznym.

Duże dodatnie wartości całkowitej entalpii dimeryzacji (w zakresie 185–720 kcal/mol) sugerują, że tworzenie rozważanych dimerów G-kwadrupleksów jest entalpowo wysoce niekorzystne, a zatem musi być ona napędzana entropowo, najprawdopodobniej przy udziale cząsteczek rozpuszczalnika (efekt hydrofobowy). Macierze oddziaływań z Rys. 5.10A pokazują, że dominującą siłą przeciwdziałającą powstawaniu dimerów jest silne odpychanie elektrostatyczne dwóch łańcuchów fosfocukrowych ( $\Delta H_{BB1-BB2} \approx 5,4\text{--}5,9 \times 10^3$  kcal/mol). To odpychanie jest skutecznie ekranowane przez wiązanie przeciwjonów potasowych ( $\Delta H_{K-BB1/BB2} +$

$\Delta H_{K-K} \approx -5,2-6,3 \times 10^3$  kcal/mol), które w stanie zdimeryzowanym mają tendencję do ob-sadzania nowo utworzonych przestrzeni o silnie ujemnym potencjale elektrostatycznym po-między łańcuchami G-kwadrupleksów. Duże wartości bezwzględne tych dwóch wkładów o podobnych wartościach wskazują, że stabilność dimerów jest efektem subtelnej równowa-gi pomiędzy konkurującymi siłami elektrostatycznymi. Ta obserwacja pośrednio tłumaczy wyznaczony eksperymentalnie wpływ stężenia soli na równowagę dimeryzacji G-kwadrup-pleksów [131] oraz potwierdza rolę kationów w stabilizacji i zwijaniu monomerów G-kwa-drupleksów [222]. Poza tym okazuje się, że dimeryzacja nie wpływa znacząco na możliwo-ści wiązania  $K^+$  w centralnych kanałach G-kwadrupleksów, które pozostają w większości zajęte przez przeciwjony. Chociaż zaobserwowałem sporadyczne wiązanie się jonów  $K^+$  w dodatkowym miejscu utworzonym między G-tetradami na interfejsie, efekt ten jest relatyw-nie niewielki, jeśli chodzi o jego wpływ na stabilność dimerów (maks. wartość  $\Delta H_{K-G1/G2} \approx -190$  kcal/mol).

Dość nieoczekiwanie odkryłem, że bardzo podobna równowaga oddziaływań elektrostatycznych oraz rola przeciwjonów w stabilizacji dimerów występuje również w stanach A-mediated (dolne trójkąty na Rys. 5.10A i Rys. B.13), które są scharakteryzowane przez zde-cydowanie większą odległość pomiędzy łańcuchami fosfocukrowymi. Niższa stabilność di-merów w stanach A-mediated (patrz Tabela 5.3) wynika z faktu, że odpychanie między fosfo-ranami nie jest kompensowane przez oddziaływania przyciągające w takim samym stopniu, jak w przypadku stanów G-mediated. W szczególności, jak widać na Rys. B.13, dimery w sta-nach A-mediated nie uzyskują dodatkowej stabilizacji zapewnionej przez oddziaływania  $\pi$  pomiędzy sąsiadującymi G-tetradami ( $\Delta H_{G1-G2} \approx -34-40$  kcal/mol w stanach G-mediated) oraz wynikającej z oddziaływań pomiędzy tyminami a łańcuchem DNA oraz adeninami a łańcuchem DNA ( $\Delta H_{T/A-BB} \approx -10-100$  kcal/mol w stanach G-mediated).

Dekompozycja energii z Rys. 5.10 pokazuje również, że pomimo podobnego wzorca kom-pensacji wkładów entalpowych, trzy możliwe mody dimeryzacji różnią się w szczególach pod względem indywidualnych wkładów do  $\Delta H$ . Z Rys. 5.10B wynika właśnie, że w domi-nującym modzie 5'-5' odpychanie między fosforanami jest zauważalnie słabsze (o około 9%) niż w modach 3'-3' i 3'-5', co może być podstawą preferencji dimeryzacji do formy 5'-5'. Aby zbadać molekularne podstawy różnic w energiach odpychania między fosforanami z sąsied-nich nici DNA, obliczyłem gęstość energii elektrostatycznej oddziaływań między fosforana-mi w funkcji odległości między nimi (Rys. 5.11A oraz Rys. B.14). Małe wartości gęstości energii występujące na małych odległościach pomiędzy fosforanami ( $P-P < 1,4$  nm) w modzie 5'-5' wskazują, że obserwowane w tym przypadku stosunkowo słabe odpychanie elektrostatycz-ne można przypisać niewielkiej liczbie szczególnie niekorzystnych kontaktów między gru-pami fosforanowymi na interfejsie dimeru. Znajduje to również odzwierciedlenie w trójwy-miarowych mapach rozkładu na Rys. 5.11B, które dla modu 5'-5' przedstawiają największą możliwą przestrzenną separację między fosforanami spośród wszystkich modów. Natomiast



RYSUNEK 5.11: A) Gęstość energii elektrostatycznej oddziaływań pomiędzy fosforanami w funkcji odległości między nimi (odległość P-P, pełny zakres odległości jest pokazany na Rys. B.16). Wstawka pokazuje skumulowaną energię odpychania między fosforanami w funkcji odległości P-P. B) Rozkład przestrzenny wszystkich grup fosforanowych na interfejsie w odległości 1 nm do dowolnej grupy fosforanowej drugiego monomeru G-kwadrupleksu. Grupy fosforanowe 3'- i 5'-końcowych G-tetrad są pokazane odpowiednio w kolorze jasno/ciemno czerwonym i niebieskim, podczas gdy pozostałe w kolorze żółtym. Analogiczne wyniki uzyskane z użyciem pola siłowego Amber-parmbsc1 zaprezentowane są na Rys. B.14

rozkłady gęstości energii dla modu 3'-3' i 3'-5' wskazują, że w tych dwóch mniej stabilnych modach dimeryzacji, liczba bliskich kontaktów między fosforanami na interfejsie jest wyraźnie

większa, co jest również przedstawione w przestrzennych rozkładach na Rys. 5.11B. Dodatkowo za pomocą równania Poissona-Boltzmanna [224] obliczyłem potencjały elektrostatyczne wokół poszczególnych monomerów G-kwadrupeksów. Potwierdzają one, że mod 5'-5' charakteryzuje się najsłabszą odpychaniem fosforan-fosforan spośród trzech rozważanych modów dimeryzacji (Rys. B.15).

Rys. 5.10A wskazuje również, że niektóre z mniej dominujących wkładów do  $\Delta H$  również sprzyjają dimeryzacji w modzie 5'-5'. Warto zauważyć, że w tym modzie oddziaływanie pomiędzy G-tetradami,  $\Delta H_{G1-G2}$ , jest o  $\sim 5$  kcal/mol bardziej korzystne niż w modach 3'-3' i 3'-5', co wynika ze wspomnianego wcześniej korzystniejszego względnego ułożenia momentów dipolowych guanin (patrz Rys. 5.9D). Dodatkowa stabilizacja dimeru w modzie 5'-5' jest uzyskiwana przez krzyżowe oddziaływania adenina-adenina ( $\Delta H_{A1-A2} \approx -18$  kcal/mol), które nie mogą występować w modzie 3'-5' oraz 3'-3'.

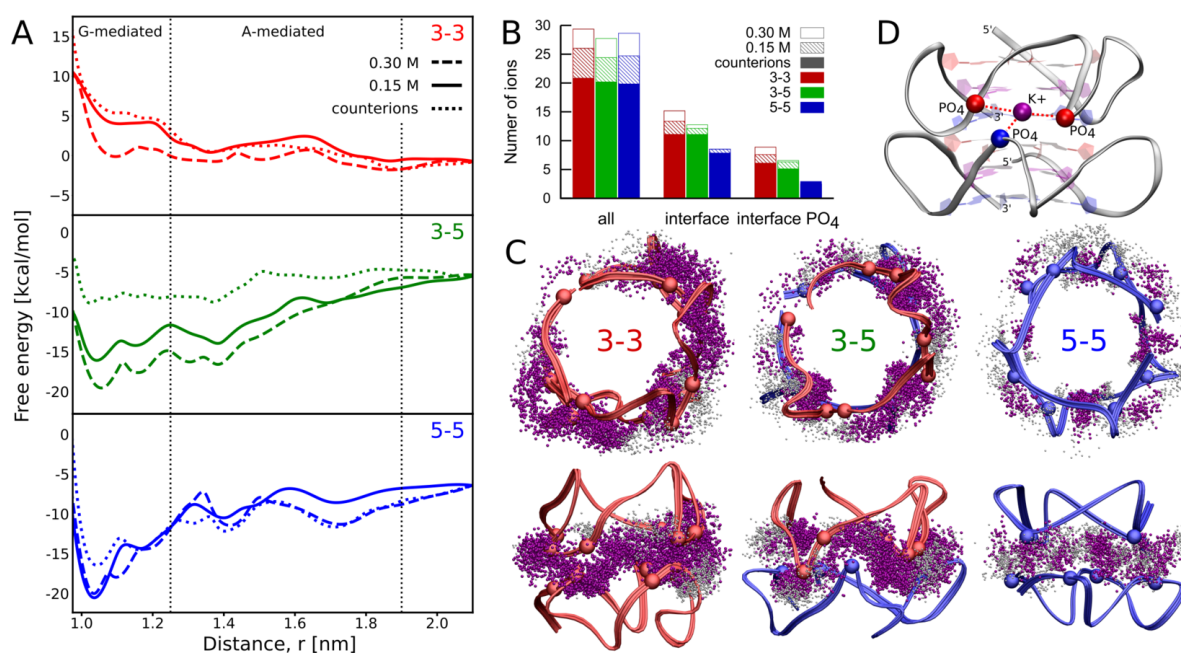
Jak pokazuje wstawka na Rys. 5.11A, całkowita energia odpychania między fosforanami nie różni się istotnie pomiędzy dimerami 3'-3' i 3'-5' ( $\Delta H_{BB1-BB2}$  na Rys. 5.10A oraz  $\Delta H_{BB-BB}$  na Rys. 5.10B). To rodzi pytanie, dlaczego pomimo tego mod 3'-5' wykazuje stabilizację większą o  $\sim 20$  kcal/mol niż mod 3'-3'. Rozwiązanie problemu sugerują gęstości energii na Rys. 5.11A, które pokazują, że asocjacja G-kwadrupeksów za pośrednictwem ich 3'-końcowych G-tetrad prowadzi do szeregu energetycznie niekorzystnych kontaktów pomiędzy grupami fosforanowymi (niezerowa gęstość w odległościach  $< 0,55$  nm). Ze względu na to, że destabilizacja dimeru 3'-3' nie może być w wystarczającym stopniu kompensowana przez oddziaływanie z przeciwnonami, różnica w stabilności pomiędzy modami 3'-3' i 3'-5' może wynikać z odmiennych właściwości tych struktur do wiązania jonów.

### 5.3.4 Modulacja równowagi dimeryzacji

Wyżej przedstawione analizy sugerują, że różnice w sposobie ekranowania elektrostatycznego odpychania grup fosforanowych przez przeciwnony  $K^+$  mogą być odpowiedzialne za znacząco wyższą stabilność dimeru 3'-5' niż dimeru 3'-3'. W celu dokładniejszego zbadania tych różnic, wyznaczyłem profile energii swobodnej tworzenia trzech form dimerycznych w obecności tylko przeciwnonów neutralizujących  $K^+$ , tj. przy efektywnym stężeniu soli wynoszącym około 0,06 M.

Profile energii swobodnej z Rys. 5.12A pokazują, że gdy stężenie  $K^+$  jest około 2,5-rza mniejsze, dimery ulegają destabilizacji o 2–7 kcal/mol w zależności od modu dimeryzacji. W zgodzie z wynikami eksperymentalnymi wskazuje to bezpośrednio na to, że obecność przeciwnonów jest niezbędna dla dimeryzacji G-kwadrupeksów [131]. Większa o 5 kcal/mol destabilizacja modu 3'-5' niż 3'-3' potwierdza, że wiązanie  $K^+$  na interfejsie jest odpowiedzialne za znaczne zwiększenie stabilności dimeru 3'-5'. Poza tym Rys. 5.12B pokazuje, że przy obniżaniu stężenia przeciwnonów faktyczna liczba jonów  $K^+$  związanych z fosforami na interfejsie zmniejsza się bardziej w przypadku 3'-3' (o 1,4 jony) niż w modzie 3'-5' (o 1,1





RYSUNEK 5.12: A) Zależność profili energii swobodnej dla tworzenia dimerów 3'-3', 3'-5' i 5'-5' w różnych stężeniach soli: tylko przeciwny K<sup>+</sup> (linia kropkowana), fizjologiczne stężenie KCl 0,15 M (linia ciągła) i 0,3 M KCl (linia kreskowana). B) Średnia liczba jonów K<sup>+</sup> w kontakcie z interfejsem dimerów: w odległości 0,7 nm od łańcucha DNA dowolnej jednostki G4 (all), w odległości 0,7 nm od obu łańcuchów DNA w tym samym czasie (interface), w odległości 0,7 nm do grupy PO<sub>4</sub> dwóch jednostek G4 w tym samym czasie (interface PO<sub>4</sub>). C) Rozkład przestrzenny jonów K<sup>+</sup> na interfejsie każdego z trzech modów dimeryzacji w stężeniu fizjologicznego KCl, wygenerowany z 1000 klatek symulacji oddalonych w czasie o 0,5 ns. Jony K<sup>+</sup> z interfejsu, które w tym samym czasie są w kontakcie z grupami PO<sub>4</sub> obu jednostek G4, są oznaczone na fioletowo. D) Przykład nowo powstałego miejsca wiązania K<sup>+</sup> utworzonego przez trzy grupy fosforanowe w modzie 3'-5'.

jony). Ze względu na większą destabilizację tego drugiego modu, wynik ten ponownie wskazuje, że wiązanie K<sup>+</sup> w modzie 3'-5' jest energetycznie korzystniejsze.

Dopiero przestrzenne rozkłady jonów K<sup>+</sup> dla trzech rozważanych interfejsów dimerów (Rys. 5.12C oraz Rys. B.17) wyjaśniają źródło różnicy w sile wiązania K<sup>+</sup>. Przede wszystkim w modzie 3'-5' jony są rozmieszczone w sposób niejednorodny oraz w większym stopniu penetrują przestrzeń interfejsu lokując się między monomerami, w konsekwencji efektywniej ekranując oddziaływania pomiędzy fosforanami. Znaczna frakcja tych jonów (około 52%) gromadzi się w 3 wyraźnie wyróżnionych miejscach, będących nowo powstałymi miejscami wiązania jonów. Owe miejsca wiązania tworzą się w wolnej przestrzeni pomiędzy jedną grupą fosforanową z jednego łańcucha DNA oraz dwoma grupami fosforanowymi z drugiego łańcucha DNA (patrzy przykład na Rys. 5.12D). Średnie obsadzenie tych miejsc wynosi niecałe 2 jony K<sup>+</sup> na jedno miejsce. Z drugiej strony mod 3'-3' charakteryzuje się bardziej jednorodnym sposobem wiązania jonów, w których to jony są najbardziej rozproszone po całym interfejsie oraz słabiej penetrują przestrzeń pomiędzy grupami fosforanowymi, co wskazuje na mniej specyficzne i słabsze wiązanie. Powodem dla którego w modzie 3'-3' jony nie wiążą się korzystnie z grupami fosforanowymi jest stosunkowo duża liczba fosforanów na

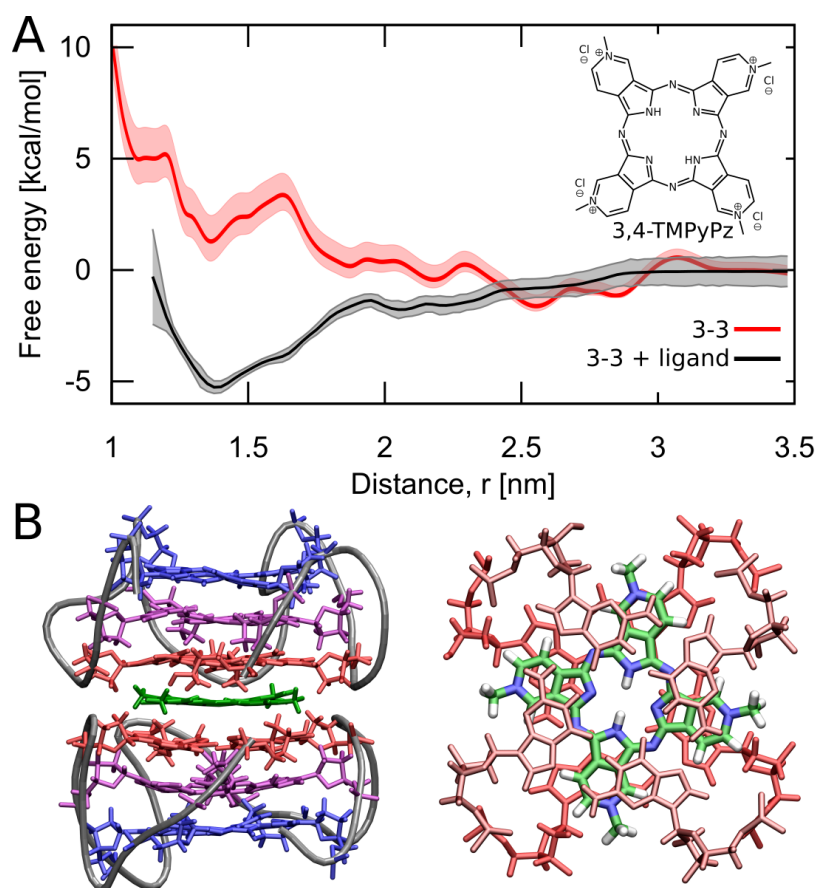
interfejsie znajdujących się blisko siebie, co w konsekwencji prowadzi do ich słabszego ekranowania. Z kolei okazuje się, że różnice w hydratacji trzech interfejsów są mniej wyraźne i tylko w przypadku 3'-3' zaobserwowałem zmniejszoną hydratację fosforanów z interfejsu w stosunku do stanu monomerycznego (Rys. B.18). Prawdopodobnie wynika to ze wspomnianych już wcześniej bezpośrednich oddziaływań fosforanów.

Profile energii swobodnej przy zwiększonym stężeniu przeciwjonów (0,30 M na Rys. 5.12) również potwierdzają wyżej opisane różnice wiązania jonów przez te dwa mniej preferowane mody dimeryzacji. Pomimo wzrostu stabilności modu 3'-5' o 3–4 kcal/mol nie zaobserwowałem znacznego wzrostu liczby jonów  $K^+$  na interfejsie (Rys. 5.12B), co wskazuje na nasycenie miejsc wiązania jonów już przy niższych stężeniach. Z kolei dimer 3'-3' pozostaje niestabilny nawet przy wyższych stężeniach soli, mimo że liczba przeciwjonów niespecyficznie związanych z interfejsem wyraźnie wzrasta.

Tymczasem mod 5'-5' cechuje się zupełnie odmienną charakterystyką wiązania jonów na interfejsie. Ze względu na to, że odległości pomiędzy grupami fosforanowymi na interfejsie 5'-5' w większości są większe niż 0,6 nm, nie tworzą one szczególnie korzystnych miejsc wiązania dla jonów  $K^+$ . W rezultacie liczba jonów obecnych na tym interfejsie jest znacznie niższa niż na dwóch pozostałych, przy czym 28% jonów  $K^+$  znajduje się w ośmiu miejscach wiążących, jednakże o niskim powinowactwie (Rys. 5.12C). Warto zauważyć, że przy niskim stężeniu soli dimer 5'-5' pozostaje wysoce stabilny  $\Delta G = -13,0$  kcal/mol (Rys. 5.12A), co pokazuje, że stosunkowo niewielka liczba jonów  $K^+$  jest wystarczająca do ekranowania odpychania fosforanów między monomerami. Ponadto, zwiększenie liczby przeciwjonów w roztworze nieznacznie zmniejsza preferencję do dimeru 5'-5', a jego największą dominację obserwuje się w niskim stężeniu, gdy jego miejsca wiązania są w zasadzie prawie nasyczone.

W celu przetestowania wniosku wskazującego, że szczególnie niekorzystne odpychanie pomiędzy fosforanami jest odpowiedzialne za destabilizację mniej preferowanych modów dimeryzacji, wyznaczyłem nowy profil energii swobodnej dimeryzacji modu 3'-3' w obecności modelowego ligandu aromatycznego jako elementu zwiększającego odległości między fosforanami. W tym celu wykorzystałem kationową pochodną, 3,4-tetrametylopirydynio-porfirazyne (3,4-TMPyPz na Rys. 5.13), która, jak wykazano, selektywnie wiąże się z ludzkim telomerowym G-kwadrupleksem DNA przez wiązanie na powierzchni G-tetrad [225, 226]. Ze względu na swój kształt, ligand 3,4-TMPyPz powinien dobrze naśladować dodatkową płaszczyznę aromatyczną między dwiema jednostkami G4-kwadrupleksów.

Rys. 5.13 pokazuje, że związek 3,4-TMPyPz znajdujący się pomiędzy agregującymi G-tetradami rzeczywiście zapewnia zwiększoną stabilność dimeru 3'-3'. Wynika to w ze zwiększonej odległości między fosforanami z interfejsu, co w konsekwencji prowadzi do zmniejszenia gęstości energii elektrostatycznej przy małych odległościach P-P. Efekt jest bardziej korzystny w porównaniu do modu 3'-3' bez ligandów lub nawet do modu 3'-5', jednak nie w takim



RYSUNEK 5.13: A) Wpływ ligandu 3,4-TMPyPz na profil energii swobodnej dla tworzenia dimeru 3'-3'. B) Reprezentatywna struktura dimeru 3'-3', w którym ligand 3,4-TMPyPz pośredniczy w tworzeniu G-kwadrupleksu. Widok z boku i góry na interfejs.

samym stopniu jak w najbardziej optymalnym modzie 5'-5' (Rys. B.19). Jednocześnie zaobserwowałem, że liczba jonów  $K^+$  na interfejsie zmniejsza się o 66% w stosunku do wersji bez liganda w modzie 3'-3' (patrz Rys. B.20), co przy nadal nieoptymalnym ułożeniu fosforanów może częściowo wyjaśniać umiarkowany wzrost w stabilności dimeru o  $\sim 6$  kcal/mol.

### 5.3.5 Podsumowanie

W tym rozdziale, stosując obliczenia typu energii swobodnej oparte na dynamice molekularnej, zbadałem molekularne determinanty dimeryzacji dwóch G-kwadrupleksów o topologii równoległej. Pokazałem, że dimery równoległych G-kwadrupleksów powstają prawie wyłącznie (99,9%) przez bezpośrednią agregację 5-końcowych G-tetrad (stan G-mediated w modzie 5'-5'), zgodnie z dostępnymi danymi eksperymentalnymi [61, 131–137]. Jednocześnie moje symulacje wskazują na to, że inne formy dimerów G-kwadrupleksów takie jak mod 3'-5' czy nawet dimery mediowane przez adeniny (stany A-mediated), są również termodynamicznie stabilne w stosunku do stanu zdysocjowanego. Jedynym wyjątkiem jest dimer

powstający poprzez agregację 3'-końcowych G-tetrad (mod 3'-3'), który okazał się niestabilny. Silna preferencja do modu 5'-5', w którym G-tetrazy od strony końców 3' są dostępne dla rozpuszczalnika i nie sprzyjają dalszej agregacji, może być przyczyną niskiej preferencji G-kwadrupeksów do tworzenia oligomerów wyższych rzędów [130, 227, 228]

Moje wyniki wskazują ponadto, że obserwowane preferencje dimeryzacji wynikają z subtelnej równowagi pomiędzy odpychaniem ujemnie naładowanych łańcuchów fosfocukrowych a korzystnym wiązaniem przeciwnonów na interfejsie dimerów G-kwadrupeksów. W szczególności odkryłem, że silna preferencja do modu 5'-5' jest efektem lokalnie optymalnego ułożenia łańcuchów DNA z największymi odległościami pomiędzy fosforanami w poprzek interfejsu. Odkrycie to również wyjaśnia wyniki eksperymentu, w którym dodatkowa grupa fosforanowa przyłączona na 5'-końcu G-kwadrupeksu o topologii równoległej zapobiega dimeryzacji w modzie 5'-5' [229]. W pozostałych dwóch mniej stabilnych modach dimeryzacji, 3'-5' i 3'-3', odpychanie elektrostatyczne między monomerami jest porównywalne i znacznie większe w porównaniu do dominującego dimeru 5'-5'. Stwierdziłem również, że znacznie wyższą stabilność modu 3'-5' można przypisać silniejszemu wiązaniu przeciwnonów w nowo powstałych miejscach wiązania utworzonych przez grupy fosforanowe na interfejsie dimerów. Innym wkładem, który znacząco promuje powstawanie dimeru 5'-5' okazały się oddziaływania pomiędzy 5'-końcowymi G-tetradami, które charakteryzuje prawie maksymalne możliwym nakładanie i sugerowane wcześniej bardzo korzystne ułożenie momentów dipolowych guanin [218].

Dokładna analiza procesów tworzenia dimerów G-kwadrupeksów przedstawiona w moich badaniach na poziomie molekularnym nasuwa potencjalne metody zmiany równowagi dimeryzacji, a tym samym otwiera nowe możliwości projektowania oligomerycznych struktur G-kwadrupeksów. Generalnie koncepcja opiera się na kontrolowaniu elektrostatycznej równowagi odpychania-przyciągania pomiędzy odpowiednimi częściami układu. W szczególności przewiduję, że poprzez podwojenie stężenia przeciwnonów  $K^+$  możliwe jest zwiększenie populację równowagowej dimeru 3'-5' (od 0,1 do ~15%), który jest szczególnie podatny na wiązanie kationów z roztworu wodnego. Co więcej, mój model wskazuje, że stabilność dimerów, w których pośredniczą G-tetrazy, można znacznie zwiększyć poprzez zmniejszenie odpychania pomiędzy fosforanami. Jak sugerują dodatkowe symulacje, zwiększenie odległości separacji między fosforanami na interfejsie poprzez związanie modelowego aromatycznego ligandu między jednostkami G-kwadrupeksów znacznie poprawia stabilność pierwotnie niestabilnego dimeru 3'-3' względem stanu monomerycznego.

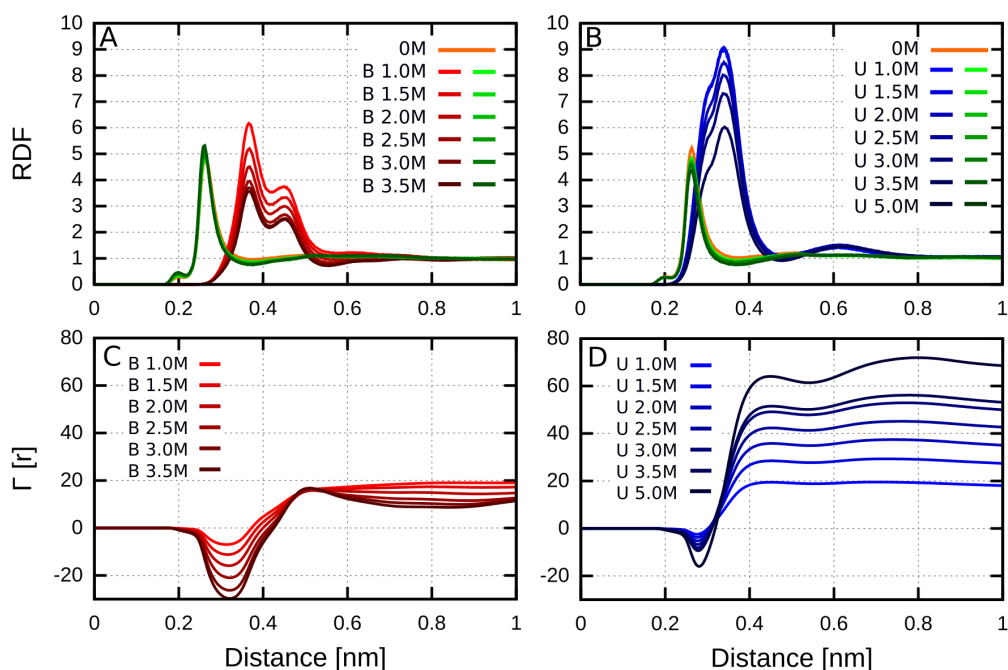
## 5.4 Wpływ denaturantów na stabilność DNA

Osmolity, czyli małowcząsteczkowe związki organiczne gromadzące się w komórkach pod wpływem stresu środowiskowego, mogą modulować stabilność biopolimerów, takich jak białka i DNA. Pomimo wielu lat badań, molekularny mechanizm ich działania zaburzającego lub stabilizacyjnego względem DNA nie został jednoznacznie ustalony. W związku z tym, aby zbadać wpływ osmolitów na DNA na poziomie molekularnym, posłużyłem się metodami symulacji dynamiki molekularnej w celu interpretacji danych spektroskopowych i kalorymetrycznych uzyskanych przez zespół prof. J. Stangreta z Katedry Chemii Fizycznej PG. W założonym celu badałem dwa modelowe osmolity: mocznik (działający denaturująco względem zarówno DNA jak i białek) oraz betainę (trimetyloglicynę, TMG; działająca denaturująco na DNA i stabilizująco na białka).

Uzyskane na Katedrze widma spektroskopowe DNA w zakresie podczerwieni sugerują, że mocznik i betaina odwadniają DNA i przesuwają równowagę konformacyjną od formy B-DNA do formy A-DNA. Efekt ten zachodzi pomimo różnych właściwości solwatacyjnych tych związków. Poniżej przedstawiłem równoważną analizę lokalnych zmian konformacyjnych DNA w różnych roztworach denaturantów. Wyizolowane widma spektroskopowe denaturantów pierwotnie wskazały na słabe specyficzne oddziaływania cząsteczek denaturantów z DNA, w szczególności dla betainy. W związku z tym za pomocą analizy radialnych funkcji rozkładu oraz współczynników preferencyjnej akumulacji postanowiłem zweryfikować tę obserwację. Opisałem również orientację oraz przestrzenne rozmieszczenie cząsteczek denaturantów wokół DNA. Analiza wiązań wodorowych i wyznaczenie energii oddziaływań z natywnym i zdenaturowanym DNA pozwoliła mi na wskazanie silnych korzystnych oddziaływań z zasadami zdysocjowanego DNA oraz zaproponowanie głównych mechanizmów odpowiedzialnych za destabilizację DNA przez mocznik i betainę.

### 5.4.1 Akumulacja denaturantów i wody wokół DNA

Dotychczasowe dane z spektroskopii w zakresie podczerwieni oraz uzyskane, z wykorzystaniem analizy chemometrycznej, widma zaburzonych denaturantów wskazują na brak silnych i specyficznych oddziaływań denaturantów z natywną strukturą DNA. Jest to zaskakująca obserwacja, ponieważ denaturanty na ogół oddziałują korzystnie i bezpośrednio z powierzchnią makrocząsteczek (np. mocznik i guanidyna z białkami). Aby zweryfikować tę obserwację wyznaczyłem radialne funkcje rozkładu (z ang. Radial Distribution Function) opisujące zmiany gęstości cząsteczek wody i denaturantów wokół cząsteczki DNA. RDF wyznaczyłem na podstawie położenia środków masy cząsteczek denaturantu lub cząsteczek wody w funkcji minimalnej odległości od powierzchni DNA. Z uwagi na fakt, że cząsteczka DNA nie ma symetrii sferycznej, oryginalne rozkłady znormalizowałem o czynnik wyznaczony jako



RYSUNEK 5.14: A) i B) Radialne funkcje rozkładu cząsteczek rozpuszczalnika względem DNA dla TMG (linie czerwone w A), mocznika (linie niebieskie w B), cząsteczek wody w roztworze denaturantu (linie zielone) i cząsteczek wody w roztworze bez denaturantu (linia pomarańczowa). Odcienie koloru zgodnie ze zmianami stężenia denaturantów. C) i D) Współczynnik preferencyjnego oddziaływania ( $\Gamma$ ) między cząsteczką denaturantu a DNA, wyznaczony dla wszystkich roztworów betainy i mocznika, w zależności od odległości separacji od powierzchni DNA. Zmniejszające się wartości  $\Gamma$  pokazują, że betaina jest silniej wykluczana, a mocznik gromadzi się na powierzchni DNA wraz ze wzrostem stężenia.

objętość powłoki dostępnej dla rozpuszczalnika w oparciu o metodologię autorstwa grupy Daggett [230, 231].

Funkcje RDF na Rys. 5.14A,B pokazują, że lokalna gęstość TMG i mocznika w pobliżu powierzchni DNA jest wyraźnie zwiększona w stosunku do gęstości w głębi roztworu (czerwona i niebieska linia). Ogólnie względne gęstości mocznika są jednak zdecydowanie wyższe niż betainy, jednak dla obu denaturantów gęstości zależą od stężenia denaturantów i ich względna gęstość zmniejsza się wraz ze wzrostem stężeń (efekt nasycenia). Powłoka solwacyjna utworzona przez cząsteczki betainy jest szersza i przesunięta w kierunku większych odległości niż w przypadku mocznika. Ma to oczywiście związek z rozmiarami obu cząsteczek. Dodatkowo betaina wykazuje zbliżoną względną gęstość cząsteczek do względnej gęstości cząsteczek wody. Dodatkowy pik dla betainy powyżej 0,4 nm sugeruje powstanie drugiej sfery solwacyjnej i możliwość pośrednich oddziaływań betainy z DNA (przez cząsteczki wody).

Funkcje RDF na Rys. 5.14A,B pokazują również, że obecność betainy nie zmienia względnej względnej gęstości cząsteczek wody w pobliżu powierzchni DNA (zielone linie). Jedynie wodne roztwory mocznika wykazują nieznaczny wpływ na względną gęstość cząsteczek wody wokół powierzchni DNA. Już przy niskich stężeniach mocznika względna gęstość wody

jest zredukowana, a dalsze zwiększenie stężenia mocznika systematycznie zmniejsza efektywną hydratację powierzchni DNA.

Następnie postanowiłem zidentyfikować molekularne źródła eksperymentalnie obserwowanej destabilizacji DNA w wyniku działania mocznika i betainy. Powszechnie uważa się, że stabilizujący lub destabilizujący wpływ osmolitów na strukturę DNA wynika z ich bezpośrednich, korzystnych oddziaływań z powierzchnią DNA lub pośrednich oddziaływań z DNA. W związku z tym, zbadałem czy mocznik i betaina są preferencyjnie wykluczone lub akumulują się przy powierzchni DNA (bezpośrednio oddziałują) oraz jak te preferencje zmieniają się wraz ze stężeniem roztworu. W tym celu obliczyłem współczynnik preferencyjnego oddziaływania  $\Gamma(r)$ , którego definicja znajduje się w Załączniku C.1.

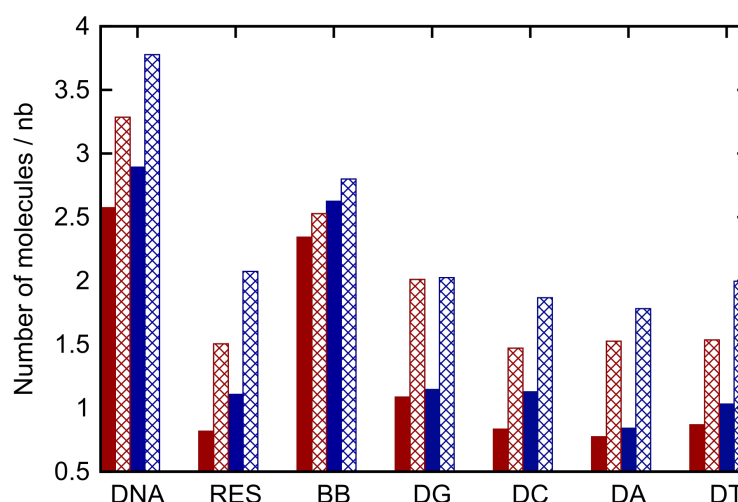
Rys. 5.14C pokazuje ujemne wartości  $\Gamma(r)$  w pierwszej sferze solwatacyjnej ( $<0,4$  nm) dla betainy, co przy niezależnej względnej gęstości wody od stężenia, wskazuje na efektywne wykluczanie cząsteczek tego denaturantu z powierzchni DNA. Największe wykluczenie TMG występuje w bezpośrednim sąsiedztwie DNA ( $\Gamma = -25$  dla  $r = 0,34$  nm przy 3,5 M). Można również zauważyć, że liczba cząsteczek betainy wykluczonych z powierzchni DNA wzrasta wraz ze wzrostem stężenia. Dopiero dla odległości większych niż 0,4 nm występuje niewielka akumulacja betainy związana z tworzoną przez nią drugą sferą solwatacyjną.

W przypadku wszystkich roztworów mocznika na Rys. 5.14D można zauważyć, że wartość  $\Gamma$  jest dodatnia w prawie całym rozważanym zakresie odległości, co wskazuje na preferencyjną akumulację denaturantu na powierzchni DNA, co jest typowym zachowaniem dla denaturantów. Dodatkowo widać, że preferencja mocznika do oddziaływania z powierzchnią DNA znacznie wzrasta wraz ze wzrostem jego stężenia, co zgadza się z wynikami badań prowadzonymi nad innymi kwasami rybonukleinowymi [232].

#### 5.4.2 Miejsca oddziaływań denaturantów w obrębie cząsteczki DNA

Aby określić miejsca w obrębie struktury DNA, które są szczególnie preferowane przez cząsteczki denaturantów, wyznaczyłem liczbę cząsteczek denaturantów oddziałujących w stanie natywnym i zdysocjowanym z wybranymi elementami strukturalnymi cząsteczki DNA. Warto zaznaczyć, że całkowita liczba cząsteczek denaturantów wokół DNA nie jest sumą dla poszczególnych elementów strukturalnych, ponieważ sfery solwatacyjne poszczególnych elementów mogą się pokrywać.

Rys. 5.15 pokazuje, że pierwsza sfera solwatacyjna DNA w stanie natywnym i zdysocjowanym zawiera więcej cząsteczek mocznika w przeliczeniu na jedną zasadę azotową niż cząsteczek betainy pomimo tego samego stężenia roztworu i różnych grubości pierwszej sfery solwatacyjnej. Dotyczy to również pierwszej sfery solwatacyjnej oddzielnie zasad azotowych (RES) i szkieletu DNA (BB). Dekompozycja strukturalna pokazuje również, że akumulacja betainy i mocznika na powierzchni DNA wynika głównie z dużo większej liczby cząsteczek

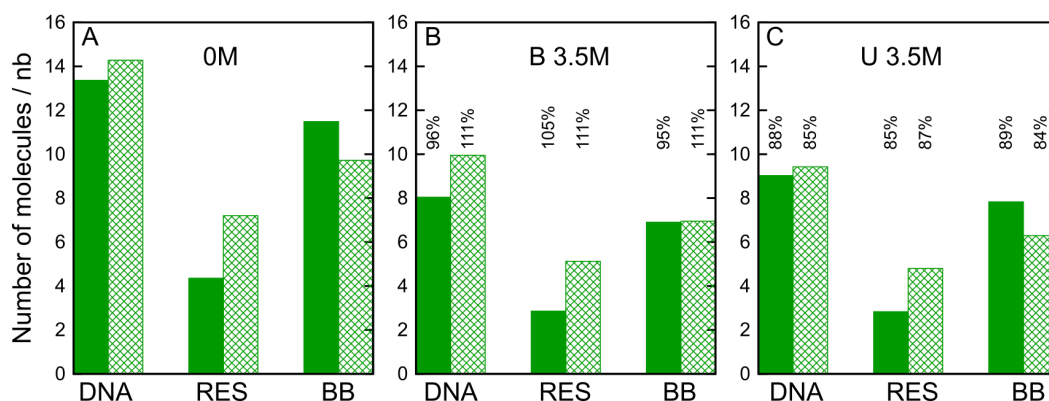


RYSUNEK 5.15: Liczba cząsteczek TMG (na czerwono) i mocznika (na niebiesko) w przeliczeniu na jeden nukleotyd w pierwszej sferze solwatacyjnej w stanie natywnym (słupki pełne) i stanie zdysocjowanym (słupki przerywane) dla całej struktury DNA (DNA), wszystkich zasad azotowych (RES), szkieletu DNA (BB), reszt guanin (DG), reszt cytozyn (DC), reszt adenin (DA) i reszt tymin (DT). Dane tylko dla 3,5 M roztworów TMG i mocznika. Wartość promienia pierwszej sfery solwatacyjnej została ustalona na podstawie radialnych funkcji rozkładu rozpuszczalników na Rys. 5.14.

denaturatów wokół szkieletu DNA (75% wszystkich cząsteczek TMG i 70% wszystkich cząsteczek mocznika). Co istotne z punktu widzenia mechanizmu destabilizacji, podczas denaturacji DNA liczba cząsteczek betainy i mocznika w pierwszej sferze solwatacyjnej zasad azotowych wzrasta odpowiednio o 185% w przypadku betainy i o 175% w przypadku mocznika. Z drugiej strony wzrost liczby cząsteczek denaturantów dla szkieletu DNA jest znacznie mniejszy (6–8%), co wskazuje na maksymalną jego solwatację już w stanie natywnym.

Następnie sprawdziłem preferencję denaturantów do oddziaływania z poszczególnymi zasadami azotowymi. Rys. 5.15 przedstawia liczbę cząsteczek denaturantów na jedną zasadę azotową (DG, DC, DA i DT) w pierwszej sferze solwatacyjnej. W stanie natywnym i zdenaturowanym cząsteczki mocznika wykazują większą akumulację dla wszystkich rodzajów zasad azotowych niż cząsteczki betainy i jednocześnie wykazują bardziej równomierną preferencję, szczególnie w stanie zdenaturowanym. Wskazuje to na faworyzowanie stanów zdenaturowanych DNA poprzez silne, ale niespecyficzne oddziaływania z eksponowanymi zasadami DNA. W przypadku betainy wyraźnie widoczna jest preferencja do guanin, która wzmacnia się w stanie zdenaturowanym DNA. Potwierdza to obserwowaną eksperymentalnie zdolność betainy do obniżania stabilności fragmentów DNA bogatych w pary guanina–cytozyna [38, 233].





RYSUNEK 5.16: Liczba cząsteczek wody na jeden nukleotyd w pierwszej sferze solwatacyjnej w stanie natywnym (pełne słupki) i stanie zdenaturowanym (przerwane słupki) dla całej struktury DNA (DNA), zasad nukleinowych (RES) i szkieletu DNA (BB). A) Dane dla roztworu wodnego bez denaturantów, B) 3,5 M roztwór TMG i C) 3,5 M roztwór mocznika. Ułamek procentowy reprezentuje stosunek między obserwowaną a oczekiwaną liczbą cząsteczek wody w pierwszej sferze hydratacyjnej (szczegóły w Załączniku C.1). Wartość promienia pierwszej sfery solwatacyjnej została ustalona na podstawie radialnych funkcji rozkładu rozpuszczalników na Rys. 5.14.

### 5.4.3 Wpływ denaturantów na rozkład wody wokół DNA

Sugeruje się, że osmolyty mogą zmieniać sfery hydratacyjne makrocząsteczek poprzez zmianę struktury wody, a tym samym lokalną gęstość wody. Jednak wpływ denaturantów na gęstość wody wokół powierzchni DNA nie jest dobrze znany. W związku z tym wyznaczyłem liczbę cząsteczek wody na jedną zasadę azotową w pierwszej sferze solwatacyjnej w stanie natywnym i zdysocjowanym. DNA podzieliłem na następujące grupy: całą cząsteczkę DNA, wszystkie zasady azotowe oraz szkielet DNA. Ponadto określiłem procentowy stosunek między obserwowaną a oczekiwaną liczbą cząsteczek wody (wynikający z lokalnych gęstości) w pierwszej sferze solwatacyjnej (szczegóły obliczeń w Załączniku C.1).

Rys. 5.16A pokazuje, że w wodnym roztworze bez denaturantów (0M) to właśnie szkielet DNA jest najsilniej hydratowanym elementem struktury DNA. Jak można oczekiwać, pod wpływem denaturacji liczba cząsteczek wody w pierwszej sferze solwatacyjnej dla całej struktury DNA wzrasta, głównie na skutek zwiększenia powierzchni eksponowanej do rozpuszczalnika zasad azotowych i w konsekwencji zwiększonej ich hydratacji. Jednocześnie obserwuje się częściową kompensację tego efektu poprzez zmniejszenie hydratacji szkieletu DNA, czego przyczyną jest bardziej zwarta struktura łańcucha DNA.

Równie spodziewanym efektem jest zmniejszenie bezwzględnej liczby cząsteczek wody wokół cząsteczki DNA w stanie natywnym po dodaniu denaturantów, odpowiednio o około 40% dla betainy i około 30% dla mocznika. Pomimo użycia tych samych stężeń denaturantów obserwuję jedynie nieznaczne zmniejszenie efektywności hydratacji DNA w roztworze betainy, w postaci 96% oczekiwanej liczby cząsteczek wody niż ma to miejsce w roztworze mocznika (88%). Sugeruje to, że betaina pomimo tego, że wypiera znaczną liczbę cząsteczek

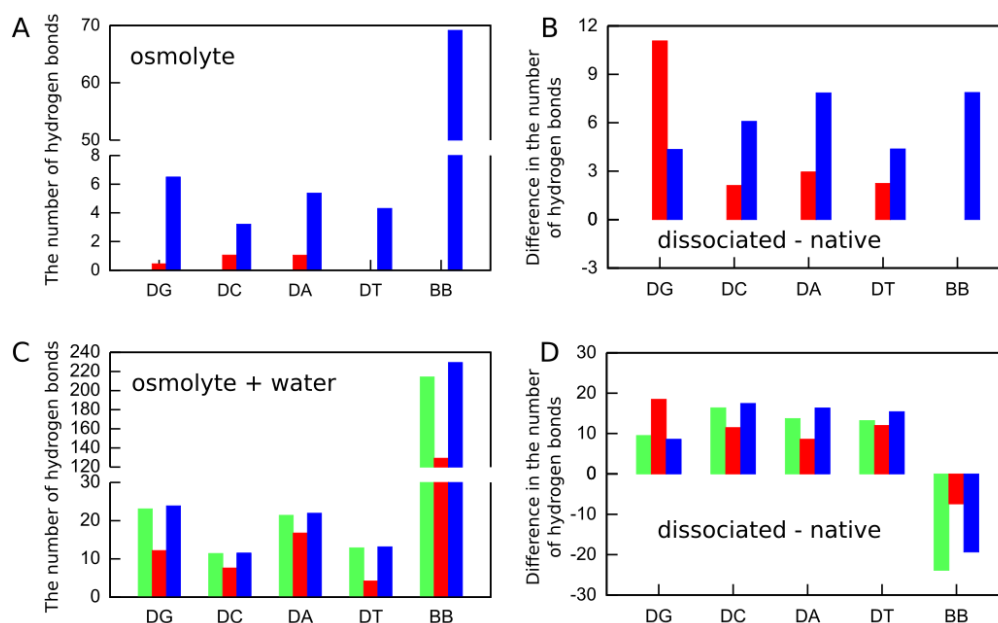
wody to nie zaburza jej struktury i gęstości. Z kolei podczas denaturacji liczba cząsteczek wody wokół DNA nie wzrasta w ten sam sposób dla obu denaturantów. W roztworze betainy widać znaczny wzrost obserwowanej liczby cząsteczek wody wokół DNA o 25% (do 111% oczekiwanej liczby cząsteczek wody), podczas gdy w roztworze mocznika jedynie o 5% (do 85% oczekiwanej liczby cząsteczek wody). Oznacza to, że w roztworze betainy DNA w stanie zdenaturowanym jest lepiej hydratowane niż wynikałoby to z lokalnych stężeń roztworu. Co ciekawe efekt ten jest głównie wynikiem zwiększenia liczby cząsteczek wody wokół zasad azotowych. W roztworach mocznika pomimo zwiększenia liczby cząsteczek wody wokół całej cząsteczki DNA sytuacja nie zmienia się i obserwowana liczba cząsteczek wody to tylko 85% oczekiwanej liczby cząsteczek wody. Wskazuje to na zdecydowanie gorsze właściwości hydratacyjne wody w roztworze mocznika.

#### 5.4.4 Wiązania wodorowe między DNA a roztworami denaturantów

Wiedza o akumulacji cząsteczek wokół powierzchni DNA nie jest wystarczająca do wyznaczenia względnej efektywności solwatacji przez roztwory osmolitów w porównaniu z roztworem wodnym pozbawionym osmolitu. W związku z tym po pierwsze wyznaczyłem liczbę wiązań wodorowych utworzonych pomiędzy wybranymi elementami strukturalnymi DNA a cząsteczkami mocznika i betainy (Rys. 5.17A) lub wszystkimi cząsteczkami roztworu (woda + denaturant) (Rys. 5.17C) w 3.5 M roztworze mocznika i 3 M roztworze betainy. Wyznaczyłem również zmiany w liczbie wiązań wodorowych podczas denaturacji DNA w tych samych roztworach (Rys. 5.17B,D).

Rys. 5.17A pokazuje, że w natywnym stanie DNA cząsteczki betainy w ogólności nie tworzą wielu wiązań wodorowych z z zasadami azotowymi DNA, co jest potwierdzeniem wcześniejszych analiz preferencji i akumulacji. Co więcej, nie zaobserwowałem żadnych wiązań wodorowych z resztami tyminowymi i, czego należało oczekiwać, ze szkieletem DNA. W przeciwieństwie do betainy, mocznik tworzy wiele wiązań wodorowych z całą powierzchnią DNA, a zwłaszcza ze szkieletem fosfocukrowym. Sugeruje to, że jego właściwości denaturujące są powiązane z możliwością tworzenia wiązań wodorowych.

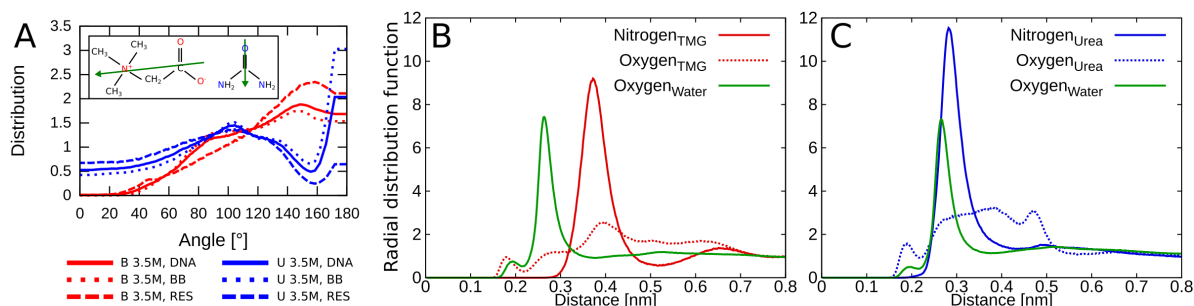
Rys. 5.17B pokazuje, że denaturacja DNA w przypadku obu roztworów znacznie zwiększa liczbę wiązań wodorowych pomiędzy cząsteczkami denaturantów a DNA. Jednak największy wzrost liczby nowych wiązań wodorowych obserwuje się pomiędzy cząsteczkami betainy a resztami guaniny. Jest to ponownie spójne z obserwowanym eksperymentalnie efektem obniżania stabilności par guanina–cytozyna w roztworach betainy. W przypadku innych zasad azotowych wzrost jest zdecydowanie mniejszy i bardziej jednorodny. Cząsteczki mocznika w stanie zdenaturowanym lepiej oddziałują z zasadami azotowymi tworząc w sumie więcej nowych wiązań wodorowych z nimi niż ze szkieletem DNA. Główny mechanizm denaturacji DNA przez mocznik polega zatem prawdopodobnie na akumulowaniu się i



RYSUNEK 5.17: A) Liczba wiązań wodorowych tworzonych przez cząsteczkę denaturantu z poszczególnymi typami zasad azotowych DNA (DG, DC, DA, DT) i ze szkieletem DNA (BB) w natywnym stanie DNA. B) Różnica w liczbie wiązań wodorowych między denaturantami a poszczególnymi typami zasad azotowych DNA (DG, DC, DA, DT) i ze szkieletem DNA (BB) odpowiadająca denaturacji (różnica między stanem zdysocjowanym a stanem natywnym). C) Całkowita liczba wiązań wodorowych pomiędzy cząsteczkami roztworu (woda + denaturant) z poszczególnymi typami zasad azotowych DNA (DG, DC, DA, DT) i ze szkieletem DNA (BB) w natywnym stanie DNA. D) Różnica między całkowitą liczbą wiązań wodorowych w stanie zdysocjowanym a natywnym. Dane dla roztworu wodnego bez denaturantów (zielone słupki), 3,5 M roztworu TMG (czerwone słupki) i 3,5 M roztworu mocznika (niebieskie słupki).

preferencyjnym tworzeniu wiązań wodorowych z zasadami, w efekcie faworyzując i wzmacniając stan zdenaturowany.

Rys. 5.17C pokazuje, że najlepiej solwatowanym elementem struktury DNA w stanie natywnym jest jego szkielet fosfocukrowy i fakt ten jest niezależny od rodzaju roztworu. Obecność cząsteczek betainy w roztworze powoduje znaczne zmniejszenie całkowitej liczby wiązań wodorowych pomiędzy DNA a roztworem. Prowadzi to do osłabienia sieci wiązań wodorowych między powierzchnią DNA a roztworem i ogólną destabilizację stanu natywnego DNA. Zmniejszenie solwatacji DNA przez betainę obserwuje się szczególnie w przypadku guanin, tymin oraz szkieletu DNA. W przypadku mocznika roztwór utrzymuje ten sam poziom solwatacji, a nawet nieco wyższy dla szkieletu DNA. Sugeruje to, że podwójna helisa DNA oddziałuje z mocznikiem tak samo korzystnie jak z wodą. Z kolei Rys. 5.17D pokazuje, że denaturacja cząsteczki DNA wiąże się ze znacznym spadkiem solwatacji szkieletu DNA dla obu roztworów, przy czym w obecności betainy spadek ten jest najmniejszy (o 5%), a dla mocznika i roztworu bez denaturantów dochodzi do 10%. Podczas denaturacji liczba nowych wiązań wodorowych tworzonych przez mocznik jest większa prawie dla wszystkich zasad azotowych niż w roztworze wodnym bez denaturantu. Ponownie potwierdza to stwierdzoną wcześniej preferencję mocznika do stanu zdenaturowanego DNA. Podczas denaturacji DNA



RYSUNEK 5.18: A) Rozkład prawdopodobieństwa kąta między osiami cząsteczek denaturantów a normalną do powierzchni cząsteczki DNA (DNA, linia ciągła), szkieletu DNA (BB, linia kropkowana) i zasad nukleinowych (RES, linia przerywana). Wstawka pokazuje atomy opisujące osie molekularne cząsteczek: atomy azotu i tlenu dla cząsteczek betainy (kolor czerwony) oraz atomy tlenu i azotu dla cząsteczek mocznika (kolor niebieski). Osie cząsteczek są przedstawione jako zielony wektor. B) i C) Radialne funkcje rozkładu dla atomów azotu betainy (czerwona linia ciągła) i cząsteczek mocznika (niebieska linia ciągła), atomów tlenu betainy (czerwona linia przerywana) i cząsteczek mocznika (niebieska linia przerywana) oraz atomów tlenu w cząsteczkach wody (zielona linia).

w obecności mocznika jedynie solwatacja guaniny pozostaje na podobnym poziomie. Roztwór betainy przejawia przeciwne zachowanie. Różnice w solwatacji zasad azotowych DNA pomiędzy stanem zdenaturowanym a natywnym dla betainy są mniejsze niż w pozostałych roztworach. Pod tym względem wyróżniają się jedynie guaniny, które tworzą wiele wiązań wodorowych z cząsteczkami betainy.

#### 5.4.5 Rozkład przestrzenny cząsteczek denaturantów wokół DNA

W celu opisu orientacji cząsteczek denaturantów w stosunku do powierzchni DNA, wyznaczyłem rozkład prawdopodobieństwa kątów między osiami cząsteczek betainy i mocznika a normalną do powierzchni DNA. Oś cząsteczki TMG została zdefiniowana przez położenie atomu azotu i środek masy atomów tlenu (patrz wstawka na Rys. 5.18A). Z kolei oś cząsteczki mocznika została zdefiniowana przez atom tlenu i środek masy atomów azotu. Taki wybór osi odzwierciedla również orientację wektorów momentów dipolowych obu cząsteczek.

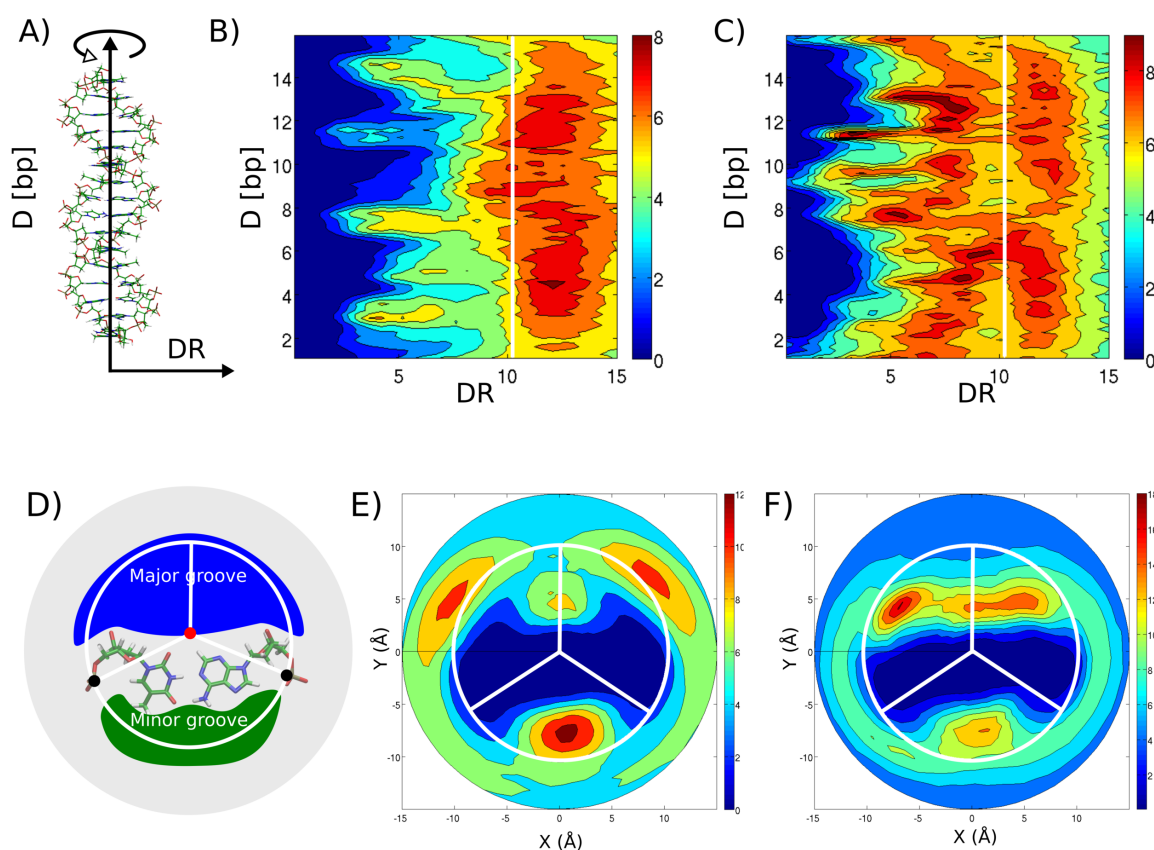
Rys. 5.18A pokazuje, że cząsteczki betainy są zorientowane w kierunku powierzchni DNA przez dodatnio naładowany fragment (tj. grupę trimetyloamoniową), natomiast ułożenie cząsteczek mocznika jest bardziej zróżnicowane. Wynik ten sugeruje, że cząsteczki betainy nie wykazują dużego potencjału do tworzenia wiązań wodorowych z DNA (przez grupy metylowe). Cząsteczki mocznika są głównie zorientowane w kierunku DNA przez grupy aminowe, ale pewna część cząsteczek ma również orientację równoległą w stosunku do powierzchni DNA (wysokie prawdopodobieństwo dla wartości kątowych bliskich  $90^\circ$ ). Istnieją również pewne różnice w ułożeniu cząsteczek denaturantów względem jedynie zasad azotowych oraz szkieletu DNA. Cząsteczki betainy wykazują jedno określone ułożenie w odniesieniu do zasad azotowych (linie przerywane na Rys. 5.18A) niż szkieletu DNA (linie kropkowane),

podczas gdy cząsteczki mocznika preferują prostopadłe ułożenie względem łańcucha DNA oraz równoległe względem powierzchni zasad azotowych.

Dla potwierdzenia tych efektów wyznaczyłem również radialne funkcje rozkładu wybranych atomów z cząsteczek denaturantów względem powierzchni DNA. Jak widać z Rys. 5.18 B,C obie cząsteczki denaturantów wykazują preferencję do oddziaływania z DNA poprzez fragmenty strukturalne z atomami azotu, co potwierdza wcześniej wyznaczone względne ułożenie tych fragmentów cząsteczek względem DNA. Teoretycznie oba denaturanty mają możliwość bezpośredniego oddziaływania z powierzchnią DNA i tworzenia wiązań wodorowych przez atomy tlenu. Natomiast radialne funkcje rozkładu na Rys. 5.18B,C potwierdzają, że mocznik wyraźnie silniej oddziałuje z DNA za pośrednictwem grupy karbonylowej niż betaina, która posiada nawet takie dwa centra oddziaływań. Poza tym mocznik wykazuje zwiększoną preferencję do oddziaływania z cząsteczką DNA poprzez atomy tlenu już na małych odległościach (w zakresie od 0,25 nm do 0,5 nm), natomiast betaina wykazuje nieznaczną preferencję do takich oddziaływań jedynie dla odległości równej około 0,42 nm. Ta obserwacja sugeruje, że część cząsteczek betainy jest wykluczana z powierzchni DNA i potencjalnie bierze udział w pośrednich oddziaływaniach z DNA przez cząsteczki wody. Te charakterystyczne sposoby ułożenia cząsteczek denaturantów mogą być częściowo odpowiedzialne za zaobserwowane eksperymentalnie nieliniowe zależności temperatury topnienia mierzone w kalorymetrii DSC.

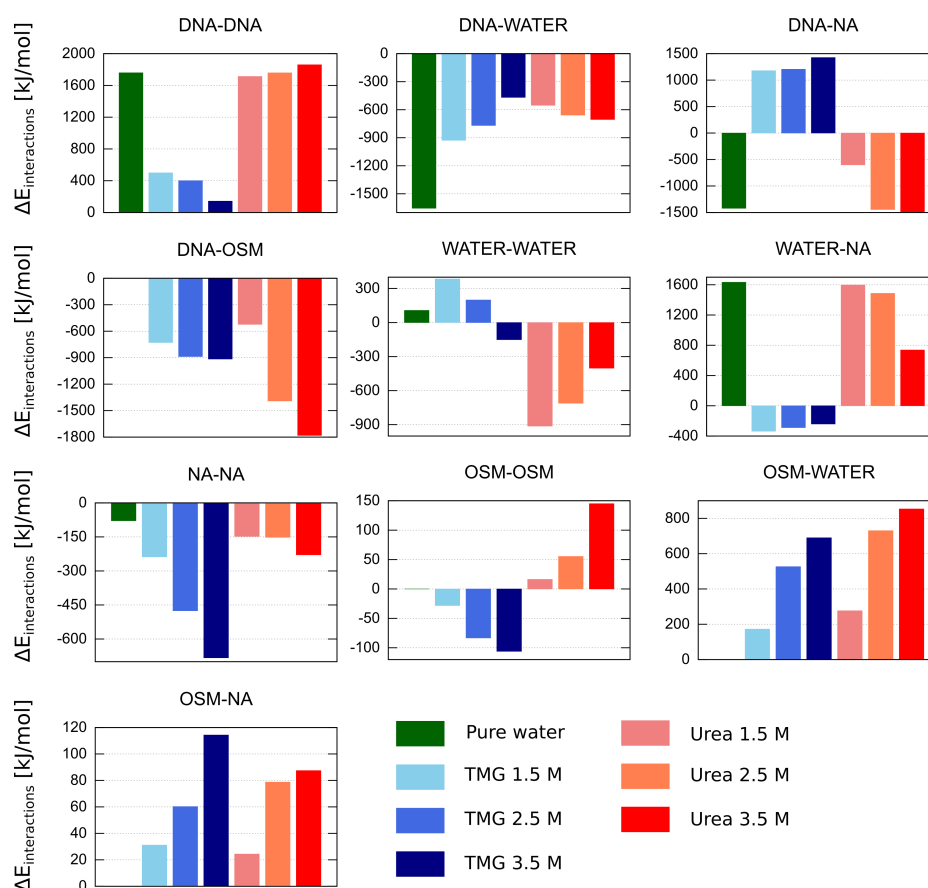
Aby dodatkowo potwierdzić opisane powyżej preferencje i ustalić dokładne rozmieszczenie cząsteczek denaturantów wokół DNA, wyznaczyłem przestrzenny rozkład cząsteczek denaturantów wokół helisy DNA, a wyniki przedstawiłem na dwóch rodzajach dwuwymiarowych map gęstości. Mapy pierwszego typu na Rys. 5.19B,C pokazują zależność gęstości cząsteczek denaturantu w funkcji odległości do osi helisy DNA (współrzędna DR) oraz wzdłuż osi helisy DNA. Położenie wzdłuż osi helisy DNA można łatwo przełożyć na położenie względem par zasad oraz pozycję w sekwencji DNA. Drugi typ map z Rys. 5.19E,F pokazuje gęstość radialną i kątową (w płaszczyźnie X–Y zasad azotowych) cząsteczek denaturantów względem głównej osi helisy DNA. Dla tego rodzaju map wartości gęstości podaje się jako średnia dla wszystkich par zasad DNA.

Jak przedstawiono na mapach gęstości radialnej (Rys. 5.19B,C) cząsteczki mocznika tworzą dwie sfery solwatacyjne, w obrębie helisy DNA oraz poza obszarem wyznaczonym przez jej promień (biała linia na Rys. 5.19B). Podczas gdy betaina tworzy jedną, ale dobrze zdefiniowaną zewnętrzną sferę solwatacyjną (duże wartości gęstości poza obszarem wyznaczonym przez promień helisy). Wynika to głównie z faktu, że cząsteczki mocznika ze względu na mniejsze rozmiary cząsteczkowe łatwiej penetrują większą i mniejszą bruzdę DNA. Powyższe obserwacje potwierdzają również mapy gęstości kątowej przedstawione na Rys. 5.19E,F.



RYSUNEK 5.19: A) Definicja odległości radialnej od osi helisy DNA (DR, obracającej się zgodnie z helisą) i odległość wzdłuż osi helisy (D, położenie względem par zasad), czyli współrzędnych helikalnych dla dwuwymiarowych map gęstości. B) i C) Dwuwymiarowy rozkład gęstości cząsteczek denaturatów, odpowiednio, dla TMG i mocznika wokół DNA we współrzędnych helikalnych. Białe linie reprezentują położenie grup fosforanowych w stosunku do osi helisy DNA i reprezentują promień helisy DNA. D) Położenie i definicja układu współrzędnych płaszczyzny X-Y na przykładzie pary zasad G-C dla drugiego typu dwuwymiarowych map gęstości. Czerwona kropka reprezentuje oś helisy DNA, czarne kropki oznaczają położenie atomów fosforu, białe koło reprezentuje promień helisy DNA, zielone pole reprezentuje obszar mniejszej bruzdy DNA, a niebieskie pole reprezentuje obszar większej bruzdy DNA. E) i F) Rozkład gęstości cząsteczek, odpowiednio, TMG i mocznika wokół DNA w 3,5 M roztworach; wartości gęstości wyznaczone jako średnie dla wszystkich par zasad, oprócz końcowych.

Cząsteczki betainy chętniej akumulują się wokół szkieletu DNA (poza obszarem wyznaczonym przez promień helisy DNA) niż wokół zasad azotowych, podczas gdy cząsteczki mocznika silniej penetrują strukturę helisy DNA. Wspomniana wcześniej preferencyjna orientacja betainy poprzez jej dodatni biegun momentu dipolowego (grupę aminową), w szczególności do grup fosforanowych DNA w mniejszej bruzdzie (Rys. 5.19) wskazuje na jej częściowy udział w stabilizacji elektrostatycznej helisy DNA. Ponadto z map gęstości radialnej na Rys. 5.19B,C można zaobserwować, podobnie jak wcześniej, preferencje cząsteczek betainy do par zasad G-C lub C-G w sekwencji DNA (pary zasad 4, 8, 12 i 15 od 5'-końca). W przypadku cząsteczek mocznika wyniki niezmiennie wskazują na brak określonej preferencji do par zasad.



RYSUNEK 5.20: Wartości energii oddziaływań między składowymi symulowanego układu (DNA, woda, jony  $\text{NA}^+$  i cząsteczki denaturantów) dla roztworu betainy (niebieskie słupki) i mocznika (czerwone słupki), obliczone jako różnica wartości energii w stanie zdenaturowanym i natywnym. Wartości ujemne oznaczają korzystny wzrost oddziaływań sprzyjających denaturacji. Kolor i jego ton wskazuje na rodzaj roztworu i stężenie denaturantów.

#### 5.4.6 Energie oddziaływania DNA z cząsteczkami roztworu

Spontaniczna denaturacja DNA jest skutkiem korzystniejszych oddziaływań pomiędzy DNA i cząsteczkami roztworu, w szczególności osmolitów, w stanie zdenaturowanym. W związku z tym postanowiłem zbadać potencjalne siły napędowe denaturacji poprzez wyznaczenie wkładów do entalpii denaturacji pochodzących od oddziaływań pomiędzy poszczególnymi elementami układu. Wkłady przedstawione na Rys. 5.20 są sumą energii elektrostatycznej oraz van der Waalsa i zostały obliczone jako uśrednione różnice entalpii w symulacjach między stanami natywnymi i zdenaturowanymi DNA.

Rys. 5.20 pokazuje, że zmiany oddziaływań wewnątrzcząsteczkowych w DNA (DNA-DNA) towarzyszące jego denaturacji w roztworze bez denaturantów i w roztworze mocznika są bardzo niekorzystne (1700–1850 kJ/mol). Z kolei w roztworze betainy ten niekorzystny wzrost energii jest zdecydowanie mniejszy (140–500 kJ/mol). Może to być wynikiem różnic konformacyjnych w stanach natywnych oraz zdenaturowanych DNA w różnych roztworach.

Równocześnie podczas denaturacji cząsteczki DNA następuje wzrost powierzchni DNA dostępnej dla roztworu, a w konsekwencji zwiększenie oddziaływań DNA z roztworem (DNA–OSM, DNA–WATER). W rzeczywistości suma oddziaływań DNA z wodą, denaturantami i przeciwjonami  $\text{Na}^+$  wskazuje na promowanie przez te oddziaływania stanu zdenaturowanego. Co więcej, denaturacja DNA w roztworach betainy prowadzi do znacznej utraty korzystnych oddziaływań DNA z jonami  $\text{Na}^+$ . Liczba jonów  $\text{Na}^+$  sąsiedztwie DNA maleje po denaturacji o około 10%, co widać na Rys. C.1 w Załączniku C.1.

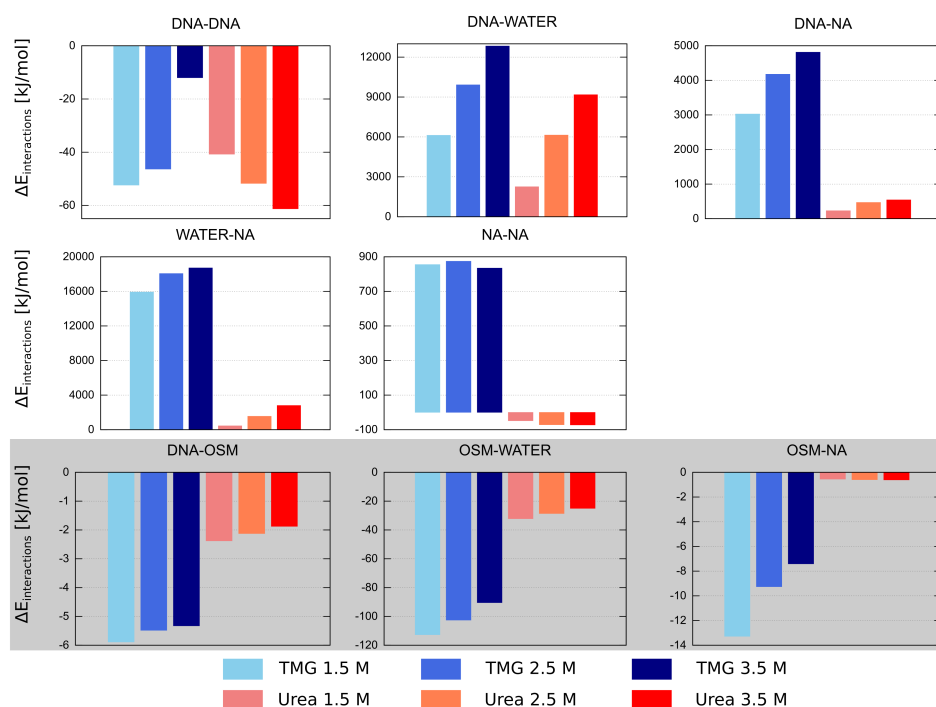
Pozostałe wkłady energetyczne dla jonów  $\text{Na}^+$  (WODA–NA, NA–NA, OSM–NA) również wskazują na silny wpływ cząsteczek betainy na zachowanie jonów  $\text{Na}^+$  w roztworze. Ze względu na ważną rolę jonów w stabilizacji natywnej struktury DNA, sugeruje to, że mechanizm denaturującego działania betainy może również polegać na wykluczaniu jonów z powierzchni DNA. Betaina dzięki dużemu momentowi dipolowemu (15,4 D) może tylko w niewielkim stopniu zastąpić kationy i stabilizować strukturę DNA.

Rys. 5.20 pokazuje także, że denaturacja w roztworach bez denaturantów i w roztworach betainy jest procesem niekorzystnym ze względu na oddziaływania pomiędzy cząsteczkami wody (WODA–WODA). Wyjątek stanowią roztwory mocznika, w których te oddziaływania sprzyjają denaturacji ( $-400$ – $900$  kJ/mol), prawdopodobnie przez korzystniejszą organizację wody w głębi roztworu. Z drugiej strony oddziaływania pomiędzy cząsteczkami denaturantów (OSM–OSM) wykazują przeciwny charakter. Tworzenie bardziej niekorzystnych oddziaływań między cząsteczkami mocznika wskazuje na wzrost kontaktu między nimi w wyniku ich agregacji wokół zdenaturowanego DNA. Natomiast cząsteczki betainy wydają się korzystniej oddziaływać ze sobą.

Wyznaczyłem również całkowitą entalpię denaturacji DNA w różnych roztworach wyrażoną jako sumę wszystkich wkładów z Rys. 5.20. Uzyskana przeze mnie wartość dla roztworu wodnego ( $22$  kJ/mol-pz) jest bardzo zbliżona do średniej wartości eksperymentalnej entalpii ( $28 \pm 2$  kJ/mol-pz) dla denaturacji DNA [234]. W przypadku roztworów betainy wartości te mieszczą się w zakresie  $12$ – $15$  kJ/mol-pz w zależności od stężenia oraz  $10$ – $13$  kJ/mol-pz dla roztworów mocznika. Biorąc pod uwagę średnią eksperymentalną entropię denaturacji ctDNA ( $79,6 \pm 4$  J/(K·mol-pz)) [234] wynika, że DNA już nawet w roztworze wodnym bez denaturantów w temperaturze  $300$  K jest tylko marginalnie stabilne ( $\Delta G \approx -1,6$  kJ/mol-bp), co zgadza się z niską temperaturą topnienia DNA bez buforu ( $51^\circ\text{C}$ ) wyznaczoną w kalorymetrii DSC.

Rys. 5.21 przedstawia różnice energii oddziaływań poszczególnych elementów układu między roztworem denaturantów i roztworem bez denaturantu. Ze względu na różną liczbę cząsteczek wody i denaturantów w systemach, zaprezentowane są tylko te wkłady do energii, które mogą być porównywane.

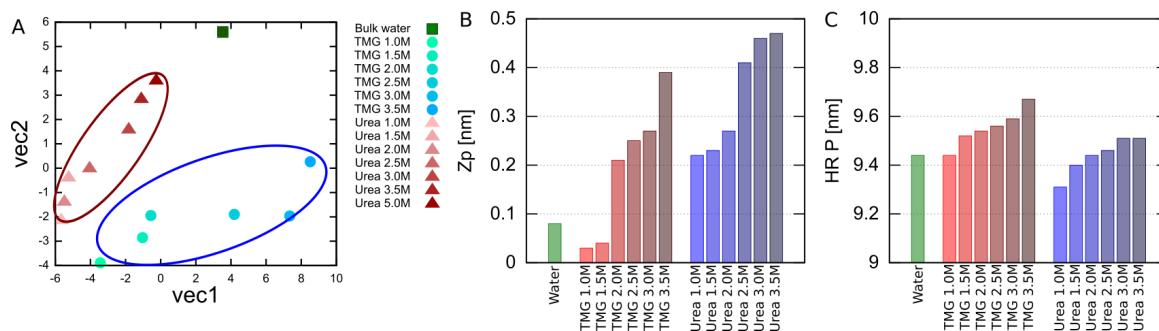




RYSUNEK 5.21: Wartości energii oddziaływań pomiędzy składowymi symulowanych układów dla roztworów TMG (niebieskie słupki) i roztworów mocznika (czerwone słupki), obliczone jako różnica wartości energii pomiędzy układem z i bez denaturanta w stanie natywnym DNA. Wartości ujemne oznaczają korzystny wzrost oddziaływań. Kolor i jego ton wskazuje na rodzaj roztworu i stężenie denaturantów. Dolny panel w szarym prostokącie zawiera wartości energii oddziaływania przeliczonej na jeden mol cząsteczek denaturantu.

Po pierwsze, Rys. 5.21 pokazuje, że dodanie denaturantów do roztworu nieznacznie zwiększa oddziaływania wewnątrzcząsteczkowe w cząsteczce DNA ( $-10$ – $60$  kJ/mol), co wynika ze zmian konformacyjnych DNA indukowanych obecnością denaturantów (późniejsza analiza). Równolegle można zaobserwować znaczną utratę oddziaływań DNA z wodą dla obu roztworów (DNA–WATER =  $2500$ – $13000$  kJ/mol) i oddziaływań DNA z jonami  $\text{Na}^+$ , szczególnie dla roztworów betainy (DNA–NA =  $3000$ – $4800$  kJ/mol). Ponadto dodanie betainy do roztworu nie jest korzystne dla oddziaływań pomiędzy samymi jonami (NA–NA =  $850$ – $900$  kJ/mol), w odróżnieniu od mocznika, którego dodanie do roztworu sprawia, że oddziaływania te stają się nieznacznie korzystniejsze (NA–NA =  $-50$ – $90$  kJ/mol). Ponadto, cząsteczki betainy powodują duże zmiany energii oddziaływań wody z jonami (WATER–NA =  $16000$ – $19000$  kJ/mol), co ponownie wskazuje na zaburzony rozkład jonów w roztworze przez ten denaturant. Okazuje się, że duże zmiany w oddziaływaniu DNA z jonami  $\text{Na}^+$  (DNA–NA =  $3000$ – $4800$  kJ/mol) można wytłumaczyć dużym spadkiem liczby jonów  $\text{Na}^+$  (o nawet  $50$ – $75$  %) wokół DNA (patrz Rys. C.1 w Załączniku C.1). Jak już wspominałem, to wykluczenie jonów i słabsze ekranowanie silnego odpychania między grupami fosforanowymi może prowadzić do znacznej destabilizacji struktury podwójnej helisy DNA.

Na koniec, w celu oceny zastosowanego przeze mnie modelu molekularnego, wyznaczyłem również entalpię bezpośrednich oddziaływań cząsteczek denaturantów z DNA, jonami



RYSUNEK 5.22: A) Serie danych z różnych symulowanych układów (wektory z wartościami poszczególnych parametrów) rozpięte na dwa pierwsze wektory własne uzyskane z analizy głównych składowych dla natywnej struktury DNA w roztworze bez denaturantów. Różne roztwory są oznaczone przez różne symbole, a ich stężenia przez ton koloru. B) Średnie wartości różnicy odległości pomiędzy rzutami położenia dwóch atomów fosforu na oś DNA ( $Z_p$ ) oraz średnie wartości promienia helisy (HR P) w różnych roztworach. Oba parametry zostały obliczone na podstawie położenia atomów fosforu DNA w różnych roztworach betajiny i mocznika.

$\text{Na}^+$  i wodą. Dolny panel na Rys. 5.21 pokazuje, że uzyskane przez mnie entalpie oddziaływań cząsteczek betajiny ( $\text{DNA-OSM} = -5,9, -5,3 \text{ kJ/mol}$ ) i cząsteczek mocznika z DNA ( $-2,4, -1,9 \text{ kJ/mol}$ ) z DNA są zgodne z wcześniejszymi wynikami otrzymanymi w kalorymetrii ITC (odpowiednio  $-2,9 \text{ kJ/mol}$  oraz  $-2,1 \text{ kJ/mol}$ ).

#### 5.4.7 Zmiany konformacyjne DNA

Jak wiadomo, struktura dwuniciowego DNA jest wrażliwa na właściwości rozpuszczalnika, w tym na aktywność obecnych w roztworze osmotów. Uzyskane w Katedrze widma spektroskopowe w zakresie podczerwieni oraz przesunięcia poszczególnych pasm na nich wskazują na zaburzenia struktury DNA przez denaturanty i przejścia konformacyjne z formy B-DNA w stronę formy A-DNA. W moich symulacjach również zaobserwowałem wyraźne zmiany konformacyjne struktury DNA wywołane obecnością modelowych denaturantów w roztworze. Aby zidentyfikować globalne oraz lokalne zmiany konformacyjne użyłem pakietu X3DNA [235], wyznaczając wybrane parametry geometryczne charakteryzujące strukturę podwójnej helisy w różnych roztworach. Następnie wszystkie serie danych dla wszystkich systemów zrzutowałem na płaszczyznę utworzoną przez dwa pierwsze wektory własne (vec1 i vec2 na Rys. 5.21A) uzyskane dla stanu natywnego w roztworze wodnym bez denaturantów przy użyciu analizy głównych składowych (z ang. Principal Component Analysis). Za pomocą tej metody ustaliłem również te parametry geometryczne, które są w największym stopniu zaangażowane w obserwowane przejścia konformacyjne DNA (opisują przynajmniej  $\sim 85\%$  całkowitej wariancji).

Położenie i wyraźne grupowanie poszczególnych serii danych (odpowiadające średnim strukturom DNA w danym roztworze) na Rys. 5.22A wskazuje, że roztwory denaturantów różnią się pomiędzy sobą, pod względem struktury przyjmowanej przez DNA. Jak można

zauważyć, pierwszy wektor własny ( $vec1$ ) najlepiej rozróżnia rodzaj obecnego denaturanta, z kolei drugi wektor własny ( $vec2$ ) różnicuje je ze względu na stężenie.

Następnie ustaliłem które z parametrów geometrycznych DNA wnoszą największy wkład do pierwszego wektora własnego. Okazało się, że spośród wszystkich parametrów największy wpływ na geometrię DNA mają dwa parametry związane z względnym położeniem grup fosforanowych ( $Zp$  i  $HR P$ ), których średnie wartości dla różnych układów przedstawiłem na Rys. 5.22B,C. Parametr  $Zp$  odnosi się do względnej pozycji wzdłuż osi helisy dwóch naprzeciwległych grup fosforanowych z sąsiadujących nici DNA. Natomiast parametr  $HR P$  to efektywny promień podwójnej helisy określony na podstawie położenia fosforanów.

Generalnie zaobserwowałem wzrost wartości parametru  $Zp$  dla struktur DNA w większości roztworów denaturantów względem struktury DNA w roztworze bez denaturantów. Oznacza to zwiększenie aksjalnej odległości między atomami fosforu w szkielecie DNA. Jedynie dla najmniejszych stężeń betainy (1 M i 1,5 M) efekt jest przeciwny. Jak pokazano wcześniej, duża wartość  $Zp$  ( $> 0,15$  nm) jest charakterystyczna dla struktur DNA o konformacji w typie A-DNA, co jest spełnione dla większości roztworów denaturantów. Ponadto zaobserwowałem także wzrost efektywnego promienia helisy ( $HR P$  na Rys. 5.22C), w szczególności dla dużych stężeń denaturantów ( $>2,5$  M), co jest także typowe dla przejść konformacyjnych w kierunku formy A-DNA (dla której promień helisy jest o około 6% większy niż w B-DNA). Na podstawie wcześniejszych analiz można stwierdzić, że to charakterystyczne przejście od formy B-DNA do formy A-DNA jest wywołane desolvatacją natywnej struktury DNA pod wpływem obecności denaturantów (zmniejszeniem liczby korzystnych wiązań wodorowych z cząsteczkami wody).

#### 5.4.8 Podsumowanie

Aby wyjaśnić mechanizmy odpowiedzialne za obserwowaną eksperymentalnie destabilizację struktury DNA przez wodne roztwory modelowych denaturantów, w moich badaniach wykorzystałem metody symulacyjne dynamiki molekularnej. Jako reprezentatywne związki denaturujące wybrałem betainę (TMG) oraz mocznik. Betaina i mocznik są denaturantami mającymi różny wpływ na stabilność białek, ale wywołującymi podobną destabilizację podwójnej helisy DNA. Analizę wyników oraz ich weryfikację prowadziłem równoległe z badaniami eksperymentalnymi prowadzonymi w Katedrze Chemii Fizycznej Politechniki Gdańskiej.

Wykonana przeze mnie analiza współczynników preferencyjnej akumulacji denaturantów wskazuje, że mocznik akumuluje się silniej wokół DNA niż TMG. Natomiast szczegółowe analizy liczby cząsteczek denaturantów wokół DNA oraz tworzonych przez nie wiązań

wodorowych z różnymi elementami strukturalnymi DNA wykazały występowanie specyficznych oddziaływań jedynie pomiędzy betainą oraz guaninami. Podczas gdy mocznik oddziałuje jednakowo silnie ze wszystkimi zasadami, w szczególności w stanie zdenaturowanym.

Wyniki obliczeń dynamiki molekularnej wskazują, że obecność betainy i mocznika w wodnym roztworze powoduje odwodnienie struktury DNA, na skutek zmniejszenia liczby cząsteczek wody przez betainę lub zmniejszenie korzystnych oddziaływań DNA z wodą, na skutek osłabienia przez mocznik struktury wody.

Ustaliłem, że oba denaturanty są preferencyjnie zorientowane dodatnimi biegunami momentów dipolowych w kierunku powierzchni DNA. Cząsteczki betainy wykazują mały potencjał do tworzenia bezpośrednich oddziaływań z DNA, zaś cząsteczki mocznika mogą w znacznym stopniu oddziaływać z DNA przez atom tlenu grupy karbonylowej. Przestrzenne rozkłady cząsteczek denaturantów względem DNA pokazały, że mocznik tworzy dwie główne warstwy solwatacyjne wokół helisy DNA. Jedną wewnętrzną (w dużej i małej bruzdzie DNA) i drugą na zewnątrz helisy DNA (w obszarze wyznaczonym przez pozycje grup fosforanowych). Betaina natomiast tworzy jedną dobrze zdefiniowaną zewnętrzną warstwę, zlokalizowaną wokół łańcucha fosfocukrowego oraz okolicach mniejszej bruzdy.

Analiza oddziaływań DNA z różnymi składnikami roztworów w stanie natywnym i zdenaturowanym dała wgląd w mechanizm destabilizacji DNA przez oba denaturanty. Wyniki powolnie wskazały na to, że betaina preferencyjnie oddziałuje z guanidyną. Wydaje się, że wpływ denaturantów na DNA wynika przede wszystkim z równowagi pomiędzy dwoma głównymi czynnikami – z jednej strony zdolnością do zaburzania ekranowania oddziaływań odpychających pomiędzy grupami fosforanowych DNA, z drugiej zaś różnego powinowactwa do zasad nukleinowych, które w stanie zdenaturowanym stają się dostępne dla cząsteczek denaturantów.

Podczas destabilizacji DNA zachodzą zauważalne zmiany konformacyjne w cząsteczce DNA, które w przypadku obu denaturantów mają charakter zmian strukturalnych z konformacji typu B-DNA w stronę konformacji typu A-DNA. Efekt ten jest bardziej zauważalny w przypadku roztworów betainy. Wymienione efekty zostały niezależnie potwierdzone przez badania spektroskopowe w zakresie podczerwieni prowadzone w Katedrze.

## 6 Ogólne podsumowanie

W przedstawionej rozprawie wykazałem, że z powodzeniem można używać symulacji dynamiki molekularnej w celu wyjaśnienia mechanizmów odpowiedzialnych za różne procesy zachodzące przy udziale biomolekuł takich jak DNA oraz w celu interpretacji danych eksperymentalnych w postaci wyznaczonych struktur cząsteczek lub charakterystycznych widm spektroskopowych cząsteczek.

Podstawowym celem moich badań było ustalenie molekularnych źródeł stabilności różnych form DNA oraz zmian konformacyjnych wywołanych różnymi czynnikami strukturalnymi oraz warunkami zewnętrznymi. W szczególności interesowały mnie niekanoniczne struktury DNA typu G-kwadrupeksów. G-kwadrupeksy powstają w obrębie sekwencji nici kwasu nukleinowego bogatych w guaniny, przyjmując charakterystyczne konformacje wyróżniające się warstwową budową utworzoną przez płaszczyzny guaninowe. Dowiedziono, że struktury G-kwadrupeksów powstają *in vivo* i są zaangażowane w regulację ważnych procesów biologicznych takich jak, replikacja i transkrypcja DNA, rekombinacja genetyczna, utrzymywanie stabilności chromosomów. Struktury i stabilność kilku różnych topologii G-kwadrupeksów wyznaczono eksperymentalnie, jednakże pomimo szeroko zakrojonych badań nie ustalono jeszcze molekularnych sił napędowych odpowiedzialnych za ich powstawanie.

Kolejnym badanym przeze mnie zagadnieniem były mechanizmy odpowiedzialne za zmiany konformacyjne zachodzące w klasycznej podwójnej helisie DNA pod wpływem obecności wybranych współrozpuszczalników. Do badań wybrałem dwa reprezentatywne związki denaturujące podwójną helisę DNA – mocznik i betainę. Znanym efektem działania obu substancji jest obniżenie stabilności podwójnej helisy DNA, jednakże ma ona różnych charakter. W przypadku betainy poziom destabilizacji struktury DNA zależy od jej sekwencji, natomiast stopień destabilizacji DNA wywoływanej przez mocznik zależy jedynie od jego stężenia. Sugeruje to, że badane denaturaty przejawiają dwa różne, ale jeszcze nieokreślone mechanizmy działania, których podstawy molekularne postanowiłem zbadać we współpracy z grupą badawczą prof. J. Stangreta z Katedry Chemii Fizycznej PG.

W przedstawionej pracy wykorzystałem w pełni atomową dynamikę molekularną, aby zbadać molekularne źródła stabilności G-kwadrupeksów występujące na różnym poziomie ich złożoności strukturalnej. Po pierwsze zbadałem proces dysocjacji i tworzenia pojedynczej G-tetrazy, czyli wspólnego motywu strukturalnego występującego w każdym G-kwadrupeksie. W tym celu wyznaczyłem profil energii swobodnej opisujący proces dysocjacji

3'-końcowej G-tetrady w pełni uformowanym telomerowym G-kwadrupleksie o topologii równoległej. Na jego podstawie wyznaczyłem dość znaczną stabilność termodynamiczną i kinetyczną badanej G-tetrady. Szczegółowo przeanalizowałem również poszczególne etapy dysocjacji G-tetrady i ustaliłem, że proces ten ma charakter sekwencyjny, pod względem zrywania wiązań wodorowych oraz kooperatywny, który dodatkowo zależy od wiązania jonu w centralnym kanale G-kwadrupleksu. Na podstawie analiz poszczególnych wkładów entalpowych ustaliłem, że sama energetyka asocjacji guanin nie tłumaczy ogólnej stabilności G-tetrady i niezbędne jest uwzględnienie oddziaływań obejmujących inne elementy układu, w szczególności szkielet fosfocukrowy. Co więcej pokazałem, że do uzyskania pełnej stabilności entalpowej konieczna jest minimalizacja energii odpychania pomiędzy silnie naładowanymi fosforanami. Ustaliłem również, że ten główny korzystny wkład entalpowy jest w dużej mierze kompensowany przez niekorzystną zmianę entropii konformacyjnej spowodowaną asocjacją guanin i optymalizacją topologii szkieletu DNA.

Kolejnym badanym elementem strukturalnym mającym duży wpływ na stabilność G-kwadrupleksów były ich pętle. Wiadome jest, że krótsze sekwencje pętli preferują tworzenie pętli śmigłowych, natomiast pętle diagonalne oraz boczne są tworzone jedynie przez dłuższe sekwencje. W celu szczegółowego przebadania wpływu długości oraz rodzaju pętli na stabilność G-kwadrupleksów przebadłem reprezentatywny zbiór modelowych struktur G-kwadrupleksów różniących się topologią oraz długością sekwencji. Do tego celu opracowałem metodę szacowania stabilności G-kwadrupleksów na podstawie czasu życia reprezentatywnego zestawu modelowych struktur G-kwadrupleksów w symulacjach równowagowych. Modelowe struktury G-kwadrupleksów zostały wygenerowane za pomocą opracowanej przeze mnie nowej procedury zwijania G-kwadrupleksów o arbitralnej topologii i sekwencji. Weryfikacja wyników uzyskanych dla znanych już struktur G-kwadrupleksów potwierdziła użyteczność moich metod. Właściwe badania już na etapie zwijania poszczególnych struktur pokazały, że nici DNA o najkrótszych możliwych sekwencjach pętli ze względów występowania zawad sterycznych i naprężeń strukturalnych nie mogą formować G-kwadrupleksów dwupłaszczyznowych o topologii innej niż równoległa. W przypadku dłuższych i symetrycznych sekwencji najbardziej stabilne okazały się G-kwadrupleksy o topologii antyrównoległej lub hybrydowej. Szczegółowa analiza rejonów pętlowych przy użyciu sekwencji niesymetrycznych wskazała, że pierwsza pętla w sekwencji jest najbardziej czuła na zmiany jej długości i do jej powstania preferowane są minimum 3 zasady azotowe. Przeprowadzone przeze mnie oszacowanie stabilności G-kwadrupleksów dwupłaszczyznowych w rezultacie pozwoliło mi to także na wyznaczenie najkrótszych sekwencji DNA mogących formować stabilne G-kwadrupleksy, co częściowo zostało potwierdzone przez grupę badawczą prof. Z. Gdaniec z Instytutu Chemii Bioorganicznej Polskiej Akademii Nauk

Wiele eksperymentów pokazało, że G-kwadrupleksy posiadające eksponowane do rozpuszczalnika tetrazy guaninowe wykazują tendencję do agregacji poprzez bezpośrednie oddziaływania między nimi. Taka dodatkowa cecha może mieć istotne znaczenie dla procesów upakowania sekwencji tworzących G-kwadrupleksów oraz może umożliwiać projektowanie struktur G-kwadrupleksów wyższych rzędów. Aby zrozumieć kolektywne molekularne siły napędzające asocjację G-kwadrupleksów, zbadałem dimeryzację dwóch telomerowych G-kwadrupleksów o topologii równoległej przy użyciu symulacji dynamiki molekularnej. Przewidziane przeze mnie energie swobodne dimeryzacji wskazują, że najbardziej preferowanym sposobem dimeryzacji jest bezpośrednio wiązanie się G-kwadrupleksów przez ich 5-końcowe G-tetrazy (mod 5'-5'), co jest jednocześnie zgodne ze wszystkimi dostępnymi danymi eksperymentalnymi. Dalsza szczegółowa analiza uzyskanych profili energii swobodnej oraz szczegółowa analiza strukturalna pokazała, że dimeryzacja G-kwadrupleksów jest w niewielkim stopniu możliwa także w modzie 3'-5' lub przy udziale dodatkowych mediujących płaszczyzn adeninowych. Przeprowadzona przeze mnie dekompozycja entalpii dimeryzacji w połączeniu z wynikami symulacji odzwierciedlających różne stężenia jonów  $K^+$  wskazują również, że zaobserwowane preferencje orientacyjne wynikają z subtelnej równowagi między odpychaniem elektrostatycznym szkieletów fosfocukrowych, a korzystnym wiązaniem przeciwjonów na interfejsie dimeryzacji. Wnioski uzyskane podczas dokładnej analizy procesów tworzenia dimerów G-kwadrupleksów pozwoliły mi również na proponowanie potencjalnych metod przesuujących równowagę dimeryzacji. W szczególności przewiduję, że podwojenie fizjologicznego stężenia jonów  $K^+$  powinno zwiększyć populację dimerów w modzie 3'-5', natomiast zwiększenie odległości separacji poprzez użycie odpowiedniego liganda mediującego, powinno znacznie poprawić stabilność dimeru w modzie 3'-3'.

Druga część moich badań miała na celu ustalenie molekularnych determinantów równowagi konformacyjnej podwójnej helisy DNA w obecności dwóch reprezentatywnych związków denaturujących – mocznika i betainy. W tym celu przeprowadziłem symulacje układów zawierających z cząsteczki DNA w stanie natywnym oraz zdenaturowanym w różnych stężeniach badanych denaturantów. Po pierwsze na podstawie analizy radialnych funkcji rozkładu oraz współczynników preferencyjnego oddziaływania ustaliłem, że mocznik silnie akumuluje się w pierwszej sferze hydratacyjnej DNA natomiast betaina jest wykluczana z powierzchni DNA do drugiej sfery hydratacyjnej. Te dwa różne efekty wpływają również niekorzystnie na rozkład wody wokół DNA. Betaina drastycznie obniża liczbę cząsteczek wody, osłabiając hydratację DNA, natomiast mocznik dodatkowo zmniejsza gęstość wody, która pozostała przy powierzchni DNA. Dalsze analizy preferencji denaturantów do miejsc oddziaływania z DNA oraz tworzenia z nimi wiązań wodorowych wskazała, że roztwory mocznika tworzą wiele wiązań wodorowych z całą zdenaturowaną formą DNA przez co stabilizują ją. Betaina natomiast wykazuje jedynie preferencję do oddziaływania z guaniną i tworzenia

większej liczby wiązań wodorowych w stanie zdenaturowanym DNA. Analizowany przeze mnie przestrzenny rozkład cząsteczek denaturantów i ich sposób ułożenia względem powierzchni DNA wskazał, że betaina ustawia się względem powierzchni DNA za pomocą jej dodatniego bieguna momentu dipolowego i najkorzystniej oddziałuje z silnie naładowanym łańcuchem fosfocukrowym DNA. Natomiast ułożenie cząsteczek mocznika i ich rozkład przestrzenny wokół DNA jest już bardziej jednorodny bez specyficznych preferencji. Na podstawie analiz oddziaływań DNA z poszczególnymi elementami roztworów odkryłem, że betaina silnie wpływa na oddziaływania jonów  $\text{Na}^+$  z całym układem. Okazało się, że betaina silnie wypiera jony  $\text{Na}^+$  z powierzchni DNA, powodując jego destabilizację. Na koniec wyznaczyłem globalne oraz lokalne zmiany konformacyjne DNA powstające w wyniku obecności w roztworze obu denaturantów oraz ich działania efektywnie osłabiającego hydratację DNA. W obu przypadkach konformacja podwójnej helisy DNA zmienia się z formy B-DNA w stronę formy A-DNA i jest to głównie wynikiem zmian względnych odległości pomiędzy fosforanami. Otrzymane wyniki pomogły w interpretacji widm spektroskopowych zaburzonych cząsteczek DNA. Na podstawie moich wyników zaproponowałem mechanizmy działania obu denaturantów. Mocznik denaturuje cząsteczkę DNA poprzez tworzenie wielu korzystnych wiązań wodorowych ze wszystkimi zasadami azotowymi w stanie zdenaturowanym. Betaina natomiast wypiera stabilizujące jony z powierzchni DNA oraz oddziałuje preferencyjnie z guaninami, w szczególności w stanie zdenaturowanym.



# A Załącznik

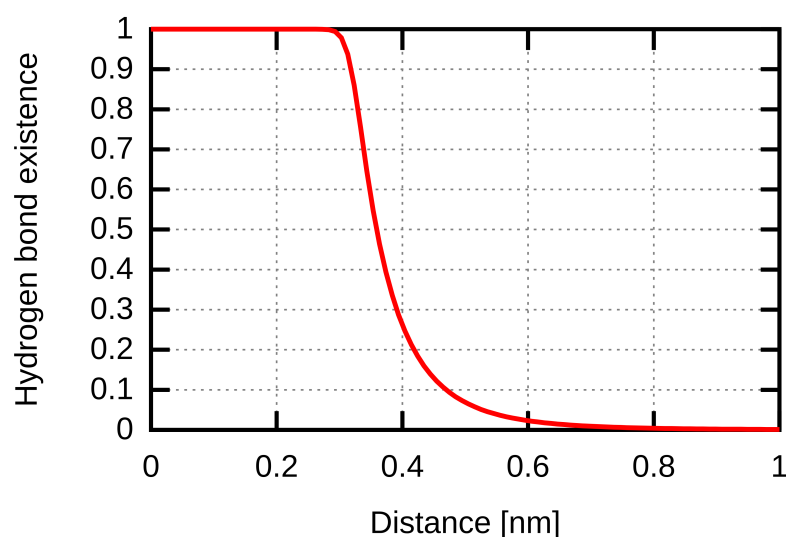
## A.1 Stabilność G-tetrazy w uformowanym G-kwadrupleksie

### A.1.1 Definicja współrzędnej opisującej dysocjację G-tetrazy ( $N_{hb}$ )

Całkowita liczba natywnych wiązań wodorowych  $N_{hb}$  została zdefiniowana jako suma funkcji aktywacji:

$$N_{hb} = \sum_{i \in A,B} s_i = \sum_{i \in A,B} \frac{1 - \left(\frac{r_i}{r_0}\right)^n}{1 - \left(\frac{r_i}{r_0}\right)^m} \quad (\text{A.1})$$

, gdzie suma przebiega po wszystkich natywnych wiązaniach tworzonych przez odpowiedni akceptor (A) i donor (D).  $s_i$  jest równe 1 gdy określona para akceptor-donor (i) jest oddalona od siebie o pewną odległość krytyczną, w przeciwnym przypadku jest równa 0. W praktyce wyrażenie  $s_i$  jest zastąpione przez funkcję aktywacji. W moim przypadku  $r_0$  było równe 0,32 nm, natomiast  $n$  oraz  $m$  odpowiednio 12 i 18. Rysunek A.1 przedstawia prawdopodobieństwo tworzenia wiązania wodorowego w zależności od odległości pomiędzy akceptorem i donorem.



RYSUNEK A.1: Funkcja aktywacji użyta do definiowania natywnych wiązań wodorowych w G-tetradzie. Prawdopodobieństwo istnienia wiązania wodorowego jest funkcją odległości między donorem i akceptorem tworzącym wiązanie.

### A.1.2 Szacowanie średniego czasu pierwszego przejścia (MFPT)

Średni czas pierwszego przejścia (z ang. Mean First Passage Time) dla dysocjacji pierwszej oraz drugiej guaniny został oszacowany na podstawie formuły zaproponowanej przez Zwanziga [236]:

$$MFPT = \int_{x_0}^{x_1} dx \exp(\beta G(x)) \cdot \frac{1}{D} \int_a^x dy \exp(-\beta G(y)) \quad (\text{A.2})$$

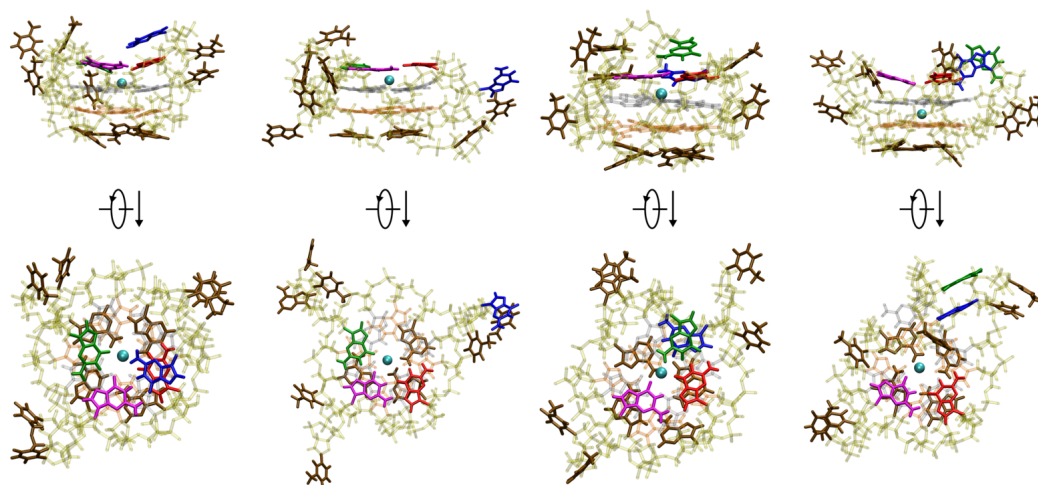
, gdzie  $G(x)$  jest profilem energii swobodnej, a  $x_0$  i  $x_1$  to granice całkowania definiujące stan początkowy oraz końcowy rozważanego procesu dysocjacji,  $\beta$  jest odwrotnością stałej temperaturowej  $(k_B T)^{-1}$ ,  $a$  jest położeniem bariery odbijającej w  $N_{hb} = 8$  (dla dysocjacji pierwszej guaniny) lub  $N_{hb} = 4$  (dla dysocjacji drugiej guaniny), a  $D$  to współczynnikiem dyfuzji dla liczby wiązań wodorowych. Współczynnik dyfuzji został oszacowany jako średnia wartość w pobliżu globalnego minimum energetycznego przy użyciu wzoru zaproponowanego przez Hummera [237]:

$$D = \frac{\sigma^2}{\tau_A} \quad (\text{A.3})$$

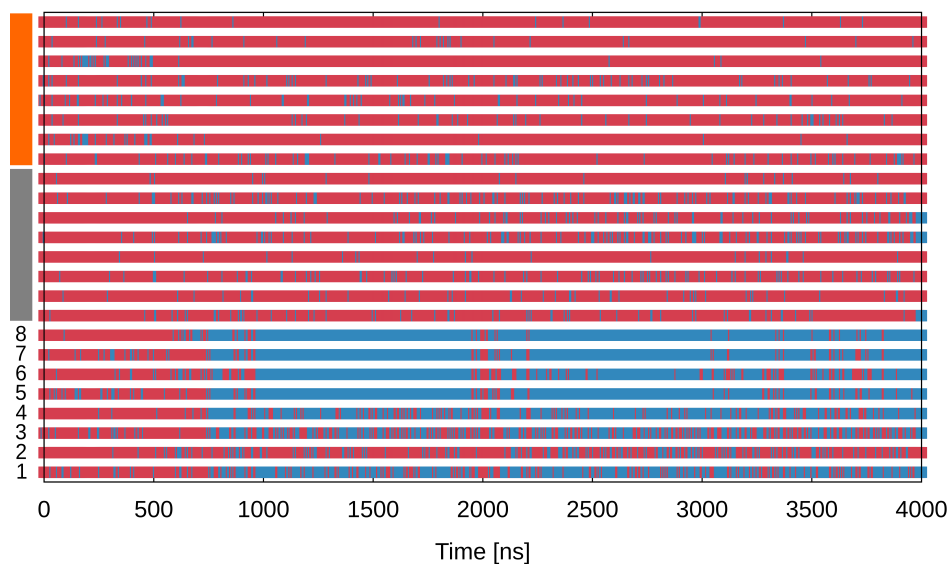
, gdzie  $\sigma^2$  jest wariancją dystrybucji liczby wiązań wodorowych, a  $\tau_A$  ich czasem autokorelacji.

### Symulacje typu energii swobodnej dla położenia jonu $K^+$ w centralnym kanale G-kwadrupeksu

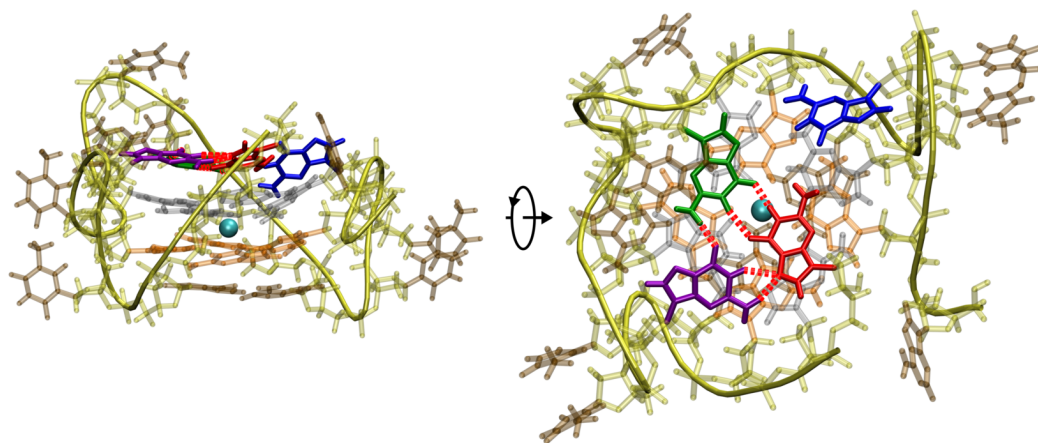
Aby wyznaczyć profil energii swobodnej dla położenia pojedynczego jonu  $K^+$  wzdłuż centralnego kanału równoległego G-kwadrupeksu w moich symulacjach zastosowałem metodę wygaszanej metadynamiki (Well-Tempered Metadynamics) [238] zaimplementowaną w programie GROMACS [164] poprzez wtyczkę PLUMED [169]. Skład systemu, zestaw parametrów pola siłowego i ogólny protokół dynamiki molekularnej były takie same jak w symulacjach dysocjacji 3'-końcowej G-tetrazy. Początkowa konfiguracja systemu z pojedynczym jonem  $K^+$  związanym w pozycji 2 kanału G-kwadrupeksu (patrz Rys. 5.2D) została wygenerowana w symulacji równowagowej. Jako współrzędną reakcji użyłem odległość między położeniem jonu a środkiem masy wszystkich atomów O6 z guanin tworzących kanał. Funkcje Gaussa związane z dodatkowym potencjałem o wysokości 0,05 kJ/mol oraz szerokości 0,01 nm były rozmieszczane co 500 kroków symulacji, a współczynnik temperaturowy ustawiłem 6. System był symulowany przez 645 ns.



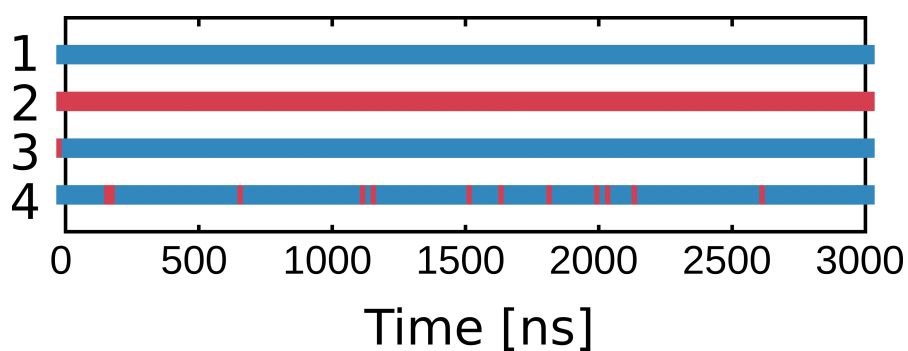
RYSUNEK A.2: Cztery typowe geometrie zdysocjowanego stanu G-tetrady pokazujące oddziaływania pomiędzy guaninami i tyminami w pętach TTA (w kolorze brązowym), sąsiednimi zdysocjowanymi guaninami oraz środkową G-tetradą (w kolorze szarym). Na górnym panelu widoczna jest również podatność reszt adeninowych do asocjacji na 5'-końcowej G-tetradzie.



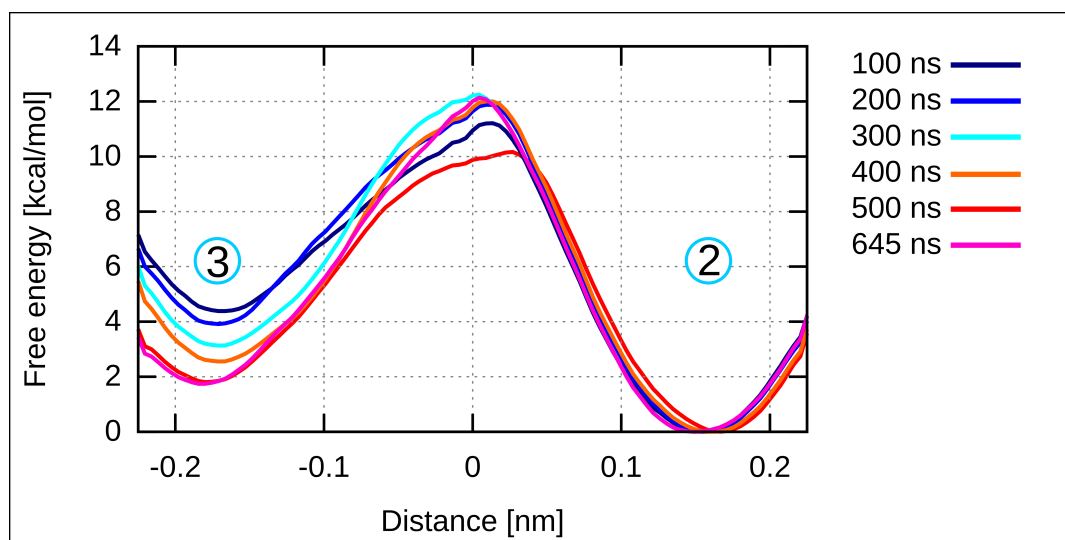
RYSUNEK A.3: Ewolucja natywnych wiązań wodorowych w dodatkowej 4  $\mu$ s symulacji dynamiki molekularnej. Po czasie 750 ns guanina IV spontanicznie dysocjuje z 3'-końcowej G-tetrady. W wyniku czego powstaje tryplet guaninowy (Rys. A.4), w którym natywne wiązania wodorowe są tworzone sporadycznie.



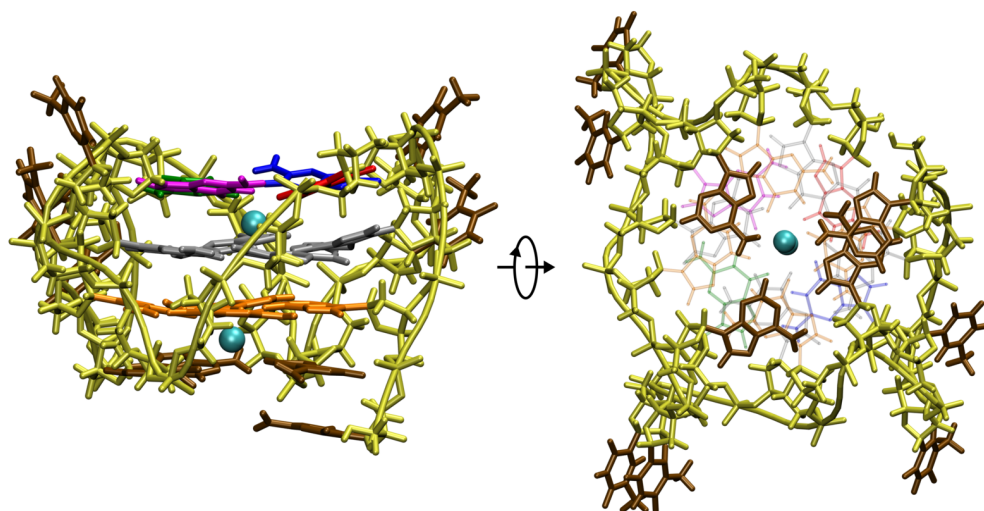
RYSUNEK A.4: Tworzenie się trypletu guaninowego w dodatkowej  $4\ \mu\text{s}$  symulacji równowagowej. Każda guanina uczestniczy w tworzeniu 3 wiązań wodorowych z dwoma pozostałymi guaninami (czerwone przerywane linie).



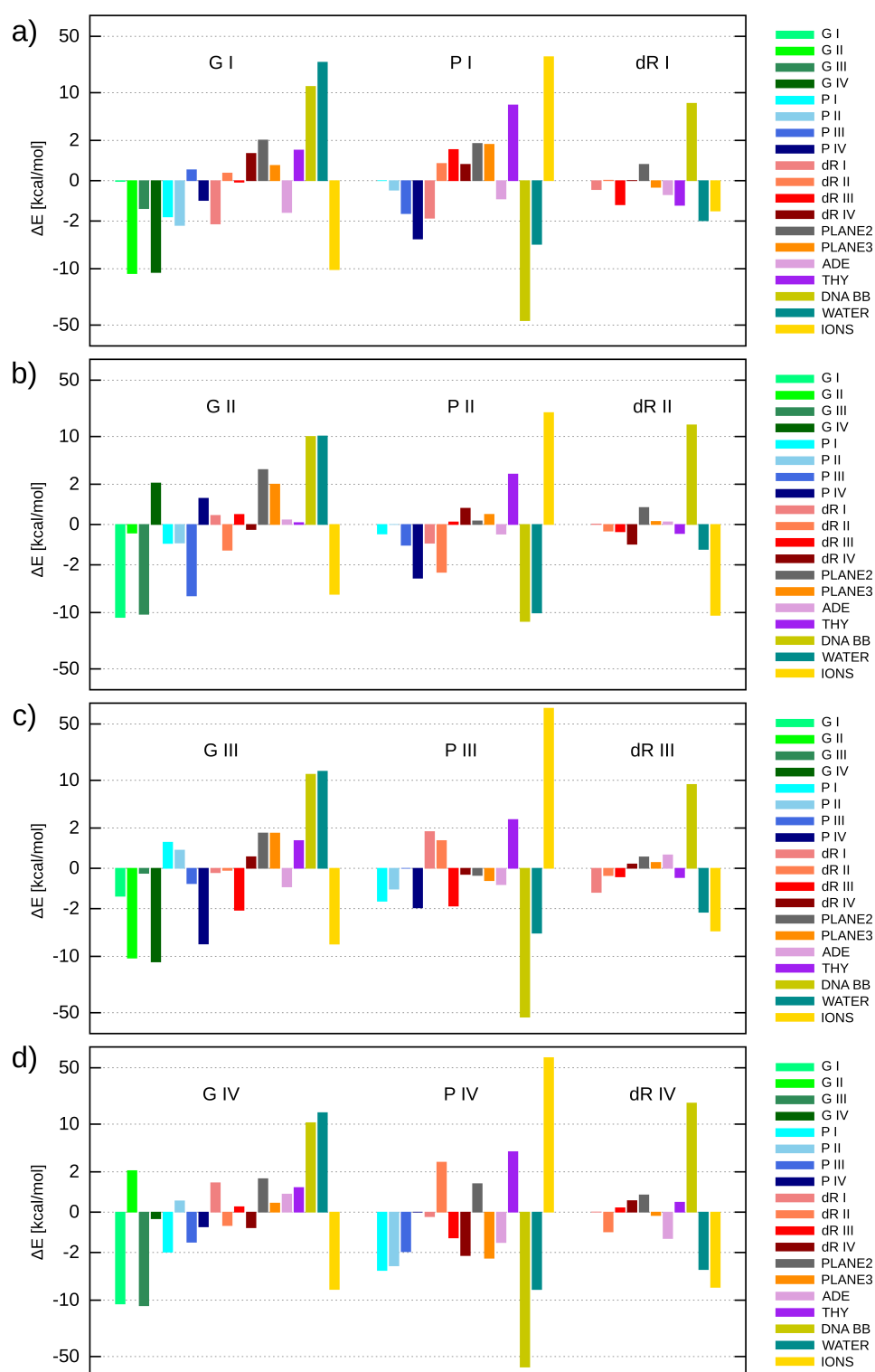
RYSUNEK A.5: Prawdopodobieństwa obsadzenia przez jony  $\text{K}^+$  4 miejsc wiązania w centralnym kanale jako funkcja czasu w dodatkowej  $4\ \mu\text{s}$  symulacji dynamiki molekularnej.



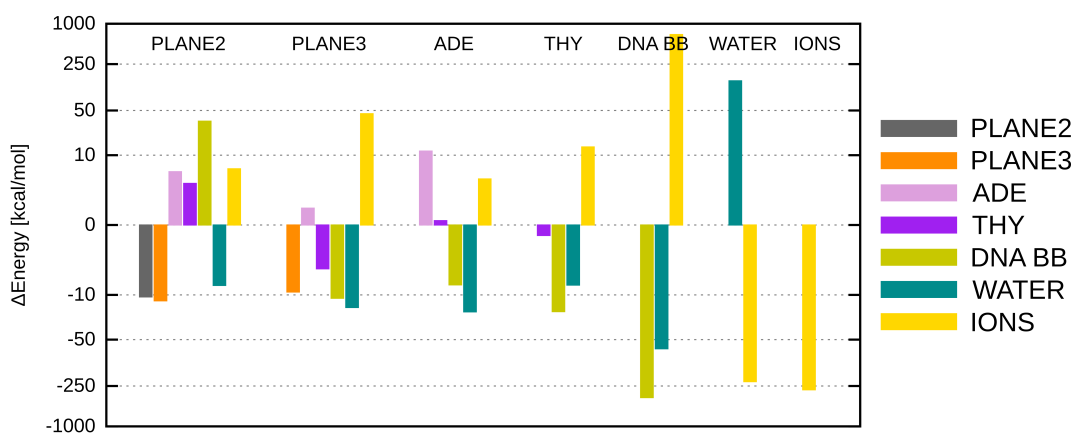
RYSUNEK A.6: Profil energii swobodnej dla położenia pojedynczego jonu  $K^+$  wzdłuż kanału G-kwadrupleksu wyznaczony co 100 ns symulacji metodą metadynamiki o łącznej długości 645 ns. Miejsca wiązania jonów 2 i 3 są zdefiniowane tak samo jak na Rys. 5.2D.



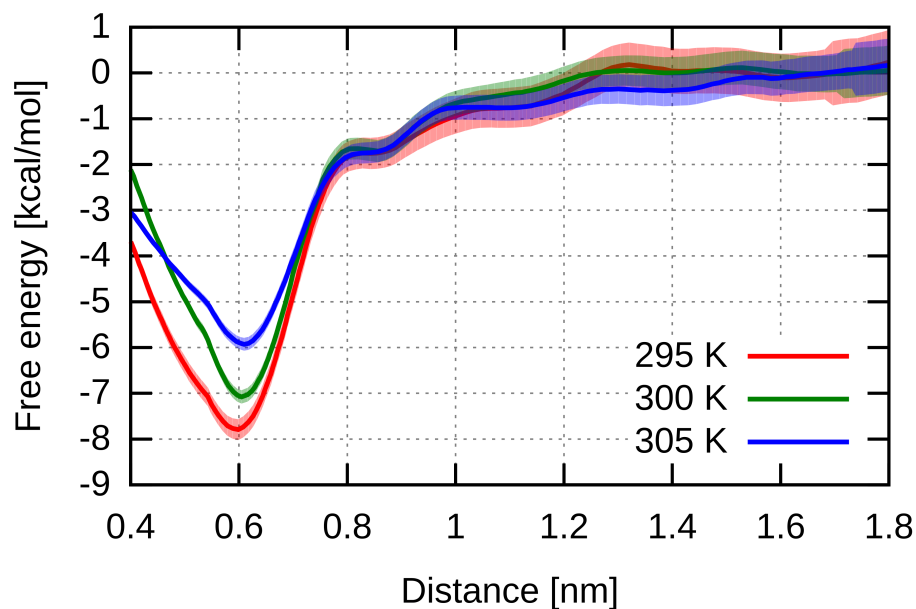
RYSUNEK A.7: Reprezentatywna struktura telomerowego G-kwadrupleksu o topologii równoległej uzyskana w symulacji pokazująca podatność adenin do asocjacji na 5'-końcowej G-tetradzie (PLANE3).



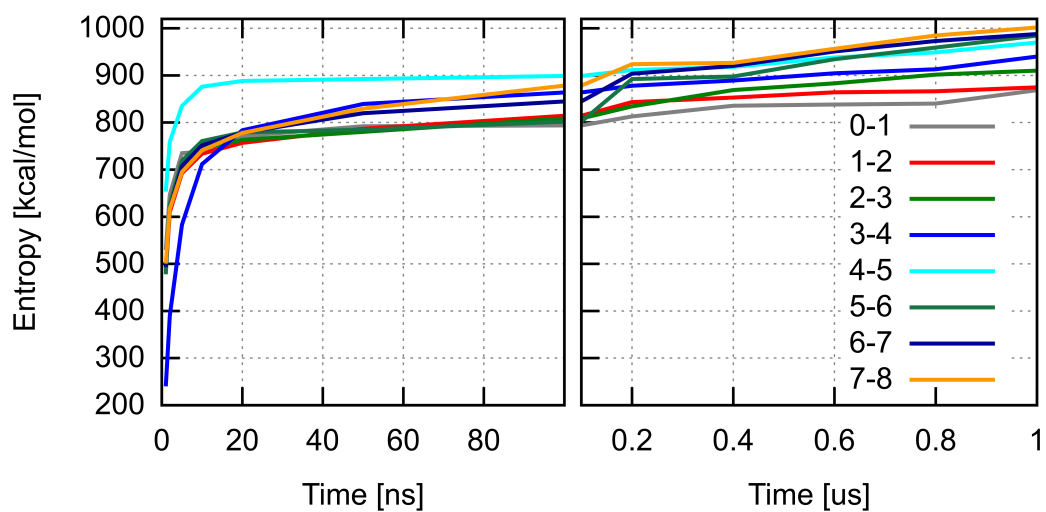
RYSUNEK A.8: Wkłady wnoszące udział do zmiany entalpii utworzenia G-tetrazy wynikające z oddziaływań między nukleotydami guaninowymi I-IV podzielonymi na guaniny (G), fosforany (P), dezoksyrybozy (dR) oraz resztę układu. Pozostałe elementy G-kwadrupleksu podzielono na: zasady środkowej G-tetrazy (PLANE2), 5-końcowej G-tetrazy (PLANE3), zasady adeninowe (ADE), tyminy (THY) oraz szkielet fosfocukrowy (DNA BB).



RYSUNEK A.9: Pozostałe wkłady wnoszące udział do zmiany entalpii utworzenia G-tetrazy z tymi samymi oznaczeniami jak na Rys. 5.3 i Rys. A.8.

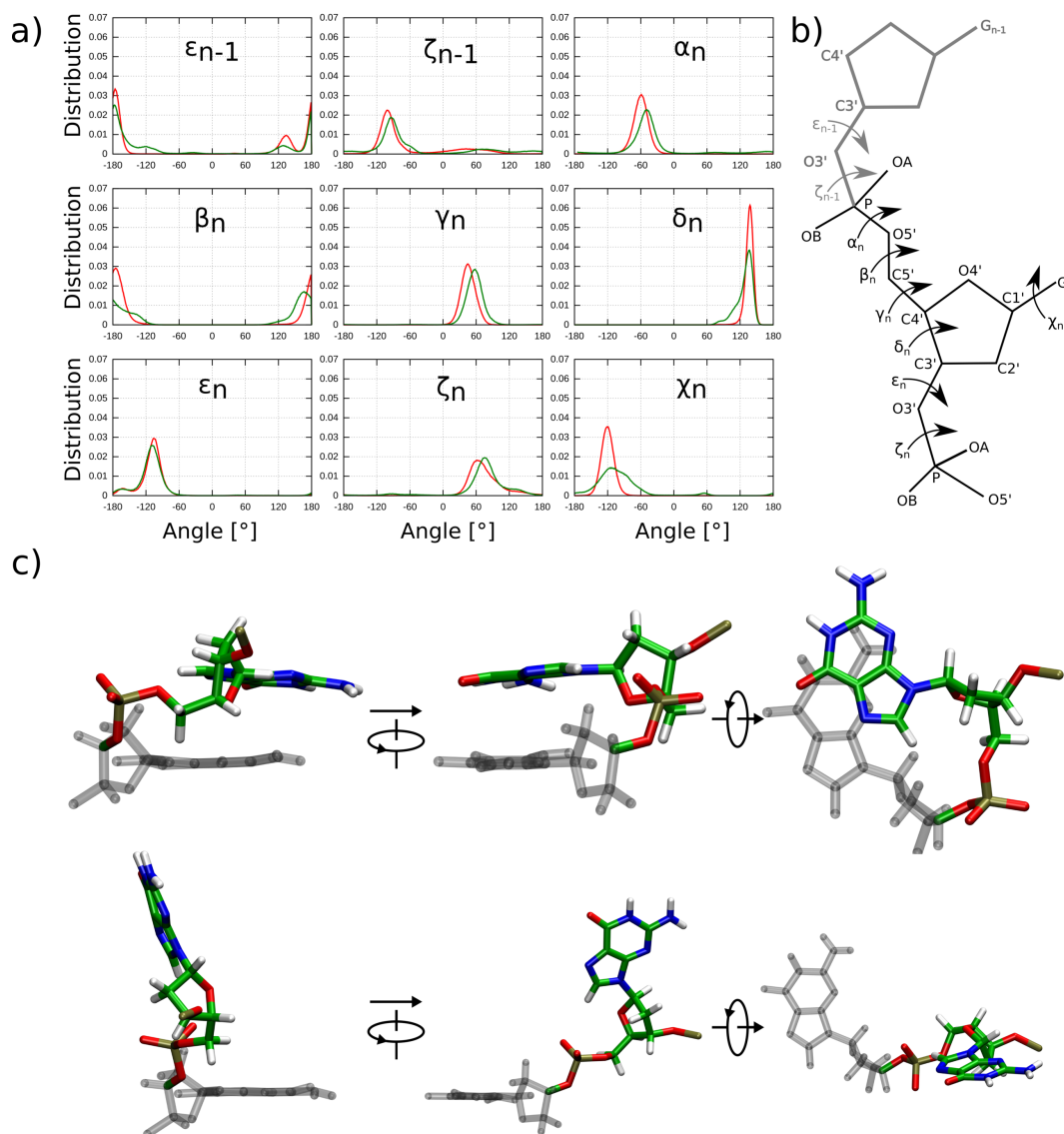


RYSUNEK A.10: Energia swobodna jako funkcja odległości między środkiem masy pojedynczej reszty guaninowej (IV na Rys. 4.1) i pozostałymi resztami guaninowymi z 3-końcowej G-tetrazy, wyznaczona w trzech temperaturach: 295 K, 300 K i 305 K.



RYSUNEK A.11: Zbieżność wartości entropii w funkcji czasu przez 8 kolejnych przedziałów wzdłuż współrzędnej  $N_{hb}$ .





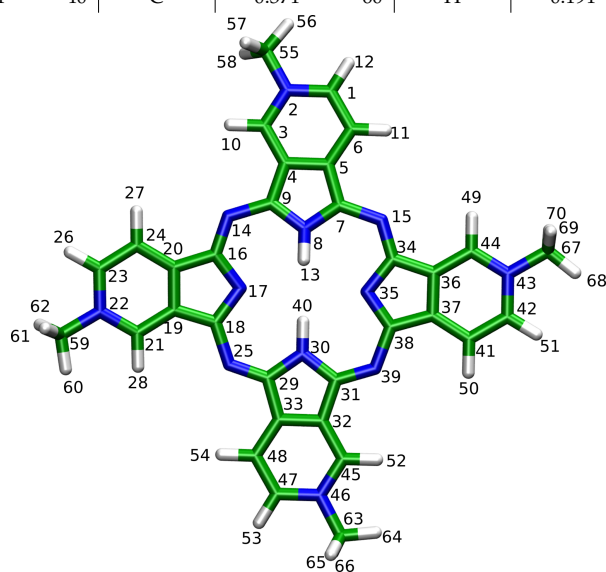
RYSUNEK A.12: A) Równowagowe rozkłady kątów dwuściennych określających geometrię nukleotydów tworzących 3'-kończącą G-tetradę w stanie natywnym (kolor czerwony) i zdysocjowanym (kolor zielony). B) Definicja kątów dwuściennych. C) Reprezentatywne ułożenia 3'-końcowej reszty guaninowej względem środkowej G-tetrady (kolor szary, przezroczysty) w stanie natywnym (górny panel) i zdysocjowanym (dolny panel).

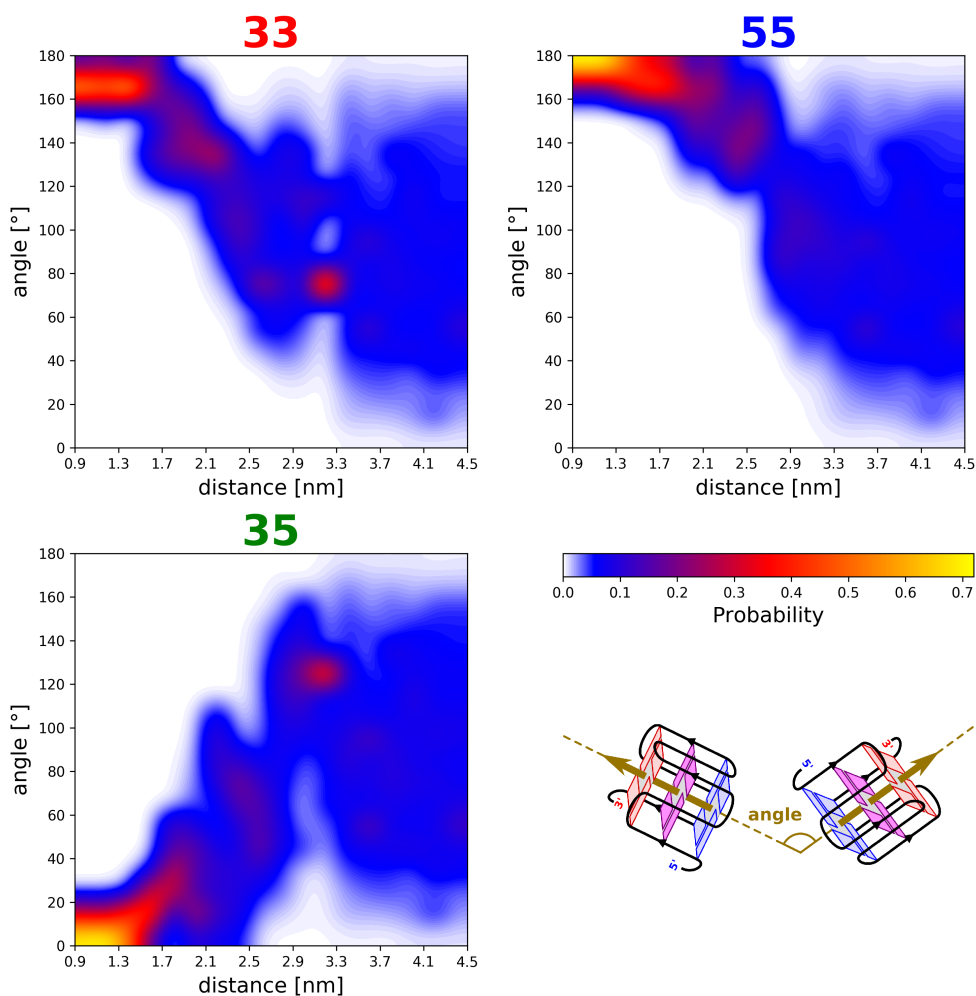
# B Załącznik

## B.1 Dimeryzacja G-kwadrupleksów o konformacji równoległej

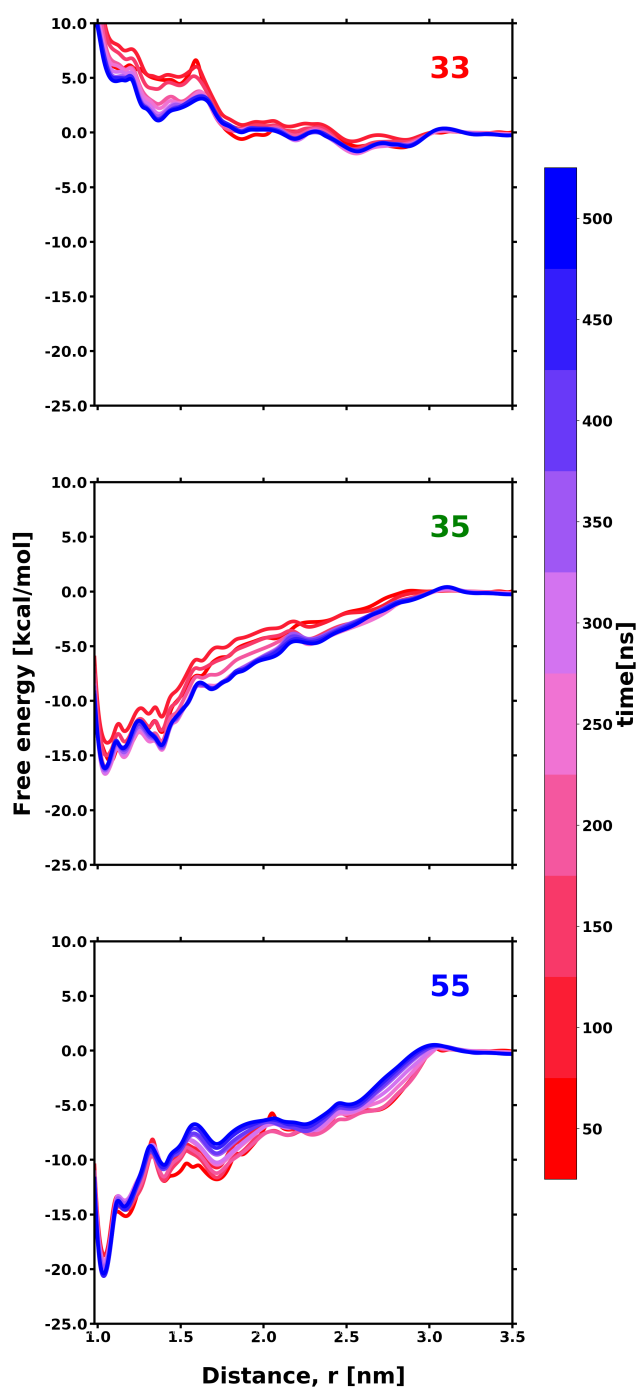
Tabela S1: Ładunki cząstkowe uzyskane dla ligandu 3,4-TMPyPz.

indeks	typ atomu	ładunek	indeks	typ atomu	ładunek	indeks	typ atomu	ładunek	indeks	typ atomu	ładunek
1	C	0.026	21	C	0.167	41	C	-0.221	61	H	0.191
2	N	0.087	22	N	0.087	42	C	0.104	62	H	0.191
3	C	0.167	23	C	0.104	43	N	0.087	63	C	-0.336
4	C	-0.294	24	C	-0.221	44	C	0.167	64	H	0.191
5	C	0.084	25	N	-0.715	45	C	0.167	65	H	0.191
6	C	-0.221	26	H	0.202	46	N	0.087	66	H	0.191
7	C	0.688	27	H	0.222	47	C	0.026	67	C	-0.336
8	N	-0.604	28	H	0.172	48	C	-0.221	68	H	0.191
9	C	0.777	29	C	0.688	49	H	0.172	69	H	0.191
10	H	0.172	30	N	-0.604	50	H	0.222	70	H	0.191
11	H	0.222	31	C	0.777	51	H	0.202			
12	H	0.202	32	C	-0.294	52	H	0.172			
13	H	0.371	33	C	0.084	53	H	0.202			
14	N	-0.715	34	C	0.777	54	H	0.222			
15	N	-0.715	35	N	-0.709	55	C	-0.336			
16	C	0.688	36	C	-0.294	56	H	0.191			
17	N	-0.709	37	C	0.084	57	H	0.191			
18	C	0.777	38	C	0.688	58	H	0.191			
19	C	-0.294	39	N	-0.715	59	C	-0.336			
20	C	0.084	40	C	0.371	60	H	0.191			

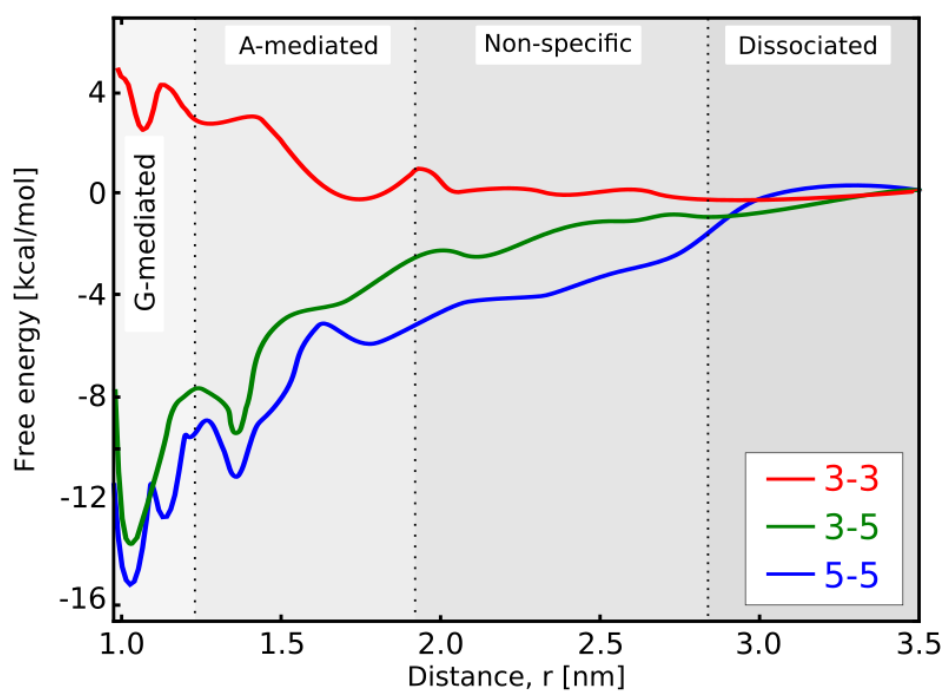




RYSUNEK B.1: Rozkłady ułożenia względem siebie dwóch G-kwadrupleksów (kątown orientacji) w trzech rozważanych modach dimeryzacji jako funkcja odległości pomiędzy środkami mas ich rdzeni guaninowych. Kąt orientacji pomiędzy G-kwadrupleksami zdefiniowałem jako kąt pomiędzy wektorami przechodzącymi przez geometryczne środki rdzeni guaninowych i jednocześnie prostopadłe do G-tetrad. Dla małych wartości odległości dimeryzacja przebiega poprzez pełną agregację i równoległe ułożenie G-tetrad z interfejsu dimeryzacji (kąt wyniósł  $\approx 180^\circ$  lub  $0^\circ$  w zależności od modu dimeryzacji). Dla większych odległości układ tworzony przez G-kwadrupleksy przyjmuje cały zakres względnej orientacji G-kwadrupleksów (kąt orientacji zawiera się w zakresie  $0-180^\circ$ ).



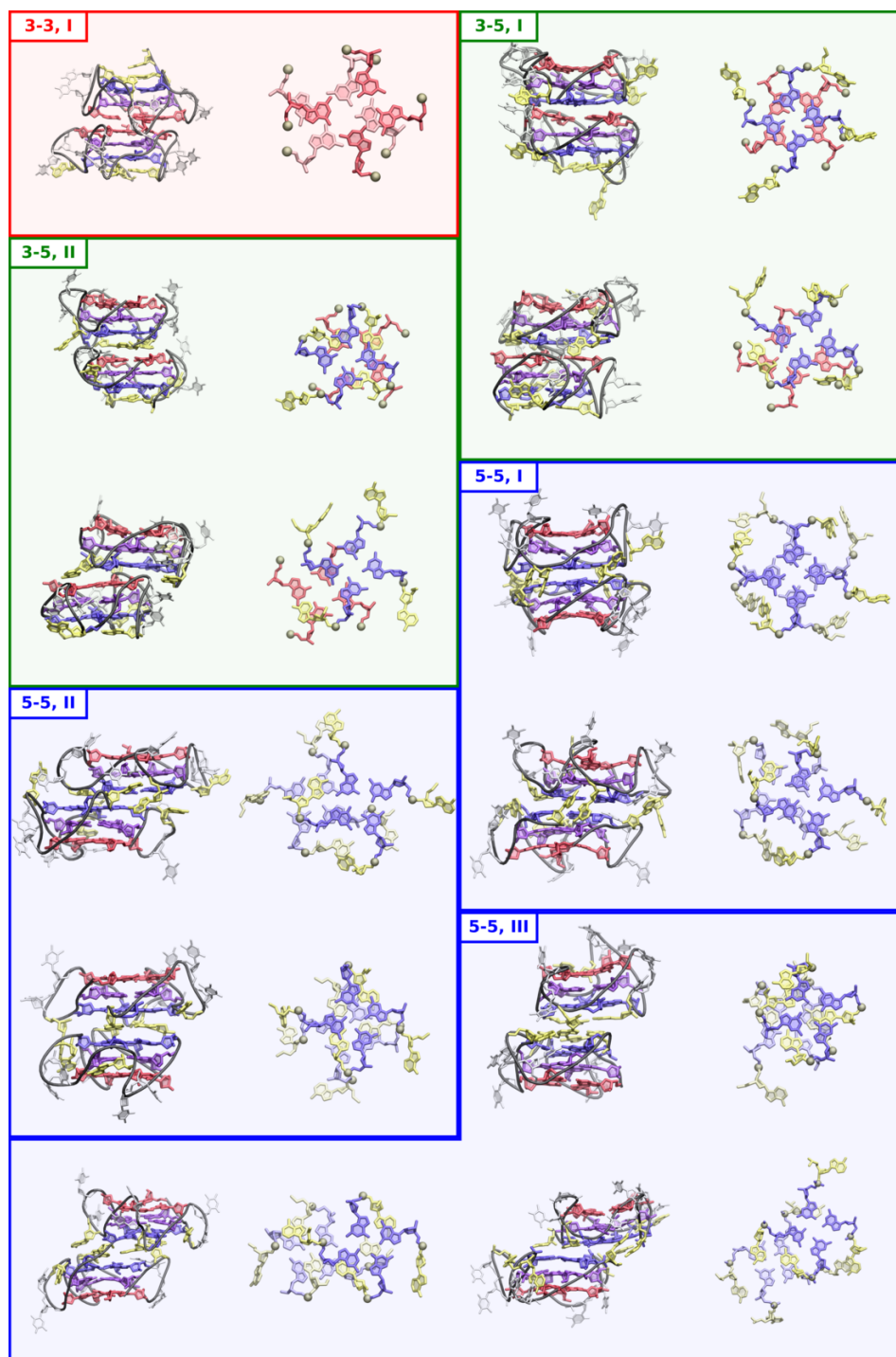
RYSUNEK B.2: Zbieżność profili energii swobodnej dimeryzacji trzech rozważanych modów dimeryzacji w czasie.



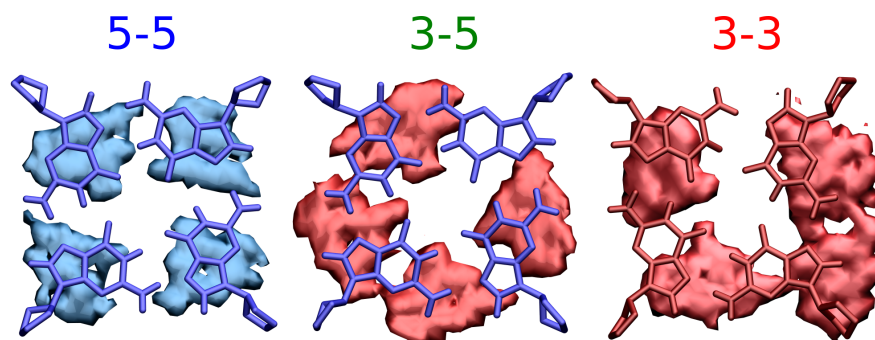
RYSUNEK B.3: Profile energii swobodnej formowania dimerów jako funkcje odległości wyznaczone dla pola siłowego parmbsc1. Pomimo tego, że tworzenie dimerów jest mniej korzystne niż w symulacjach dla pola siłowego CHARMM36, kolejność preferencji została zachowana.



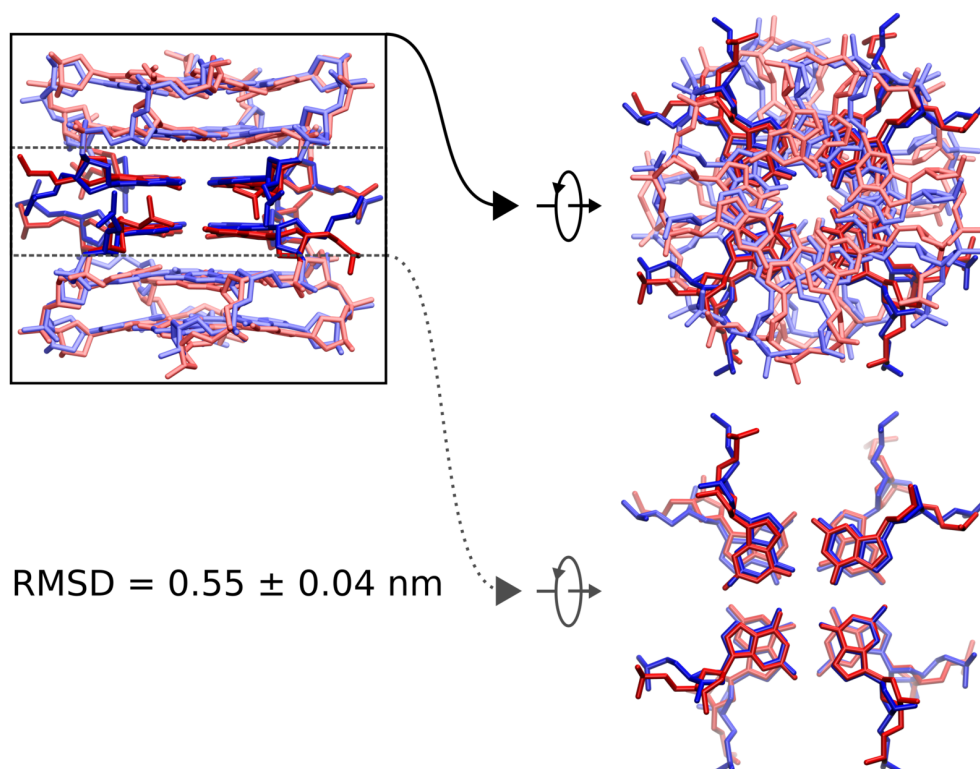
RYSUNEK B.4: Równowagowy rozkład powierzchni kontaktu pomiędzy guaninami oraz adeninami uczestniczącymi w tworzeniu interfejsu dla każdego stanu zdefiniowanego na Rys. 5.8B. Różne kolory zostały użyte dla rozróżnienia różnych klastrów z danego minimum.



RYSUNEK B.5: Widok z boku oraz góry na interfejsy dimeryzacji dla wszystkich stanów dime-rycznych zdefiniowanych dla modu 5'-5', 3'-5' oraz 3'-3'.

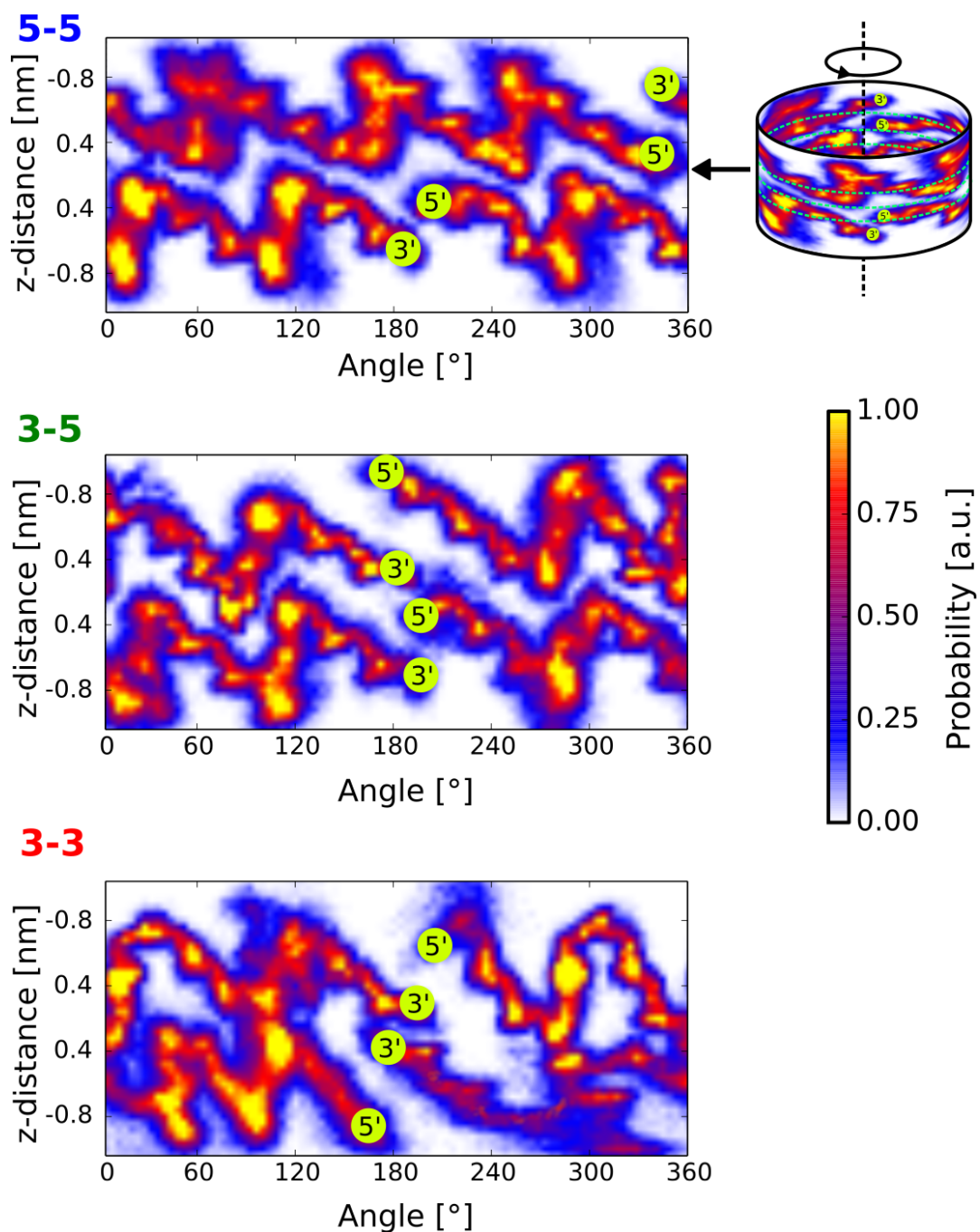


RYSUNEK B.6: Rozkład przestrzenny zasad guaninowych na interfejsie dimeryzacji dla modu 5'-5', 3'-5' oraz 3'-3'.

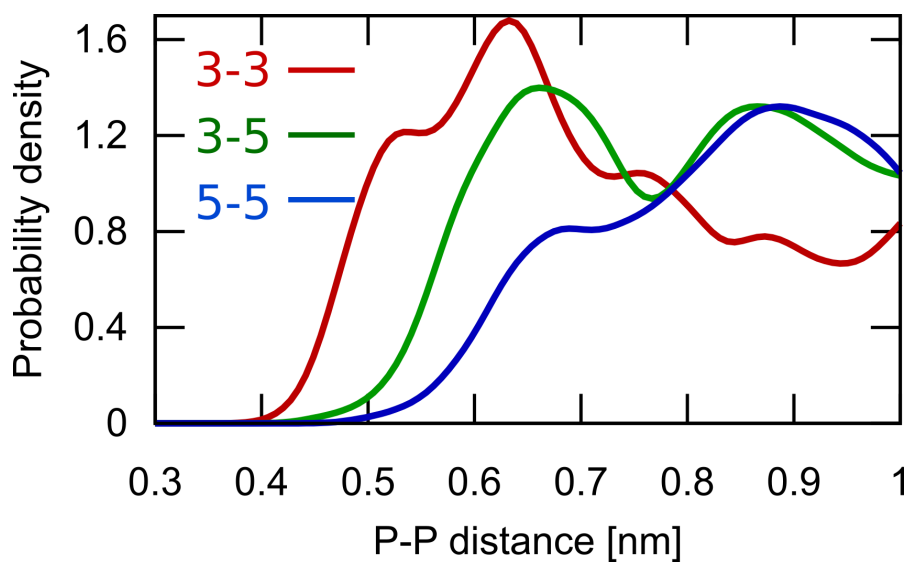


RYSUNEK B.7: Porównanie strukturalne modu 5'-5' przewidzianego przez moje symulacje wraz z interfejsem dimeryzacji występującym w strukturze telomerowego G-kwadrupleksu o topologii równoległej uzyskanego za pomocą krystalografii rentgenowskiej (PDB 1KF1). Struktury zostały nałożone na siebie na podstawie położenia grup fosforanowych. Średnia wartość RMSD dla ciężkich atomów z obu struktur wynosi 0,55 nm.

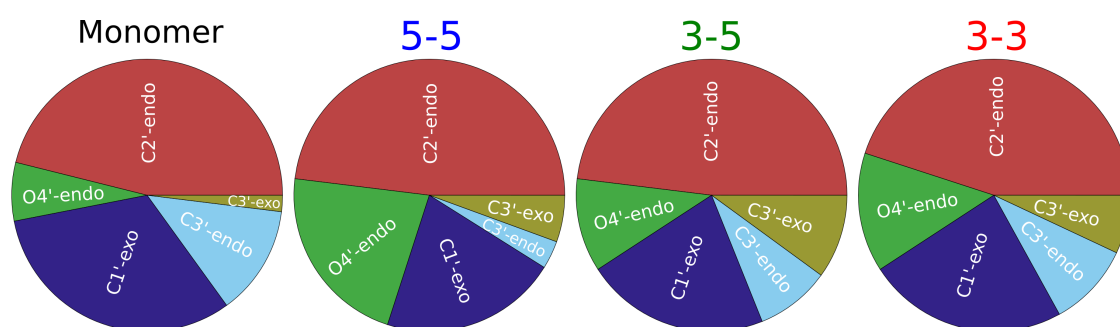




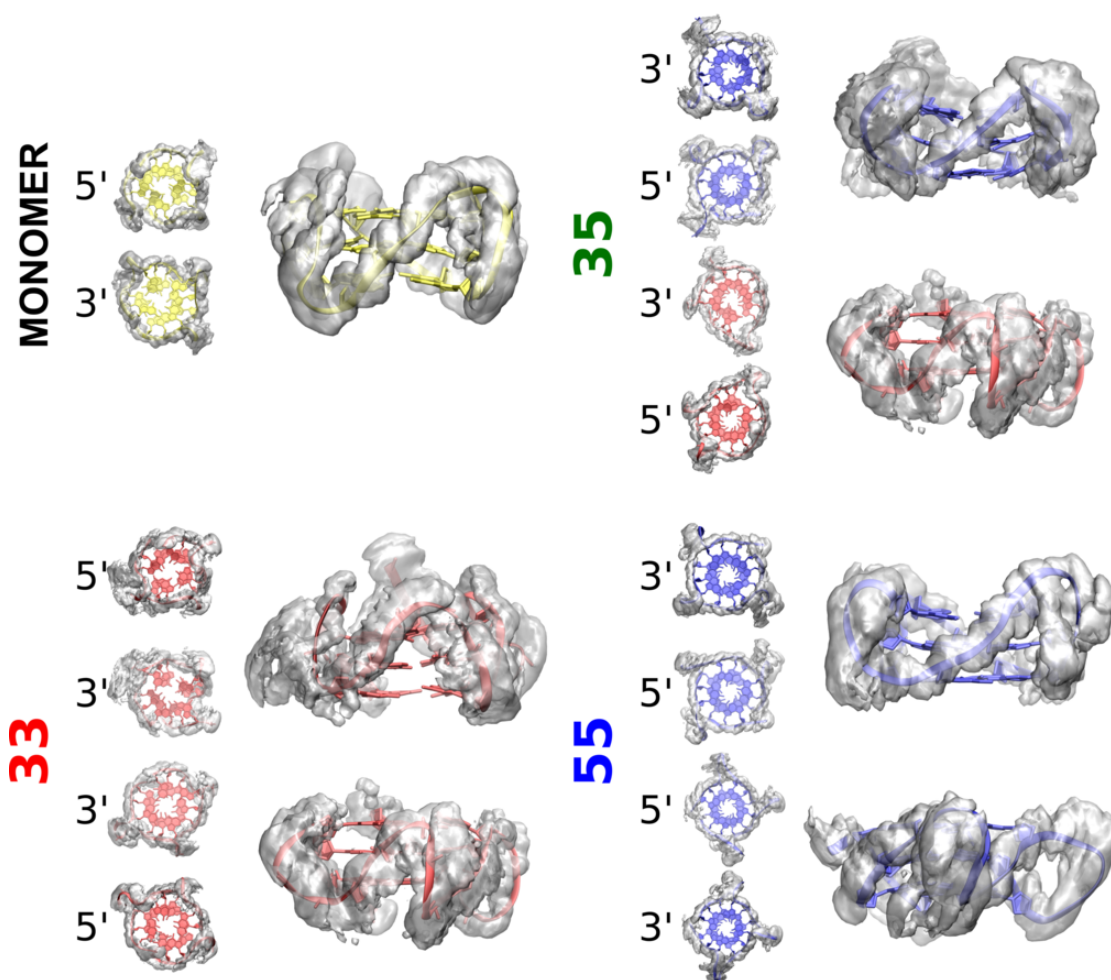
RYSUNEK B.8: Względne ułożenie nici łańcucha fosfocukrowego dwóch agregujących G-kwadru-pleksów w 3 rozważanych modach dimeryzacji. Względne ułożenie przedstawiłem jako projekcję położenia atomów łańcucha na zakrzywioną powierzchnię walca o średnicy odpowiadającej dime-rowni G-kwadru-pleksów.



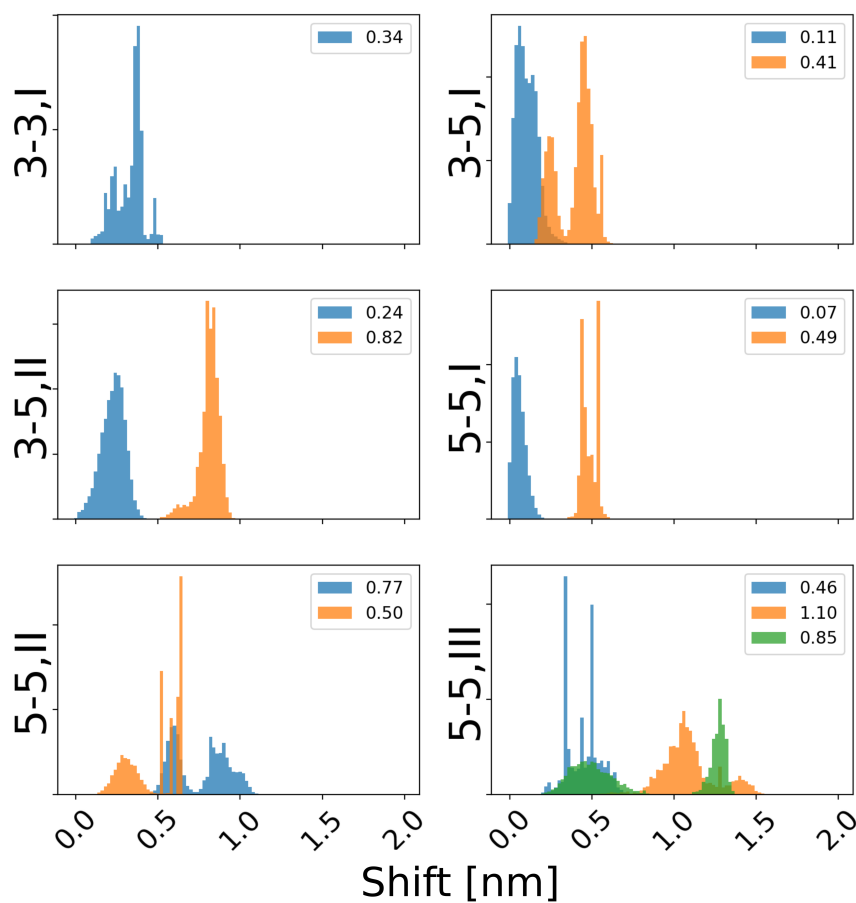
RYSUNEK B.9: Rozkład odległości pomiędzy najbliższymi sąsiednimi parami fosforanów z dwóch łańcuchów G-kwadrupleksów.



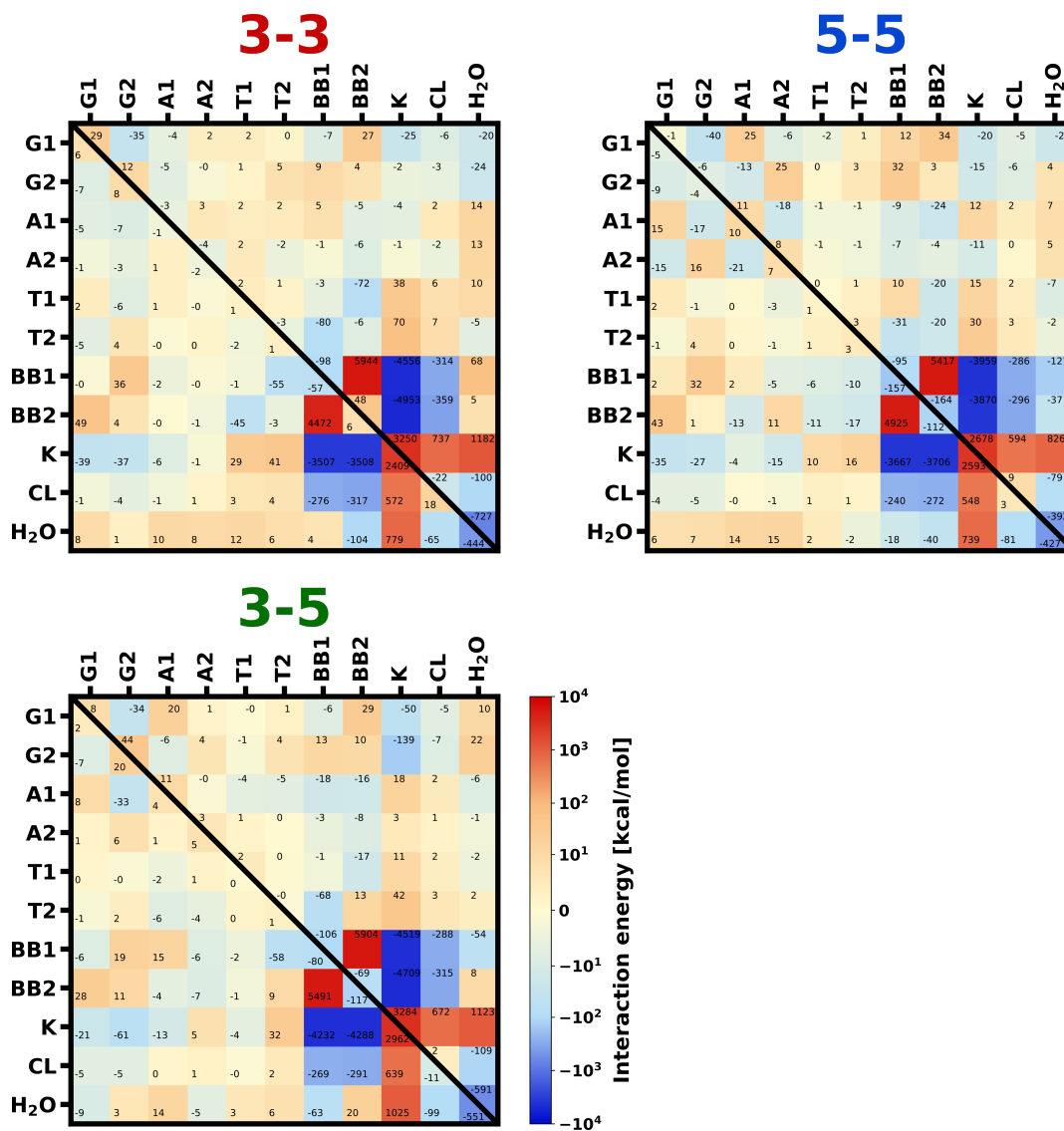
RYSUNEK B.10: Porównanie procentowe populacji konformacji pierścienia deoksyrybozy pomiędzy stanem monomerycznym oraz na 3 interfejsach dimerów.



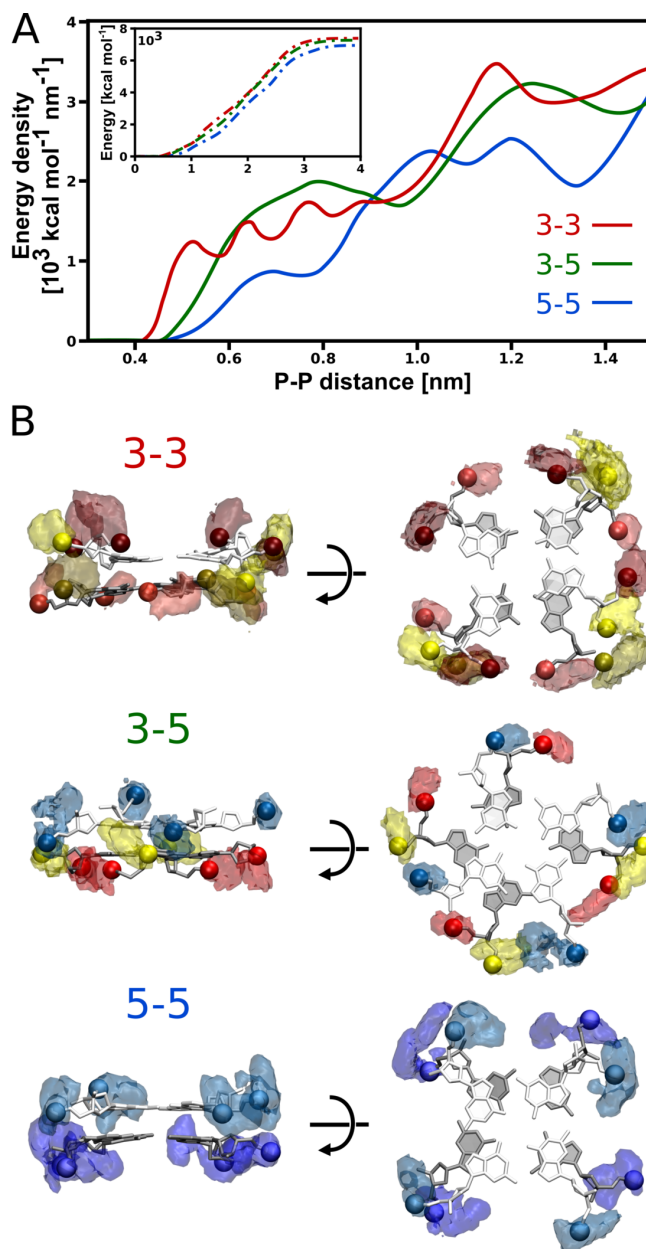
RYSUNEK B.11: Zakres fluktuacji strukturalnych łańcucha fosfocukrowego monomerów oraz dimerów G-kwadrupeksów. Izopowierzchnia przedstawia 90% zakresu maksymalnych fluktuacji. Dla każdego monomeru G-kwadrupeksu przedstawiłem rzut z boku, góry oraz dołu (od 5'- i 3'-końcowej G-tetrazy).



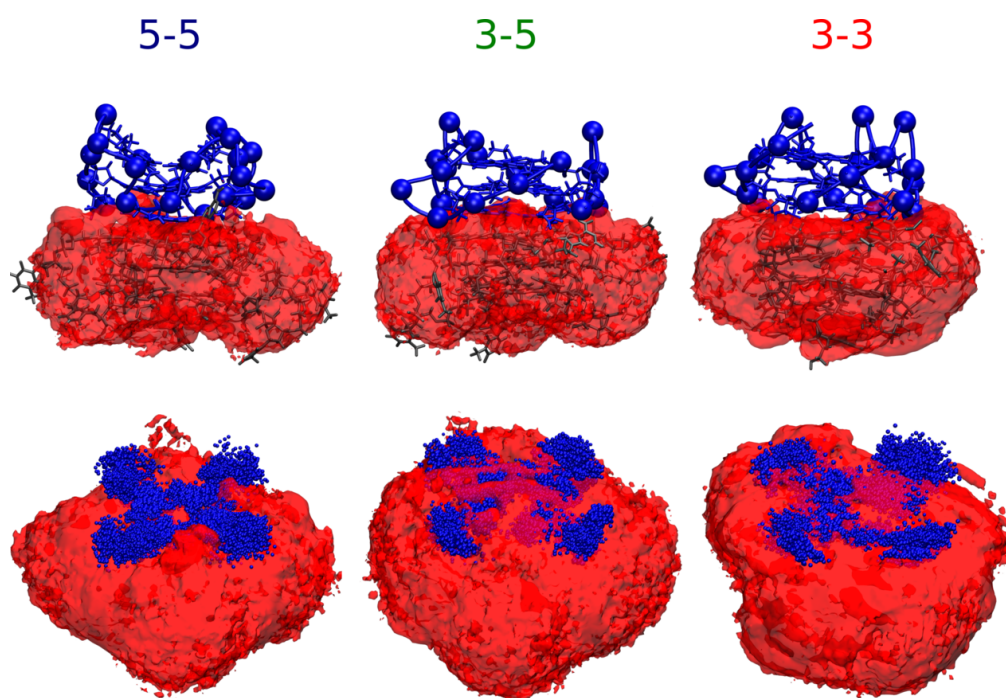
RYSUNEK B.12: Rozkład równowagowy względnego przesunięcia G-kwadrupleksów. Przesunięcie zdefiniowałem jako odległość pomiędzy osiami G-kwadrupleksów zrzutowanych na płaszczyznę interfejsu dimeryzacji. Osie G-kwadrupleksów wyznaczyłem jako wektory łączące centra mas dwóch zewnętrznych G-tetrad w poszczególnych G-kwadrupleksach. Różne kolory zostały użyte dla rozróżnienia różnych klastrów z danego minimum.



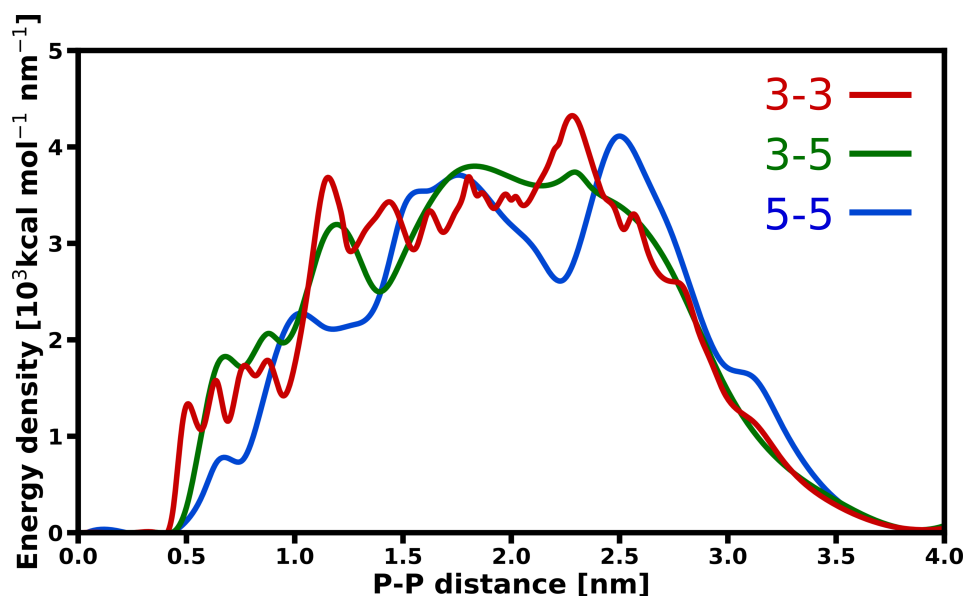
RYSUNEK B.13: A) Poszczególne wkłady do entalpii dimeryzacji ( $\Delta H$ ) wyznaczone jako zmiany energii oddziaływań pomiędzy poszczególnymi elementami strukturalnymi układu: guaniny (G), adeniny (A), tyminy (T), łańcuch fosfocukrowy DNA (BB), kationy potasu (K), aniony chlorkowe (CL) i woda (H<sub>2</sub>O) w stanie G-mediated (górny trójkąt macierzy) i w stanie A-mediated (dolny trójkąt macierzy); poszczególne monomery G-kwadrupleksów oznaczono jako 1 i 2. Dla przejrzystości wartości numeryczne energii zostały zaokrąglone do jedności.



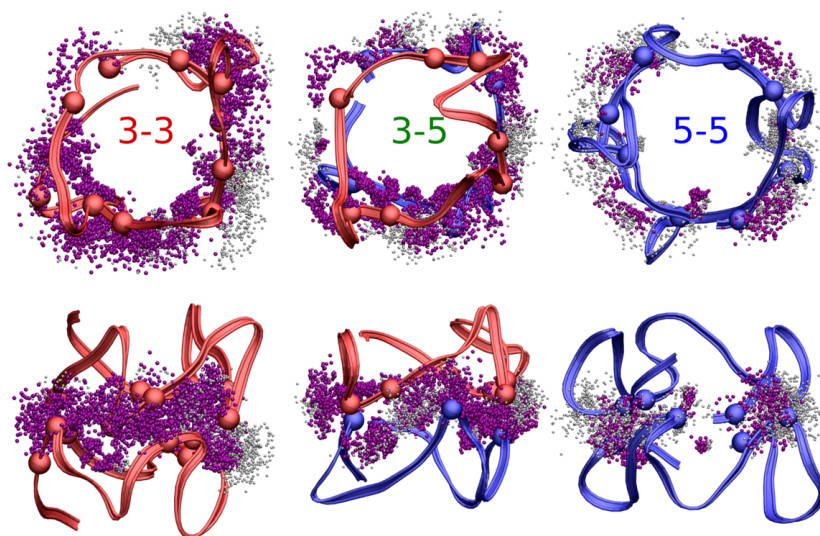
RYSUNEK B.14: A) Gęstość energii elektrostatycznej dla par oddziaływań między fosforanami w funkcji odległości między nimi (odległość P-P) dla pola siłowego parmbcs1. Wstawka pokazuje skumulowaną energię odpychania między fosforanami w funkcji odległości P-P. B) Rozkład przestrzenny wszystkich grup fosforanowych na interfejsie w odległości 1 nm do dowolnej grupy fosforanowej drugiego monomeru G-kwadupleksu. Grupy fosforanowe 3'- i 5'-końcowych G-tetrad są pokazane odpowiednio w kolorze czerwonym i niebieskim, podczas gdy pozostałe w kolorze żółtym.



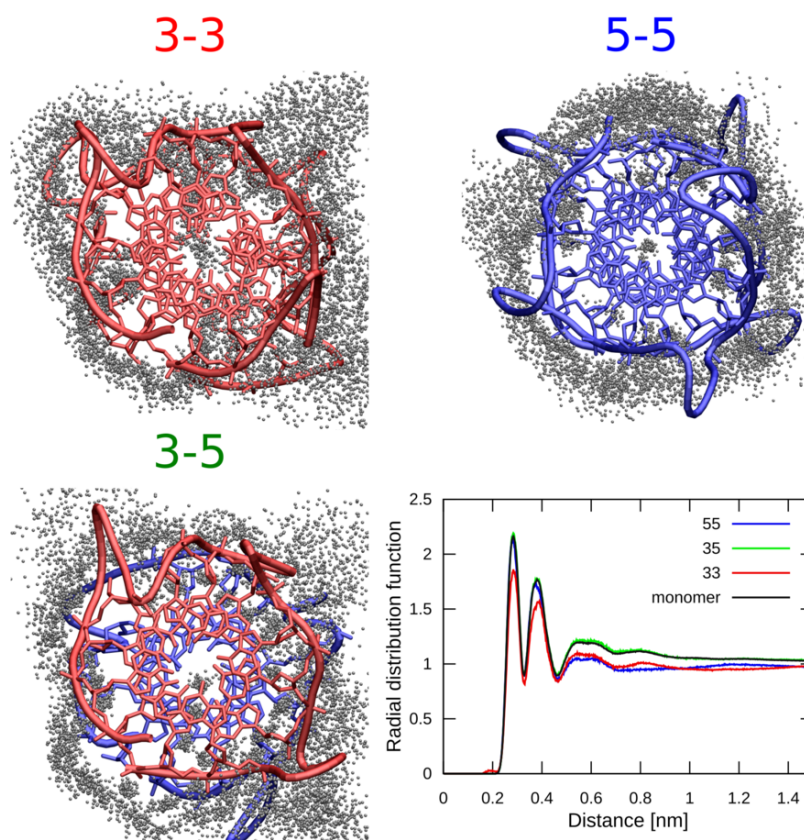
RYSUNEK B.15: Rozkład potencjału elektrostatycznego ( $-10 \text{ kT/e}$ ) wygenerowanego przez pojedynczy G-kwadrupleks w każdym z 3 modów dimeryzacji. Rozkład uzyskałem poprzez rozwiązanie liniowego równania Poissona-Boltzmana przy użyciu programu Adaptive Poisson-Boltzmann Solver (APBS) [239]. Stężenie dodatnich i ujemnych jonów o promieniu  $0,2 \text{ nm}$  ustawiłem na  $150 \text{ mM}$ . Użyłem siatki obliczeniowej o rozmiarze  $120^3$  i gęstości  $0,069 \text{ nm}$ . Zastosowaliśmy warunki brzegowe Dirichleta z wartością potencjału granicznego określoną z modelu Debye-Huckel dla pojedynczej kuli z ładunkiem punktowym, dipolowym i kwadropolowym. Niebieskie kule w górnym panelu reprezentują położenia grup fosforanowych względem drugiej jednostki G-kwadrupleksu. Niebieskie kropki w dolnym panelu pokazują rozkład wszystkich ciężkich atomów z G-tetrazy sąsiedniej jednostki G-kwadrupleksu.



RYSUNEK B.16: Pełny zakres odległości dla gęstość energii elektrostatycznej dla par oddziaływań między fosforanami w funkcji odległości między nimi (odległość P-P)

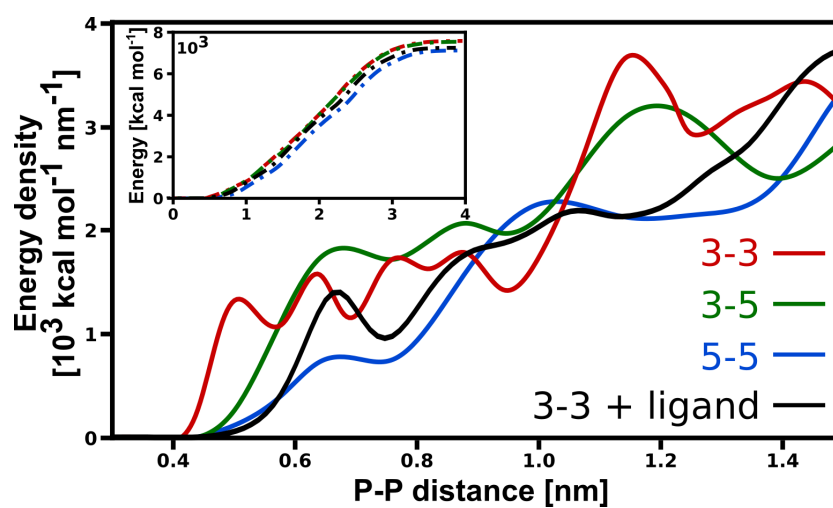


RYSUNEK B.17: Przestrenny rozkład jonów  $K^+$  na interfejsie każdego z trzech modów dime-ryzacji w stężeniu fizjologicznym KCl w polu siłowym parmbcs1, wygenerowany z 1000 klatek symulacji oddalonych od siebie co 0,5 ns. Jony  $K^+$  z interfejsu, które w tym samym czasie są w kontakcie z grupami  $PO_4$  obu jednostek G4-kwadrupeksów, są oznaczone na fioletowo.

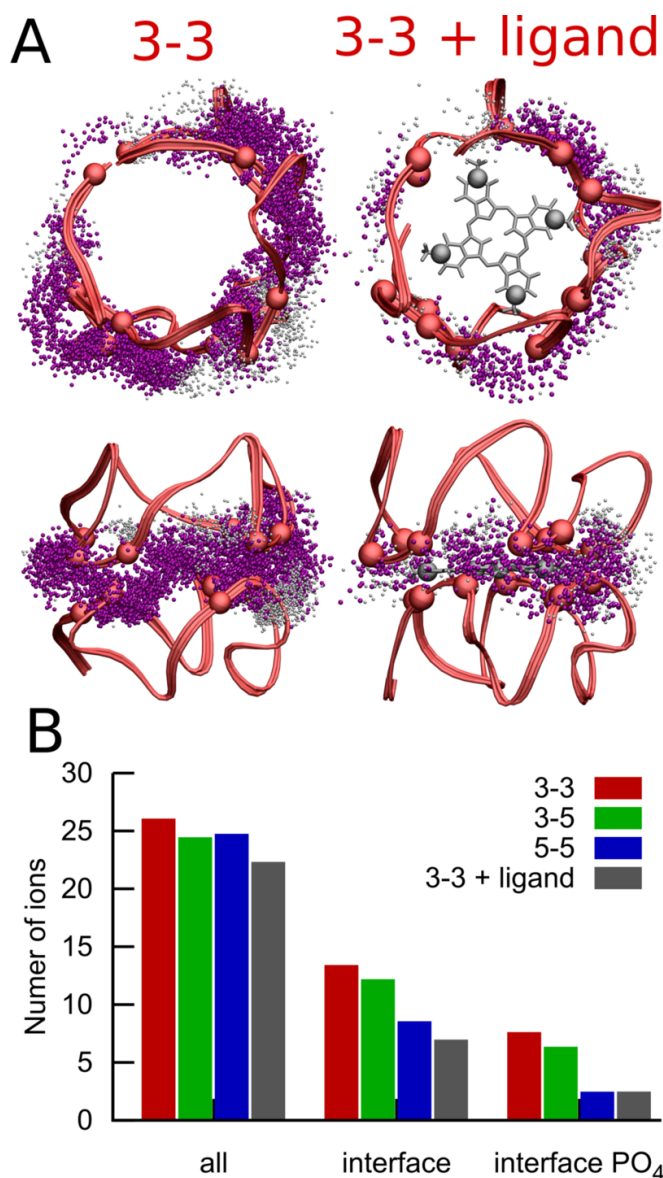


RYSUNEK B.18: Przestrenny rozkład cząsteczek wody na interfejsie każdego z trzech modów dime-ryzacji w stężeniu fizjologicznym KCl, wygenerowany z 100 klatek symulacji oddalonych od siebie co 5 ns. Wykres na dolnym panelu przedstawia radialne funkcje rozkładu cząsteczek wody względem grup fosforanowych pochodzących z zewnętrznych G-tetrad (monomer) lub G-tetrad zaangażowanych w tworzenie dimeru.





RYSUNEK B.19: Efekt obecności liganda 3,4-TMPyPz na interfejsie pomiędzy G-tetradami na gęstość energii elektrostatycznej dla par oddziaływań między fosforanami w funkcji odległości między nimi (odległość P-P). Wstawka pokazuje skumulowaną energię odpychania między fosforanami w funkcji odległości P-P.



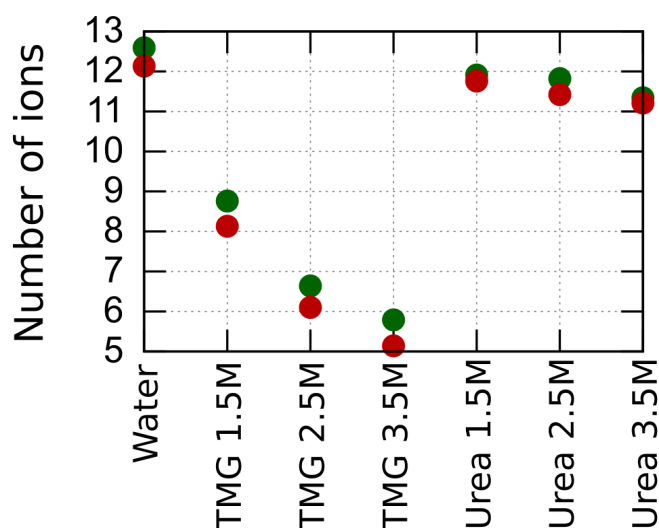
RYSUNEK B.20: A) Efekt obecności liganda 3,4-TMPyPz na interfejsie pomiędzy G-tetradami na przestrzenny rozkład jonów  $K^+$  na interfejsie 3'-3' modu dimeryzacji w stężeniu fizjologicznym KCl, wygenerowany z 1000 klatek symulacji oddalonych od siebie co 0,5 ns. Jony  $K^+$  z interfejsu, które w tym samym czasie są w kontakcie z grupami  $PO_4$  obu jednostek G4, są oznaczone na fioletowo. B) Średnia liczba jonów  $K^+$  w kontakcie z interfejsem dimeru: w odległości 0,7 nm od łańcucha DNA dowolnej jednostki G4 (all), w odległości 0,7 nm od obu łańcuchów DNA w tym samym czasie (interface), w odległości 0,7 nm do grupy  $PO_4$  dwóch jednostek G4 w tym samym czasie (interface  $PO_4$ ).

# C Załącznik

## C.1 Wpływ denaturantów na DNA

TABELA C.1: Skład cząsteczkowy wszystkich układów symulowanych użytych w badaniach.

System	Liczba cząsteczek wody	Liczba cząsteczek denaturantu	Ułamek molowy
0 M	12186	—	0
B 1.0 M	10266	257	0.021
B 1.5 M	9295	385	0.038
B 2.0 M	8397	514	0.055
B 2.5 M	7534	641	0.085
B 3.0 M	6696	769	0.115
B 3.5 M	6243	842	0.137
U 1.0 M	11296	257	0.028
U 1.5 M	10866	385	0.035
U 2.0 M	10443	514	0.049
U 2.5 M	10015	641	0.064
U 3.0 M	9588	769	0.080
U 3.5 M	9362	842	0.090
U 5.0 M	8013	1282	0.160



RYSUNEK C.1: Średnia liczba jonów sodu w pierwszej sferze hydratacji powierzchni DNA w różnych roztworach mocznika i betainy. Roztwory w natywnym stanie DNA są oznaczone zielonym kółkiem, a roztwory z denaturowanym stanem DNA są oznaczone czerwonym kółkiem.

### C.1.1 Definicja odległości minimalnej

$$MinDist = \frac{\beta}{\log \sum_i \exp(\frac{\beta}{s_i})} \quad (C.1)$$

, gdzie  $s_i$  to poszczególne odległości pomiędzy parami atomami fosforów z dwóch osobnych nici DNA. Suma przebiega po wszystkich możliwych parach odległości. Wartość  $\beta$  ustaliłem na 5. Taka metoda wyznaczania odległości pomiędzy nićmi DNA powoduje, że istotne są tylko najmniejsze wartości odległości.

### C.1.2 Współczynnik preferencyjnego oddziaływania

Współczynnik preferencyjnego oddziaływania  $\Gamma$  jest zdefiniowany według wzoru:

$$\Gamma_c(r) = \langle N_o(r) - \frac{N_o^b}{N_w^b} N_w(r) \rangle, \quad (C.2)$$

gdzie  $N_o(r)$  i  $N_w(r)$  to odpowiednio liczba cząsteczek denaturantu i wody w odległości  $r$  od DNA oraz  $N_o^b$  i  $N_w^b$  to liczba cząsteczek denaturantu i wody w całej objętości rozpuszczalnika.  $\langle \rangle$  oznacza uśrednianie po całej trajektorii MD. Ogólnie dodatnie wartości  $\Gamma_c$  wskazują na preferencyjną akumulację denaturantu lub wykluczenie wody z powierzchni białka, podczas gdy wartości ujemne oznaczają wykluczenie denaturantu lub akumulację wody.

### C.1.3 Oczekiwana liczba cząsteczek wody w pierwszej sferze hydratacji

W sytuacji, w której denaturanty nie zmieniają struktury wody, stosunek liczby cząsteczek wody w pierwszej sferze hydratacji w roztworze osmolitu ( $N_{o-DNA}$ ) i liczby cząsteczek wody w roztworze bez osmolitu ( $N_{w-DNA}$ ) powinien być zachowany i być równy stosunkowi pomiędzy liczbą cząsteczek wody w całej objętości roztworu denaturantu ( $N_o$ ) a liczbą cząsteczek wody w całej objętości roztworu ( $N_w$ ), co jest opisane następującą formułą:

$$\frac{N_{o-DNA}}{N_{w-DNA}} = \frac{N_o}{N_w} \quad (C.3)$$

Znając wartości  $N_o$ ,  $N_w$  i  $N_{o-DNA}$  można wyznaczyć oczekiwaną liczbę cząsteczek wody (gęstość wody niewykluczonej) w pierwszej sferze hydratacji w roztworze osmolitu ( $N_{o-DNA}$ ) i porównaj ją z wartością uzyskaną z rzeczywistej symulacji.

## D Udział w innych projektach badawczych

Biore także czynny udział w innych projektach badawczych częściowo związanych z tematyką mojego doktoratu. Część z nich obejmuje badania z zakresu oddziaływań małowcząsteczkowych związków organicznych (osmolitów) z pewnymi modelowymi białkami, czego efektem są dwie publikacje, których jestem współautorem.

Poza tym nawiązałem współpracę z Instytutem Chemii Bioorganicznej Polskiej Akademii Nauk w Poznaniu w ramach, której prowadzę badania nad oddziaływaniami pewnych nowo zaprojektowanych ligandów ze strukturami hybrydowych G-kwadrupleksów RNA. Modelowy cel molekularny w postaci jednoniciowej struktury RNA wyróżnia się tym, że tworzony jest przez równoległy G-kwadrupleks połączony z helisą typu A-RNA. Badane ligandy to nowe pochodne związków zaprojektowanych na podstawie znanych już ligandów takich jak Phen-DC3[240] oraz o-BMVC[241]. Analiza i interpretacja uzyskanych do tej pory danych pozwoliła na wskazanie potencjalnych miejsc wiązania ligandów do G-kwadrupleksu oraz wyznaczenie względnych preferencji wiązania pomiędzy różnymi pochodnymi.

Kolejnym projektem w którym brałem udział dotyczył badań nad nowymi związkami o działaniu przeciwnowotworowym zaburzającymi funkcje telomerów. Mój wkład do projektu polega na ustaleniu wpływu pewnych elementów strukturalnych w pochodnych triazolokrydonu C-1305 [242] na sposób i efektywność interkalacji do sekwencji telomerowej DNA. Projekt finansowany przez Narodowe Centrum Badań i Rozwoju w ramach grantu *Profilaktyka i leczenie chorób cywilizacyjnych - STRATEGMED*. Na końcu rozprawy doktorskiej umieściłem również pełny spis wypracowanego przeze mnie dorobku naukowego.

# E Dorobek naukowy

## Publikacje naukowe:

1. **M. Kogut**, C. Kleist, J. Czub, *Molecular dynamics simulations reveal the balance of forces governing the formation of a guanine tetrad - a common structural unit of G-quadruplex DNA*, Nucleic Acids Research 2016, 44(7), 3020-3030
2. B. Adamczak, M. Wieczór, **M. Kogut**, J. Stangret, J. Czub, *Molecular basis of the osmolyte effect on protein stability: lesson from the mechanical unfolding of lysozyme*, Biochemical Journal 2016, 473(20), 3705-3724
3. B. Adamczak, **M. Kogut**, J. Czub, *Effect of osmolytes on the thermal stability of proteins: replica exchange simulations of Trp-cage in urea and betaine solutions*, Physical Chemistry Chemical Physics 2018, 20(16), 11174-11182
4. P.W. Rakowska, **M. Kogut\***, J. Czub, J. Stangret, *Effect of osmolytes of different type on DNA behavior in aqueous solution. Experimental and theoretical studies*, Journal of Molecular Liquids 2018, 271, 186-201
5. **M. Kogut**, C. Kleist, J. Czub, *Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization examined in atomic detail*, PLOS Computational Biology, **w recenzji**

## Doniesienia konferencje:

1. **M. Kogut**, J. Czub, *Molecular basis of the stability of G-quadruplexes with human telomeric sequence - molecular dynamics simulations study*, From Computational Biophysics to Systems Biology, Gdańsk, 25-27.05.2014
2. **M. Kogut**, J. Czub, *Molecular basis of the stability of parallel G-quadruplex - molecular dynamics simulations study*, 3rd Annual CCP-BioSim Conference: Frontiers of Biomolecular Simulation, Edynburg, 21-23.05.2014
3. **M. Kogut**, M. Wieczór, A. Tobiszewski, J. Czub, *Molecular basis of the stability of G-quadruplexes - molecular dynamics simulation study*, 9-th European Biophysics Congress, Lizbona, 13-17.07.2013

## F Streszczenie w języku polskim

Kwas deoksyrybonukleinowy jest jednym z najistotniejszych elementów każdej żywej komórki ze względu na rolę głównego nośnika informacji genetycznej. Posiada on instrukcje potrzebne do wzrostu, budowy, funkcjonowania oraz namnażania się wielu żywych organizmów, w tym także wirusów. Okazuje się, że w zależności o warunków środowiskowych, DNA może przyjmować wiele różnych struktur drugorzędowych, z których większość stanowi formy istotne biologicznie. W związku z tym w mojej pracy doktorskiej użyłem symulacji molekularnych w celu zbadania różnych aspektów wpływających na stabilność termodynamiczną oraz zmiany konformacyjne różnych form DNA, w tym G-kwadrupeksów. G-kwadrupeksy to aktywne biologicznie struktury DNA. Ze względu na szereg funkcji jakie potencjalnie pełnią w komórkach (uważa się, że są zaangażowane m.in. w regulację replikacji i transkrypcji oraz utrzymanie stabilności genomu), są one obecnie celem wielu badań jak i potencjalnym przedmiotem terapii nowotworowych. W szczególności zbadalem molekularne determinanty pojawiające się na różnym poziomie złożoności strukturalnej, a odpowiedzialne za stabilność niekanonicznych struktur DNA typu G-kwadrupeksów. Kolejnym badanym aspektem był wpływ współrozpuszczalników na zmiany konformacyjne oraz oddziaływania stabilizujące strukturę helisy DNA. W przedstawionej pracy opisałem mechanizmy odpowiedzialne za denaturację klasycznej podwójnej helisy DNA wywołanej obecnością w roztworze dwóch modelowych denaturantów, mocznika i betainy. Ta część badań była prowadzona w ramach współpracy eksperymentalnej w Katedrze Chemii Fizycznej Politechniki Gdańskiej.

## G Streszczenie w języku angielskim

Deoxyribonucleic acid is one of the most important elements of every living cell due to the role of the main carrier of genetic information. It has the instructions needed to grow, build, function and multiply many living organisms and viruses. It turns out that depending on environmental conditions, DNA can take on many different secondary structures, most of which are biologically significant. Therefore, in my dissertation, I used molecular simulations to study various aspects affecting thermodynamic stability and conformational changes in various DNA forms with a special emphasis on the so-called G-quadruplexes. G-quadruplexes are biologically active DNA structures. Due to a number of functions that they potentially perform in cells (it is believed that they are involved, inter alia, in the regulation of replication and transcription, and maintaining genome stability), they are currently the subject of numerous studies as well as a potential target for cancer therapies. In particular, I studied the molecular determinants at different levels of structural complexity and responsible for the stability of non-canonical G-quadruplex type DNA structures. Another aspect studied was the effect of co-solvents on conformational changes and stabilizing interactions of the DNA double helix. In the presented work I described the mechanisms responsible for the denaturation of the classic DNA double helix caused by the presence in the solution of two model denaturants, urea and betaine. This part of the research was conducted as part of the experimental cooperation at the Department of Physical Chemistry of the Gdańsk University of Technology.



# Bibliography

1. Watson, J. D., Crick, F. H., *et al.* Molecular structure of nucleic acids. *Nature* **171**, 737–738 (1953).
2. Erfurth, S. C., Bond, P. J. & Peticolas, W. L. Characterization of the A–B transition of DNA in fibers and gels by laser Raman spectroscopy. *Biopolymers: Original Research on Biomolecules* **14**, 1245–1257 (1975).
3. Belmont, P., Constant, J.-F. & Demeunynck, M. Nucleic acid conformation diversity: from structure to function and regulation. *Chemical Society Reviews* **30**, 70–81 (2001).
4. Bacolla, A. & Wells, R. D. Non-B DNA conformations, genomic rearrangements, and human disease. *Journal of Biological Chemistry* **279**, 47411–47414 (2004).
5. Herr, W. Diethyl pyrocarbonate: a chemical probe for secondary structure in negatively supercoiled DNA. *Proceedings of the National Academy of Sciences* **82**, 8009–8013 (1985).
6. Rich, A. DNA comes in many forms. *Gene* **135**, 99–109 (1993).
7. Wang, A. H.-J. *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680 (1979).
8. Rich, A., Nordheim, A. & Wang, A. H.-J. The chemistry and biology of left-handed Z-DNA. *Annual review of biochemistry* **53**, 791–846 (1984).
9. Herbert, A. ADAR and Immune Silencing in Cancer. *Trends in cancer* (2019).
10. Yancey, P. H., Blake, W. R. & Conley, J. Unusual organic osmolytes in deep-sea animals: adaptations to hydrostatic pressure and other perturbants. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **133**, 667–676 (2002).
11. Kinne, R. K. The role of organic osmolytes in osmoregulation: from bacteria to mammals. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* **265**, 346–355 (1993).
12. Yancey, P. H. Compatible and counteracting solutes: protecting cells from the Dead Sea to the deep sea. *Science progress* **87**, 1–24 (2004).
13. Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D. & Somero, G. N. Living with water stress: evolution of osmolyte systems. *Science* **217**, 1214–1222 (1982).

14. Nakano, S.-i., Miyoshi, D. & Sugimoto, N. Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids. *Chemical Reviews* **114**, 2733–2758 (2014).
15. Hong, J. *et al.* Preferential interactions of glycine betaine and of urea with DNA: implications for DNA hydration and for effects of these solutes on DNA stability. *Biochemistry* **43**, 14744–14758 (2004).
16. Record Jr, M. T., Zhang, W. & Anderson, C. F. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Advances in protein chemistry* **51**, 281 (1998).
17. Klump, H. & Burkart, W. Calorimetric measurements of the transition enthalpy of DNA in aqueous urea solutions. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **475**, 601–604 (1977).
18. Texter, J. Nucleic acid-water interactions. *Progress in biophysics and molecular biology* **33**, 83–97 (1979).
19. Westhof, E. Water: an integral part of nucleic acid structure. *Annual review of biophysics and biophysical chemistry* **17**, 125–144 (1988).
20. Saenger, W. Structure and dynamics of water surrounding biomolecules. *Annual review of biophysics and biophysical chemistry* **16**, 93–114 (1987).
21. Chalikian, T. V., Plum, G. E., Sarvazyan, A. P. & Breslaver, K. J. Influence of Drug Binding on DNA Hydration: Acoustic and Densimetric Characterizations of Netropsin Binding to the Poly(dAdT)·Poly(dAdT) and Poly(dA)·Poly(dT) Duplexes and the Poly(dT)·Poly(dA)·Poly(dT) Triplex at 25°C. *Biochemistry* **33**, 8629–8640 (1994).
22. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proceedings of the National Academy of Sciences* **82**, 1376–1380 (1985).
23. Miyoshi, D. & Sugimoto, N. Molecular crowding effects on structure and stability of DNA. *Biochimie. Targeting DNA Part 1* **90**, 1040–1051 (2008).
24. Saenger, W. *Principles of nucleic acid structure* (Springer Science & Business Media, 2013).
25. Franklin, R. E. & Gosling, R. G. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallographica* **6**, 673–677 (1953).
26. Falk, M., Hartman, K. A. & Lord, R. C. Hydration of deoxyribonucleic acid. III. A spectroscopic study of the effect of hydration on the structure of deoxyribonucleic acid. *Journal of the American Chemical Society* **85**, 391–394 (1963).
27. Ivanov, V. I. *et al.* The detection of B-form/A-form junction in a deoxyribonucleotide duplex. *Biophysical journal* **71**, 3344–3349 (1996).

28. Wang, Y., Thomas, G. A. & Peticolas, W. L. A duplex of the oligonucleotides d(GGGGG-TTTTT) and d(AAAAACCCCC) forms an A to B conformational junction in concentrated salt solutions. *Journal of Biomolecular Structure and Dynamics* **6**, 1177–1187 (1989).
29. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
30. Mohr, S. C., Sokolov, N. V., He, C. M. & Setlow, P. Binding of small acid-soluble spore proteins from *Bacillus subtilis* changes the conformation of DNA from B to A. *Proceedings of the National Academy of Sciences* **88**, 77–81 (1991).
31. Arnott, S. & Selsing, E. Structures for the polynucleotide complexes poly(dA) · poly(dT) and poly(dT) · poly(dA) · poly(dT). *Journal of molecular biology* **88**, 509–521 (1974).
32. Pilet, J. & Brahms, J. Dependence of B–A conformational change in DNA on base composition. *Nature* **236**, 99–100 (1972).
33. Ivanov, V., Minchenkova, L. E., Schyolkina, A. K. & Poletayev, A. I. Different conformations of double-stranded nucleic acid in solution as revealed by circular dichroism. *Biopolymers* **12**, 89–110 (1973).
34. Pohl, F. M. Polymorphism of a synthetic DNA in solution. *Nature* **260**, 365–366 (1976).
35. Nishimura, Y., Torigoe, C. & Tsuboi, M. Salt induced B–A transition of poly(dG).poly(dC) and the stabilization of A form by its methylation. *Nucleic acids research* **14**, 2737–2748 (1986).
36. Fujimoto, S. & Yang-Xin, Y. Effect of electrolyte concentration on DNA A–B conformational transition: An unrestrained molecular dynamics simulation study. *Chinese Physics B* **19**, 088701 (2010).
37. Flett, F. & Interthal, H. Separation of DNA Oligonucleotides Using Denaturing Urea PAGE. English. *Methods in Molecular Biology* **1054** (ed Makovets, S.) 173–185 (2013).
38. Henke, W., Herdel, K., Jung, K., Schnorr, D. & Loening, S. A. Betaine Improves the PCR Amplification of GC-Rich DNA Sequences. *Nucleic Acids Research* **25**, 3957–3958 (1997).
39. Ohishi, H., Tozuka, Y., Da-Yang, Z., Ishida, T. & Nakatani, K. The rare crystallographic structure of d(CGCGCG)<sub>2</sub>: the natural spermidine molecule bound to the minor groove of left-handed Z-DNA d(CGCGCG)<sub>2</sub> at 10°C. *Biochemical and biophysical research communications* **358**, 24–28 (2007).
40. Ivar, B. Untersuchungen über die Guanylsäure. *Biochem. Biochemische Zeitschrift* **26**, 293–231 (1910).
41. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 2013 (1962).
42. Arnott, S., Chandrasekaran, R. & Marttila, C. M. Structures for polyinosinic acid and polyguanylic acid. *Biochemical Journal* **141**, 537–543 (1974).

43. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco Jr, I. & Blackburn, E. H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine· guanine base pairs. *Cell* **51**, 899–908 (1987).
44. Williamson, J. R., Raghuraman, M. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871–880 (1989).
45. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proceedings of the National Academy of Sciences* **98**, 8572–8577 (2001).
46. Shafer, R. H. & Smirnov, I. Biological aspects of DNA/RNA quadruplexes. *Biopolymers: Original Research on Biomolecules* **56**, 209–227 (2000).
47. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic acids research* **35**, 406–413 (2006).
48. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research* **33**, 2901–2907 (2005).
49. Todd, A. K. Bioinformatics approaches to quadruplex sequence location. *Methods* **43**, 246–251 (2007).
50. Huppert, J. Hunting G-quadruplexes. *Biochimie* **90**, 1140–1148 (2008).
51. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends in cell biology* **19**, 414–422 (2009).
52. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic acids research* **36**, 5482–5515 (2008).
53. Pagano, B., Mattia, C. A. & Giancola, C. Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes. *International journal of molecular sciences* **10**, 2935–2957 (2009).
54. Neidle, S. The structures of quadruplex nucleic acids and their drug complexes. *Current opinion in structural biology* **19**, 239–250 (2009).
55. Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-length-dependent folding of G-quadruplexes. *Journal of the American Chemical Society* **126**, 16405–16415 (2004).
56. Risitano, A. & Fox, K. R. Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic acids research* **32**, 2598–2606 (2004).
57. Rachwal, P. A., Findlow, I. S., Werner, J. M., Brown, T. & Fox, K. R. Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic acids research* **35**, 4214–4222 (2007).
58. Guedin, A., Alberti, P. & Mergny, J.-L. Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic acids research* **37**, 5559–5567 (2009).

59. Patel, D. J., Phan, A. T. & Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic acids research* **35**, 7429–7455 (2007).
60. Webba da Silva, M. Geometric formalism for DNA quadruplex folding. *Chemistry–A European Journal* **13**, 9738–9745 (2007).
61. Parkinson, G. N., Lee, M. P. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876 (2002).
62. Dai, J., Carver, M. & Yang, D. Polymorphism of human telomeric quadruplex structures. *Biochimie* **90**, 1172–1183 (2008).
63. Karsisiotis, A. I., O’Kane, C. & da Silva, M. W. DNA quadruplex folding formalism—a tutorial on quadruplex topologies. *Methods* **64**, 28–35 (2013).
64. Marušič, M., Šket, P., Bauer, L., Viglasky, V. & Plavec, J. Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic acids research* **40**, 6946–6956 (2012).
65. Shklover, J., Weisman-Shomer, P., Yafe, A. & Fry, M. Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression. *Nucleic acids research* **38**, 2369–2377 (2010).
66. Laughlan, G. *et al.* The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* **265**, 520–524 (1994).
67. Wang, Y. & Patel, D. J. Guanine residues in d(T<sub>2</sub>AG<sub>3</sub>) and d(T<sub>2</sub>G<sub>4</sub>) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **31**, 8112–8119 (1992).
68. Krishnan-Ghosh, Y., Liu, D. & Balasubramanian, S. Formation of an interlocked quadruplex dimer by d(GGGT). *Journal of the American Chemical Society* **126**, 11009–11016 (2004).
69. Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] G-tetraplex. *Structure* **1**, 263–282 (1993).
70. Phan, A. T. & Patel, D. J. Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *Journal of the American Chemical Society* **125**, 15021–15027 (2003).
71. Dai, J. *et al.* An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *Journal of the American Chemical Society* **128**, 1096–1098 (2006).

72. Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. & Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *Journal of molecular biology* **273**, 171–182 (1997).
73. Haider, S. M., Parkinson, G. N. & Neidle, S. Structure of a G-quadruplex–ligand complex. *Journal of molecular biology* **326**, 117–125 (2003).
74. Dexheimer, T. S., Sun, D. & Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *Journal of the American Chemical Society* **128**, 5404–5415 (2006).
75. Zhang, D.-H. & Zhi, G.-Y. Structure monomorphism of RNA G-quadruplex that is independent of surrounding condition. *Journal of biotechnology* **150**, 6–10 (2010).
76. Schultze, P., Hud, N. V., Smith, F. W. & Feigon, J. The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the *Oxytricha nova* telomere repeat oligonucleotide d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>). *Nucleic acids research* **27**, 3018–3028 (1999).
77. Miyoshi, D. & Sugimoto, N. Molecular crowding effects on structure and stability of DNA. *Biochimie* **90**, 1040–1051 (2008).
78. Miyoshi, D., Matsumura, S., Nakano, S.-i. & Sugimoto, N. Duplex dissociation of telomere DNAs induced by molecular crowding. *Journal of the American Chemical Society* **126**, 165–169 (2004).
79. Miyoshi, D., Nakao, A. & Sugimoto, N. Molecular crowding regulates the structural switch of the DNA G-quadruplex. *Biochemistry* **41**, 15017–15024 (2002).
80. Kan, Z.-y. *et al.* Molecular crowding induces telomere G-quadruplex formation under salt-deficient conditions and enhances its competition with duplex formation. *Angewandte Chemie International Edition* **45**, 1629–1632 (2006).
81. Heddi, B. & Phan, A. T. Structure of human telomeric DNA in crowded solution. *Journal of the American Chemical Society* **133**, 9824–9833 (2011).
82. Zheng, K.-w., Chen, Z., Hao, Y.-h. & Tan, Z. Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic acids research* **38**, 327–338 (2009).
83. Kan, Z.-y. *et al.* G-quadruplex formation in human telomeric (TTAGGG)<sub>4</sub> sequence with complementary strand in close vicinity under molecularly crowded condition. *Nucleic acids research* **35**, 3646–3653 (2007).
84. Miyoshi, D., Nakamura, K., Tateishi-Karimata, H., Ohmichi, T. & Sugimoto, N. Hydration of Watson-Crick base pairs and dehydration of Hoogsteen base pairs inducing structural polymorphism under molecular crowding conditions. *Journal of the American Chemical Society* **131**, 3522–3531 (2009).

85. Wells, R., Collier, D., Hanvey, J., Shimizu, M. & Wohlrab, F. The chemistry and biology of unusual DNA structures adopted by oligopurine. oligopyrimidine sequences. *The FASEB Journal* **2**, 2939–2949 (1988).
86. Simonsson, T. G-quadruplex DNA structures - Variations on a theme. *Biol.Chem.* **382**, 621–628 (2001).
87. Georgiades, S. N., Abd Karim, N. H., Suntharalingam, K. & Vilar, R. Interaction of metal complexes with G-quadruplex DNA. *Angewandte Chemie International Edition* **49**, 4020–4034 (2010).
88. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic acids research* **34**, 5402–5415 (2006).
89. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90**, 1149–1171 (2008).
90. Guschlbauer, W., Chantot, J.-F. & Thiele, D. Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA. *Journal of Biomolecular Structure and Dynamics* **8**, 491–511 (1990).
91. Miyoshi, D., Nakao, A., Toda, T. & Sugimoto, N. Effect of divalent cations on antiparallel G-quartet structure of d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>). *FEBS letters* **496**, 128–133 (2001).
92. Risitano, A. & Fox, K. R. Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry* **42**, 6507–6513 (2003).
93. Rachwal, P. A., Brown, T. & Fox, K. R. Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry* **46**, 3036–3044 (2007).
94. Huang, C.-H., Lin, Y.-S., Yang, Y.-L., Huang, S.-w. & Chen, C. W. The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Molecular microbiology* **28**, 905–916 (1998).
95. Rawal, P. *et al.* Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome research* **16**, 644–655 (2006).
96. Palumbo, S. L. *et al.* A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic acids research* **36**, 1755–1769 (2008).
97. Lim, K. W. *et al.* Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *Journal of the American Chemical Society* **132**, 12331–12342 (2010).
98. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic acids research* **34**, 3887–3896 (2006).

99. Murchie, A. I. & Lilley, D. M. Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures. *Nucleic acids research* **20**, 49–53 (1992).
100. Howell, R., Woodford, K., Weitzmann, M. & Usdin, K. The Chicken-Globin Gene Promoter Forms a Novel Cinched Tetrahelical Structure. *Journal of Biological Chemistry* **271**, 5208–5214 (1996).
101. Catasti, P., Chen, X., Moyzis, R. K., Bradbury, E. M. & Gupta, G. Structure–function correlations of the insulin-linked polymorphic region. *Journal of molecular biology* **264**, 534–545 (1996).
102. Arthanari, H. & Bolton, P. H. Functional and dysfunctional roles of quadruplex DNA in cells. *Chemistry & biology* **8**, 221–230 (2001).
103. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic acids research* **33**, 2908–2916 (2005).
104. Neumann, A. A. & Reddel, R. R. Telomere maintenance and cancer? look, no telomerase. *Nature Reviews Cancer* **2**, 879 (2002).
105. Petracek, M. E., Lefebvre, P. A., Silflow, C. D. & Berman, J. Chlamydomonas telomere sequences are A+T-rich but contain three consecutive G-C base pairs. *Proceedings of the National Academy of Sciences* **87**, 8222–8226 (1990).
106. Oka, Y. & Thomas Jr, C. The cohering telomeres of Oxytricha. *Nucleic acids research* **15**, 8877–8898 (1987).
107. Lundblad, V. & Blackburn, E. H. An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* **73**, 347–360 (1993).
108. Yang, Q., Zheng, Y.-L. & Harris, C. C. POT1 and TRF2 cooperate to maintain telomeric integrity. *Molecular and cellular biology* **25**, 1070–1080 (2005).
109. Giannone, R. J. *et al.* The protein network surrounding the human telomere repeat binding factors TRF1, TRF2, and POT1. *PLoS One* **5**, e12407 (2010).
110. Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
111. Neidle, S. & Parkinson, G. N. The structure of telomeric DNA. *Current opinion in structural biology* **13**, 275–283 (2003).
112. Bekaert, S., De Meyer, T. & Van Oostveldt, P. Telomere attrition as ageing biomarker. *Anticancer research* **25**, 3011–3021 (2005).
113. Phan, A. T. & Mergny, J.-L. Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. *Nucleic acids research* **30**, 4618–4625 (2002).



114. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature structural & molecular biology* **12**, 847 (2005).
115. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry* **5**, 182 (2013).
116. Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nature chemical biology* **8**, 301 (2012).
117. Zhang, N., Phan, A. T. & Patel, D. J. (3+ 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *Journal of the American Chemical Society* **127**, 17277–17285 (2005).
118. Rujan, I. N., Meleney, J. C. & Bolton, P. H. Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic acids research* **33**, 2022–2031 (2005).
119. Ambrus, A. *et al.* Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Research* **34**, 2723–2735 (Jan. 2006).
120. Lim, K. W. *et al.* Sequence variant (CTAGGG)<sub>n</sub> in the human telomere favors a G-quadruplex structure containing a G·C·G·C tetrad. *Nucleic Acids Research* **37**, 6239–6248 (2009).
121. Schultze, P., Macaya, R. F. & Feigon, J. Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *Journal of molecular biology* **235**, 1532–1547 (1994).
122. Marathias, V. M. & Bolton, P. H. Structures of the potassium-saturated, 2:1, and intermediate, 1:1, forms of a quadruplex DNA. *Nucleic acids research* **28**, 1969–1977 (2000).
123. Marathias, V. M. *et al.* Determination of the number and location of the manganese binding sites of DNA quadruplexes in solution by EPR and NMR in the presence and absence of thrombin. *Journal of molecular biology* **260**, 378–394 (1996).
124. Heddi, B., Martin-Pintado, N., Serimbetov, Z., Kari, T. M. A. & Phan, A. T. G-quadruplexes with (4n-1) guanines in the G-tetrad core: formation of a G-triad water complex and implication for small-molecule binding. *Nucleic acids research* **44**, 910–916 (2016).
125. A double chain reversal loop and two diagonal loops define the architecture of unimolecular DNA quadruplex containing a pair of stacked G(syn)·G(syn)·G(anti)·G(anti) tetrads flanked by a G·(T-T) triad and a T·T·T triple<sup>11</sup>Edited by M. F. Summers. *Journal of Molecular Biology* **310**, 181–194 (2001).

126. Hu, L., Lim, K. W., Bouaziz, S. & Phan, A. T. Giardia Telomeric Sequence d(TAGGG)<sub>4</sub> Forms Two Intramolecular G-Quadruplexes in K<sup>+</sup> Solution: Effect of Loop Length and Sequence on the Folding Topology. *Journal of the American Chemical Society* **131**. PMID: 19874015, 16824–16831 (2009).
127. Lim, K. W. *et al.* Structure of the human telomere in K<sup>+</sup> solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *Journal of the American Chemical Society* **131**, 4301–4309 (2009).
128. Kankia, B. *et al.* Stable domain assembly of a monomolecular DNA quadruplex: implications for DNA-based nanoswitches. *Biophysical journal* **110**, 2169–2175 (2016).
129. Saintomé, C., Amrane, S., Mergny, J.-L. & Alberti, P. The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic acids research* **44**, 2926–2935 (2016).
130. Smargiasso, N. *et al.* G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *Journal of the American Chemical Society* **130**, 10208–10216 (2008).
131. Do, N. Q. & Phan, A. T. Monomer–Dimer Equilibrium for the 5′–5′ Stacking of Propeller-Type Parallel-Stranded G-Quadruplexes: NMR Structural Study. *Chemistry-A European Journal* **18**, 14752–14759 (2012).
132. Matsugami, A. *et al.* An intramolecular quadruplex of (GGA)<sub>4</sub> triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction 1. *Journal of molecular biology* **313**, 255–269 (2001).
133. Do, N. Q., Lim, K. W., Teo, M. H., Heddi, B. & Phan, A. T. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic acids research* **39**, 9448–9457 (2011).
134. Kuryavyi, V., Cahoon, L. A., Seifert, H. S. & Patel, D. J. RecA-binding pilE G4 sequence essential for pilin antigenic variation forms monomeric and 5′ end-stacked dimeric parallel G-quadruplexes. *Structure* **20**, 2090–2102 (2012).
135. Adrian, M. *et al.* Structure and conformational dynamics of a stacked dimeric G-quadruplex formed by the human CEB1 minisatellite. *Journal of the American Chemical Society* **136**, 6297–6305 (2014).
136. Nicoludis, J. M. *et al.* Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *Journal of the American Chemical Society* **134**, 20446–20456 (2012).
137. Meier, M. *et al.* Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic acids research* **46**, 5319–5331 (2018).
138. Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *The FEBS journal* **277**, 1107–1117 (2010).

139. Yu, H.-Q., Miyoshi, D. & Sugimoto, N. Characterization of structure and stability of long telomeric DNA G-quadruplexes. *Journal of the American Chemical Society* **128**, 15461–15468 (2006).
140. Bryan, T. M. & Cech, T. R. Telomerase and the maintenance of chromosome ends. *Current opinion in cell biology* **11**, 318–324 (1999).
141. Collie, G. W. *et al.* Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. *Journal of the American Chemical Society* **134**, 2723–2731 (2012).
142. Huang, X.-X. *et al.* Two cationic porphyrin isomers showing different multimeric G-quadruplex recognition specificity against monomeric G-quadruplexes. *Nucleic acids research* **42**, 8719–8731 (2014).
143. Parkinson, G. N., Cuenca, F. & Neidle, S. Topology conservation and loop flexibility in quadruplex–drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex–drug complexes. *Journal of molecular biology* **381**, 1145–1156 (2008).
144. Bai, L.-P., Hagihara, M., Jiang, Z.-H. & Nakatani, K. Ligand binding to tandem G quadruplexes from human telomeric DNA. *ChemBioChem* **9**, 2583–2587 (2008).
145. Campbell, N. H., Parkinson, G. N., Reszka, A. P. & Neidle, S. Structural basis of DNA quadruplex recognition by an acridine drug. *Journal of the American Chemical Society* **130**, 6722–6724 (2008).
146. Haider, S. M. & Neidle, S. *A molecular model for drug binding to tandem repeats of telomeric G-quadruplexes* 2009.
147. Bazzicalupi, C., Ferraroni, M., Bilia, A. R., Scheggi, F. & Gratteri, P. The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic acids research* **41**, 632–638 (2012).
148. Langley, D. R. Molecular Dynamic Simulations of Environment and Sequence Dependent DNA Conformations: The Development of the BMS Nucleic Acid Force Field and Comparison with Experimental Results. *Journal of Biomolecular Structure and Dynamics* **16**. PMID: 10052609, 487–509 (1998).
149. Sen, S., Andreatta, D., Ponomarev, S. Y., Beveridge, D. L. & Berg, M. A. Dynamics of Water and Ions Near DNA: Comparison of Simulation to Time-Resolved Stokes-Shift Experiments. *Journal of the American Chemical Society* **131**. PMID: 19191698, 1724–1735 (2009).
150. Besler, B. H., Merz Jr, K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry* **11**, 431–439 (1990).
151. Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **110**, 1657–1666 (1988).

152. Soares, T. A. *et al.* An improved nucleic acid parameter set for the GROMOS force field. *Journal of computational chemistry* **26**, 725–737 (2005).
153. Hart, K. *et al.* Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory. Comput.* **8**, 348–362 (2011).
154. Galindo-Murillo, R. *et al.* Assessing the current state of AMBER force field modifications for DNA. *Journal of chemical theory and computation* **12**, 4114–4127 (2016).
155. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926–935 (1983).
156. Leach, A. R. & Leach, A. *Molecular modelling: principles and applications* (Pearson education, 2001).
157. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952–962 (1992).
158. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation* **4**, 116–122 (2008).
159. Den Otter, W. K. & Briels, W. J. The calculation of free-energy differences by constrained molecular-dynamics simulations. *The Journal of chemical physics* **109**, 4139–4146 (1998).
160. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **13**, 1011–1021 (1992).
161. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. Multi-dimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry* **16**, 1339–1350 (1995).
162. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926–935 (1983).
163. Foloppe, N. & MacKerell Jr., A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry* **21**, 86–104 (2000).
164. 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).
165. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).

166. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).
167. Williams, T., Kelley, C. & many others. *Gnuplot 4.4: an interactive plotting program* <http://gnuplot.sourceforge.net/>. 2010.
168. Inkscape Project. *Inkscape* version 0.28.1. Mar. 4, 2011.
169. Bonomi, M. *et al.* PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **180**, 1961–1972 (2009).
170. Darve, E. & Pohorille, A. Calculating free energies using average force. *The Journal of Chemical Physics* **115**, 9169–9183 (2001).
171. Hénin, J., Fiorin, G., Chipot, C. & Klein, M. L. Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. *Journal of Chemical Theory and Computation* **6**, 35–47 (2010).
172. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).
173. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **103**, 4613–4621 (1995).
174. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327–341 (1977).
175. Phan, A. T., Kuryavyi, V., Luu, K. N. & Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K<sup>+</sup> solution. *Nucleic acids research* **35**, 6517–6525 (2007).
176. Lim, K. W. *et al.* Sequence variant (CTAGGG)<sub>n</sub> in the human telomere favors a G-quadruplex structure containing a G·C·G·C tetrad. *Nucleic acids research* **37**, 6239–6248 (2009).
177. Stadlbauer, P. *et al.* Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic acids research* **43**, 9626–9644 (2015).
178. Bian, Y. *et al.* Atomistic picture for the folding pathway of a hybrid-1 type human telomeric DNA G-quadruplex. *PLoS computational biology* **10**, e1003562 (2014).
179. Kogut, M., Kleist, C. & Czub, J. Molecular dynamics simulations reveal the balance of forces governing the formation of a guanine tetrad—a common structural unit of G-quadruplex DNA. *Nucleic acids research* **44**, 3020–3030 (2016).

180. Tawani, A., Mishra, S. K. & Kumar, A. Structural insight for the recognition of G-quadruplex structure at human c-myc promoter sequence by flavonoid Quercetin. *Scientific reports* **7**, 3600 (2017).
181. Zhou, J.-K., Yang, D.-Y. & Sheu, S.-Y. The molecular mechanism of ligand unbinding from the human telomeric G-quadruplex by steered molecular dynamics and umbrella sampling simulations. *Physical Chemistry Chemical Physics* **17**, 12857–12869 (2015).
182. Akhshi, P., Acton, G. & Wu, G. Molecular dynamics simulations to provide new insights into the asymmetrical ammonium ion movement inside of the  $[d(G_3T_4G_4)]_2$  G-quadruplex DNA structure. *The Journal of Physical Chemistry B* **116**, 9363–9370 (2012).
183. Akhshi, P. & Wu, G. Umbrella sampling molecular dynamics simulations reveal concerted ion movement through G-quadruplex DNA channels. *Physical Chemistry Chemical Physics* **19**, 11017–11025 (2017).
184. Ivani, I. *et al.* Parmbsc1: a refined force field for DNA simulations. *Nature methods* **13**, 55 (2016).
185. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **31**, 671–690 (2010).
186. Frisch, M. J. *et al.* *Gaussian 16 Revision B.01* Gaussian Inc. Wallingford CT. 2016.
187. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126**, 014101 (2007).
188. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52**, 7182–7190 (1981).
189. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. Multi-dimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry* **16**, 1339–1350 (1995).
190. Deng, N., Wickstrom, L., Cieplak, P., Lin, C. & Yang, D. Resolving the ligand-binding specificity in c-MYC G-quadruplex DNA: absolute binding free energy calculations and SPR experiment. *The Journal of Physical Chemistry B* **121**, 10484–10497 (2017).
191. Lee, M. S. & Olson, M. A. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophysical journal* **90**, 864–877 (2006).
192. Colasanti, A. V., Lu, X.-J. & Olson, W. K. Analyzing and building nucleic acid structures with 3DNA. *Journal of visualized experiments: JoVE* (2013).
193. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926–935 (1983).

194. Foloppe, N. & MacKerell Jr, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of computational chemistry* **21**, 86–104 (2000).
195. Adamczak, B., Wieczór, M., Kogut, M., Stangret, J. & Czub, J. Molecular basis of the osmolyte effect on protein stability: a lesson from the mechanical unfolding of lysozyme. *Biochem. J.* **473**, 3705–3724 (2016).
196. Islam, B. *et al.* Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale. *Nucleic acids research* **41**, 2723–2735 (2013).
197. Wei, D., Husby, J. & Neidle, S. Flexibility and structural conservation in a c-KIT G-quadruplex. *Nucleic acids research* **43**, 629–644 (2014).
198. Islam, B. *et al.* Extended molecular dynamics of a c-kit promoter quadruplex. *Nucleic acids research* **43**, 8673–8693 (2015).
199. Cang, X., Šponer, J. & Cheatham III, T. E. Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic acids research* **39**, 4499–4512 (2011).
200. Šponer, J. *et al.* Relative stability of different DNA guanine quadruplex stem topologies derived using large-scale quantum-chemical computations. *Journal of the American Chemical Society* **135**, 9785–9796 (2013).
201. Lech, C. J., Heddi, B. & Phan, A. T. Guanine base stacking in G-quadruplex nucleic acids. *Nucleic acids research* **41**, 2034–2046 (2012).
202. Ghimire, C. *et al.* Direct quantification of loop interaction and  $\pi$ - $\pi$  stacking for G-quadruplex stability at the submolecular level. *Journal of the American Chemical Society* **136**, 15537–15544 (2014).
203. Heddi, B. & Phan, A. T. Structure of Human Telomeric DNA in Crowded Solution. *Journal of the American Chemical Society* **133**, 9824–9833 (2011).
204. Fujimoto, T., Nakano, S.-i., Sugimoto, N. & Miyoshi, D. Thermodynamics-hydration relationships within loops that affect G-quadruplexes under molecular crowding conditions. *The Journal of Physical Chemistry B* **117**, 963–972 (2012).
205. Miyoshi, D., Nakao, A. & Sugimoto, N. Structural transition from antiparallel to parallel G-quadruplex of d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>) induced by Ca<sup>2+</sup>. *Nucleic acids research* **31**, 1156–1163 (2003).
206. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research* **34**, 5402–5415 (2006).
207. Gray, R. D., Trent, J. O. & Chaires, J. B. Folding and unfolding pathways of the human telomeric G-quadruplex. *Journal of molecular biology* **426**, 1629–1650 (2014).

208. Gray, R. D., Petraccone, L., Trent, J. O. & Chaires, J. B. Characterization of a K<sup>+</sup>-induced conformational switch in a human telomeric DNA oligonucleotide using 2-aminopurine fluorescence. *Biochemistry* **49**, 179–194 (2009).
209. Bončina, M., Lah, J., Prislán, I. & Vesnaver, G. Energetic basis of human telomeric DNA folding into G-quadruplex structures. *Journal of the American Chemical Society* **134**, 9657–9663 (2012).
210. Sengar, A., Heddi, B. & Phan, A. T. Formation of G-quadruplexes in poly-G sequences: structure of a propeller-type parallel-stranded G-quadruplex formed by a G15 stretch. *Biochemistry* **53**, 7718–7723 (2014).
211. Andricioaei, I. & Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *The Journal of Chemical Physics* **115**, 6289–6292 (2001).
212. Yu, Z. *et al.* Click chemistry assisted single-molecule fingerprinting reveals a 3D biomolecular folding funnel. *Journal of the American Chemical Society* **134**, 12338–12341 (2012).
213. Li, W., Hou, X.-M., Wang, P.-Y., Xi, X.-G. & Li, M. Direct measurement of sequential folding pathway and energy landscape of human telomeric G-quadruplex structures. *Journal of the American Chemical Society* **135**, 6423–6426 (2013).
214. Bamberg, E. *et al.* The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Research* **39**, 5768–5775 (2011).
215. Nagatoishi, S., Isono, N., Tsumoto, K. & Sugimoto, N. Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie* **93**, 1231–1238 (2011).
216. Haider, S., Parkinson, G. N. & Neidle, S. Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophysical journal* **95**, 296–311 (2008).
217. Cang, X., Šponer, J. & Cheatham III, T. E. Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis. *J. Am. Chem. Soc.* **133**, 14270–14279 (2011).
218. Lech, C. J., Heddi, B. & Phan, A. T. Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res.* **41**, 2034–2046 (2012).
219. Varizhuk, A. M. *et al.* Polymorphism of G4 associates: from stacks to wires via interlocks. *Nucleic Acids Res.* **46**, 8978–8992 (2018).
220. Martadinata, H. & Phan, A. T. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K<sup>+</sup> solution. *Journal of the American Chemical Society* **131**, 2570–2578 (2009).



221. Martadinata, H., Heddi, B., Lim, K. W. & Phan, A. T. Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry* **50**, 6455–6461 (2011).
222. Islam, B. *et al.* Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale. *Nucleic acids research* **41**, 2723–2735 (2013).
223. Beauchamp, K. A. *et al.* MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory. Comput.* **7**, 3412–3419 (2011).
224. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Science* **27**, 112–128 (2018).
225. Gonçalves, D. P., Rodriguez, R., Balasubramanian, S. & Sanders, J. K. Tetramethylpyridiniumporphyrazines—a new class of G-quadruplex inducing and stabilising ligands. *Chemical Communications*, 4685–4687 (2006).
226. Monchaud, D. *et al.* “One Ring to Bind Them All”—Part I: The Efficiency of the Macrocyclic Scaffold for G-Quadruplex DNA Recognition. *Journal of nucleic acids* **2010** (2010).
227. Víglaský, V., Tlučková, K. & Bauer, L. The first derivative of a function of circular dichroism spectra: biophysical study of human telomeric G-quadruplex. *Eur. Biophys. J.* **40**, 29–37 (2011).
228. Collie, G. W. *et al.* Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc.* **132**, 9328–9334 (2010).
229. Uddin, M. K., Kato, Y., Takagi, Y., Mikuma, T. & Taira, K. Phosphorylation at 5' end of guanosine stretches inhibits dimerization of G-quadruplexes and formation of a G-quadruplex interferes with the enzymatic activities of DNA enzymes. *Nucleic Acids Res.* **32**, 4618–4629 (2004).
230. Beck, D. A., Alonso, D. O. & Daggett, V. A microscopic view of peptide and protein solvation. *Biophysical chemistry* **100**, 221–237 (2002).
231. Saladino, G. *et al.* Metadynamics study of a  $\beta$ -hairpin stability in mixed solvents. *Journal of the American Chemical Society* **133**, 2897–2903 (2011).
232. Miner, J. C. & García, A. E. Equilibrium denaturation and preferential interactions of an RNA tetraloop with urea. *The Journal of Physical Chemistry B* **121**, 3734–3746 (2017).
233. Rees, W. A., Yager, T. D., Korte, J. & Hippel, P. H. V. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32**. PMID: 8418834, 137–144 (1993).
234. Vaitiekunas, P., Crane-Robinson, C. & Privalov, P. L. The energetic basis of the DNA double helix: a combined microcalorimetric approach. *Nucleic acids research* **43**, 8577–8589 (2015).

235. Lu, X.-J. & Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research* **31**, 5108–5121 (2003).
236. Zwanzig, R. Diffusion in a rough potential. *Proceedings of the National Academy of Sciences* **85**, 2029–2030 (1988).
237. Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* **7**, 34 (2005).
238. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **100**, 020603 (2 2008).
239. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* **98**, 10037–10041 (2001).
240. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M.-P. & Phan, A. T. Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angewandte Chemie International Edition* **53**, 999–1002 (2014).
241. Tseng, T.-Y. *et al.* Fluorescent probe for visualizing guanine-quadruplex DNA by fluorescence lifetime imaging microscopy. *Journal of biomedical optics* **18**, 101309–101309 (2013).
242. Lemke, K. *et al.* Induction of unique structural changes in guanine-rich DNA regions by the triazoloacridone C-1305, a topoisomerase II inhibitor with antitumor activities. *Nucleic acids research* **33**, 6034–6047 (2005).