

Molecular Packing Interaction in DNA Crystals

Amen Shamim ^{1,2}, Nazia Parveen ¹, Vinod Kumar Subramani ^{1,*}  and Kyeong Kyu Kim ^{1,*} 

¹ Department of Precision Medicine, Institute for Antimicrobial Resistance Research and Therapeutics, Sungkyunkwan University School of Medicine, Suwon 16419, Korea; amenshamim@gmail.com (A.S.); naziaahmad.nm@gmail.com (N.P.)

² Center of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad 38040, Pakistan

* Correspondence: mailvinod@skku.edu (V.K.S.); kyeongkyu@skku.edu (K.K.K.); Tel.: +82-31-299-6136 (V.K.S.); Fax: +82-31-299-6159 (V.K.S.)

Received: 31 October 2020; Accepted: 26 November 2020; Published: 28 November 2020



Abstract: DNA crystallography provides essential structural information to understand the biochemical and biological functions of oligonucleotides. Therefore, it is necessary to understand the factors affecting crystallization of DNA to develop a strategy for production of diffraction-quality DNA crystals. We analyzed key factors affecting intermolecular interactions in 509 DNA crystals from the Nucleic Acid Database and Protein Databank. Packing interactions in DNA crystals were classified into four categories based on the intermolecular hydrogen bonds in base or backbone, and their correlations with other factors were analyzed. From this analysis, we confirmed that hydrogen bonding between terminal end and mid-region is most common in crystal packing and in high-resolution crystal structures. Interestingly, $P2_12_12_1$ is highly preferred in DNA crystals in general, but the $P6_1$ space group is relatively abundant in A-DNA crystals. Accordingly, $P2_12_12_1$ contains more terminal end-mid-region interactions than other space groups, confirming the significance of this interaction. While metals play a role in the production of a good crystal in B-DNA conformation, their effect is not significant in other conformations. From these analyses, we found that packing interaction and other factors have a strong influence on the quality of DNA crystals and provide key information to predict crystal growth of candidate oligonucleotides.

Keywords: DNA; crystal structures; crystal packing; crystallography; terminal end; base; sugar-phosphate backbone

1. Introduction

Biomacromolecules and their interactions are essential for biochemical reactions and maintenance of cellular homeostasis. Thus, examining interactions at the atomic level is necessary for comprehensive understanding of biological processes in cells. The crystallographic approach provides a snapshot of atomic details of biomacromolecules at the highest resolution and are considered the best way to investigate their structure, conformation, interactions, and function. However, the crystallographic approach is hindered by the difficulty of achieving diffraction-quality crystal. Production of highly ordered crystals of biomolecules is the biggest challenge with this method.

Double-stranded DNAs typically form a right-handed B-form double-helical structure containing Watson–Crick base-pairs [1]. However, DNA duplexes adopt various conformations such as A- and Z-DNA [2,3]. In addition, single-stranded DNA can be folded into a variety of structures such as hairpin, triplex, G-quadruplex, and I-motif [4–8]. To understand the structure and function of DNA, the crystallographic approach is the best available choice. However, crystallization is an important challenge due to the presence of negatively charged phosphate groups, coupled with high solvent content and the dynamic flexible nature of the molecule [9–11]. In addition, the structurally similar phosphate groups present all along a DNA molecule allow non-specific DNA–DNA interactions,

which are the main factors hindering crystallization of DNA. For forming a diffraction-quality crystal, DNA molecules must be tightly packed with low conformational flexibility, which can be achieved by various factors in the crystallization conditions. Therefore, understanding molecular interactions among DNA in the crystals under crystallization conditions and factors affecting crystal quality allows for the design of experiments for growing diffraction-quality crystals and structure determination.

Currently, a vast number of oligodeoxynucleotides (ODNs) have been successfully crystallized, and the structures are determined by X-ray diffraction analyses [12]. This structural information enables understanding of fine details of the conformation of DNA molecules and their interactions with other ligands such as small molecules and metals. Further, these DNA structures could hold key information on the factors that affect successful DNA packing from solution into highly ordered lattice. Therefore, it is key to study crystal structures and their packing interactions to understand how DNAs are packed into crystals to increase the success rate of growing diffraction-quality DNA crystals for structure determination. In this study, we develop a JAVA-based program to classify DNA crystal packing interactions (DXPI), and DNA structures available in the Protein Databank (PDB) are analyzed. Furthermore, correlations between DXPI and other factors that affect crystal formation and diffraction quality such as conformation, sequence length, resolution, symmetry, and metal/ligands are also analyzed. Based on this analysis, we propose key factors and packing interactions in the achievement of diffraction-quality DNA crystal.

2. Methods

2.1. Data Collection and Extraction

Annotated information on all three-dimensional DNA structures listed in the Nucleic Acid Database (NDB) [13] was extracted. The NDB database comprises information on the sequence, structural features, function, and experimental methods. Based on DNA conformation, structures deduced from crystals were categorized as A-, B-, and Z-DNA. Further information about resolution, symmetry and metal/ligand were extracted from the Protein Data Bank (PDB) [14].

2.2. Crystal Packing

To investigate crystal packing interactions and test the stability of individual crystal packing contacts, we used the PISA program [15]. To investigate the same for all the extracted DNA-only crystal structures, we automated the process using a short program designed in scripting language (Shell and AWK language) to generate symmetry mates as a preparation for PISA (Protein Interfaces, Surfaces and Assemblies) analyses. The script is available at (<https://github.com/Amenshamim/Crystal-packing.git>). For the script, we defined intermolecular contacts closer than 4.0 Å as a pair of symmetry-equivalent mates.

2.3. DNA Crystal Packing Type

To define the packing type, we designed a program in JAVA with NetBeans IDE (Oracle) that required the Java Runtime Environment (JRE) on a computer. We executed this program with an Intel(R) Xeon(R) CPU E5-2640 v4 (2.40GHz) 10 core processor system, (HPC KOREA, Mannyon-dong, Seo-gu, Daejeon, South Korea). This program simultaneously analyzed multiple PISA interface outputs of DNA crystal structures in minutes to generate the packing types that are classified into Types 1, 2, 3, and 4 based on the primary DNA crystal packing interaction (DXPI) categories listed in Table 1. Our program is available on github (<https://github.com/Amenshamim/Crystal-packing.git>).

Table 1. DNA crystal packing interaction.

Intermolecular Interacting Sites	Moiety Involved in Hydrogen Bond	Interaction Category	Hydrogen Bond
Terminal end-Terminal end	Base - Base	C1	N-H—O or N
Terminal end-Terminal end	Sugar-Phosphate Base-Phosphate Sugar-Base	C2	3'OH—OP1 or OP2 5'OH—OP1 or OP2 N-H—OP1 or OP2 3'OH—O or N
Mid-region-Terminal end	Sugar-Phosphate Base-Phosphate	C3	3'OH—OP1 or OP2 5'OH—OP1 or OP2 N-H—OP1 or OP2
Mid-region-Terminal end	Sugar-Phosphate	C4	3'OH—OP1 or OP2 5'OH—OP1 or OP2

2.4. Packing Interactions and Factor Analyses

All the factors were extracted and tabulated for the listed 509 DNA-only crystal structures (Table S1). The results were grouped by packing type, conformation, sequence length, resolution, symmetry, and metal/ligand. Primary analyses involved Principal Component Analyses (PCA), which revealed no clear separation between the various factors investigated. However, a Pearson's correlation matrix was drawn to reveal any relationship between these factors (Table S2). Further, each individual factor was manually analyzed with respect to either packing type or conformation (B-, A-, or Z-DNA). Percentage calculations were performed to assess either the overall proportion of the observed individual features (Overall %) or within a particular factor type under consideration (either Row % or Column %). The results of these analyses were tabulated and are shown in the main (Tables 2–7) and supplementary tables (Tables S3–S11) of this study.

Table 2. Number and proportion of DNA structures belonging to different packing types and their conformations.

Packing Type	Conformation	Structures in NDB (Count)	Structures in NDB (%)
Type 1	A-DNA	2	0.4%
	B-DNA	27	5.3%
	Z-DNA	4	0.8%
Type 1 Total		33	6.5%
Type 2	A-DNA	32	6.3%
	B-DNA	49	9.6%
	Z-DNA	32	6.3%
Type 2 Total		113	22%
Type 3	A-DNA	95	18.7%
	B-DNA	228	44.8%
	Z-DNA	35	6.9%
Type 3 Total		358	70.3%
Type 4	A-DNA	0	0.0%
	B-DNA	0	0.0%
	Z-DNA	5	1.0%
Type 4 Total		5	1.0%
Grand Total		509	100.0%

Table 3. Number of DNA structures of different sequence lengths, their packing types, and conformations.

Sequence Length	Packing Type				Grand Total	Conformation			Grand Total
	Type 1	Type 2	Type 3	Type 4		B-DNA	A-DNA	Z-DNA	
12	13	7	209	0	229	217	10	2	229
10	9	46	84	0	139	66	72	1	139
6	2	35	44	5	86	12	10	64	86
8	0	17	9	0	26	0	26	0	26
14	0	0	7	0	7	0	7	0	7
7	5	0	1	0	6	2	0	4	6
4	0	4	2	0	6	1	0	5	6
9	2	2	2	0	6	2	4	0	6
11	2	0	0	0	2	2	0	0	2
20	0	1	0	0	1	1	0	0	1
13	0	1	0	0	1	1	0	0	1
Grand Total	33	113	358	5	509	304	129	76	509

Table 4. Number of DNA structures belonging to different resolution ranges, their packing types, and conformations.

Resolution (Å)	Packing Type				Grand Total	Conformation			Grand Total
	Type 1	Type 2	Type 3	Type 4		B-DNA	A-DNA	Z-DNA	
<1.4	3	16	62	1	82	38	18	26	82
1.4–1.6	1	20	60	1	82	41	29	12	82
1.6–1.8	6	20	49	1	76	28	29	19	76
1.8–2.0	5	17	48	1	71	44	21	6	71
2.0–2.2	5	11	47	0	63	47	13	3	63
2.2–2.4	4	11	34	0	49	41	7	1	49
2.4–2.6	3	8	35	1	47	35	9	3	47
2.8–3.0	4	4	9	0	17	13	2	2	17
2.6–2.8	1	5	11	0	17	13	1	3	17
>3.0	1	1	3	0	5	4	0	1	5
Grand Total	33	113	358	5	509	304	129	76	509

Table 5. Number of DNA structures belonging to different space groups, their packing types, and conformations.

Symmetry	Packing Type				Grand Total	Conformation			Grand Total
	Type 1	Type 2	Type 3	Type 4		B-DNA	A-DNA	Z-DNA	
P 2 ₁ 2 ₁ 2 ₁	17	30	277	4	328	210	69	49	328
P 6 ₁	0	18	9	0	27	3	24	0	27
P 3 ₂ 2 1	1	10	10	0	21	8	10	3	21
H 3	8	3	8	0	19	18	1	0	19
P 1 2 ₁ 1	1	9	6	0	16	3	3	10	16
C 1 2 1	0	7	5	0	12	11	0	1	12
C 2 2 2 ₁	0	6	6	0	12	4	6	2	12
P 1	0	4	5	0	9	9	0	0	9
P 3 ₂	3	4	1	0	8	4	0	4	8
P 4 ₁ 2 ₁ 2	1	0	5	0	6	2	4	0	6
P 4 ₃	0	2	4	0	6	2	4	0	6
P 6 ₅	0	3	2	0	5	2	0	3	5
P 3 ₁	0	2	2	0	4	4	0	0	4
P 6	0	4	0	0	4	4	0	0	4

Table 5. Cont.

Symmetry	Packing Type				Grand Total	Conformation			Grand Total
	Type 1	Type 2	Type 3	Type 4		B-DNA	A-DNA	Z-DNA	
P 3 ₂ 1 2	0	2	1	0	3	3	0	0	3
P 4 ₃ 2 ₁ 2	0	1	2	0	3	1	1	1	3
P -1	0	3	0	0	3	3	0	0	3
P 2 ₁ 2 ₁ 2	0	1	1	0	2	1	1	0	2
P 6 ₅ 2 2	0	1	1	0	2	1	1	0	2
P 6 ₁ 2 2	0	0	2	0	2	0	2	0	2
P 4 ₁	0	0	2	0	2	0	2	0	2
P 1 1 2 ₁	0	1	0	1	2	0	0	2	2
P 4 ₁ 2 2	0	0	2	0	2	2	0	0	2
B 2 2 ₁ 2	0	1	1	0	2	1	0	1	2
P b c a	0	0	1	0	1	1	0	0	1
P 1 2 ₁ /n 1	0	0	1	0	1	1	0	0	1
P 1 2 ₁ /c 1	0	0	1	0	1	1	0	0	1
I 2 3	1	0	0	0	1	1	0	0	1
C 2 2 2	1	0	0	0	1	1	0	0	1
P 3 ₁ 2 1	0	0	1	0	1	1	0	0	1
I 2 2 2	0	0	1	0	1	1	0	0	1
I 4 ₁ 2 2	0	1	0	0	1	1	0	0	1
P 3	0	0	1	0	1	0	1	0	1
Grand Total	33	113	358	5	509	304	129	76	509

Table 6. Number of DNA structures containing different metal/ligands, their packing types, and conformations. Full list is provided in Table S9.

Metal / Ligand	Packing Type				Grand Total	Conformation			Grand Total
	Type 1	Type 2	Type 3	Type 4		B-DNA	A-DNA	Z-DNA	
(blank)	5	53	119	5	182	72	76	34	182
MG	6	16	91	0	113	91	14	8	113
CA	9	5	10	0	24	21	2	1	24
SPM	1	4	15	0	20	4	11	5	20
NT	2	3	7	0	12	12	0	0	12
HT	0	0	12	0	12	12	0	0	12
NCO	1	4	5	0	10	3	1	6	10
NA	0	5	5	0	10	4	6	0	10
BA	0	0	7	0	7	2	2	3	7
K	0	0	7	0	7	4	3	0	7
CO	1	4	1	0	6	6	0	0	6
CU	0	4	1	0	5	0	0	5	5
SR	0	0	5	0	5	2	3	0	5
ZN	1	2	2	0	5	2	2	1	5
MN	1	2	1	0	4	1	1	2	4
DAP	1	0	3	0	4	4	0	0	4
RB	0	0	3	0	3	1	2	0	3
IA	0	0	3	0	3	3	0	0	3
NI	3	0	0	0	3	3	0	0	3
DMY	2	0	1	0	3	3	0	0	3
HT1	0	0	3	0	3	3	0	0	3
IB	0	0	3	0	3	3	0	0	3
RO2	0	0	2	0	2	2	0	0	2
PTN	0	0	2	0	2	1	0	1	2
NRU	0	2	0	0	2	0	0	2	2
IPY	0	2	0	0	2	2	0	0	2
CL	0	1	1	0	2	1	1	0	2

Table 6. Cont.

Metal / Ligand	Packing Type				Grand Total	Conformation			Grand Total
	Type 1	Type 2	Type 3	Type 4		B-DNA	A-DNA	Z-DNA	
HG	0	0	2	0	2	2	0	0	2
TNT	0	0	2	0	2	2	0	0	2
BBZ	0	0	2	0	2	2	0	0	2
ILT	0	0	2	0	2	2	0	0	2
BRN	0	0	2	0	2	2	0	0	2
Others	0	6	39	0	45	32	5	8	45
Grand Total	33	113	358	5	509	304	129	76	509

Table 7. Number of DNA structures containing different percentage range of solvent content, their packing types, and conformations. Full list is provided in Table S10.

Solvent Content (%)	DXPI Type				Grand Total	Conformation			Grand Total
	Type 1	Type 2	Type 3	Type 4		B-DNA	A-DNA	Z-DNA	
70–80	0	1	1	0	2	0	2	0	2
60–70	3	4	15	0	22	13	9	0	22
50–60	6	29	37	0	72	44	25	3	72
40–50	15	37	209	0	261	203	52	6	261
30–40	9	20	61	0	90	35	38	17	90
20–30	0	14	27	5	46	4	2	40	46
10–20	0	2	1	0	3	0	0	3	3
0–10	0	0	2	0	2	0	0	2	2
(blank)	0	6	5	0	11	4	1	6	11
Grand Total	33	113	358	5	509	303	129	77	509

2.5. Solvent Content

Using the molecular weights, unit cell parameters and the space group from the PDB records, the Matthews coefficient (V_M), and the solvent content were calculated [16,17]. For analyses the results were grouped into ranges starting from 0% to 80% with increments of 10 units i.e., 0–10, 10–20, 20–30, 30–40, 40–50, 50–60, 60–70, and 70–80 (Table 7 and Table S10).

2.6. Structural Visualization

Structures were visualized and rendered in CCP4mg suite [18], and interactions within packed crystals were analyzed in Visual Molecular Dynamics tool (VMD) (Version 1.9.4) [19].

3. Results

3.1. Data Extraction

In order to investigate molecular packing in DNA crystals, we collected DNA structure information from the Nucleic Acid Data Base (NDB), comprising a total of 7724 structures containing DNA, of which 1916 had only DNA structures, 5365 were DNA-Protein complexes, and 443 were DNA-Drug complex structures. Among the DNA-only structures, 760 were determined by nuclear magnetic resonance (NMR) and 1156 by X-ray crystallography. For analysis of packing interactions in DNA crystals, we searched DNA-only crystal structures in Protein Databank, however only 851 structures were available. After removing the structures of single-stranded DNAs, 509 double-stranded DNA structures comprising 304 B-DNA, 129 A-DNA, and 76 Z-DNA structures were analyzed (Tables S1 and S3, Figure 1).

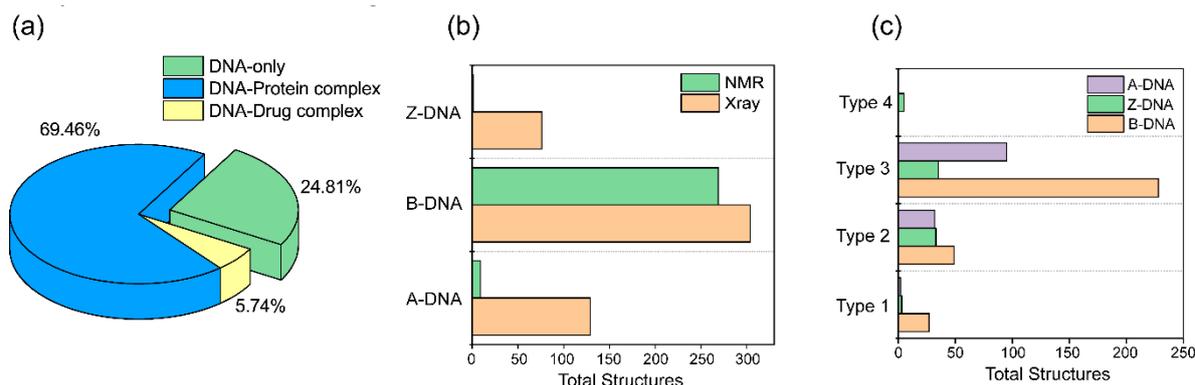


Figure 1. DNA crystal structures and their packing interactions. (a) Distribution of the reported DNA crystal structures deposited in the Nucleic Acid Database (NDB). DNA structures are classified by the binding partners; no ligand (DNA-only), protein, and drug, (b) DNA-only crystal structures in NDB are analyzed based on their conformations and methods for structure determination. (c) DNA-only crystal structures are analyzed based on their packing types and conformations.

3.2. Packing Interaction

To understand the nature of packing interactions important to crystal quality, we analyzed DNA structures and their packing interactions inside crystals. For this purpose, we first classified the DNA crystal packing interactions (DXPI) into four categories (C1, C2, C3, and C4) based on hydrogen bonds (H-bonds) among symmetry equivalent molecules in the crystal lattice (Table 1). First, we looked into H-bonds between terminal-ends of double-stranded oligodeoxynucleotides (dsODNs) and divided them into two categories, C1 and C2. C1 represents the intermolecular H-bonds between two bases, and thus it contains NH-N and NH-O bonds in base pairs (Figure 2a). C2 contains H-bonds between the terminal phosphate and sugar or base in the terminal end of the symmetry equivalent molecules (Figure 2b). Then, intermolecular H-bonds between the terminal end and mid-region of DNA can be defined as C3 and C4 depending on the involvement of base interaction. Accordingly, C3 contains intermolecular H-bonds between the terminal phosphate and sugar or base in the mid-region nucleotides (Figure 2c), or between the terminal end sugar or base and the mid-region phosphate. C4 interaction comprises the phosphate and sugar H-bond interaction between the terminal end and mid-region of DNA (Figure 2d).

3.3. Packing Types

From our observations, we defined DXPI into four categories according to the moiety (base, sugar, and phosphate) involved in H-bonds and their locations (terminal end and mid-region) among symmetry equivalent dsODNs (Table 1 and Figure 2). Based on these interaction categories, we grouped 509 DNA crystal structures into four packing types (Figure 3). We first searched C1 and C2 interactions by checking for the presence of H-bonds between two terminal ends. Then, the packing type of crystal structures is defined as Type 1 when C1 interaction is observed. If no C1 interaction is found, the packing type of crystal structures are defined as Type 2. If there are no H-bonds among the terminal ends, or no C1 or C2 interaction, the packing type can be defined as either Type 3 or Type 4 depending on the presence of C3 or C4 interaction (Figure 3). Type 4 has intermolecular hydrogen bonds between backbone and terminal 3'OH or 5'OH without the involvement of a base. However, such interactions were extremely rare (1%), found only in five Z-DNA structures. This implies in crystal packing interactions, intermolecular hydrogen bonds with a base are inevitable (99%). Accordingly, crystal structure with packing Type 3 and 4 contains only C3 and C4 interaction, respectively. However, crystal structures with packing Type 1 and Type 2 possibly contain C3 or C4 interaction between symmetry equivalent molecules.

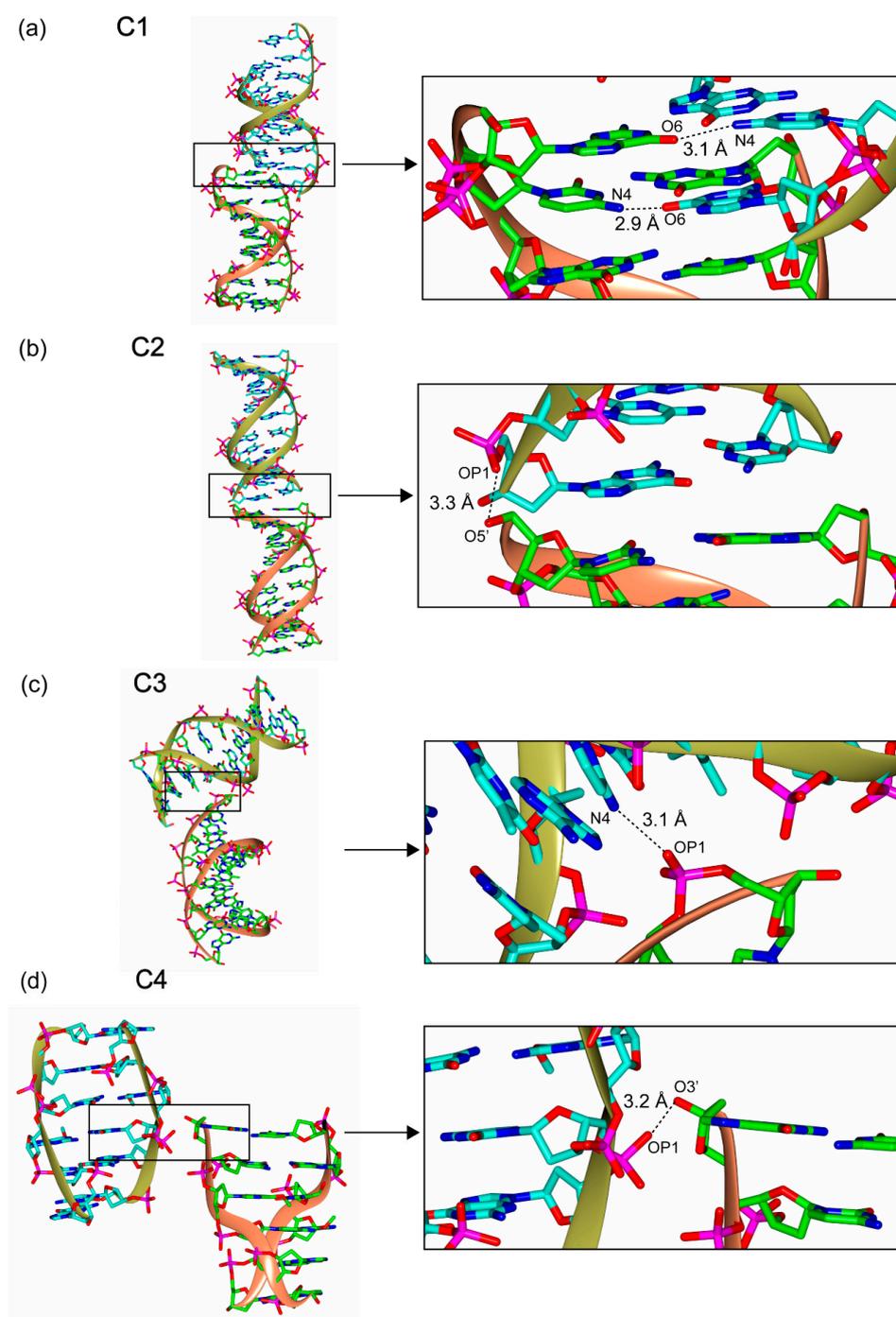


Figure 2. Categories of DNA crystal packing interactions (DXPIs). Four categories of DXPI are defined based on the intermolecular H-bonds among the atoms in the terminal end and mid-region. The following representative structures and their interactions are displayed to explain the intermolecular H-bond in each category; (a) C1: intermolecular H-bond between bases (O6 vs N4) present in the terminal ends of symmetry equivalent double-stranded oligodeoxynucleotides (dsODNs) (PDB ID 1ENN), (b) C2: intermolecular H-bond between the terminal end phosphate (OP1) and terminal end sugar (HO5') (PDB ID 3IXN), (c) C3: intermolecular H-bond between the terminal end phosphate (OP1) and mid-region base (N4) (PDB ID 3IXN), and (d) C4: intermolecular H-bond between mid-region phosphate (OP1) and terminal end sugar (HO3') (PDB ID 1D24). The atoms involved in the intermolecular H-bond are labeled with respective bond lengths in block dotted lines. The backbones (coral) and bases (cyan) of the reference structures and their symmetry equivalents (backbone: yellow, and base: green) are represented in ribbon and stick diagrams. The detailed definition of each interaction category is explained in Table 1.

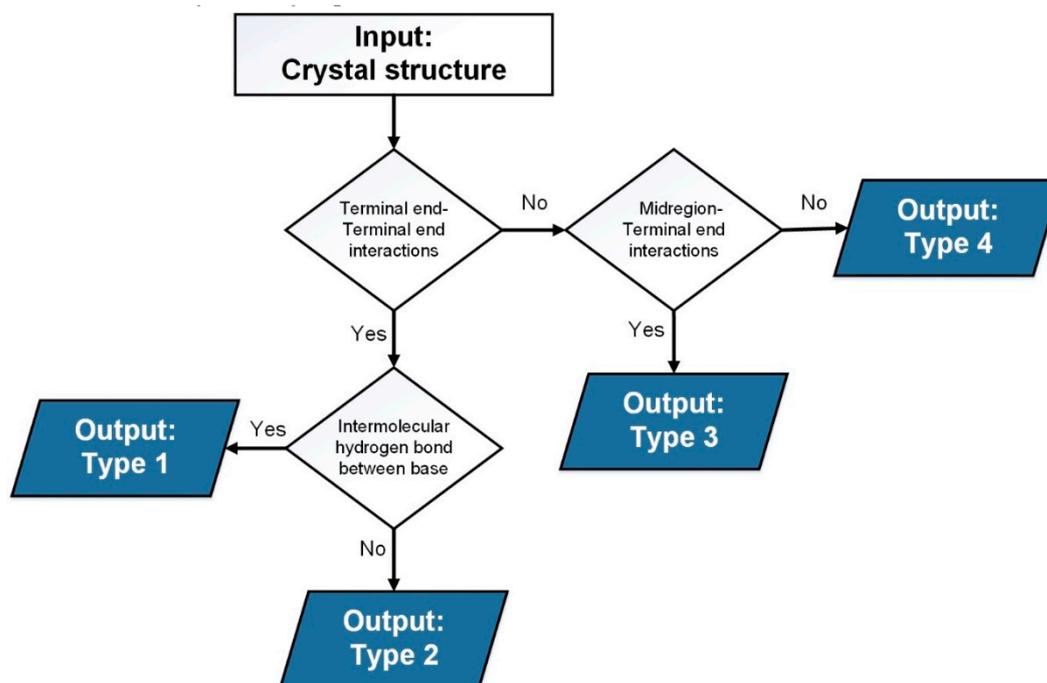


Figure 3. Logic flow for determining crystal packing type of DNA crystal structures. This flowchart explains how crystal packing types are defined by taking stepwise decisions on the input crystal structures. The first input decision is made based on the presence and absence of terminal end-terminal end interactions. When the answer is Yes, the packing type of DNA structure must be either Type 1 or Type 2. However, if the answer is No, the types would be Type 3 or Type 4. The second input decision is made based on the presence and absence of intermolecular H-bonds between bases. If there are H-bonds between base and base, the answer is Yes and the packing type is Type 1. Otherwise, the Type 2 packing type is assigned. The third decision is made based on the category of the intermolecular H-bond between mid-region and terminal end. If base atoms are involved in the intermolecular H-bond (Yes), Type 3 is assigned. If the answer is No, Type 4 is assigned.

We automated our analysis with a custom-designed JAVA-based program (see Materials and Methods) and investigated 509 DNA crystal structures to search DXPI and define their packing types. In addition, we investigated other factors that could influence DXPI and crystal quality. For this purpose, conformation of DNA, sequence length, resolution, symmetry (space group), metal/ligands, and solvent content of 509 structures were extracted from the Protein Data Bank (PDB).

3.4. Correlation Among Factors Affecting Crystal Quality

The categorized DNA crystal packing types were combined and analyzed for correlation using Pearson's test (Table S2). We found that sequence composition was highly correlated with resulting conformation of DNA in the crystals ($r = 0.83$), and that sequence length weakly correlated with symmetry of the crystal ($r = 0.51$). Similarly, percentage solvent content showed a weak correlation ($r = 0.49$) with DNA conformation. The remaining factors did not show significant association, which highlights their independent roles in crystal packing. Principal Component Analysis (PCA) revealed no clear separation (data not shown) between these factors, which implies that all factors contribute to crystal packing and symmetry. We further analyzed each factor and its contribution to or correlation with other factors.

3.4.1. Packing vs. Other Factors

Most DNA-only crystals are of the canonical B-conformation (304/509 or 59.7%), followed by A-DNA (129/509 or 25.3%) and Z-DNA (76/509 or 14.9%) (Figure 1, Table 2 and Table S3). Among these,

Type 3 packing dominates at 70.3%, followed by Type 2 at 22%, Type 1 at 6.5%, and Type 4 at 1%. However, Type 2 packing is relatively equally distributed across the three conformations investigated, A (43.4%), B (28.3%), and Z (28.3%). Type 1 and Type 3 packing dominates B-DNA structures (81.8% and 63.7% respectively), while Type 4 is found only in 1% of total structures and is exclusive to Z-DNA crystals (5/509). In both B- and A-DNA structures, Type 3 packing dominates at 75% and 73.6%, respectively, while Type 2 (42.1%) and Type 3 (46.1%) packings in the crystals with Z-DNA conformations are abundant to grow highly ordered crystals.

A total of 70.3% of structures pack with Type 3, with 41% of them involving a dodecamer and 16.5% a decameric dsODN. In addition, about half of crystals with Type 3 (54.4%) packing occur in the orthorhombic $P2_12_12_1$ symmetry (Table S8). The frequency of Type 3 packing is followed by Type 2 packing at 22% incidence (Table S3).

3.4.2. Sequence vs. Other Factors

Overall, the frequency of dodecamer dsODN dominates (45%) the list of reported DNA structures, followed by 10 bp (27.3%) and 6 bp (16.9%) dsODN. Further analyses revealed that the popular choice of dsODN length for crystallizing B-DNA is 12 bp (71.4%) starting with the classic B-DNA structure reported by Dickerson group [1], while that for A-DNA is 10 bp (55.8%) and 6 bp (84.2%) dsODN for Z-DNA (Table 3 and Table S4).

Influence of the sequence composition across packing types was additionally investigated. Interestingly, most dsODNs used in crystals contain cytosine base (73.3%) in their 5' end. In the second position, guanine base is relatively dominant (63.1%). However, at the 3' end, the sequences end most commonly with guanine (48.6%) or cytosine (33.6%). While the majority of all packing types have sequences beginning with cytosine in their 5' ends, Type 1 packing comprises structures with a similar distribution of sequences beginning with either cytosine (45.5%) or guanine (45.5%) (Table S5).

Furthermore, we looked specifically at bases involved in the inter-strand interaction (Table S6), and learned that only Type 1 packing structures have base-base intermolecular packing interactions and that guanine bases are predominantly present in the terminal end (75.8%). We also investigated the bases involved in Type 3 packing structures and found that 73.5% of crystals contain the intermolecular interaction between the terminal base and the mid-region of symmetry equivalent DNA molecules. The remaining 26.5% of crystals harbor intermolecular packing via hydrogen bonds between the terminal phosphates or sugars and the mid-region bases. In this packing, 64.6% of terminal bases involved in the intermolecular interaction are guanine base, and the second most frequently found base is cytosine (Table S6).

3.4.3. Resolution vs. Other Factors

The extracted 509 DNA-only structures were categorized based on resolution into 10 groups starting with the group $< 1.4 \text{ \AA}$ comprising structures with the highest atomic resolution. This was followed by groups that constitute structures in the resolution ranges between 1.4–1.6 \AA , 1.6–1.8 \AA , and so on to those comprising low-resolution structures beyond 3.0 \AA . We found that 61.1% of the structures had a resolution quality higher than 2.0 \AA , which signifies the quality of these crystals and their highly ordered arrangement, which is essential in studying the minute differences among conformations of highly flexible DNA molecules. Among the three DNA conformations, most of the Z-DNA structures showed the highest resolution: 34.2% of the Z-DNA structures diffract greater than 1.4 \AA . In comparison, only 12.5% of B-DNA and 14% of A-DNA structures were solved in this high-resolution range.

With regard to the type of packing observed, among all the structures investigated across all the resolution ranges, Type 3 showed the highest prevalence at 70.3%, followed by Type 2 (22.2%). A similar trend was seen across resolution ranges. Of particular mention is that the structures that resolved poorly with resolution lower than 2.8 \AA showed a shift in packing type from Type 3 (59%) to either Type 2 (25.6%) or Type 1 (15.4%). This indicates that Type 3 packing contributes much to a higher

ordered lattice growth and hence a high-resolution structure. There is no clear relationship between metal/ligand regarding influence on the resolution of DNA structures, implying that other factors influence the outcome (Table S9). However, among the structures with Mg, the presence of metal coordination was seen in 66.4% of those with higher than 1.4 Å resolution (Table S9). However, the same is not true for the other metals or ligands observed. The incidence of DNA length in our dataset shows the highest number of 12 bp dsODNs, followed by 10 bp and 6 bp lengths. Most had a resolution higher than 2.0 Å (Table S9). Interestingly, 31.4% of the 6 bp dsODN crystal structures grouped in the highest resolution range of < 1.4 Å, while 19.4% and 10.5% of 10 bp and 12 bp, respectively, dsODN crystal structures diffract higher than 1.4 Å (Table 4 and Table S4). This result suggests shorter length DNAs can be crystalized with high quality.

3.4.4. Symmetry vs. Other Factors

The orthorhombic $P2_12_12_1$ is the most common space group found in macromolecular crystals. The same is true across and within the three conformations of DNA investigated. The $P2_12_12_1$ space group accounts for 64.4% of all DNA structures included in this study and more than half of the solved structures of each conformation. $P6_1$ is the second most abundant space group found in DNA crystal (5.3%). $P2_12_12_1$ packing is dominated by B-conformation (64.0%), while $P6_1$ packing is the most frequently found in A-conformation (89.0%) (Table 5 and Table S8). Accordingly, B-DNA is crystalized mostly in $P2_12_12_1$ (69.0%) since the second most prevalent space group (H3) only covers 5.9% of B-DNA crystals. However, A-DNA is crystalized in $P2_12_12_1$ (53.5%) and $P6_1$ (18.6%). In the case of Z-DNA, $P2_12_12_1$ (64.5%) and $P12_11$ (13.2%) are the most prevalent space groups. Within the $P2_12_12_1$ space group, 57.3% belong to DNA of 12 bp in length, followed next by 10 bp ODN at 25.6% and 6 bp ODN at 13.7% (Table 3, Table 5 and Tables S4 and S8). Interestingly, among the high incidence of symmetry observed in our dataset, $P2_12_12_1$ shows an 84.5% occurrence of Type 3 packing, whereas Type 2 packing dominates the $P6_1$ symmetry (66.7%).

3.4.5. Metal/Ligands vs. Other Factors

One-third of all DNA-only structures (36%) form well-ordered crystals without the need for metal coordination. However, among metal coordinated or ligand binding structures, it is predominantly magnesium ion (Mg^{2+} , 22.2%) that facilitates DNA packing stability. Also, metal coordination is not mandatory to grow highly ordered crystals of non-B conformations of A-DNA (58.9%) and Z-DNA (44.7%). However, Mg^{2+} is an important metal required to stabilize the canonical B-DNA conformation (30%) against no metal condition (23.7% for no metal). In structures that are stabilized by Type 1 packing interactions, calcium coordination is preferred (27.3%), while Mg^{2+} is found in 15.2%, the same as structures stabilized without a metal ion. In the case of Type 2, 3, and 4 packings, non-metal-coordinated packing is found in 46.9%, 33.2%, and 100%, respectively (Table 6 and Table S9).

3.4.6. Solvent vs. Other Factors

Among the solvent content range, 40–50% is predominant in the DNA crystals as 51.3% of all the structures have a solvent content of 40–50%, which is also true when the packing type is considered; 45.5%, 32.7%, and 58.4% of crystals belonging to Types 1, 2, and 3, respectively, contain 40–50% solvent content (Table S10). Across DNA conformations, the incidences containing 40–50% solvent content in B-, A-, and Z-DNAs are 67.0%, 40.3%, and 7.8%, respectively. However, if considering the solvent content lower than 40%, the occurrences in B-, A-, and Z-conformations are 12.9%, 31.1%, and 80.5%, respectively. These analyses imply that DNAs with B-conformation crystallize under high solvent condition while Z-DNA prefers the lowest solvent content among the three dsDNA conformations investigated in this study.

The solvent content of crystals is proportional to the length of dsDNA in crystals as the most 12 bp dsODNs are crystalized with solvent content in the range of 40–50%, while crystals with 6 bp dsODNs contain the lower solvent content as 20–30% solvent content is the highest occurrence in the

crystal containing 6 bp dsDNA. Solvent content analysis also reveals that crystals in $P2_12_12_1$ symmetry contain low solvent contents (60.7% in 40–50% and 20.7% in 30–40%) while those in the second most populous space group ($P6_1$) have a higher solvent content (40.7% in both 40–50%, and 50–60%) (Table S9). Interestingly, we observed that the metal-containing crystals have lower solvent content than those without metal; incidences containing the solvent content in the range of 40–50% are 48.4%, 60.2%, and 62.5% for no-metal, magnesium ion, and calcium ion conditions, respectively. These results suggest that metal and $P2_12_12_1$ space groups are required for the formation of the high-quality crystals represented by the low solvent contents.

3.5. Comparison of Molecular Interactions in The Preferred Space Groups

We confirmed that $P2_12_12_1$ is most favorable in DNA crystals and thus investigated which factors affect this preference. For this purpose, we also analyzed two more favorable space groups ($P6_1$ and $P3_221$). To evaluate which packing interactions are dominant in those space groups, we counted the number of crystal structures belonging to the following factors; conformations (A, B, and Z), Types (1, 2, and 3), and Space groups ($P2_12_12_1$, $P6_1$ and $P3_221$) (Table S11). In the case of B-DNA, Type 3 is the most important interaction in $P2_12_12_1$ since Type 3 packing is highly dominant. However, in the case of Z-DNA, both Type 2 and Type 3 packing seems to contribute to the crystal formation when we count their prevalence. In the case of A-DNA, Type 3 seems to be more dominant than Type 2 in $P2_12_12_1$, but Type 2 is more significant than Type 3 in $P6_1$ space group. To visualize their packing, we provide the crystal structures and their packing belonging to the following criteria, B-DNA-Type 3- $P2_12_12_1$ (Figure 4a); A-DNA-Type 2- $P2_12_12_1$ (Figure 4b); A-DNA-Type 2- $P6_1$ (Figure 4c), A-DNA-Type 3- $P2_12_12_1$ (Figure 4d), Z-DNA-Type 2- $P2_12_12_1$ (Figure 4e), and Z-DNA-Type 3- $P2_12_12_1$ (Figure 4f). It is noteworthy that C3 interactions are also observed in Figure 4b,c,e, suggesting C3 interactions are the most important packing interactions in the DNA crystals.

4. Discussions

X-ray crystallography has allowed key findings in biology and its processes at high resolution. Therefore, crystallization of biomolecules is a prerequisite for this process, though diffraction-quality crystals are challenging to grow. The same is true for DNA crystallography due to the presence of negatively charged phosphate groups, coupled with high solvent content and the dynamic flexible nature of the molecule [9–11]. Analysis inside the crystals of all reported DNA structures may provide important clues to understand the principle for the DNA crystallization, which provides hints for successful growth of highly ordered DNA crystals. For that purpose, an initial screening showed that hydrogen bond interactions drive key DNA crystal packing interactions (DXPI). We then classified the DNA crystal structure as Types 1, 2, 3, and 4 based on their hydrogen bond interaction among the symmetry equivalent molecules. We extracted all key factors affecting crystal quality from all DNA-only crystals in Protein Data Bank and investigated the DXPI categories and packing types by automation using a custom-designed JAVA-based program. We further examined their relationships with several key factors of X-ray crystal structures - conformation, sequence length, resolution, symmetry, and presence of metal/ligand coordination.

Our investigation of 509 DNA-only crystal structures found that sequence composition was highly correlated with resulting conformation of the DNA ($r = 0.83$), and sequence length was correlated with the symmetry of the crystal ($r = 0.51$). Results of similar analyses on other factors did not show significance. Therefore, we analyzed these factors and their contributions to crystal packing. Several key findings emerged from our analyses. When DNA-only crystals are successfully grown, more than 60% diffract to resolutions higher than 2.0 Å, which is partially attributed to the size of the molecules used in crystallization since the lengths of most DNA used in this study are less than 12 bp. However, since the diffraction limit is also affected by the beam brilliance and detector setting, correlation between the diffraction resolution and other factors must be carefully considered. In particular, crystal quality can be also interpreted by DXPI that invariably stabilizes the dsODN packing through various hydrogen

bonds, which results in a highly ordered crystal lattice. Recent molecular dynamics simulation work of DNA crystals suggests that the flexibility of the terminal end of DNA is highly stabilized in a crystal [20]. The work suggests that stabilization of the terminal end of DNA is key to achieve diffraction-quality crystals. Consistently, we observed that terminal ends (5' OH, 3' OH, or terminal base) are always involved in the crystal packing. Our investigation also unravels that 99% of DNA-only crystals are packed by the base-mediated inter-molecular contacts with symmetry equivalent DNA molecules (Table 2), suggesting that nucleobases not only play an important role in maintaining the structure and function of the double stranded DNA, but also are important to maintain intermolecular contacts for crystallization.

In this study, we found that $P2_12_12_1$ space group symmetry is the most common in DNA-only crystals (64.0%). In the protein crystals, $P2_12_12_1$ space group is also highly present near 30% [21,22]. However, while $P2_1$ is also abundant in protein crystals, no $P2_1$ space group is found in DNA crystals. An entropic model indicates that $P2_12_12_1$ and $P2_1$ in protein crystals are the least restrictive to packing (rigid body degrees of freedom) and can be packed in a larger number of ways with a preference to screw axis over rotation axis [22]. However, in the case of small organic molecules, space group preference is explained based on the molecular packing [23]. Considering the extremely low space group frequency (3.9%) for tetragonal space groups, DNA crystals seem to be similar to organic molecule crystals rather than protein crystals. Accordingly, space group preference might be explained by the packing interaction. Indeed, when the packing interaction is investigated for crystal structures belonging to the highly abundant packing types and space groups (Table S11 and Figure 4), we found many intermolecular contacts, supporting that crystal packing could be one of the important factors for the space group preference.

Metals are an essential component for the stabilization and crystallization of nucleic acids [24], and Mg^{2+} has been widely used for this purpose. Accordingly, in many cases, high resolution structures contain metals bound to DNA molecules (Table S9). We found that 64.2% of DNA crystals contain metals/ligands. However, this is not true when the DNA crystallizes in a non-B-DNA conformation, like A- or left-handed Z-DNA since 44.7% and 58.9% of Z- and A-DNAs, respectively, are crystallized in the absence of metals/ligands (Table S9). Alternatively, the effect of metals can be explained by their roles in stabilizing the DNA structure and packing interaction in a crystal lattice since we found that intermolecular packing is also achieved by the metal coordination among the mid-region phosphates in addition to the terminal-end interaction. Therefore, 12 bp dsODN shows higher metal dependency than 6 bp dsODN; 32.8% of crystals with 12 bp DNA contain the Mg^{2+} coordination, while only 9.1% 6 bp-crystals have the Mg^{2+} coordination (Table S9).

It is possible that current results and analyses might be insufficient to explain the principle of DNA crystallization and their packing interaction due to the lack of comprehensiveness of key factors extracted in this study. For example, several factors affecting crystallization such as buffer composition, precipitants, additives, and pH have not been considered possibly because some of them are not fully identified due to their low occupancy. This makes it especially impossible to know the identity and quantity of these components that finally end up in crystals [20], probably a result of crowded packing environment. Indeed, the packing contacts mediated by buffer and other excluded components would make for an interesting and challenging study in future. For this reason, our investigations in this study were exclusively restricted to only the direct inter molecular contacts made by DNA with symmetry equivalent molecules. Results from this study should aid in making successful crystallization strategies such as use of metals for the crystallization of lengthy dsODNs. In addition, this study also confirmed that the nature of dsODNs such as length, conformation, and existence of overhang must be considered for designing crystallization experiments. Therefore, we expect that the current study is a starting point for a comprehensive understanding of the principle of DNA crystallization and developing a better crystallization strategy, which contributes to the development of nucleic acid structural biology.

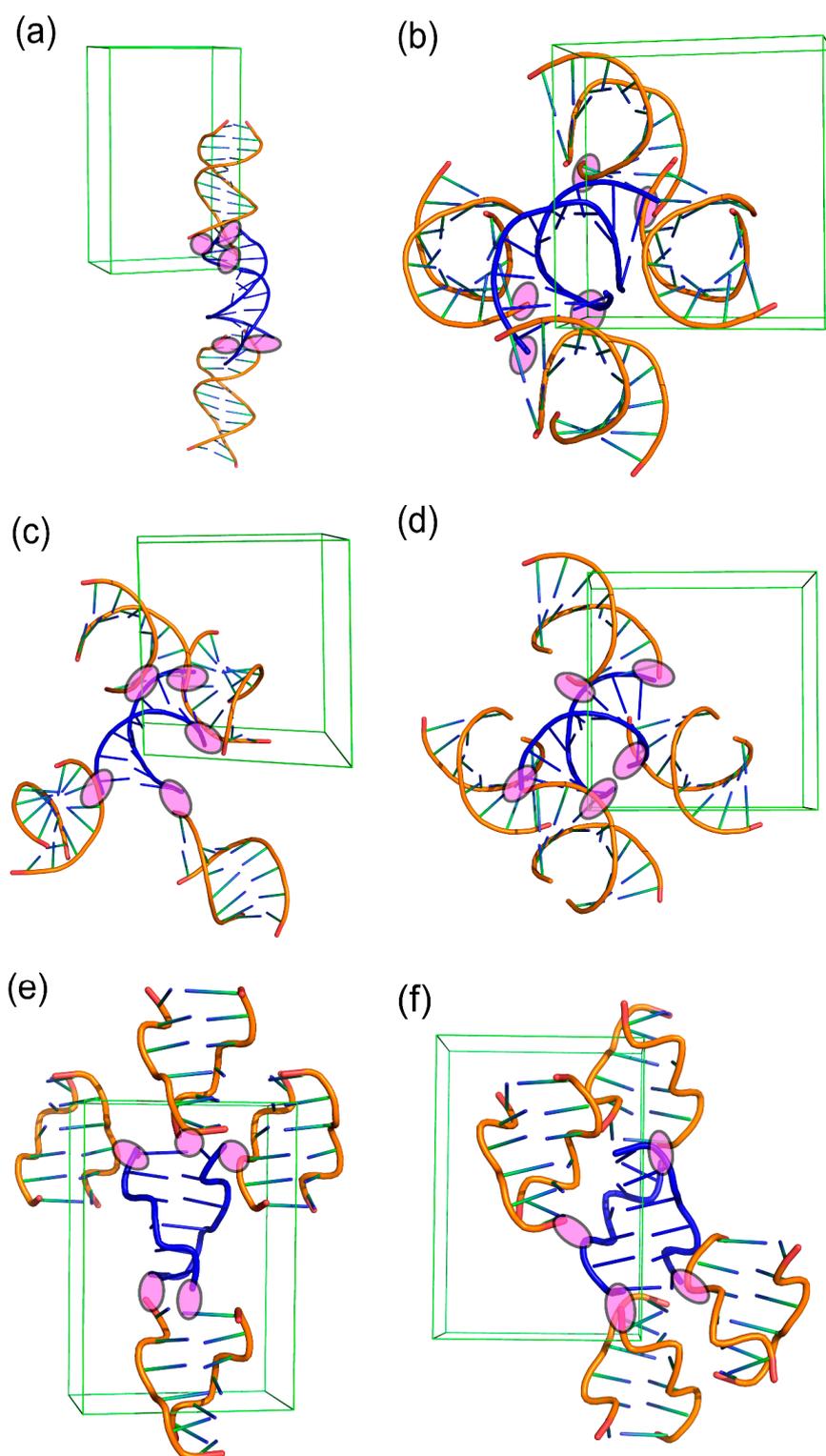


Figure 4. The packing interactions found in various symmetry, packing type, and conformation. Representative DNA structures belonging to the groups with following combinations, conformation-packing type-space group, are used for presenting structures and intermolecular H-bond interaction in each group, a unit cell is also shown. (a) B-DNA-Type 3- $P2_12_12_1$ (PDB ID 388D), (b) A-DNA-Type 2- $P2_12_12_1$ (PDB ID 321D), (c) A-DNA-Type 2- $P6_1$ (PDB ID 1D91), (d) A-DNA-Type 3- $P2_12_12_1$ (PDB ID 371D), (e) Z-DNA-Type 2- $P2_12_12_1$ (PDB ID 133D), and (f) Z-DNA-Type 3- $P2_12_12_1$ (PDB ID 6BST). The representative structure in each group is shown in blue and its symmetry equivalent molecules in coral. The intermolecular H-bonds are shown as magenta filled ellipses.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4352/10/12/1093/s1>, Table S1: Entire Dataset; Table S2: Correlation matrix; Table S3: Packing type vs. Conformation; Table S4: Sequence length vs. Packing type, Conformation, Resolution; Table S5: Packing type vs. Sequence composition; Table S6: Packing type vs. Interacting bases and entity; Table S7: Resolution vs. Packing type, Conformation; Table S8: Symmetry vs. Packing type, Conformation, Sequence length, Resolution; Table S9: Metal/Ligand vs. Packing type, Conformation, Sequence Length, Resolution; Table S10: Solvent content vs. Packing type, Conformation, Sequence length, Resolution, Symmetry, Metal/Ligand, Sequence; Table S11: Packing type vs. Populous Symmetry title.

Author Contributions: A.S., K.K.K. designed the work. A.S. conceived and performed the computational work and A.S., N.P., V.K.S. interpreted the results. A.S., V.K.S., K.K.K. wrote the paper. V.K.S. and K.K.K. proofread the manuscript. K.K.K. provides the computer resources. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Research Foundation of Korea funded by the Ministry of Science and ICT (2019R1A2C2089148 and 2020R1A4A1018019 to KK) and by Overseas Ph.D. Scholarships under ADP, University of Agriculture, Faisalabad, to AS.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Drew, H.R.; Wing, R.M.; Takano, T.; Broka, C.A.; Tanaka, S.; Itakura, K.; Dickerson, R.E. Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2179–2183. [[CrossRef](#)]
2. Arnott, S.; Hukins, D. Optimised parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504–1509. [[CrossRef](#)]
3. Ravichandran, S.; Subramani, V.K.; Kim, K.K. Z-DNA in the genome: From structure to disease. *Biophys. Rev.* **2019**, *11*, 383–387. [[CrossRef](#)] [[PubMed](#)]
4. Choi, J.; Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011**, *40*, 5893–5909. [[CrossRef](#)] [[PubMed](#)]
5. Zeraati, M.; Langley, D.B.; Schofield, P.; Moye, A.L.; Rouet, R.; Hughes, W.E.; Bryan, T.M.; Dinger, M.E.; Christ, D. I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **2018**, *10*, 631–637. [[CrossRef](#)] [[PubMed](#)]
6. Parveen, N.; Shamim, A.; Cho, S.; Kim, K.K. Computational Approaches to Predict the Non-canonical DNAs. *Curr. Bioinform.* **2019**, *14*, 470–479. [[CrossRef](#)]
7. Spiegel, J.; Adhikari, S.; Kendrick, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2020**, *2*, 123–136. [[CrossRef](#)]
8. Ravichandran, S.; Ahn, J.-H.; Kim, K.K. Unraveling the Regulatory G-Quadruplex Puzzle: Lessons From Genome and Transcriptome-Wide Studies. *Front. Genet.* **2019**, *10*, 1002. [[CrossRef](#)]
9. Zheng, J.; Birktoft, J.J.; Chen, Y.; Wang, T.; Sha, R.; Constantinou, P.E.; Ginell, S.L.; Mao, C.; Seeman, N.C. From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nat. Cell Biol.* **2009**, *461*, 74–77. [[CrossRef](#)]
10. Zhang, W.; Szostak, J.W.; Huang, Z. Nucleic acid crystallization and X-ray crystallography facilitated by single selenium atom. *Front. Chem. Sci. Eng.* **2016**, *10*, 196–202. [[CrossRef](#)]
11. Saenger, W. *Principles of Nucleic Acid Structure*; Springer: Berlin/Heidelberg, Germany, 1984. [[CrossRef](#)]
12. Egli, M. Nucleic acid crystallography: Current progress. *Curr. Opin. Chem. Biol.* **2004**, *8*, 580–591. [[CrossRef](#)] [[PubMed](#)]
13. Narayanan, B.C.; Westbrook, J.; Ghosh, S.; Petrov, A.I.; Sweeney, B.; Zirbel, C.L.; Leontis, N.B.; Berman, H.M. The Nucleic Acid Database: New features and capabilities. *Nucleic Acids Res.* **2013**, *42*, D114–D122. [[CrossRef](#)] [[PubMed](#)]
14. Burley, S.K.; Berman, H.M.; Christie, C.; Duarte, J.M.; Feng, Z.; Westbrook, J.; Young, J.; Zardecki, C. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* **2018**, *27*, 316–330. [[CrossRef](#)] [[PubMed](#)]
15. Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007**, *372*, 774–797. [[CrossRef](#)] [[PubMed](#)]
16. Weichenberger, C.X.; Rupp, B. Ten years of probabilistic estimates of biocrystal solvent content: New insights via nonparametric kernel density estimate. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2014**, *70*, 1579–1588. [[CrossRef](#)]
17. Matthews, B. Solvent content of protein crystals. *J. Mol. Biol.* **1968**, *33*, 491–497. [[CrossRef](#)]

18. McNicholas, S.; Potterton, E.; Wilson, K.S.; Noble, M.E.M. Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr.* **2011**, *67*, 386–394. [[CrossRef](#)]
19. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
20. Kuzmanic, A.; Dans, P.D.; Garcia-Lopez, A. An In-Depth Look at DNA Crystals through the Prism of Molecular Dynamics Simulations. *Chem* **2019**, *5*, 649–663. [[CrossRef](#)]
21. Padmaja, N.; Ramakumar, S.; Viswamitra, M.A. Space-group frequencies of proteins and of organic compounds with more than one formula unit in the asymmetric unit. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1990**, *46*, 725–730. [[CrossRef](#)]
22. Wukovitz, S.W.; Yeates, T.O. Why Protein Crystals Favor Some Space-Groups over Others. *Nat. Struct. Biol.* **1995**, *2*, 1062–1067. [[CrossRef](#)]
23. Schwarzenbach, D. Acta Crystallographica Section A: Foundations of Crystallography. *Acta Crystallogr. Sect. A Found. Crystallogr.* **2008**, *64*, 167. [[CrossRef](#)]
24. Gao, Y.-G.; Sriram, M.; Wang, A.H.-J. Crystallographic studies of metal ion—DNA interactions: Different binding modes of cobalt(II), copper(II) and barium(II) to N7 of guanines in Z-DNA and a drug-DNA complex. *Nucleic Acids Res.* **1993**, *21*, 4093–4101. [[CrossRef](#)] [[PubMed](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).