

The Volume Changes of Unfolding of dsDNA.

George I Makhatadze^{1}, Calvin R. Chen¹, Irine Khutsishvili², and Luis A. Marky²*

¹ *Departments of Biological Sciences, Chemistry and Chemical Biology, and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180, United States.*

² *Department of Pharmaceutical Sciences, University of Nebraska Medical Center, 986025 Nebraska Medical Center, Omaha, Nebraska 68198-6025, United States.*

* Corresponding author: George I. Makhatadze, CBIS 3244A, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180. Phone: (518) 276-4417; Fax: (518) 276-2955; E-mail: makhag@rpi.edu.

Keywords: double stranded DNA; hydrostatic pressure; stability; volume change; molecular dynamics simulations; pressure perturbation calorimetry.

Abstract

High hydrostatic pressure can have profound effects on the stability of biomacromolecules. The magnitude and direction (stabilizing or destabilizing) of this effect is defined by the volume changes in the system, ΔV . Positive volume changes will stabilize the starting native state, while negative volume changes will lead to the stabilization of the final unfolded state. For the DNA double helix, experimental data suggested that when the thermostability of dsDNA is below 50°C, increase in hydrostatic pressure will lead to destabilization, i.e. helix-to-coil transition has negative ΔV . In contrast, the dsDNA sequences with the thermostability above 50°C, showed positive ΔV values and were stabilized by hydrostatic pressure. In order to get insight into this switch in the response of dsDNA to hydrostatic pressure as a function of temperature first we further validated this trend using experimental measurements of ΔV for 10 different dsDNA sequences using pressure perturbation calorimetry. We also developed a computational protocol to calculate the expected volume changes of dsDNA unfolding, which was benchmarked against the experimental

set of 50 ΔV values that included, in addition to our data, the values from the literature. Computation predicts well the experimental values of ΔV . Such agreement between computation and experiment lends credibility to the computation protocol and provides molecular level rational for the observed temperature dependence of ΔV that can be traced to the hydration. Difference in the ΔV value for A/T vs G/C base pairs is also discussed.

Significance

There is an apparent paradox in how dsDNA responds to high hydrostatic pressure at low and high temperatures. At low temperatures, the dsDNA structure is destabilized by pressure while at high temperatures dsDNA is stabilized by the pressure. To this end, we first validated the exiting experimental observations by performing additional experiments. We then followed it with extensive computational modeling that not only reproduces well the experimental data but also identifies critical role of changes in hydration upon dsDNA unfolding as a molecular rational for the apparent paradox.

Introduction

Nucleic acids are one of the most essential biomolecules for life. They function to store the genetic code of the cell, i.e. the programming that dictates when genes are turned on and off, what proteins the cell can express, and is the medium of propagation of information across generations. Certain ribonucleic acid structures even possess enzymatic activity. Termed ribozymes, they possess the ability to aid in the synthesis of proteins (1), duplication of nucleotide sequences (2), cleavage of polynucleotide chains (3), and many other enzymatic activities crucial for early life. The most well-known ribozyme is the ribosome which catalyzes the synthesis of proteins (1). The ability of RNA to carry out both storage of genetic information and enzymatic activity lends credence to the RNA world hypothesis, which posits that RNA was the first complex macromolecule formed by natural chemical reactions occurring on early earth (4, 5). Due to the historical relevance of nucleic acids being progenitor building blocks of life, which may have evolved at extreme temperatures and pressures, it is important to understand the biophysical characteristics that allowed early nucleic acids to remain stable and functional at high temperatures and pressures. To address that topic, we first need to understand the effect of pressure on the stability of nucleic acids, which is directly related to the volume changes upon unfolding $(d\Delta G/dP)_T = \Delta V$. Thus, the response of the system to changes in pressure is driven by Le Chatelier's principle: if volume changes upon unfolding are positive, increase in hydrostatic pressure will lead

to an increase in stability, while if the changes are negative, the stability will decrease with the increase in hydrostatic pressure. Previous studies using poly-deoxynucleotides with homo (poly(dA).poly(dT), and alternating sequences (poly(AT).poly(dAT) and poly(dGC).poly(dGC)) have shown that volume changes upon unfolding strongly depend on temperature (6-10). When combined with results of mixed deoxy/ribo polynucleotides, poly(dA).poly(rU), poly(rA).poly(dT), and poly(dA).poly(rU), values of ΔV appears to be negative at low temperatures and, after crossing zero at $\sim 50^\circ\text{C}$, becomes positive at high temperatures (11). This trend has not been further validated for heterogeneous sequences and no molecular explanation has been provided as to why there is a sign change in the volume change of the helix-coil transition of nucleic acid double helices.

Here we have measured the volume changes of unfolding of 10 different dsDNA templates using pressure perturbation calorimetry. We have also developed a computational protocol to calculate the volume changes upon unfolding of dsDNA. These computations were benchmarked against 50 experimental data points on volume changes of dsDNA unfolding. The analysis allowed us to attribute the temperature dependence of volume changes to the change of hydration upon dsDNA unfolding.

Materials and Methods

Oligodeoxynucleotides and Pressure perturbation Calorimetry

All DNA molecules were synthesized by IDT (Coralville, IA), purified by reverse-phase HPLC and desalted on a G-10 Sephadex column before being lyophilized to dryness. The concentration of the oligonucleotides was determined from absorbance measurements taken at 260 nm and 90°C using the molar absorptivity coefficients listed in **Table S1**, and were calculated using previously reported procedures (12). In the case of non-self-complementary duplexes, the average values of the two strands were used. Buffer solutions used in this study consisted of 10 mM sodium phosphate (NaPi) at pH 7.0 adjusted with salt to 0.1 M NaCl. All chemicals used in this study were reagent grade and used without further purification. A VP-DSC differential scanning calorimeter from Malvern MicroCal (Northampton, MA), equipped with a pressure perturbation accessory (PPC), was used to measure the temperature dependence of the expansion coefficient, $\alpha(T)$. The total volume changes upon unfolding of the DNA oligonucleotide were calculated from the temperature integration of $\alpha(T)$. Prior to the PPC experiment, a DSC melt is

carried out to determine the temperature range and temperature step to be used in the PPC experiment. A sample solution with concentration of duplex DNA in the range 0.1-0.2 mM is allowed to equilibrate against the same buffer solution, at constant temperature and external pressure. The external pressure is then increased by ~50 psi, causing heat to be absorbed differentially by the sample and reference cells. These heats, ΔQ , are obtained from integration of the compression and decompression peaks resulting from switching the external pressure on and off at particular temperatures determined by the DSC curve. The resulting values of ΔQ are then used to calculate $\alpha(T)$, using relationships described elsewhere (13, 14). Volume changes were obtained by integrating the $\alpha(T)$ using Origin software routines supplied by the manufacturer (see **Figure S1**). The density of the solutions of the PPC experiment is measured with an Anton Paar (Graz, Austria) DMA densitometer in the differential mode, using two 602-M micro cells, each with a volume of ~150 μL . The reference cell is filled with water while the measuring cell is filled with solution or buffer. The density, ρ , is calculated from the oscillation period T of the cell using the following relationship: $\rho = AT^2 + B$, where A and B are constants determined from calibrating densities (and periods) of water and air, as described elsewhere (15).

Native State Simulations of Nucleic Acids

The B-form helical conformations were generated from sequence using the computational toolkit 3DNA (16). Molecular dynamics simulations were performed using these starting nucleic acid helices. Each structure was placed in a dodecahedral box with such dimensions that there was a 1 nm distance from the structure to the edges. Structures were then solvated with TIP3P water, neutralized with 0.1 M excess NaCl, and passed through 1,000 steps of steepest descent energy minimization, 2 ns of NVT equilibration, 2 ns of NPT equilibration, and 200 ns of NPT production. A 2 fs time step was used for all equilibration and production runs. We used the Parrinello–Rahman (17) pressure control with a 2 ps relaxation time and a compressibility of $4.6 \times 10^{-5} \text{ atm}^{-1}$ and v-scale temperature coupling (18). LINCS (19) and SETTLE (20) algorithms were used to constraint high-frequency bond vibrations that allowed the use of a 2 ps integration step. The electrostatic interactions were modelled by the smooth particle mesh Ewald method (21), using a $75 \times 75 \times 75$ grid, with fourth-order charge interpolation and a real space cutoff of 1.0 nm. Two different force-fields were used to benchmark the effect of force-field parameters on the resulting volumetric properties of the native ensemble: charmm27 (22), and amber99bsc1 (23). Three different excess salt concentrations (0 M, 0.1 M, and 0.2 M NaCl) were initially used for selected

subset of sequences and found that they have a negligible contribution, in terms of volume and surface areas, on the native nucleic acid structures. Thus, the majority of simulations were done in 0.1 M NaCl. Coordinates for dsDNA were extracted from the production trajectory every 1 ns resulting in 200 pdb structures. Benchmarking runs using larger number of pdb files for analysis (saving more frequently, i.e. every 100 ps or 200 ps) did have an effect on the computed averaged properties. Errors were estimated using standard deviations of block averaging.

Simulations of the Nucleic Acid Unfolded State

The unfolded state of nucleic acids were simulated using the structure-based model (SBM) generator SMOG (24). The structure files of these nucleic acids were generated using 3DNA. The resulting coordinates in pdb format were used as an input into SMOG v 2.0 and were processed with the all-atom and shadow contact models (24, 25). The resulting outputs were GROMACS-compatible coordinate (gro) and topology (top) files. The output topology files were stripped of all pairwise interactions, leaving only the Lennard-Jones, bond, angle, dihedral, and improper contributions to the energy function. Binary run input files (tpr) files were generated from the coordinate and topology files. The SBM simulations were run in GROMACS 4.6.7 at a reduced temperature of 1.25 ($T_{\text{red}} = T_{\text{GROMACS}}/k_B$ so in GROMACS temperature units $T_{\text{red}} = 1.25 = T_{\text{GROMACS}} = 150 \text{ K}$) for 10^9 steps with a time step of 0.2 fs. The simulation conditions used included: no solvent, no periodic boundaries, charges, and a short-range neighbor list and van der Waals cut-off of 1.2 nm. Equilibration of radius of gyration was seen after 5×10^6 to 5×10^8 steps, depending on the sequence length. After equilibration, 200 pdb structures every 1.25×10^6 steps were extracted, resulting in a total of 200 structures per chain per nucleic acid. Benchmarking runs using a larger number of pdb files for analysis (saving more frequently, i.e. every 1.25×10^5 runs) did have an effect on the computed averaged properties. Errors were estimated using standard deviations of block averaging. The extracted structures underwent energy minimization for 1,000 steps in explicit TIP3P solvent and excess of 0.1 M NaCl using Charmm27 or Amber99bsc1 force-fields, with dielectric constants of 80 in GROMACS 4.6.3. This step was required to be able to add hydrogen atoms and adjust bond length, consistent with the simulations of the native state. The applicability of such modeling of the unfolded single stranded deoxypolynucleotides was validated by comparison of the computed values of radii of gyration, R_g , with the experimentally-measured ones (26). There is a good correlation between the R_g values. Importantly, similar dependence of

R_g on polydeoxynucleotide length is observed. In particular, calculated values of R_g have the same scaling factor as the experimental R_g (see **Figure S2**).

Results and Discussions

Volume Definitions

The volume changes of dsDNA upon unfolding in aqueous solution, ΔV_{Tot} , is defined as a difference between volumes of two unfolded single stranded DNA, V_{U1} and V_{U2} , and the volume of the native ssDNA, V_N : $\Delta V_{Tot} = V_{U1} + V_{U2} - V_N$. From the practical point of view, one can use a hypothetical thermodynamics cycle shown in **Figure 1**. This cycle allows to separately account for the volume changes associated with the interactions of the initial and final states with solvent water, i.e. hydration volumes $\Delta V_{Hyd} = V_{Hyd,U1} + V_{Hyd,U2} - V_{Hyd,N}$, and the volume changes associated with the conformational changes, i.e. void volumes $\Delta V_{void} = V_{void,U1} + V_{void,U2} - V_{void,N}$.

Solvent-excluded volume, V_{SE} , is the volume enclosed by the molecular surface (MS) of a molecule. Molecular surface is defined as the surface traced by the center of a probe of 1.4 Å rolled over a molecule with all atoms represented by van der Waals spheres. The void volume, V_{void} , is defined as the difference between the solvent-excluded volume and van der Waals volume of a molecule. All volume calculations are done using the ProteinVolume program, described previously and freely available at gmlab.bio.rpi.edu (27). It is important to note that the ProteinVolume program was originally written to calculate volumes of proteins. However, the name is a misnomer because during the calculations no distinction is made whether the structure file supplied in pdb format is for a protein, DNA, RNA or any other molecule.

Parameterization of Hydration Coefficients

The hydration volume, V_{Hyd} , of a solute is the volume change in the system upon transfer of a solute from gas phase into water (28). It is equal to the partial molar volume of the solute in aqueous solution, V_{aq} , minus the solvent-excluded volume, V_{SE} . It has been demonstrated previously that hydration thermodynamics can be parametrized in terms of corresponding surface areas (29-32). Such parametrization for volume has been recently done for protein functional groups (28, 33) but not for nucleic acids. To this end, a set of 35 model compounds (purines, pyrimidines, sugars, nucleobases, nucleosides and nucleotides, see **Table S2**) related to nucleic

acids with available experimental V_{aq} values (11, 34-38) was analyzed. The PDB structure of each model compound was generated using the CORINA webserver (39) and the V_{SE} calculated using ProteinVolume (27). Molecular surface area (MS) per atom for all model compound structures was analytically calculated using the MSMS software package (40). Each structure had its MS broken down into carbon, nitrogen, oxygen, and hydrogen surface areas. MS for hydrogen atoms was combined with the surface area of the corresponding heavy atom. The MS area for carbon atoms was defined as non-polar (MS_{NP}). MS of nitrogen and/or oxygen was treated as polar and assigned to either MS_{PolSug} or $MS_{PolBase}$, depending on whether it was part of a nucleobase or sugar moiety, respectively. MS for phosphate and the corresponding bound oxygen atoms was treated as MS_{Phos} . V_{Hyd} values for 35 model compounds were fitted to the following equation:

$$V_{Hyd}(T) = MS_{NP} \cdot k_{NP}(T) + MS_{PolSug} \cdot k_{PolSug}(T) + MS_{PolBase} \cdot k_{PolBase}(T) + MS_{Phos} \cdot k_{Phos} + a \quad (1)$$

where a is the hydration volume when MS_{tot} is zero, and thus corresponds to a volume of a water molecule (28, 33), and the individual coefficients $k_{NP}(T)$, $k_{PolSug}(T)$, $k_{PolBase}(T)$, and k_{Phos} corresponding to the volume of hydration of a unit of non-polar, polar sugar, polar base and phosphate areas, respectively, are linearly dependent on temperature as:

$$k_{NP}(T) = A_{NP} \cdot T(^{\circ}C) + B_{NP} \quad (2)$$

$$k_{PolSug}(T) = A_{PolSug} \cdot T(^{\circ}C) + B_{PolSug} \quad (3)$$

$$k_{PolBase}(T) = A_{PolBase} \cdot T(^{\circ}C) + B_{PolBase} \quad (4)$$

$$k_{Phos}(T) = B_{Phos} \quad (5)$$

The coefficients of the fits are reported in **Table S3**. The comparison of experimental and fitted values at 25°C showing quality of the fit is presented in **Figure 2**. A more detailed comparison of experimental and fitted values at all temperatures is given in the **Table S2**, again supporting the applicability of equations 1-5. It has to be noted that the experimental data on partial molar volumes of model compounds in aqueous solutions is available for the temperature range 18-55 °C and appear to be increasing linearly with the increase in temperature in this range (11, 34-38). Such an increase in the partial molar volume for different classes of model compounds has been

well documented in the literature dated back to the 1980 (see e.g. (41-43)). Importantly, this leads to a positive slope of the temperature dependence of V_{Hyd} because $V_{Hyd} = V_{aq} - V_{SE}$ and V_{SE} for small molecular weight compounds is essentially the van der Waals volume and thus is constant. The solute molecule upon dissolution, will take up the volume occupied by bulk water plus the volume that will be required by the rearrangement of water around the solvent. Since the volume of the solute molecule does not change, the hydration volume is defined as the changes in volume of water around solute relative to the bulk water. So the positive values of V_{Hyd} suggest that the volume of water around solute is larger than in the bulk. The positive slope of the temperature dependence of V_{Hyd} (which actually is the expansivity) suggests that volume occupied by water around solute increases faster with temperature, than the volume in the bulk. Recent studies of V_{aq} provided measurements into high temperature range (44, 45). These studies suggest a non-linear dependence of volume on temperature, particularly at high temperatures where the V_{aq} levels off and becomes independent of temperature (44-46). This leveling of V_{aq} will translate into leveling of the ΔV_{Hyd} at high temperature, which should be taken into consideration and will be discussed later.

Additivity Analysis of the Properties of Unfolded State

There are five different structural properties required for characterizing the volumetric or surface properties of poly-deoxynucleotides. The volumetric properties include the void volume, V_{void} . The surface properties are the non-polar molecular surface, MS_{NP} , the polar surface of the bases, $MS_{PolBase}$, the polar surface of the ribose, $MS_{PolSugar}$, and the polar surface of the phosphodiester bonds, MS_{Phosp} .

Each of these five properties for a set of 194 unique sequences ranging in length between 8 and 200 nucleotides (designated as LearnSetU) were fit to a simple additivity scheme described by the following equation:

$$Property = P_{term} + \sum_{X=A,T,C,G} P_X \cdot N_X \quad (6)$$

where N_X is the number of A, T, C or G deoxynucleotides in the sequence, P_X is the additive contribution of each of the four deoxynucleotides to a given property of a single-stranded polydeoxyoligonucleotide, and the P_{term} represents the sum of contributions of the 5'- and 3'-ends.

The values computed from the direct molecular dynamics simulations agree well with the fit using simple additivity as described by equation 6 (see **Tables S4** and **S5**, and **Figures S3-S4**). For example, the range of computed values for V_{void} for the LearnSetU is between 225 and 6577 Å³, with a max standard deviation of 20 Å³. The residual of the fit using equation 6 is between -22 and 22 Å³, well within the computed max standard deviation for V_{void} . The fit residuals are between -1.7% and +1.1% of the directly computed values of V_{void} . Similar results are obtained for other relevant properties (see **Tables S4** and **S5**, and **Figures S2-S3**). To further test the validity of the P_X coefficients obtained from the LearnSetU, we tested their applicability of equation 6 to the TestSet. The TestSetU consists of unique (i.e. different for the sequences included in the LearnSetU) ~120 sequences of length between 4 and 80 nucleotides (see **Tables S4** and **S5**, and **Figures S2-S3**). The maximum residuals for predicting values of V_{void} for the TestSetU using equation 6 are between -25 and 15 Å³, again well within the standard deviation of direct individual calculations. Importantly, the results are similar for two different force-fields, amber99bsc1 and charmm27, albeit with each force field requiring its own set of P_X coefficient (see **Table S5**). These findings altogether support the notion that the properties of the unfolded state of polydeoxynucleotides closely follow the simple additivity rules as described by eq. 6, and thus the parameters of the fit can be used to predict the properties of unfolded single-stranded polydeoxynucleotides of *any sequence and length*.

Additivity Analysis of the Properties of Native dsDNA

Two data sets have been generated for the analysis of the properties of the native dsDNA structure. The LearnSetN consists of 90 unique dsDNA sequences ranging in length between 8 and 80 base pairs (see **Table S6**). The results of analysis performed on the LearnSetN were further tested on the TestSetN that consists of 60 unique dsDNA sequences ranging in length between 9 and 120 base pairs (see **Table S6**).

The volumetric properties (void volume, V_{void}) and the surface properties (the non-polar molecular surface, MS_{NP} , the polar surface of the bases, MS_{PolBase} , the polar surface of the ribose, MS_{PolSugar} , and the polar surface of the phosphodiester bonds, MS_{Phosp}) obtained using simulations for dsDNA sequences from LearnSetN were analyzed. We first used a simple additivity of contributions from each of the two types of base pairs:

$$Property = P_{Ends} + \sum_{X=d(A/T),d(G/C)} P_X \cdot N_X \quad (7)$$

where N_X is the number of dA/T or dG/C base pairs in the sequence, P_X is the additive contribution of each of the two types of base pairs to a given property of dsDNA. P_{Ends} represents the contribution of the dsDNA ends. The fit of the volumetric or surface properties for dsDNA using such simple additivity (eq. 7) did not yield to a uniform distribution of the residual (see **Tables S6 and S7**, and **Figures S4-S5**). This effect was particularly evident for MS_{NP} , which is related in part to the extent of the non-polar surface formed by the methyl groups of the thymine base in repeating sequences. For example, the homopolymer poly(dA).poly(dT) will have the largest MS_{NP} surface while the alternating poly(dAT)poly(dTA) will have the smallest, and poly(dAATT).poly(dTTAA) somewhere in between. This suggests the existence of more intricate relationships between the sequence and the actual properties of interest. Application of the simple additivity calculations (eq. 7) to the TestSetN, consisting of a set of 60 unique dsDNA sequences ranging in length between 9 and 120 base pairs, further substantiated this conclusion (see **Figures S4-S5**).

We thus turn to a nearest-neighbor (NN) model that has been very successful in predicting other thermodynamic properties of dsDNA such as enthalpy, entropy, and melting temperature of the helix-to-coil transition in dsDNA (47):

$$Property = P_{Ends} + \sum P_{NN} \cdot N_{NN} \quad (8)$$

where N_{NN} is the number of nearest-neighbor types in the sequence, P_{NN} is the contribution of each type of nearest-neighbor to a given property of dsDNA, and P_{Ends} accounts for the contribution of the dsDNA ends. There are in total ten P_{NN} parameters (see **Table S8**), and this model significantly improves the distribution of the residuals of the fit (see **Figures S4-S5**). We note that the NN fit to eq. 8 for the LearnSetN data obtained using the amber99bsc1 force field was better than the fit using charmm27 data. The range of residuals for V_{void} was -19 to +43 Å³ and -64 to +73 Å³, respectively. However, this difference disappeared when the P_{NN} parameters, obtained from the corresponding force fields, were applied to TestSetN values. The range of residuals for V_{void} was 13 to +42 Å³ and -26 to +19 Å³ for amber99bsc1 and charmm27, respectively. Overall, it appears that simple additivity (eq. 7) better reproduces the values for shorter sequences and non-repetitive sequences while NN-additivity (eq. 8) works better for longer repetitive sequences.

Calculation of Volume Changes for dsDNA unfolding

The volume change upon unfolding of dsDNA as a function of temperature is defined as the difference in volumes of the two individual unfolded chains ($V_{U1} + V_{U2}$) and the volume of the native dsDNA (V_N):

$$\Delta V_{Tot}(T) = V_{U1}(T) + V_{U2}(T) - V_N(T) = (V_{U1,void} + V_{U2,void} - V_{N,void}) + (V_{U1,Hyd}(T) + V_{U2,Hyd}(T) - V_{N,Hyd}(T)) \quad (9)$$

The values of $V_{void,U1}$, $V_{void,U2}$ and $V_{void,N}$ can be calculated directly for structural ensembles obtained from simulations for the two unfolded chains and double stranded native DNA, as described in Materials and Methods, or using additivity schemes described above (eq. 6 and eqs. 7 or 8). The structural ensembles obtained from direct simulations are also used to calculate the average surface properties of each of these ensembles such as the non-polar molecular surface, MS_{NP} , the polar surface of the bases, $MS_{PolBase}$, the polar surface of the ribose, $MS_{PolSugar}$, and the polar surface of the phosphodiester bonds, MS_{Phosp} . These parameters are independent of temperature. Alternatively, these properties can be calculated using additivity schemes described above (eqs. 6 and 7 or 8). The surface properties of each of these ensembles are used to compute the values $V_{Hyd,U1}(T)$, $V_{Hyd,U2}(T)$ and $V_{Hyd,N}(T)$ at a given temperature using equation 1.

Comparison with Experimental Data.

The experimental set of values for the volume changes upon dsDNA unfolding, consists of 50 data points. We have measured the volume changes upon unfolding of 10 different dsDNA sequences using pressure perturbation calorimetry (see Materials and Methods). In addition, the survey of the published literature (11, 48, 49) for volume changes obtained using either pressure-induced unfolding at a constant pressure or temperature-induced unfolding at different pressure yielded 41 other data points (see **Table S9**). This includes both relatively short heterogeneous sequences as well as long homogeneous poly-deoxynucleotides, such as poly(dA).poly(dT) and poly(dGC).poly(dCG) (11). The experimentally measured values, ΔV_{exp} , expressed per mole of base pairs, as a function of temperature are shown in **Figure 3A**. There is an apparent trend whereby the ΔV_{exp} are negative at lower temperatures but then become positive

at temperatures above $\sim 50^{\circ}\text{C}$. Importantly, this trend is well reproduced by the calculated values of ΔV_{Tot} for the same sequences at the melting temperatures observed in the experiments. The exception are the predictions for poly(dGC).poly(dCG) that have very high melting temperatures ($100-114^{\circ}\text{C}$): the predicted values are larger than experimental for these sequences. This overestimate is probably due to linear extrapolation of hydration term into this temperature range. If we assume that the hydration volume dependence on temperature is leveling off at temperatures above 90°C as suggested by the experimental measurements of expansivity (46), then the agreement between experiment and calculations for these highly stable sequences is expected to be better (compare dashed and dotted lines in **Figure 3A**).

Overall, the predicted values correlate well with the experimentally measured values (see **Figure 3B**) and this is largely independent of the force field that was used for the computation or whether the direct computation or additivity schemes were used (see **Figures S7-S9**). Such good correspondence between experimental and computed values allows us to address two interrelated questions:

1. What defines the changes in sign of ΔV_{Tot} at $\sim 40-50^{\circ}\text{C}$? To answer this question, we should look at the temperature dependences of the individual contributions to ΔV_{Tot} , namely ΔV_{Void} and ΔV_{Hyd} . The values of ΔV_{Void} do not appear to depend on the temperature. This conclusion is based on the simulations of dsDNA and unfolded DNA at different temperatures, which shows that V_{Void} for both these states remains constant within the standard deviations of the calculations (see **Figure S10**). This is because dsDNA maintains native structure over a broad range of temperature. The corresponding ΔV_{Void} values also remain independent of temperature and negative in sign. The values of ΔV_{Hyd} on the other hand are positive and increase with an increase in temperature. Thus, it appears that at lower temperatures the positive values of ΔV_{Hyd} are not enough to overcome the negative ΔV_{Void} values, making ΔV_{Tot} negative at temperatures below $\sim 40^{\circ}\text{C}$. As the temperature increases, the positive values of ΔV_{Hyd} also increase and overcome the negative ΔV_{Void} values at temperatures above $\sim 50^{\circ}\text{C}$. As a result, ΔV_{Tot} becomes positive at temperatures above $\sim 50^{\circ}\text{C}$.

2. Is there a difference in the values of ΔV_{Tot} for dA.dT vs dG.dC base pairs? And if yes, what is the reason for such a difference? It has been noted before that the volume changes upon unfolding of AT-rich dsDNA sequences are always larger than that of GC-rich dsDNA when compared at the same temperature. This trend is reproduced by the calculations presented here

(see **Figure 3A**) and thus allows us to gain insight into the origins of these differences (see **Figure 4**). There is a small difference in the ΔV_{Void} values between an AT base pair and a GC base pair, $-34 \text{ \AA}^3$ versus $-32 \text{ \AA}^3$, respectively. However, the difference in the ΔV_{Hyd} further amplifies the difference in ΔV_{Tot} : at 20°C the ΔV_{Hyd} for an AT base pair is $30 \text{ \AA}^3$, which is somewhat larger than $24 \text{ \AA}^3$ for a GC base pair. Thus, differences in hydration between AT and GC base pairs contribute to the larger values of ΔV_{Tot} per AT base pair.

Concluding Remarks

The discussion above highlights the importance of the hydration term as a dominant factor in defining the sign and absolute values of ΔV_{Tot} . At low temperatures, independent of the sequence, dsDNA will be pressure unstable because ΔV_{Tot} is negative, i.e. increase in hydrostatic pressure will lead to unfolding. At higher temperatures, also independent of the sequence, ΔV_{Tot} will become positive, and dsDNA structure will be stabilized by an increase in hydrostatic pressure. More importantly, these effects will be dependent on the sequence only in a relatively narrow temperature range between $\sim 40^\circ\text{C}$ and $\sim 50^\circ\text{C}$. In this temperature range, AT-rich sequences will have positive ΔV_{Tot} while GC-rich sequences will have ΔV_{Tot} that is negative and thus destabilized by pressure.

We hope that the framework for calculating the volume changes upon unfolding of dsDNA presented here will be used by others. To facilitate this, we have launched a free web-service that allows computing volume changes upon unfolding of standard dsDNA of any sequence at a given temperature (<https://gmlab.bio.rpi.edu/DNAVolume.html>).

Supporting Material

Supporting material can be found online at

□ Supplementary Figures S1-S9 and Supplementary Tables S1, S3, S5, S7-S9 (PDF) □
Supplementary Tables S2, S4 and S6 (Excel)

Author Contributions

G.I.M. devised the project, the main conceptual ideas and proof outline. C.R.C. and G.I.M. developed the computational formalism that was extended and analyzed by G.I.M. I.K. performed

experiments and together with L.A.M. analyzed the data. G.I.M. wrote paper with the help of L.A.M., C.R.C. and I.K.

Declaration of Interests

The authors declare no competing interest.

Acknowledgments

This work was supported by grants CHEM/CLP-1803045 (to G.I.M.) and MCB-1912587 (to L.A.M) from the U.S. National Science Foundation (NSF). This work used the Extreme Science and Engineering Discovery Environment (XSEDE) comet (SDSC) and stampede2 (TACC) using allocation TG-MCB140107, which is supported by the U.S. National Science Foundation grant number ACI-1548562. Additional computational resources were provided by the Center for Computational Innovations at RPI.

REFERENCES

1. Cech, T. R. 2000. Structural biology. The ribosome is a ribozyme. *Science* 289:878-879.
2. Wochner, A., J. Attwater, A. Coulson, and P. Holliger. 2011. Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332:209-212.
3. Zaug, A. J., and T. R. Cech. 1986. The Tetrahymena intervening sequence ribonucleic acid enzyme is a phosphotransferase and an acid phosphatase. *Biochemistry* 25:4478-4482.
4. Condie, K. C. 2021. Earth as an evolving planetary system. Academic Press.
5. Gilbert, W. 1986. Origin of life: The RNA world. *Nature* 319:618-618.
6. Wu, J. Q., and R. B. Macgregor, Jr. 1995. Pressure dependence of the helix-coil transition temperature of poly[d(G-C)]. *Biopolymers* 35:369-376.
7. Wu, J. Q., and R. B. Macgregor, Jr. 1993. Pressure dependence of the melting temperature of dA.dT polymers. *Biochemistry* 32:12531-12537.
8. Chalikian, T. V., J. Volker, G. E. Plum, and K. J. Breslauer. 1999. A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques. *Proc. Natl. Acad. Sci. U. S. A.* 96:7853-7858.
9. Rentzeperis, D., D. W. Kupke, and L. A. Marky. 1993. Volume changes correlate with entropies and enthalpies in the formation of nucleic acid homoduplexes: differential hydration of A and B conformations. *Biopolymers: Original Research on Biomolecules* 33:117-125.

10. Marky, L. A., and R. B. Macgregor, Jr. 1990. Hydration of dA.dT polymers: role of water in the thermodynamics of ethidium and propidium intercalation. *Biochemistry* 29:4805-4811.
11. Chalikian, T. V., and K. J. Breslauer. 1998. Volumetric properties of nucleic acids. *Biopolymers* 48:264-280.
12. Cantor, C. R., M. M. Warshaw, and H. Shapiro. 1970. Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers: Original Research on Biomolecules* 9:1059-1077.
13. Carr, C. E., I. Khutsishvili, and L. A. Marky. 2018. Energetics, ion, and water binding of the unfolding of AA/UU base pair stacks and UAU/UAU base triplet stacks in RNA. *The Journal of Physical Chemistry B* 122:7057-7065.
14. Schweiker, K. L., and G. I. Makhataдзе. 2009. Use of pressure perturbation calorimetry to characterize the volumetric properties of proteins. *Methods Enzymol.* 466:527-547.
15. Kankia, B. I., A. M. Soto, N. Burns, R. Shikiya, C. S. Tung, and L. A. Marky. 2002. DNA oligonucleotide duplexes containing intramolecular platinated cross-links: Energetics, hydration, sequence, and ionic effects. *Biopolymers: Original Research on Biomolecules* 65:218-227.
16. Li, S., W. K. Olson, and X. J. Lu. 2019. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* 47:W26-W34.
17. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* 52:7182-7190.
18. Pronk, S., S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845-854.
19. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463-1472.
20. Miyamoto, S., and P. A. Kollman. 1992. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952-962.
21. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103:8577-8593.
22. Brooks, B. R., C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczero, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. 2009. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 30:1545-1614.
23. Ivani, I., P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, and A. Balaceanu. 2016. Parmbsc1: a refined force field for DNA simulations. *Nature methods* 13:55-58.
24. Noel, J. K., M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic, and P. C. Whitford. 2016. SMOG 2: A versatile software package for generating structure-based models. *PLoS Comput. Biol.* 12:e1004794.

25. Noel, J. K., P. C. Whitford, and J. N. Onuchic. 2012. The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B* 116:8692-8702.
26. Sim, A. Y., J. Lipfert, D. Herschlag, and S. Doniach. 2012. Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering. *Physical Review E* 86:021901.
27. Chen, C. R., and G. I. Makhatadze. 2015. ProteinVolume: calculating molecular van der Waals and void volumes in proteins. *BMC Bioinformatics* 16:1-6.
28. Chen, C. R., and G. I. Makhatadze. 2017. Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nat Commun.* 8:14561-14561.
29. Makhatadze, G. I., and P. L. Privalov. 1993. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* 232:639-659.
30. Lopez, M. M., and G. I. Makhatadze. 1998. Solvent isotope effect on thermodynamics of hydration. *Biophys. Chem.* 74:117-125.
31. Privalov, P. L., and G. I. Makhatadze. 1993. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* 232:660679.
32. Makhatadze, G. I., and P. L. Privalov. 1994. Energetics of interactions of aromatic hydrocarbons with water. *Biophys. Chem.* 50:285-291.
33. Chen, C. R., and G. I. Makhatadze. 2017. Molecular Determinants of Temperature Dependence of Protein Volume Change upon Unfolding. *J. Phys. Chem. B* 121:8300-8310.
34. Lee, A., and T. V. Chalikian. 2001. Volumetric characterization of the hydration properties of heterocyclic bases and nucleosides. *Biophys. Chem.* 92:209-227.
35. Buckin, V. A. 1988. Hydration of nucleic bases in dilute aqueous solutions. Apparent molar adiabatic and isothermal compressibilities, apparent molar volumes and their temperature slopes at 25 degrees C. *Biophys. Chem.* 29:283-292.
36. Chalikian, T. V. 1998. Ultrasonic and Densimetric Characterizations of the Hydration Properties of Polar Groups in Monosaccharides. *J. Phys. Chem. B* 102:6921-6926.
37. Kishore, N., and J. C. Ahluwalia. 1990. Partial Molar Heat Capacities and Volumes of Some Nucleic Acid Bases, Nucleosides and Nucleotides in Aqueous Urea Solutions at 298.15 K. *J. Chem. Soc. Faraday Trans.* 86:905-910.
38. Kishore, N., and J. C. Ahluwalia. 1990. Partial Molar Heat Capacity and Volumes of Transfer of Nucleic Acid Bases, Nucleosides and Nucleotides from Water to Aqueous Solutions of Sodium and Calcium Chloride at 25°C. *J. Solution Chem.* 19:51-64.
39. Sadowski, J., J. Gasteiger, and G. Klebe. 1994. Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Model.* 34:1000-1008.
40. Sanner, M. F., A. J. Olson, and J. C. Spehner. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305-320.
41. Durchschlag, H. 1986. Specific volumes of biological macromolecules and some other molecules of biological interest. In *Thermodynamic data for biochemistry and biotechnology*. Springer. 45-128.
42. Hoiland, H. 1986. Thermodynamic data for Biochemistry and Biotechnology. *Chap. 2 Partial Molar Volumes of Biochemical Model Compounds in Aqueous Solutions*:17-44.
43. Cabani, S., P. Gianni, V. Mollica, and L. Lepori. 1981. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Solution Chem.* 10:563-595.

44. Makhatadze, G., and P. Privalov. 1989. Heat capacity of alcohols in aqueous solutions in the temperature range from 5 to 125 C. *J. Solution Chem.* 18:927-936.
45. Makhatadze, G. I., V. N. Medvedkin, and P. L. Privalov. 1990. Partial molar volumes of polypeptides and their constituent groups in aqueous solution over a broad temperature range. *Biopolymers*.
46. Lin, L.-N., J. F. Brandts, J. M. Brandts, and V. Plotnikov. 2002. Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Anal. Biochem.* 302:144-160.
47. SantaLucia Jr, J., and D. Hicks. 2004. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 33:415-440.
48. Dubins, D. N., and R. B. Macgregor, Jr. 2004. Volumetric properties of the formation of double stranded DNA: a nearest-neighbor analysis. *Biopolymers* 73:242-257.
49. Lin, M. C., and R. B. Macgregor, Jr. 1997. Activation volume of DNA duplex formation. *Biochemistry* 36:6539-6544.

Figure Legends

Figure 1. Hypothetical thermodynamic cycle for computing the volume changes of unfolding of dsDNA in aqueous solution, ΔV_{Tot} . Considering that the volume is a state function, the native dsDNA can be transferred into the gas phase. This process will be accompanied by changes in the volume of hydration of the native state $V_{Hyd,N}$. The dsDNA, in the gas phase, is then unfolded into two single-stranded DNA molecules. On this step, the relevant volume change for the overall thermodynamic cycle will be represented by ΔV_{void} . Finally, the two unfolded single-stranded DNA molecules will be transferred back from gas phase into the aqueous solution. This process will be accompanied by the change in the interactions between solvent water and DNA and corresponds to the volume of hydration of unfolded state, $V_{Hyd,U1}$ and $V_{Hyd,U2}$. Thus the net change in volume in aqueous solution will be: $\Delta V_{Tot} = \Delta V_{void} + (V_{Hyd,U1} + V_{Hyd,U2} - V_{Hyd,N})$.

Figure 2. Comparison of experimental and fitted (using equation 1) values of hydration volume for model compounds at 25°C. The correlation coefficient $R^2=0.994$ and the slope is 0.99 ± 0.01 . Coefficients of the fit are listed in **Table S3**, and comparison of the experimental and fitted values at other temperatures are given in **Table S2**.

Figure 3. Comparison of experimental and computed values for volume change upon unfolding of dsDNA. **Panel A.** Volume change upon double stranded DNA unfolding, ΔV_{exp} , for over 50

different experimental data points compared with the value computed at the same temperature as the experiment. Red symbols are values computed from direct simulations (amber99bsc1 forcefield), while green symbols are for the polynucleotides, with values computed using simple additivity contributions derived for amber99bsc1 force-field. Analogous plots for the nearestneighbor (NN) and/or charmm27 force-field show similar trends (see **Figures S7, S8, and S9**). Error bars for each point are smaller than the symbols (please see **Table S9** for actual estimated errors). **Panel B.** Dependence of volume change upon dsDNA unfolding, ΔV_{Tot} , on temperature. Blue circles - experimental data; red triangles - values computed from direct simulations using amber99bsc1 force-field; green triangles - values computed using simple additivity contributions derived for amber99bsc1 force-field (see **Tables S5 and S7**). The solid lines are calculated temperature dependences of volume changes using linear extrapolation of V_{Hyd} on temperature for AT (red) and GC (green) base pairs. The dashed lines show expected temperature dependencies if V_{Hyd} levels off at high temperatures.

Figure 4. Contribution from changes in void volume, and hydration to the total volume changes of unfolding of AT or GC base pairs, at 20°C and 60°C. Gray bars - ΔV_{void} ; blue bars - ΔV_{hyd} ; dark-red bars - ΔV_{Tot} for AT base pair; dark-green bars - ΔV_{Tot} for GC base pair.

Figure 1

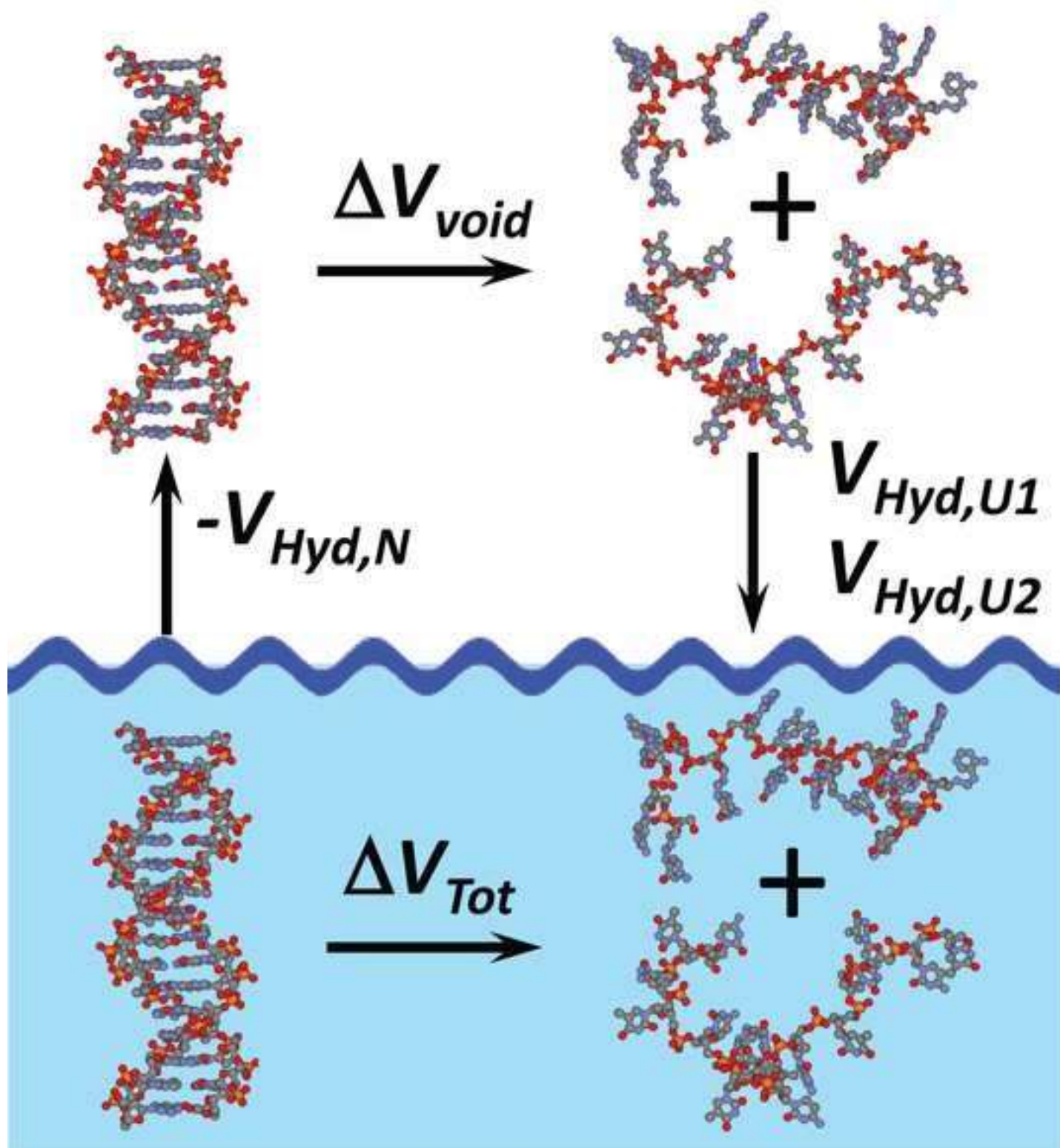


Figure2

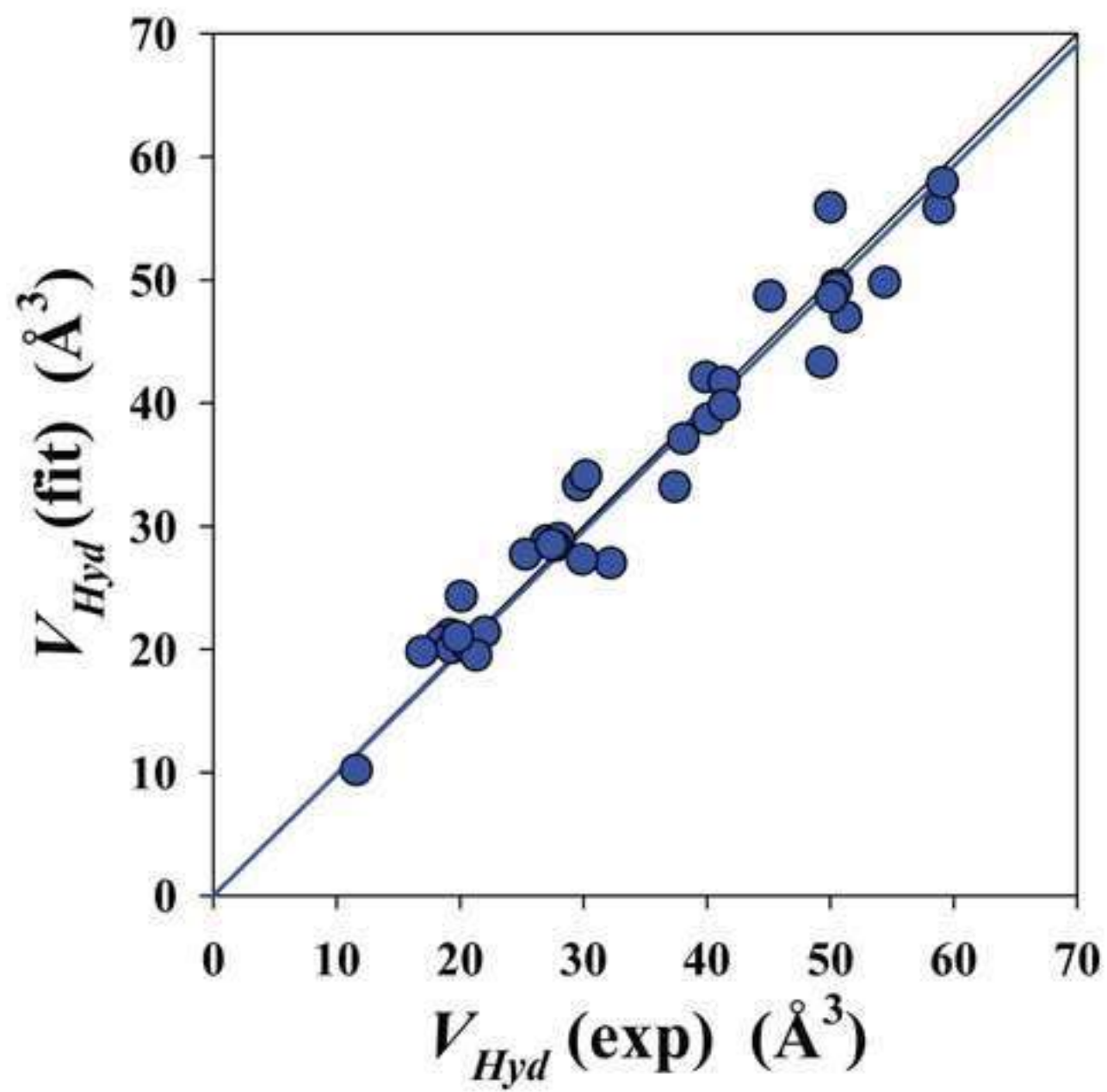


Figure3

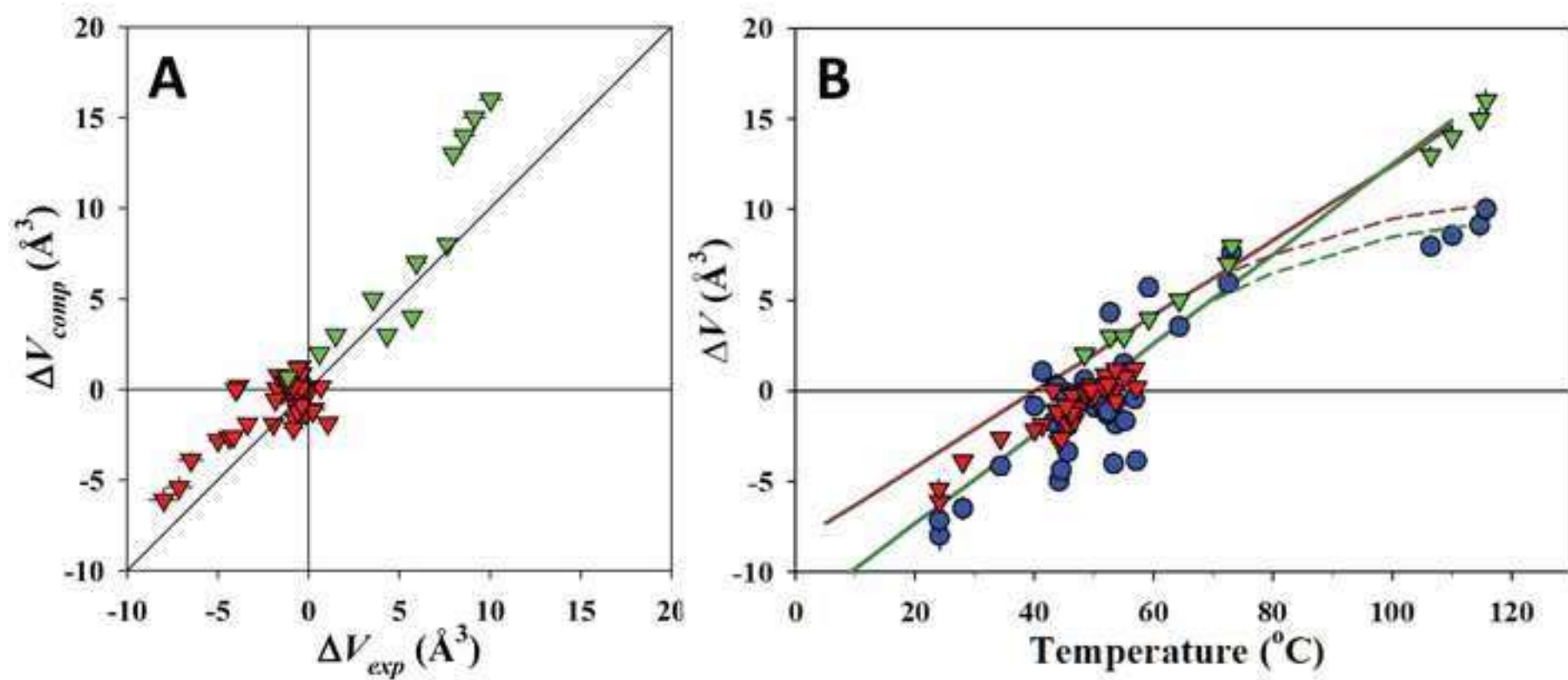
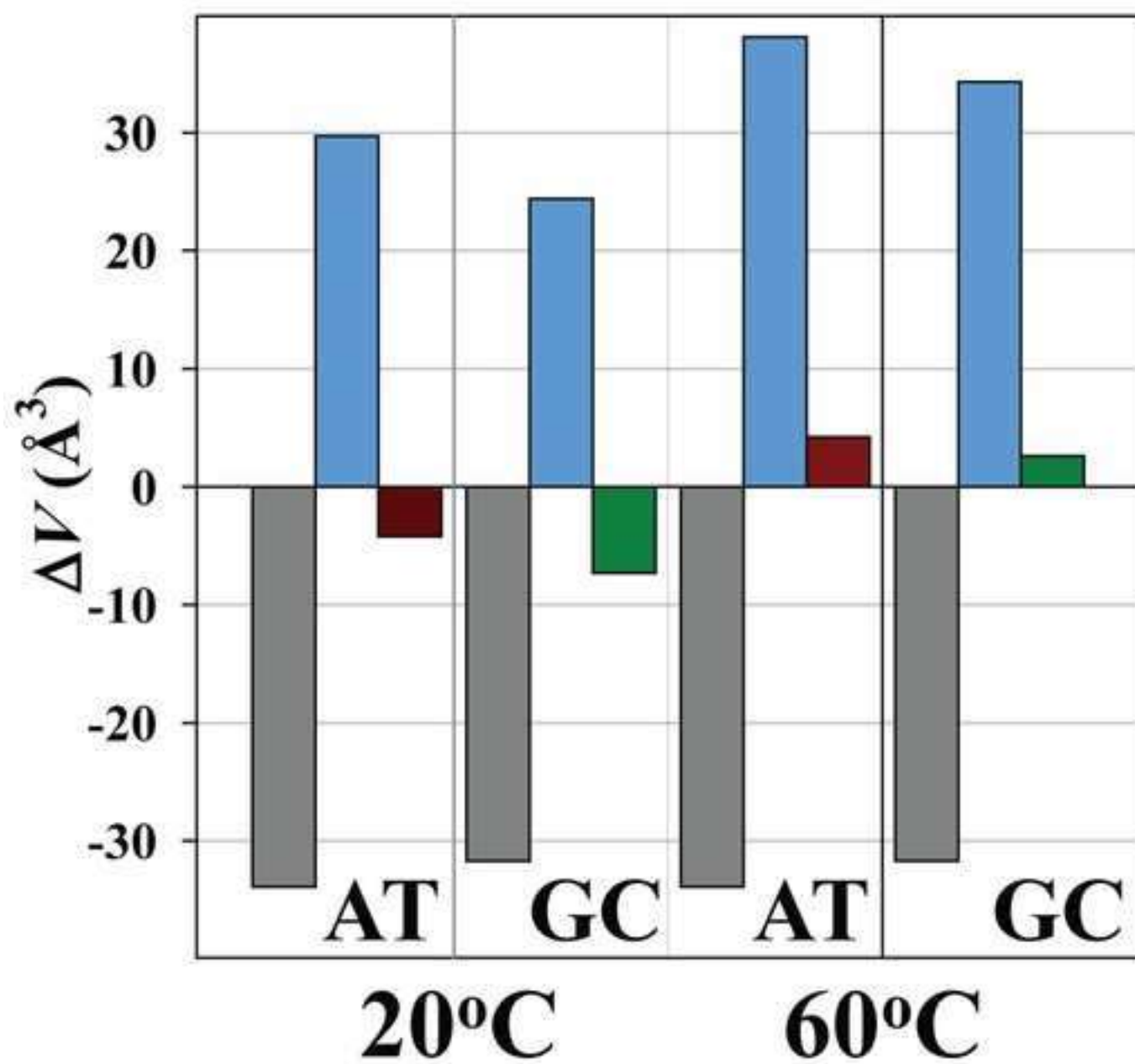
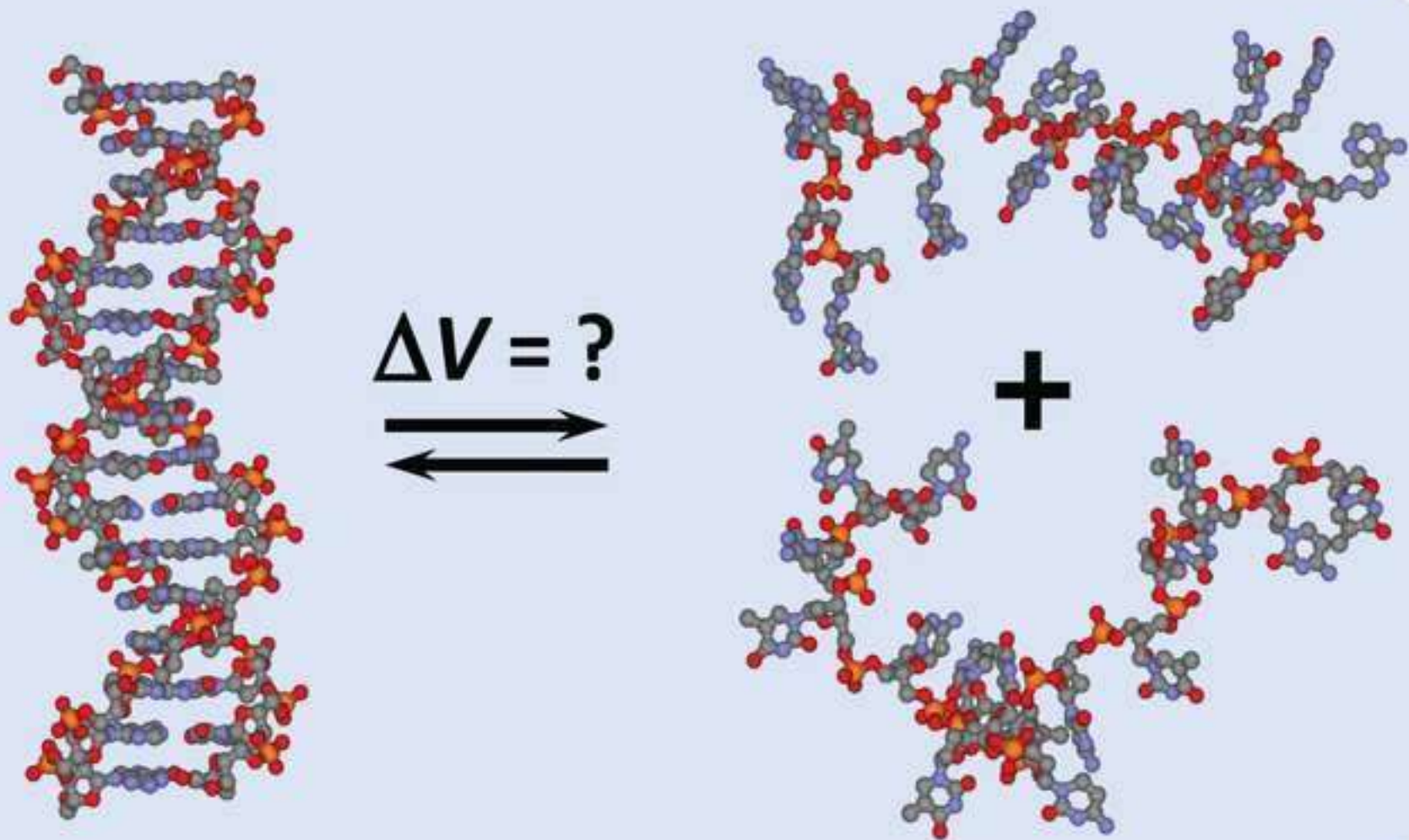


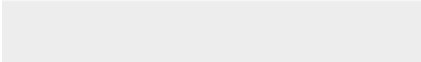
Figure4

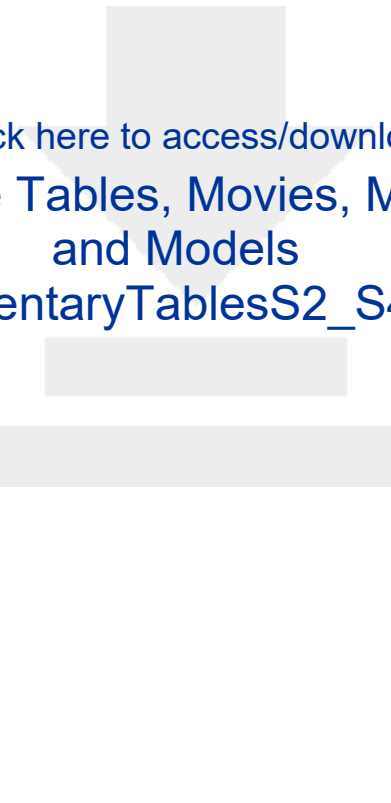




[Click here to access/download](#)
Supplemental Information
DNA_Volume-v.7.0_BJ-SI.pdf

Supplemental Large Tables, Movies, MATLAB, and Data and
Models





[Click here to access/download](#)

Supplemental Large Tables, Movies, MATLAB, and Data
and Models

SupplementaryTablesS2_S4_S6.xlsx

