PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

PCCP

RSCPublishing

ARTICLE

Kinetics of DNA duplex formation: A-tracts *versus* AT-tracts

Cite this: DOI: 10.1039/x0xx00000x

Jean Ann Wyer,*[a] Mads Bejder Kristensen,[a] Nykola C. Jones,[a] Søren Vrønning Hoffmann[a] and Steen Brøndsted Nielsen[a]

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

The hybridisation and melting of DNA strands are critical steps in many biological processes, but still a deeper understanding of the kinetics is lacking. This is evident from the absence of a clear correlation between rate constants for duplex formation and the number of bases in the strand or the sequence. Here we have probed differences between formation times of A-tracts and AT-tracts by studying complementary model strands mainly comprised of adenine (A) and thymine (T) in stopped-flow (SF) experiments. These strands are relevant as DNA replication begins in regions with large numbers of AT base pairs. Interpretation of our results is aided by secondary-structure modelling where both the fractions of the different types of structures and the number of paired bases in the lowest-energy ones are determined. The model is based on calculations of free energies using fixed values for enthalpies and entropies associated with base pairing and a stochastic sampling of possible structures. We find that the strand length affects rates: The activation energy for the formation of short (16-base pairs) A-tracts is larger than that for longer ones (20-base pairs). Activation energies for the formation of AT-tracts is an order of magnitude larger, and larger for shorter strands than for long ones. These higher activation energies are in agreement with the fact that the fraction of unpaired bases in the constituent AT-tract strands is less than in those which comprise the A-tracts. That the pre-structures of the single strands significantly affect rates is also used to rationalise results for two pairs of complementary 12-mer strands that have the same bases but in different sequence; we report here similar activation energies as earlier reported and that these are strongly sequence dependent. Finally, we demonstrate that SF can be coupled with the measurement of circular dichroism (CD) in the vacuum ultraviolet (VUV) region, taking advantage of a synchrotron radiation facility, and that CD is useful to probe geometrical structures in the VUV where the absorption by DNA is high. Though this work is preliminary, our initial results suggest that the strands align prior to the formation of base pairs.

Introduction

Two crucial stages of many important biological operations in cells are the separation of two DNA strands from each other and the reunification of them, processes which are only somewhat understood.[1-4] A fundamental molecular level comprehension of pairing between strands is of interest across a broad range of fields, from the fundamental, where such information could be valuable in order to quantify the dynamics and energetics of processing genetic information, to the applied, in disciplines such as DNA-based nanotechnology. Indeed, hybridisation, in which an oligonucleotide probe recognises and binds to its complementary target, is the most vital component of DNA- and RNA-based biosensor technology.[5] Furthermore, quantitative kinetic and thermodynamic aspects of duplex formation are relevant for the successful design of analytical or biomedical applications such as the polymerase chain reaction, where a piece of DNA is replicated many times in order to produce millions of copies of a particular DNA sequence,[6] or *in situ* hybridisation.[7] Despite this, there are relatively few quantitative studies on DNA-duplex formation and melting.

It has been shown that for the particular duplexes studied, DNA-duplex formation follows second-order reaction kinetics.[4, 5, 8, 9] In general, the hybridisation is initiated by the

formation of a nucleus consisting of a few base pairs (the rate-limiting step), followed by the rapid pairing of remaining bases.[2, 10, 11] Previous results are reviewed in the following paragraph.

Rate constants of 10^7 $M^{-1}s^{-1}$ were obtained for recombination of DNA 10-mers using temperature-jump (T-jump) measurements,[12, 13] while Williams *et al.*[14] found the rates of duplex formation for the self-complementary strand 5´-GCA TGC-3´ (where A, T, C and G represent adenine, thymine, cytosine and guanine, respectively) depended on the salt concentration and were on the order of 10^6 $M^{-1}s^{-1}$. For the association of complementary oligonucleotides and poly(Adenine+Uracil), second-order rate constants for duplex formation were also measured with values ranging in magnitude from 10^5 to 10^7 $M^{-1}s^{-1}$.[10, 15-18] Similar rate constants were seen for the formation of a duplex between DNA oligonucleotides and complementary locked nucleic acid oligomers.[19, 20] At the single-molecular level, rate constants for duplex formation within a protein nanopore were consistent with values derived from macroscopic solution studies.[21] Surface plasmon resonance studies gave the rate constant for the association of a 15-mer strand with its complement to be 7.4×10^5 $M^{-1}s^{-1}$.[22] Rate constants for 22-mer duplex formation were found to be on the order of $10^2 - 10^4$ M^{-1} s^{-1} using denaturation-renaturation thermal hysteresis.[23] Some work has been done on the kinetics of molecular beacons (hairpins labeled with a fluorescent dye at one end and a fluorescence quencher at the other) binding to DNA targets with rates on the order of 10^1 to 10^4 M^{-1} s^{-1} measured.[24, 25] Using Förster Resonance Energy Transfer (FRET) a rate constant of 5.7×10^5 M^{-1} s^{-1} was reported for the hybridisation of the 16-mer 5´-GTA AAA CGA CGG CCA G-3´ with its complement,[8] while Rode *et al.*[26] obtained rate constants on the order of 10^5 $M^{-1}s^{-1}$ for the hybridisation of DNA 15-mers with complementary 15-mers or 9-mers. Morrison and Stols[27] used FRET to study complementary 10-mers (5´-TTG GTG ATC C-3´ and 5´-GGA TCA CCA A-3´) and 20-mers (5´-AGA TTA GCA GGT TTC CCA CC-3´ and 5´-GGT GGG AAA CCT GCT AAT CT-3´) and measured activation energies of 41.6 ± 3.3 kJ mol^{-1} and 68.6 ± 7.5 kJ mol^{-1}, respectively. However, as Chen *et al.*[28] found the activation energy for hybridisation between a DNA 20-mer and its complementary strand using FRET to be 18 ± 4.2 kJ mol^{-1} and 7.1 ± 6.3 kJ mol^{-1} depending on the FRET pair used for monitoring, care must be taken when comparing results from different studies. The rates of duplex formation for two 8-mer pairs (5´-CAC GGC TC-3´:3´-GTG CCG AG-5´ and 5´-CAC AGC AC-3´:3´-GTG TCG TG-5´) were determined using stopped-flow (SF) spectroscopy to be on the order of 10^6 $M^{-1}s^{-1}$.[3] While Gao *et al.*[5] measured hybridisation rate constants on the order of 10^5 $M^{-1}s^{-1}$ for DNA 25-mers. Finally, the kinetics of 12-mer DNA oligomers was investigated in work by Carrillo-Nava *et al.*[4] Their quantitative work involved studying two pairs of 12-mer DNA oligonucleotides (Pair **I**: 5´-TAG GTC AAT ACT-3´ and its complementary strand 3´-ATC CAG TTA TGA-5´, and pair **II**: 5´-ATC CTC AAT ACT-3´ and its complementary strand 3´-TAG GAG TTA TGA-5´). The pairs of sequences differed from one another only in the first four bases where in the first strand of pair **I** these were TAGG and in the first strand of pair **II** they were the complementary ATCC. Hence, in the two duplexes the number of AT and GC base pairs were the same. Remarkably different reaction rate constants were, however, measured (0.562×10^6 $M^{-1}s^{-1}$ and 1.591×10^6 $M^{-1}s^{-1}$). The observation of dramatically different activation parameters for duplex formation between such similar 12-mers led them unable to conclude anything regarding the formation dynamics.

In Table 1 and Fig. 1 we have summarised the previous results. For fully complementary strands, rate constants differ by up to five orders of magnitude!

In this work we probe the differences in the formation times of A-tracts and AT-tracts. Sequences with many adenine and thymine base pairs are particularly interesting to study as DNA replication commences in these regions due to the fact that A-T base pairs require less energy than G-C base pairs to melt (two *versus* three hydrogen bonds per base pair). The exact sequence of these replication origins differs, but they tend to be AT-rich regions containing repeated sequences and stretches of poly-A (and corresponding poly-T on the complementary strand).[29-35] The presence of poly-A introduces a bent conformation which allows the DNA to more easily interact with replication-initiating proteins.[31, 34] Differences in the formation kinetics of A-tracts (A_n:T_n) and AT-tracts ($(AT)_n$:$(TA)_n$) might occur due to the larger conformational rigidity associated with A_n:T_n regions (A-tracts) in comparison with AT-tracts due to additional cross-strand (CS) interactions.[36, 37] A CS interaction occurs between an adenine and a thymine in the adjacent base pair, on the opposite strand, specifically between an adenine N-6 amine and the thymine O-4 or between the electropositive C-2-H group of adenine and thymine O-2, and causes stiffening of the DNA helix.[36-41]

In order to examine the pairing between sequences of A_n and T_n and between sequences of $(AT)_n$ and $(TA)_n$ we have studied different pairs of complementary strands mainly comprised of adenine and thymine bases (see Table 2). As strands composed purely of AT-tracts are self-complementary, a series of guanine and cytosine bases were introduced at either end of the strands. These bases also prevent the formation of a "slipped structure" which can form when two complementary strands form a duplex that is not the optimal structure (Fig. 2). This was done for all strands in order to keep the number of A-T and G-C base pairs constant for comparison purposes. Furthermore, we study the effect of strand length on the duplex-formation times by looking at both short and long strands (comprised of sixteen and twenty bases, respectively), and use secondary-structure predictions (Fig. 2) calculated using the program UNAFold[42] to explain observed differences. This program calculates free energies based on enthalpies and entropies associated with forming a GC base pair or an AT base pair, partition-function calculations, and a stochastic sampling of structures. Our experimental approach is rapid mixing of two solutions, each containing one of the relevant single strands, and then following the absorption change over time at a particular wavelength (260 nm), *i.e.*, this is an SF kinetics experiment. Furthermore, we repeated the measurements made by Carrillo-Nava *et al.*[4] on two pairs of complementary strands to ensure our experimental setup and method are correct. These results will be discussed first and fresh insights into the large differences in hybridisation times seen between the two pairs based on secondary-structure predictions are provided.

Finally, we present synchrotron radiation circular dichroism (CD) measurements of the formation of AT-tracts in the vacuum-ultraviolet (VUV) region, again from SF experiments. CD is an absorption spectroscopy technique which is based on measuring the difference in absorption of left-hand and right-hand circularly polarised light by chiral molecules such as DNA and can be used to study structural changes. Previous work has shown CD to be a highly useful technique to identify duplex formation and especially so in the VUV region.[43-46]

Hence, it is of particular importance to carry out experiments in the VUV region where there are very strong electronic transitions, and here we take advantage of the high photon fluxes from the ASTRID synchrotron radiation facility in Aarhus. These are, to our knowledge, the first SF VUV-CD measurements on DNA-duplex formation.

Experimental

DNA strands were purchased from DNA Technology A/S Aarhus (purified by high-performance liquid chromatography (strands **A_S-a**, **A_S-b**, **A_L-a**, **A_L-b**, **AT_S-a**, **AT_S-b**, **AT_L-a**, **AT_L-b**), ethanol precipitation (strands **Ia**, **Ib**, **IIa** and **IIb**), or reverse-phase fast cartridge purification (for CD experiments)) and prepared in a 100-mM NaCl (absorbance measurements) or 100-mM NaF (for CD measurements where, in order to measure at low wavelengths, chloride ions in the buffer should be avoided) and 10-mM phosphate buffer (pH 7). See Table 2 for a complete list of nucleobase sequences. Complementary strands were prepared with equal concentrations, the values of which were chosen so that their absorbances at 260 nm were around one. These values were checked by measuring the absorbances and using the extinction coefficients of the appropriate unfolded single strands (see Table 3).[47, 48] Absorption measurements were performed using an RX2000 rapid mixing stopped-flow instrument from Applied Photophysics, UK (dead-time: 8 ms; path length: 1 cm) in combination with a Thermo Evolution 300 spectrometer. A lens with a 100 mm focal length was used to focus the light from the spectrometer through the SF cell. As the difference in absorbance between single strands and duplexes is large at 260 nm (Supporting Information), the change in absorbance at this wavelength, after mixing equal volumes of complementary strands, was monitored over time. Experiments were conducted at a range of temperatures between 10°C and 35°C and typically repeated at least three times at each temperature. Measurements on the short strands were repeated after 10 months and results were consistent with the initial ones. Reaction rates were determined by fitting the averaged datasets with a second-order rate equation or the initial data with a straight line (see Results and Discussion for further details) and activation energies determined from Arrhenius plots.

CD spectra were measured at the CD1 beam line at the ASTRID storage ring facility at Aarhus University, Denmark.[49, 50] Steady-state measurements were made in a quartz suprasil cylindrical cell (Hellma type 121.000) with a path length of 0.05 cm, while rate measurements were made using a home-built SF apparatus equipped with a quartz suprasil flow cell (Hellma type 170.000) also with a path length of 0.05 cm.

Temperature-dependent absorbance spectra of single strands were measured in a 1 cm path length quartz cuvette, using a peltier thermostatted cell holder accessory in the Evolution 300 bench-top spectrometer. Temperature-dependent CD spectra for the single strands were measured on the AU-UV beam line on ASTRID2, Aarhus University using the same apparatus for CD measurement as for the CD1 beam line described earlier and in references 49 and 50. Using this CD setup it is possible to simultaneously measure an absorbance spectrum, which can then be compared with the results from the bench-top spectrometer.[51, 52]

The software package UNAFold,[42] which combines free-energy minimisation, partition-function calculations and stochastic sampling to predict folding and hybridisation of DNA strands, was used to calculate the fractions of DNA forms present in our samples and the structures of these forms. Input information in the programme are DNA and Na^+ concentrations and for the structure calculations also the temperature.

Results and discussion

The kinetics of duplex formation from 12-mers

First, we investigated the hybridisation of the same complementary 12-mers as those in the work by Carrillo-Nava *et al.*[4] over a range of temperatures in order to test our experimental method. The results can be found in the Supporting Information, and we only summarise our main findings here.

As in the previous work we find that the duplex-formation times are clearly different with pair **II** being faster than pair **I**. Rate constants at each temperature were determined using two different methods, 1: fitting the data with a second-order model and 2: fitting the initial data with a linear function (see below for further details on the fitting processes). From Arrhenius plots, the activation energies for the formation of pair **I** are found to be 44.5 ± 7.2 kJ mol^{-1} and 36.2 ± 8.9 kJ mol^{-1} for methods 1 and 2, respectively, while that reported previously was 42.3 ± 4.8 kJ mol^{-1}, which is in between our two values.[4] In the case of pair **II** we find activation energies of 17.6 ± 5.5 kJ mol^{-1} and 15.7 ± 3.7 kJ mol^{-1} for methods 1 and 2, respectively, in good agreement with that reported previously (20.5 ± 1.5 kJ mol^{-1}).[4] It should be noted that for pair **II** we have left out the results at the highest temperature (35 °C) as the melting temperature of the duplex is 38.6°C (calculated using the method described in Ref. [53] and references therein). Hence not all of the strands at 35°C form duplexes, which implies that the models do not describe the data well at this temperature (melting is not accounted for). The result is non-linearity in the Arrhenius plot at high temperatures. Such a non-linear trend in an Arrhenius plot is not unprecedented for strands containing secondary structure. In a work on the duplex formation of a series of 21-mer and 33-mer strands,[54] the authors propose that at high temperatures the rate-limiting step is controlled by the entropy of nucleation, resulting in a negative activation energy, while the rate-limiting step at lower temperatures is dominated by the enthalpy of melting of secondary structure giving a positive activation energy. In view of this, in the next section we will discuss calculated secondary structures of these strands.

In conclusion, we consider the agreement to the previous work to be rather good and, therefore, our method to be validated.

Modelling the folding and hybridisation of the DNA strands involved in pair I and pair II

Simulated percentages of the different DNA forms (see caption of Fig. 2 for definitions) were calculated for pair **I** using UNAFold (Supporting Information). We find that the initial samples are comprised of unfolded single strands and folded single strands (hairpins), while the number of homo-duplexes are insignificant. Hairpins are more dominant at the lower experimental temperatures. After mixing, the predominant DNA form is hetero-duplexes at all temperatures. However, as the melting temperature is approached, the number of hairpins and unfolded strands increases, so that at 35°C the percentage of hetero-duplexes formed is only 81%. The program was also used to calculate the minimum-energy structures at room temperature. For the hetero-duplexes the majority of bases are expected to couple, while for both types of homo-duplexes and

both types of hairpin only two base pairs form (Supporting Information).

For pair **II** the predominant form of DNA prior to mixing is unfolded single strands at all measured temperatures. Hairpins only account for 16% of the **IIa** strands at 10°C and 4% at 35°C, while for the **IIb** strands the values are 44% and 8%, respectively. The percentages of homo-duplexes are negligible for both strands. The calculated minimum-energy structures at room temperature reveal that the majority of bases are expected to couple in the hetero-duplexes, while for both types of homo-duplexes and both types of hairpin only two base pairs form.

The presence of hairpins in the solutions prior to mixing will probably affect the duplex-formation times as the formation of a duplex from hairpins is different from the formation of a duplex from unfolded single strands due to the additional barrier which must be overcome (breakage of the base-pairs forming the hairpin). Indeed, previous experiments on a series of 25-mer strands[5] and a series of 21-mer and 33-mer strands[54] showed that strands containing large amounts of secondary structure hybridised by a more complex mechanism than traditional two-state duplex formation, with duplex formation likely involving fast nucleation followed by slow partial strand displacement. There are two possible schemes for the formation of duplexes from complementary pairs of hairpins, one of which involves the total denaturation of the hairpins to unfolded single strands prior to duplex formation, while the other involves the formation of an intermediate "kissing" complex followed by the pairing of all bases.[5, 54, 55] Hairpins comprise a larger proportion of the DNA single strands involved in pair **I** than pair **II**, thus giving a possible explanation for the lower activation energy seen for pair **II**.

The pairs **I** and **II** were chosen in the previous work as simple models for single strands and duplexes. However, as our analysis has shown, it is necessary also to consider hairpins (and maybe also homo-duplexes). In reality, it is very difficult to find two complementary strands that by themselves do not undergo self-folding (hairpin structures) or homo-duplex formation.

Formation of A-tracts

First we studied the 16-mer A-tract pair $\mathbf{A_S}$ (see Table 2). Fig. 3 shows the simulated percentages of DNA in the different forms for the pair and its component single strands. As shown in the figure, prior to mixing the $\mathbf{A_S}$**-a** and $\mathbf{A_S}$**-b** strands are found as unfolded single strands and as hairpins. Hairpins are dominant at lower temperatures (83% and 66% at 10°C, respectively), while unfolded single strands are dominant at higher temperatures (77% and 85% at 35°C, respectively). Homo-duplexes are negligible at all temperatures. After mixing, hetero-duplex formation is 100% over the entire experimental temperature range, contrasting with the case for pairs **I** and **II** discussed above.

The calculated minimum-energy structures at room temperature for $\mathbf{A_S}$**-a** and $\mathbf{A_S}$**-b** are shown in Fig. 4. It can be seen that all of the bases in the single strands pair to form hetero-duplexes, while for homo-duplexes and hairpins only two base pairs are expected to form.

As all of the strands are expected to eventually form hetero-duplexes, the reverse process where a duplex melts into its two component single strands may be disregarded: The rate constant for melting is too low to be of importance. Thus, if the process of two single strands (*a* and *b*) associating to form a double strand (*ds*), $a + b \rightarrow ds$, follows a second-order rate law similar to that observed for previously measured strands[4] and the concentration of each of the single strands are the same ($[a] = [b]$), then the concentration of *a*, $[a]$, can be described as follows after simple integration of $\mathrm{d}[ds] / \mathrm{d}t = -\mathrm{d}[a] / \mathrm{d}t = k\,[a]\,[b] = [a]^2$ from time zero (initial concentration $[a]_0$) to time *t*:[56]

$$[a] = \frac{[a]_0}{1 + k\,t\,[a]_0} \qquad (Eq.\,1)$$

where *k* is the rate constant. This is the simplest description not involving any tertiary encounters for the duplex formation. It then follows that the equation to fit our absorption spectra is:

$$Absorbance = (\varepsilon_a + \varepsilon_b - \varepsilon_{ds})\,l\,\frac{[a]_0}{1+k\,t\,[a]_0} + \varepsilon_{ds}\,l\,[a]_0 \qquad (Eq.\,2)$$

where ε_a, ε_b, ε_{ds} are the extinction coefficients of the *a* and *b* single strands and the duplex, respectively, and *l* is the path length. The Arrhenius equation relates the rate constants to the activation energy of a process, E_a:

$$\ln k = -\frac{E_a}{R\,T} + \ln \mathcal{A} \qquad (Eq.\,3)$$

where *R* is the gas constant, *T* is the absolute temperature and $\mathcal{A}$ is the pre-exponential factor. Hence, the activation energies are found by fitting the data in Arrhenius plots with a straight line.

The absorbance changes at 260 nm associated with duplex formation between $\mathbf{A_S}$**-a** and $\mathbf{A_S}$**-b** *versus* time are shown in Fig. 5. It was found that second-order fits to the full data sets did not accurately describe the data at early times (see Fig. 6). As the important information is contained in the initial part of the time-dependent absorbance spectra, it was decided to limit the data ranges to be fit. Accordingly, changes of 70% of the maxima were fit instead, and it was found that such fits better described the initial data. As the presence of hairpins will affect our concentration measurements (folded strands have a different extinction coefficient than unfolded strands), concentrations were determined from the absorbance measurements at late times when all of the strands had formed duplexes. Note, the presence of any uncoupled strands resulting from unequal initial concentrations of strand *a* and *b* will affect the accuracy of this calculation.

As an alternative, the rate constants were also calculated from the initial slopes of the absorbance data: At very short times the concentrations do not change significantly, and a Taylor expansion of *Eq.* 1 about $t = 0$ gives,

$$[a] = [a]_0 - k\,[a]_0{}^2\,t \qquad (Eq.\,4)$$

The absorbance is then,

$$Absorbance = -(\varepsilon_a + \varepsilon_b - \varepsilon_{ds})\,l\,k\,[a]_0{}^2\,t + (\varepsilon_a + \varepsilon_b)\,l\,[a]_0 \qquad (Eq.\,5)$$

The advantage of this model is that any back reaction certainly does not matter as the concentration of the duplex is close to zero at initial times. However, it relies on fewer data points (only the linear part close to time zero).

The resulting data from both methods were plotted in an Arrhenius plot (Fig. 7), and values for the activation energies were found to be 9.6 ± 4.9 kJ mol^{-1} and 13.2 ± 6.4 kJ mol^{-1} for the data obtained by fitting the data with Eq. 2 and Eq. 5, respectively.

In a previous work on a similar 7-mer pair (5´-CAA AAG-3´and 5´-CTT TTG-3´), Nelson *et al.*[57] found the activation energy to be -2.09 ± 8.37 kJ mol^{-1}. Unlike the $\mathbf{A_S}$**-a** and $\mathbf{A_S}$**-b** pair presented here, these strands cannot form hairpins or homo-duplexes, therefore such a negative activation energy is in agreement with the previously mentioned theory

that a lack of secondary structure in the constituent single strands results in negative activation energies.[54]

Effect of length on duplex-formation kinetics

To study how the inclusion of four extra bases on each single strand affected the duplex-formation kinetics, the longer (20-mer) A-tract pair $\mathbf{A_L}$ (see Table 2) was examined. Previous experiments found that the second-order rate constants for DNA were proportional to the square root of the length and explained that this non-linear relationship arises as it is more difficult for longer strands to interpenetrate and find complementary sites.[11] Here, we find that the rate constants are indeed larger for the longer strands (see Fig. 7).

Modelling results are shown in Fig. 3 and indicate that the $\mathbf{A_L}$**-a** and $\mathbf{A_L}$**-b** strands form hairpins and unfolded single strands similar to the shorter A-tract strands. Again hairpins are dominant at lower temperatures (76% and 56% at 10°C, respectively), while unfolded single strands are dominant at higher temperatures (84% and 90% at 35°C, respectively). After mixing, hetero-duplex formation is 100% over the entire experimental temperature range. If we consider the calculated minimum-energy structures for $\mathbf{A_L}$**-a** and $\mathbf{A_L}$**-b** (Fig. 4), we can see that, similar to the case of $\mathbf{A_S}$, all of the bases in the single strands couple to form hetero-duplexes, while for the homo-duplexes and hairpins only two base pairs are expected to form. However as the $\mathbf{A_L}$ strands are longer than the $\mathbf{A_S}$ ones, the percentage of unpaired bases in each is different: 80% of the bases are not paired in the longer-strand hairpins while 75% of the bases are free in the shorter ones. Considering the percentage of hairpins in the solution and the fraction of unpaired bases in the minimum-energy structures at each temperature for these strands, this means that overall 87% and 97% of bases in the long strands are not paired at 10°C and 35°C, respectively, while for the shorter strands these values are lower (81% and 95%).

The number of free bases in a strand should affect the formation rates if the duplex formation involves fast nucleation followed by slow partial strand displacement, as found for a series of 25-mer strands:[5] More free bases increases the chance for formation of the initial nucleus.[54] Indeed, we found the activation energy for the formation of $\mathbf{A_L}$ to be 3.9 ± 4.4 kJ mol^{-1} and 7.5 ± 4.5 kJ mol^{-1} using methods 1 and 2, respectively, which are lower than the values found for the shorter duplex $\mathbf{A_S}$ (9.6 ± 4.9 kJ mol^{-1} and 13.2 ± 6.4 kJ mol^{-1} using methods 1 and 2, respectively). In apparent contrast with our results, it has previously been shown that the activation energy of duplex formation between complementary pairs of hairpins comprised of repetitive CXG sequences (where X is one of the nucleobases) increased with length.[55] However, unlike our strands the fractions of unpaired bases in the hairpins they studied decreased with length.

Formation of AT-tracts

In order to assess how duplexes comprised of AT-tracts form relative to those comprised of A-tracts, we studied the DNA pairs $\mathbf{AT_S}$ (16-mer) and $\mathbf{AT_L}$ (20-mer). Even though a sequence of guanines and cytosines were appended to either end of the strands in order to prevent them from being self-complementary, our modelling results show that a significant fraction of strands still form homo-duplexes and hairpins (Fig. 8). Furthermore, an increased proportion of bases are coupled in the homo-duplexes for these strands relative to the A-tract strands (Fig. 4). Overall the percentages of uncoupled bases in the long single strands (accounting for the fraction of uncoupled bases in the different DNA forms and the percentage of each form in solution) are 19% and 55% at 10°C and 35°C, respectively, while for the shorter strands these values are 24% and 83%. The increased number of coupled bases in these strands is likely to increase the formation time of the hetero-duplexes as fewer bases are available to form an initial nucleus, and the barrier for formation is higher.

Time courses for the formation of these duplexes are shown in Fig. 9. As expected, the reaction rates increase with increasing temperature for both pairs. However, the spectral shapes for the lower two temperatures of both pairs are disparate, showing an initial increase in absorbance before it duly falls. This increase we take to be due to the rate of dissolution of base-pairs in the hairpins and homo-duplexes being faster than the rate of formation of new base pairs between the strands that form hetero-duplexes.

A comparison between Figs. 5 and 9 shows that the formation times for the AT-tract strands are significantly longer than those for the A-tract strands. This could be due to the following reasons:

1. Reduced number of coupled bases in the A-tract precursors relative to the AT-tract ones: Duplex-formation times are likely reduced if the chance to form an initial nucleus is increased. Furthermore, the barrier for formation is lower if fewer base pairs need to be broken in order to form the hetero-duplexes.
2. Increased number of nucleation sites on the A-tracts relative to the AT-tracts: The long chains of poly(A) in the A-tracts provide numerous nucleation sites to initially bind the complementary strands together. Subsequently, the minimum-energy configuration corresponding to a correctly zippered hetero-duplex could be found. Less potential nucleation sites are available in the AT-tracts due to the alternating sequence of nucleobases.
3. Additional CS interactions found in A-tract regions in comparison with AT-tracts: Perhaps duplex formation is expedited due to the more rigid A-tracts which may provide a more favorable conformation for the completion of duplex formation.

In order to quantify the observed differences, we also fit our AT-tract data with Eq. 2 (Fig. 10). According to our modelling, 97% and 98% of the short and long AT-tract strands form hetero-duplexes at 35°C after mixing. As the percentage of strands which do not form hetero-duplexes is minimal, our model which ignores DNA melting is used to analyse the data. However, due to the unusual profiles of the time-dependent absorption spectra at the lowest temperatures (*vide supra*), the model could not be used to fit these data. Activation energies for the formation of $\mathbf{AT_S}$ were found to be 221 ± 39 kJ mol^{-1} and 195 ± 40 kJ mol^{-1} using methods 1 and 2, respectively, while the values for the formation of $\mathbf{AT_L}$ were 117 ± 28 kJ mol^{-1} and 39 ± 18 kJ mol^{-1} (Fig. 11). As the fraction of unpaired bases in the AT-tract strands are less than in the A-tract strands, it is not surprising that the activation energies are higher.

The effect of sequence on the hybridisation kinetics of self-complementary 12-mer DNA strands (5′-A_6T_6-3′, 5′-$(A_3T_3)_2$-3′, 5′-$(A_2T_2)_3$-3′, 5′-$(AT)_6$-3′, 5′-T_6A_6-3′) containing adenines and thymines was previously studied by Zuo *et al.*[9] In this work the kinetics of the hairpin-to-duplex transition was measured using salt-jump kinetics measurements on five different strands in a 0.01 M PIPES (piperazine-N,N′-bis(2-ethanesulfonic acid)), 1 mM EDTA (ethylenediaminetetraacetic acid), pH 7.0

buffer with 0.4 M NaCl. The rate constants were found to be on the order of 10^5 $M^{-1}s^{-1}$, and are smaller than those for duplex formation from non-self-complementary strands. This is in agreement with the results of Gao *et al.*,[5] who found that the rate constants decrease with increasing secondary structure, and our results where we found lower rate constants for the formation of AT-tracts relative to A-tract strands (see Table 1). In their work, Zuo *et al.*[9] report the activation energy for formation of the duplex between two self-complementary 5´-A_6T_6-3´ strands to be 130 kJ mol^{-1}. It is difficult to compare this result directly with ours due to the experiment being performed under different conditions. However, it is noted that this activation energy is higher than what we measured for our A-tract strands and lies between the values obtained for our short and long AT-tracts.

In order to try and understand if the differences seen between the A-tract and AT-tract formation are entirely due to the presence of folded structures in the precursors, we consider two extreme situations. In both situations we assume that the unfolded strands are reactive but the folded strands (hairpins and homo-duplexes) are not. In situation 1 the transition from folded strands to unfolded strands is instantaneous, and the proportion of folded strands is greater than unfolded strands; hence at early times the concentrations of unfolded strands are constant. Therefore, we can fit our data at early times (linear fit as $d[ds] / dt = k [a]_0^2$ and therefore $[ds] = k [a]_0^2 t$ after integration) and calculate the rate constants using the concentrations of unfolded DNA determined from UNAFold simulations. Using this method, we determine the activation energy of $\mathbf{AT_S}$ to be 68 ± 14 kJ mol^{-1}. In situation 2 we assume that the transition from folded strands to unfolded strands is so slow that it can be ignored. Thus, fitting the data at early times with a second-order model and using the concentration of unfolded DNA determined from our simulations, an activation energy of 81 ± 20 kJ mol^{-1} was obtained for $\mathbf{AT_S}$. The values obtained using both methods are lower than those determined previously, but are nonetheless higher than all values obtained for the A-tracts. This suggests that even without the presence of base pairs in the precursors, the formation rate of AT-tracts is different to that of A-tracts, and that it is slower.

Circular dichroism studies of AT-tracts

Here we present very early results on CD studies of DNA strands in the VUV. Steady-state CD spectra are shown in Fig. 12. It can be seen that there is a significant difference between the single strands and the duplexes in the VUV region (< 200 nm). Therefore, the SF kinetics experiments were carried out at a wavelength of 192 nm where the difference is greatest. Higher concentrations of DNA were used for the CD measurements due to the smaller path length of the cell (0.05 cm). At these higher concentrations, the percentages of unfolded single strands, folded single strands and homo-duplexes at 25°C according to UNAFold calculations are 12%, 53% and 35% for the $\mathbf{AT_S}$**-a** strands and 19%, 48% and 34% for the $\mathbf{AT_S}$**-b** strands, respectively. After mixing, 100% of the strands formed hetero-duplexes. Time-dependent absorption spectra for the formation of $\mathbf{AT_S}$ are shown in Fig. 12. It is evident that the CD signal changes at a faster rate than the absorbance signal (the rate of change of absorbance is only 63% that of CD). As CD measures structural changes whereas absorbance measures base pairing, this difference could indicate that the strands line up prior to the formation of base pairs. This idea is to be investigated further in the future.

Melting of hairpins and/or homo-duplexes

To support our conclusion, and also the predictions of the UNAfold program, that some strands are engaged in base-pair interactions prior to mixing with their complementary strands, we recorded absorbance and CD spectra at different temperatures for separate solutions of $\mathbf{AT_s}$**-a** and $\mathbf{AT_s}$**-b** single strands. Indeed, we find that the absorbance at 260 nm increases by almost 20% from 10°C to 60°C as expected when base pairs are broken, and the largest change occurs at 30-35 °C (Supporting Information). Likewise, CD spectra depend on temperature: CD signals at 191 nm (–), 249 nm (+), and 272 nm / 275 nm (–) either increased (+) or decreased (–) between 25°C and 50°C (Supporting Information). The temperature range we observe largest spectral changes is in full accordance with the predictions of hairpin melting and the formation of unfolded single strands (*cf.*, Figure 8A and 8B).

Conclusions

The differences between the formation times of A-tracts and AT-tracts were studied using SF spectroscopy on complementary model strands. Secondary-structure modelling was used to determine the fractions of unfolded single strands, hairpins, homo-duplexes and hetero-duplexes in our solutions and the lowest-energy structures of each form. This allowed us to determine the percentage of paired bases in the structures which assisted our interpretation of the observed trends. Rate constants were determined by fitting time-dependent absorption spectra with either a second-order model (method 1) or fitting the initial data with a linear model (method 2). Activation energies for the formation of $\mathbf{A_S}$ were determined to be 9.6 ± 4.9 kJ mol^{-1} and 13.2 ± 6.4 kJ mol^{-1}, respectively. Lower activation energies were seen for the formation of the longer $\mathbf{A_L}$ strands (3.9 ± 4.4 kJ mol^{-1} and 7.5 ± 4.5 kJ mol^{-1}, respectively) in agreement with the theory that the number of free bases in a strand affects the formation rates. Correspondingly, higher activation energies were seen for the formation of AT-tracts, whose components have fewer unpaired bases. We emphasize that we have not accounted for hairpin structures and homoduplexes in our kinetics description which may have affected the obtained rates of formation and activation energies. While the effect of secondary structure is a common problem for the association kinetics of DNA strands, it is particularly so for AT strands. Finally, initial VUV-CD SF measurements on the formation of DNA duplexes indicate an alignment of the component strands prior to the formation of duplexes.

Acknowledgements

We gratefully acknowledge support from Lundbeckfonden and Carlsbergfondet (grant number 2013_01_0030).

Table 1: Rate constants for DNA-duplex formation from fully complementary strands.[a]

Number of bases	*Sequence of strand 1*	*Rate constant ($M^{-1}s^{-1}$)*	*Reference*
6	5´-GCA TGC-3´	0.33×10^6	14
7	5´-CAA AAA G-3´	8.3×10^6	57
8	5´-CAC AGC AC-3´	7.3×10^6	3
8	5´-CAC GGC TC-3´	7×10^6	3
8	5´-CAG GAG CA-3´	2.1×10^7	19
8	5´-GGT GAA TG-3´	1.3×10^7	21
8	5´-TAC GTG GA-3´	2.2×10^7	21
10	5´-CAG GTC ACA G-3´	5.4×10^7	12
10	5´-TAG GTT ATA A-3´	1.2×10^7	12
10	5´-TAG GTT ATA A-3´	2.1×10^7	13
10	5´-TGA CTG ATG C-3´	1.5×10^6	20
10	5´-TTG GTG ATC C-3´	8.1×10^2	27
12	5´-AAA AAA TTT TTT-3´	3.0×10^5	9
12	5´- AAA TTT AAA TTT-3´	2.0×10^5	9
12	5´-AAG AAA GAA AAG-3´	4.4×10^6	18
12	5´- AAT TAA TTA ATT-3´	3.1×10^5	9
12	5´-ATA TAT ATA TAT -3´	5.2×10^5	9
12	5´-ATC CTC AAT ACT-3´	1.6×10^6	4
12	5´-ATC CTC AAT ACT-3´	$(1.2, 0.47) \times 10^6$ [c]	This work
12	5´-$(CAG)_4$-3´	2.7×10^6	55
12	5´-$(CCG)_4$-3´	2.0×10^6	55
12	5´-TAG GTC AAT ACT-3´	0.56×10^6	4
12	5´-TAG GTC AAT ACT-3´	$(4.2, 2.4) \times 10^5$ [d]	This work
12	5´-TTT TTT AAA AAA-3´	3.4×10^5	9
15	5´-ATC TAG ACA CTG GTA-Fam-3´[b]	4.7×10^5	26
15	5´-CAG GTC ACA GAT CAT-Fam-3´[b]	1.9×10^5	26
15	5´-TTT TTT TGT TTT TTT-3´	7.4×10^5	22
16	5´-GCG $(AA)_5$ GCG-3´	$(14, 6.6) \times 10^5$ [e]	This work
16	5´-GCG $(AT)_5$ GCG-3´	$(8.3, 7.6) \times 10^3$ [f]	This work
20	5´-AGA TTA GCA GGT TTC CCA CC-3´	6.6×10^2	27
20	5´-GCG $(AA)_7$ GCG-3´	$(19, 8.4) \times 10^5$ [g]	This work
20	5´-GCG $(AT)_7$ GCG-3´	$(2.5, 9.3) \times 10^3$ [h]	This work
22	5´-AAA GGA AAA AAA AAA GAA AAA A-3´	2.3×10^2	23
22	5´-AAA GGA GGA AAA AAA GAA AAA A-3´	9.9×10^2	23
22	5´-AAA GGA GGA GAA GAA GAA AAA A-3´	1.7×10^3	23
22	5´-AGA GGA GGA GAA GAA GAG GAG A-3´	4.9×10^4	23
25	5´-AGA TCA GTG CGT CTG TAC TAG CAC A-3´	$(20, 4.8) \times 10^4$ [i]	5
25	5´-AGA TCA GTG CGT CTG TAC TAG CAG T-3´	7.2×10^5	5
25	5´-GTT GTC AAG ATG CTA CCG TTC AGA G-3´	1.2×10^6	5
30	5´-$(CAG)_{10}$-3´	3.4×10^5	55
30	5´-$(CCG)_{10}$-3´	1.0×10^6	55
45	5´-$(CAG)_{15}$-3´	9.5×10^5	55
45	5´-$(CCG)_{15}$-3´	6.6×10^5	55

[a] See individual references for details.
[b] Fam = 6-carboxyfluorescein. In the complementary sequence tetramethylrhodamine is at the 5´ end.
[c] Results from method 1 and method 2, respectively. Temperature was 20.2 ºC.
[d] Results from method 1 and method 2, respectively. Temperature was 20 ºC.
[e] Results from method 1 and method 2, respectively. Temperature was 20.6 ºC.
[f] Results from method 1 and method 2, respectively. Temperature was 20.3 ºC.
[g] Results from method 1 and method 2, respectively. Temperature was 20.2 ºC.
[h] Results from method 1 and method 2, respectively. Temperature was 20.3 ºC.
[i] The fast and slow components, respectively.

Table 2: The DNA strands studied in this work.

Pair	*DNA strand 1*	Sequence	*DNA strand 2*	*Sequence*	*Number of bases*
I	**Ia**	5´-TAG GTC AAT ACT-3´	**Ib**	3´-ATC CAG TTA TGA-5´	12
II	**IIa**	5´-ATC CTC AAT ACT-3´	**IIb**	3´-TAG GAG TTA TGA-5´	12
A_S	**A_S-a**	5´-GCG $(AA)_5$ GCG-3´	**A_S-b**	3´-CGC $(TT)_5$ CGC-5´	16
A_L	**A_L-a**	5´-GCG $(AA)_7$ GCG-3´	**A_L-b**	3´-CGC $(TT)_7$ CGC-5´	20
AT_S	**AT_S-a**	5´-GCG $(AT)_5$ GCG-3´	**AT_S-b**	3´-CGC $(TA)_5$ CGC-5´	16
AT_L	**AT_L-a**	5´-GCG $(AT)_7$ GCG-3´	**AT_L-b**	3´-CGC $(TA)_7$ CGC-5´	20

Table 3: Extinction coefficients for the unfolded single strands and the duplexes formed between them.[47, 48]

DNA Strand 1	ε (M^{-1} cm^{-1})	*DNA Strand 2*	ε (M^{-1} cm^{-1})	*DNA Duplex*	ε (M^{-1} cm^{-1})
Ia	120600	**Ib**	123200	**I**	192358
IIa	114300	**IIb**	128500	**II**	191569
A_S-a	176200	**A_S-b**	129000	**A_S**	243702
A_L-a	224200	**A_L-b**	161400	**A_L**	301308
AT_S-a	164700	**AT_S-b**	157900	**AT_S**	257596
AT_L-a	208900	**AT_L-b**	202100	**AT_L**	321155

Table 4: Activation energies for the formation of duplexes.

Pair	E_a (kJ mol^{-1})	
	Method 1: Fit to a 70% drop in absorbance	*Method 2: Linear fit to the initial data*
I	44.5 ± 7.2	36.2 ± 8.9
II	7.5 ± 6.4	4.3 ± 4.9
A_S	9.6 ± 4.9	13.2 ± 6.4
A_L	3.9 ± 4.4	7.5 ± 4.5
AT_S	221 ± 39	195 ± 40
AT_L	117 ± 28	39 ± 18

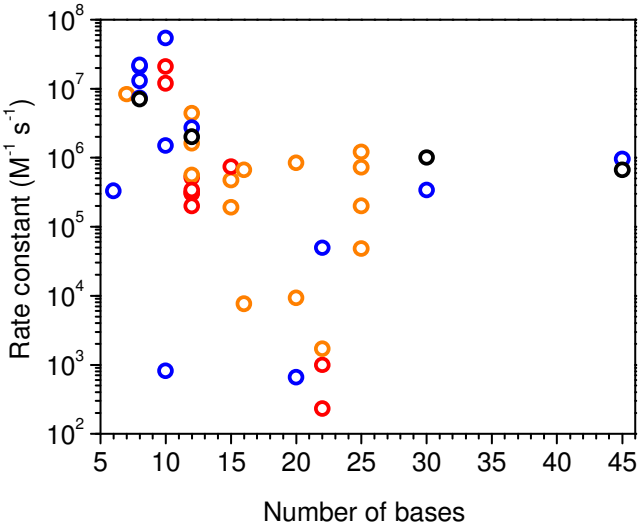


Fig 1: Rate constants for DNA-duplex formation from fully complementary strands as a function of the number of bases in each of the constituent single strands. The fraction of G and C bases in each strand are indicated in color: Red: 0 – 0.24, orange: 0.25 – 0.49, blue: 0.50 – 0.74, black: 0.75 – 1. See Table 1 for more details.

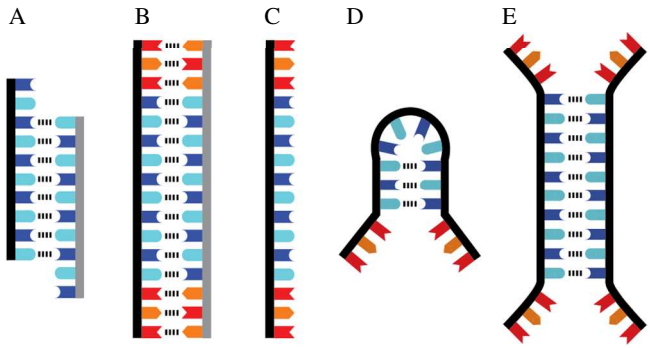


Fig 2: Schematic of possible DNA structures. (A) A slipped hetero-duplex, (B) a hetero-duplex, (C) an unfolded single strand, (D) a hairpin, (E) a homo-duplex. Hydrogen bonds are indicated with dashed lines. A hetero-duplex is defined as where a DNA strand pairs with a complementary strand. An unfolded single strand is defined as a strand in which none of the nucleobases have formed a pair with another. The strand will, however, assume a helical structure.[58-61] A folded single strand or hairpin is defined as a single strand in which some of the nucleobases have formed pairs with other bases on the same

strand. A homo-duplex is defined as a strand which partially hybridises with another strand of the same type.

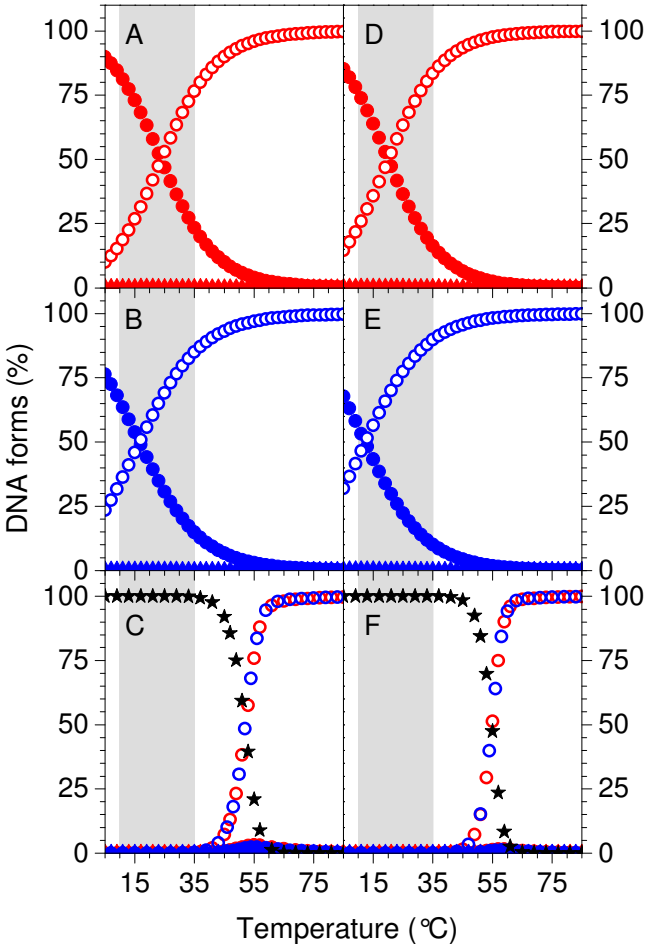


Fig 3: Simulated percentages of DNA in the different forms for strands $\mathbf{A_S}$**-a**, $\mathbf{A_S}$**-b**, $\mathbf{A_L}$**-a** and $\mathbf{A_L}$**-b** in 0.1-M Na^+ solutions as a function of temperature obtained using the UNAFold software package. Concentrations were (A) 7.0 μM of $\mathbf{A_S}$**-a**, (B) 7.0 μM of $\mathbf{A_S}$**-b**, (C) 3.5 μM of $\mathbf{A_S}$**-a** with 3.5 μM of $\mathbf{A_S}$**-b**, (D) 5.0 μM of $\mathbf{A_L}$**-a**, (E) 5.0 μM of $\mathbf{A_L}$**-b**, and (F) 2.5 μM of $\mathbf{A_L}$**-a** with 2.5 μM of $\mathbf{A_L}$**-b**. Data for the unfolded single strands (○), folded single strands (hairpins) (●), homo-duplexes (▲) and hetero-duplexes (★) are plotted for $\mathbf{A_S}$**-a** (red symbols) and $\mathbf{A_S}$**-b** (blue symbols) in the left column and $\mathbf{A_L}$**-a** (red symbols) and $\mathbf{A_L}$**-b** (blue symbols) in the right. The experimental range is highlighted in grey.

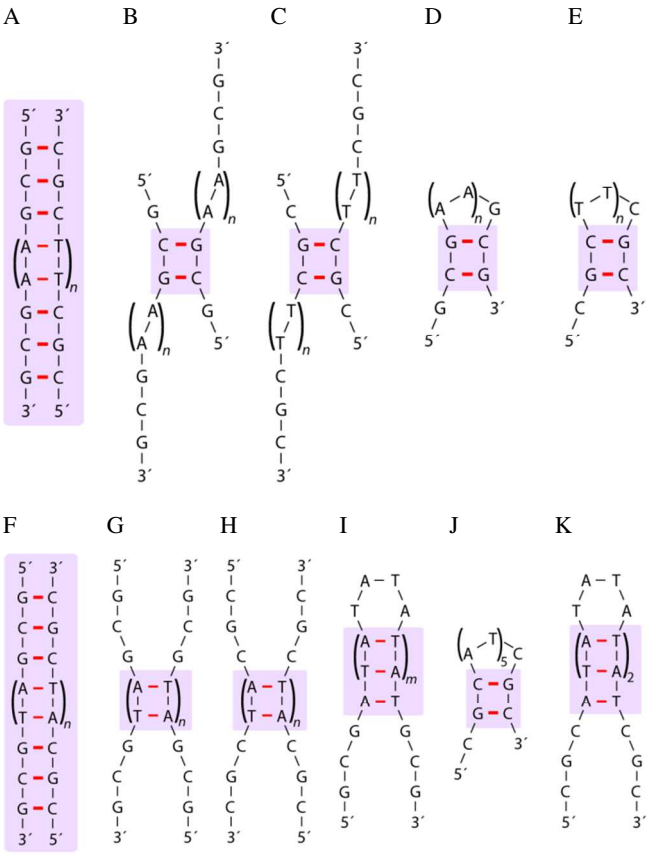


Fig 4: Calculated minimum energy structures at room temperature (25°C) for (A) the hetero-duplex between $\mathbf{A_S}$**-a** and $\mathbf{A_S}$**-b** (n=5) and $\mathbf{A_L}$**-a** and $\mathbf{A_L}$**-b** (n=7); (B) the homo-duplex for $\mathbf{A_S}$**-a** (n=5) and $\mathbf{A_L}$**-a** (n=7); (C) the homo-duplex for $\mathbf{A_S}$**-b** (n=5) and $\mathbf{A_L}$**-b** (n=7); (D) the hairpin of $\mathbf{A_S}$**-a** (n=5) and $\mathbf{A_L}$**-a** (n=7); (E) the hairpin of $\mathbf{A_S}$**-b** (n=5) and $\mathbf{A_L}$**-b** (n=7); (F) the hetero-duplex between $\mathbf{AT_S}$**-a** and $\mathbf{AT_S}$**-b** (n=5) and $\mathbf{AT_L}$**-a** and $\mathbf{AT_L}$**-b** (n=7); (G) the homo-duplex for $\mathbf{AT_S}$**-a** (n=5) and $\mathbf{AT_L}$**-a** (n=7); (H) the homo-duplex for $\mathbf{AT_S}$**-b** (n=5) and $\mathbf{AT_L}$**-b** (n=7); (I) the hairpin of $\mathbf{AT_S}$**-a** (m=1) and $\mathbf{AT_L}$**-a** (m=2); (J) the hairpin of $\mathbf{AT_S}$**-b**; (K) the hairpin of $\mathbf{AT_L}$**-b**. Note: regions of duplex formation are highlighted and the subscripts associated with the brackets denote the number of times the enclosed sequence is repeated.

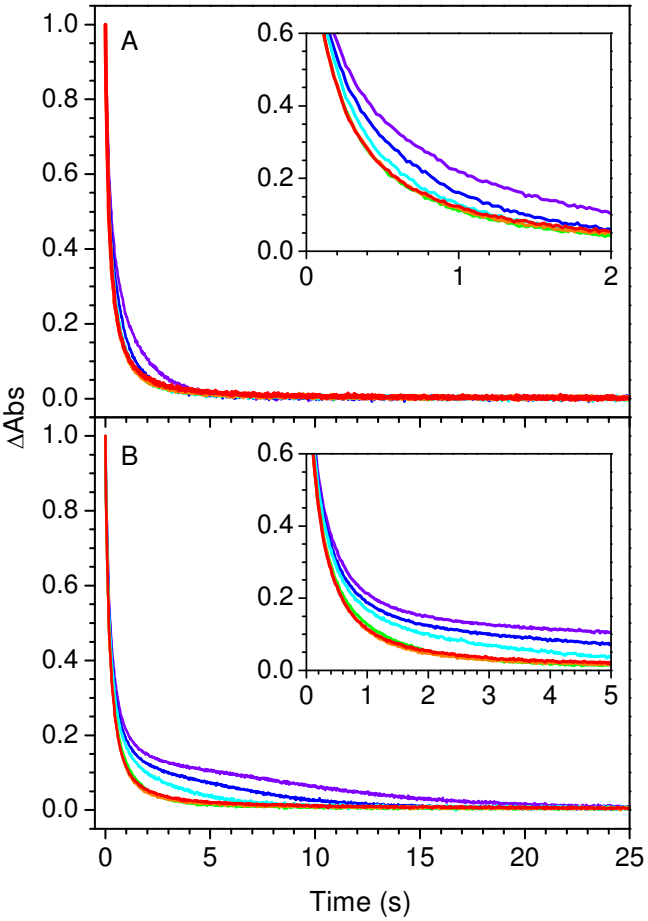


Fig 5: Time course of the absorbance changes (scaled to 1) at 260 nm associated with duplex formation between (A) 3.2 μM of strands $\mathbf{A_S}$**-a** and $\mathbf{A_S}$**-b** at 11°C (violet), 15°C (blue), 20°C (cyan), 25°C (green), 30°C (orange) and 35°C (red) and (B) 2.6 μM of strands $\mathbf{A_L}$**-a** and $\mathbf{A_L}$**-b** at 11°C (violet), 16°C (blue), 20°C (cyan), 25°C (green), 30°C (orange) and 34°C (red).

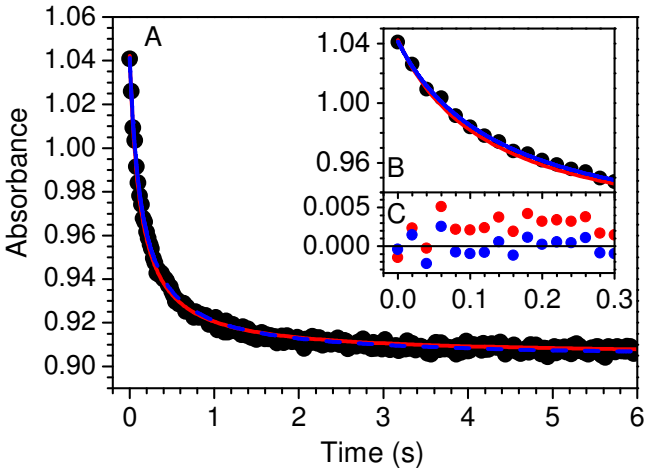


Fig 6: (A) Time course of the absorbance at 260 nm associated with duplex formation between 3.7 μM of strands $\mathbf{A_S}$**-a** and $\mathbf{A_S}$**-b** at 35ºC. A second-order equation was fit to the full data set (red line) and that corresponding to an absorbance change equal to 70% of the total (the first 0.3 s) (blue line). The dotted line shows the fit to the first 0.3 s extended over the full timescale. (B) An expanded view of A. (C) The residuals (data values minus fit values) of the fits shown in A and B. It can be seen that the fit to the reduced data set better describes the initial data where the important information is found.

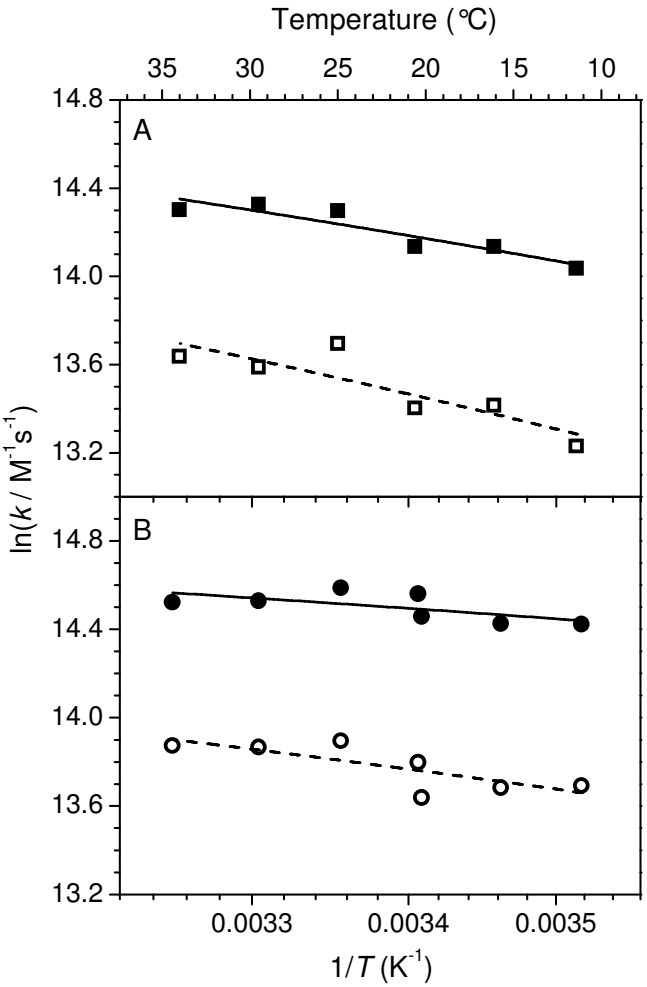


Fig 7: Arrhenius plots for the formation of (A) the duplex $\mathbf{A_s}$ and (B) the duplex $\mathbf{A_L}$. The activation energies for the formation of $\mathbf{A_S}$ and $\mathbf{A_L}$ were found to be 9.6 ± 4.9 kJ mol^{-1} and 3.9 ± 4.4 kJ mol^{-1} for method 1 and 13.2 ± 6.4 kJ mol^{-1} and 7.5 ± 4.5 kJ mol^{-1} for method 2, respectively. Solid black symbols: values for *k* found from second-order fits to a 70% drop in absorbance (method 1). Open black symbols: values for *k* found from linear fits to the initial data (method 2).

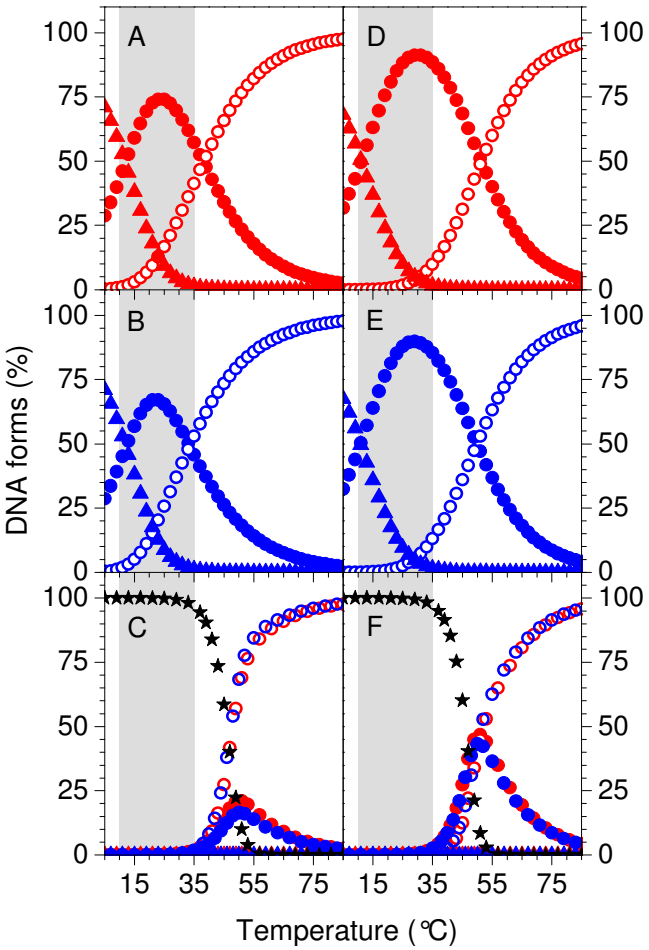


Fig 8: Simulated percentages of DNA in the different forms for strands **AT_S-a**, **AT_S-b**, **AT_L-a** and **AT_L-b** in 0.1-M Na^+ solutions as a function of temperature obtained using the UNAFold software package. Concentrations were (A) 7.0 μM of **AT_S-a**, (B) 7.0 μM of **AT_S-b**, (C) 3.5 μM of **AT_S-a** with 3.5 μM of **AT_S-b**, (D) 5.0 μM of **AT_L-a**, (E) 5.0 μM of **AT_L-b**, and (F) 2.5 μM of **AT_L-a** with 2.5 μM of **AT_L-b**. Data for the unfolded single strands (○), folded single strands (hairpins) (●), homo-duplexes (▲) and hetero-duplexes (★) are plotted for **AT_S-a** (red symbols) and **AT_S-b** (blue symbols) in the left column and **AT_L-a** (red symbols) and **AT_L-b** (blue symbols) in the right. The experimental range is highlighted in grey.

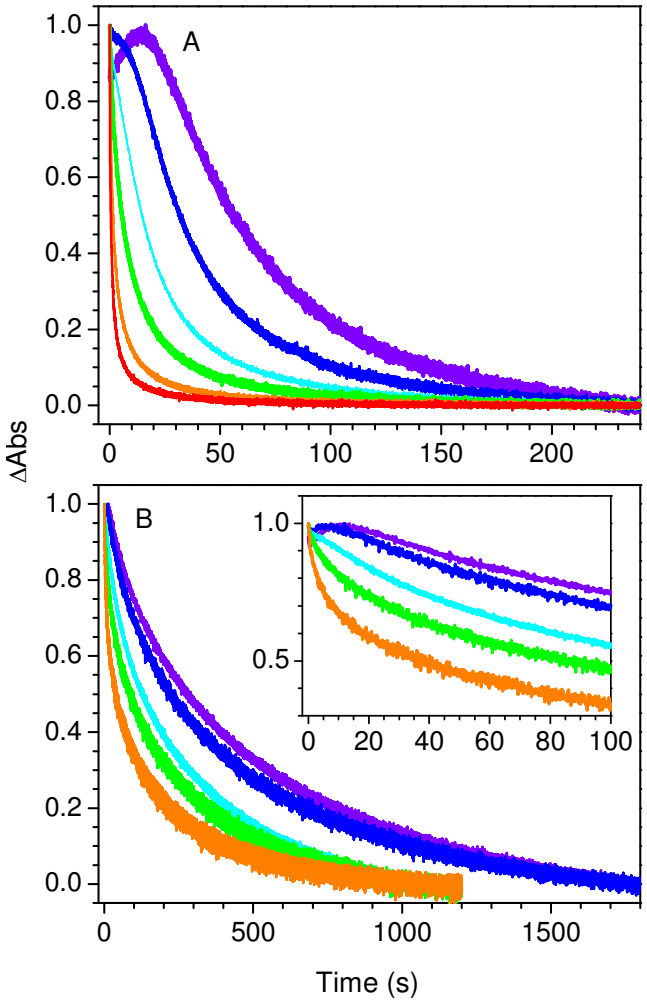


Fig 9: Time course of the absorbance change (scaled to 1) at 260 nm associated with duplex formation between (A) 3.2 μM of strands **AT_S-a** and **AT_S-b** at 11°C (violet), 15°C (blue), 20°C (cyan), 25°C (green), 30°C (orange) and 35°C (red) and (B) 2.6 μM of strands **AT_L-a** and **AT_L-b** at 11°C (violet), 16°C (blue), 20°C (cyan), 25°C (green), 30°C (orange).

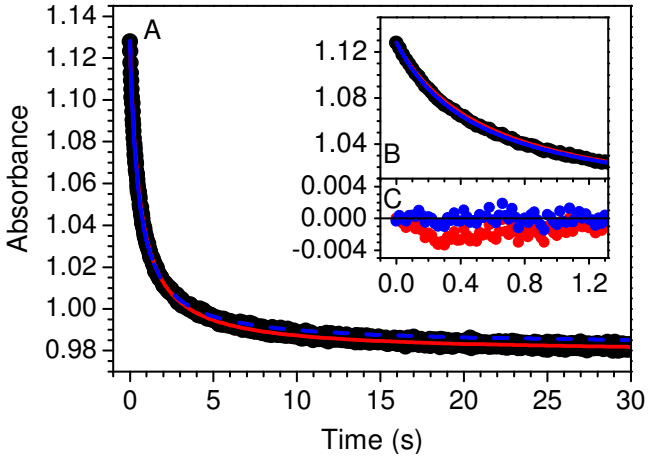


Fig 10: (A) Time course of the absorbance at 260 nm associated with duplex formation between 3.4 μM of strands **AT_S-a** and **AT_S-b** at 35ºC. A second-order model equation was fit to the full data set (red line) and that corresponding to an absorbance change equal to 70% of the total (the first 1.32 s) (blue line). The dashed blue line shows the fit to the first 1.32 s extended over the full timescale. (B) An expanded view of A. (C) The residuals of the fits shown in A and B. It can be seen that the fit to the reduced data set better describes the initial data where the important information is found.

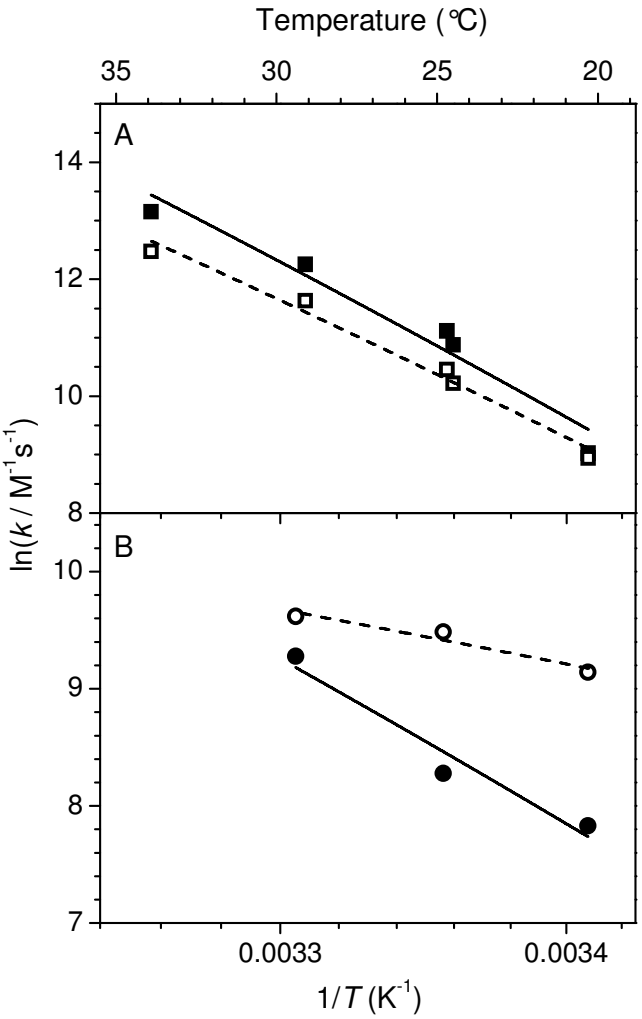


Fig 11: Arrhenius plots for the formation of (A) the duplex $\mathbf{AT_s}$ and (B) the duplex $\mathbf{AT_L}$. The activation energies for the formation of $\mathbf{AT_S}$ and $\mathbf{AT_L}$ were found to be 221 ± 39 kJ mol^{-1} and 117 ± 28 kJ mol^{-1} for method 1 and 195 ± 40 kJ mol^{-1} and 39 ± 18 kJ mol^{-1} for method 2, respectively. Solid black symbols: values for *k* found from second-order fits to a 70% drop in absorbance (method 1). Open black symbols: values for *k* found from linear fits to the initial data (method 2).

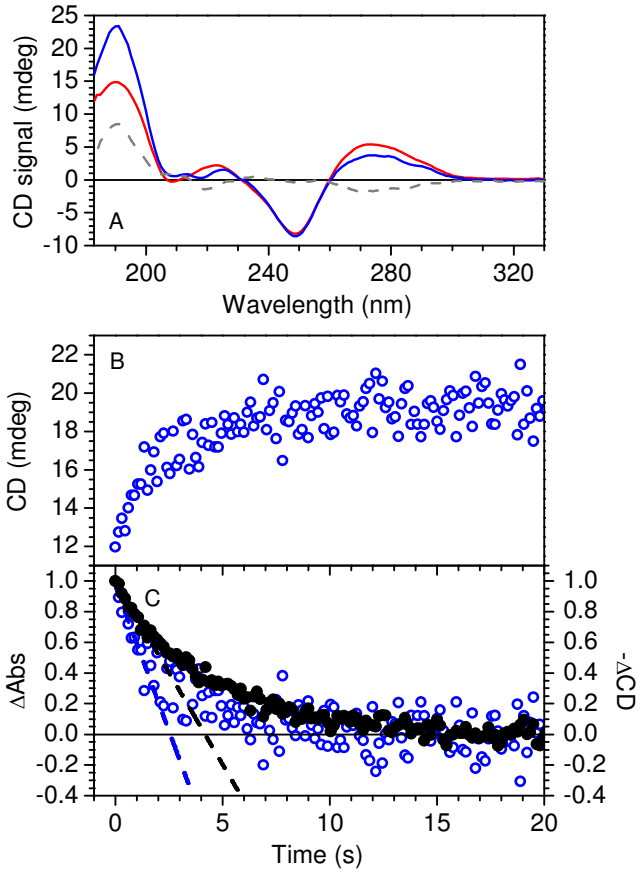


Fig 12: (A) The sum of the CD signals for the single strands **AT$_S$-a** (51 μM) and **AT$_S$-b** (50 μM) in a 10-mM phosphate buffer (pH 7.0), 100-mM NaF solution (red line) and the CD signal of the duplex formed between them (blue line). The difference between the aforementioned spectra is shown as a grey dashed line. (B) Time course of the CD signal at 192 nm associated with duplex formation between 26 μM of strand **AT$_S$-a** and 25 μM of strand **AT$_S$-b** (25°C). (C) The absorbance changes (scaled to 1) monitored over time for the same reaction (black filled circles). For comparison purposes, the negative of the change in CD signal (scaled to 1) is also shown in B (open blue circles). Linear fits to the initial data points are plotted as dashed lines.

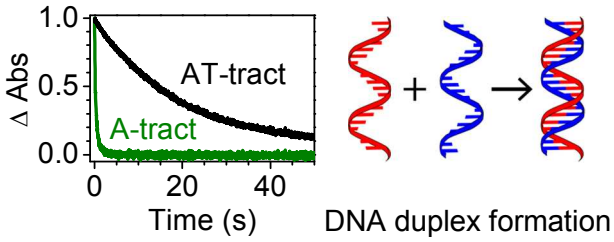
Notes and references

[a] Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. E-mail: jeanannwyer@gmail.com

Electronic Supplementary Information (ESI) available: Absorption spectra of single strands and duplexes; Kinetics data for pairs I and II, simulated percentages of forms present, and calculated hetero-duplex structures; Absorbance and CD data for **AT$_s$-a** and **AT$_s$-b** single strands at different temperatures. See DOI: xxx

1. E. B. Starikov and B. Norden, *J. Phys. Chem. B*, 2009, **113**, 4698-4707.
2. Y. Yin and X. S. Zhao, *Acc. Chem. Res.*, 2011, **44**, 1172-1181.
3. B. Rauzan, E. McMichael, R. Cave, L. R. Sevcik, K. Ostrosky, E. Whitman, R. Stegemann, A. L. Sinclair, M. J. Serra and A. A. Deckert, *Biochemistry*, 2013, **52**, 765-772.
4. E. Carrillo-Nava, Y. Mejia-Radillo and H. J. Hinz, *Biochemistry*, 2008, **47**, 13153-13157.
5. Y. Gao, L. K. Wolf and R. M. Georgiadis, *Nucleic Acids Res.*, 2006, **34**, 3370-3377.
6. J. M. S. Bartlett and D. Stirling, *Methods Mol. Biol.*, 2003, **226**, 3-6.
7. P. Wu, S. Nakano and N. Sugimoto, *Eur. J. Biochem.*, 2002, **269**, 2821-2830.
8. K. M. Parkhurst and L. J. Parkhurst, *Biochemistry*, 1995, **34**, 285-292.
9. E. T. Zuo, F. A. Tanious, W. D. Wilson, G. Zon, G. S. Tan and R. M. Wartell, *Biochemistry*, 1990, **29**, 4446-4456.
10. D. Porschke and M. Eigen, *J. Mol. Biol.*, 1971, **62**, 361-381.
11. J. G. Wetmur and N. Davidson, *J. Mol. Biol.*, 1968, **31**, 349-370.
12. X. B. Gu, S. Nakano and N. Sugimoto, *Chem. Commun.*, 2007, 2750-2752.
13. T. Ohmichi, H. Nakamuta, K. Yasuda and N. Sugimoto, *J. Am. Chem. Soc.*, 2000, **122**, 11286-11294.
14. A. P. Williams, C. E. Longfellow, S. M. Freier, R. Kierzek and D. H. Turner, *Biochemistry*, 1989, **28**, 4283-4291.
15. M. E. Craig, D. M. Crothers and P. Doty, *J. Mol. Biol.*, 1971, **62**, 383.
16. D. Porschke, O. C. Uhlenbec and F. H. Martin, *Biopolymers*, 1973, **12**, 1313-1335.
17. P. D. Ross and J. M. Sturtevant, *J. Am. Chem. Soc.*, 1962, **84**, 4503.
18. S. Wang, A. E. Friedman and E. T. Kool, *Biochemistry*, 1995, **34**, 9774-9784.
19. U. Christensen, N. Jacobsen, V. K. Rajwanshi, J. Wengel and T. Koch, *Biochem. J.*, 2001, **354**, 481-484.
20. U. Christensen, *Biosci. Rep.*, 2007, **27**, 327-333.
21. S. Howorka, L. Movileanu, O. Braha and H. Bayley, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 12996-13001.
22. M.-H. Hou, S.-B. Lin, J.-M. P. Yuann, W.-C. Lin, A. H.-J. Wang and L.-S. Kan, *Nucleic Acids Res.*, 2001, **29**, 5121-5128.
23. M.-C. Lin and R. B. Macgregor, *Biochemistry*, 1997, **36**, 6539-6544.
24. A. Tsourkas, M. A. Behlke and G. Bao, *Nucleic Acids Res.*, 2002, **30**, 5168-5174.
25. A. Tsourkas, M. A. Behlke, S. D. Rose and G. Bao, *Nucleic Acids Res.*, 2003, **31**, 1319-1330.
26. A. B. Rode, T. Endoh, H. Tateishi-Karimata, S. Takahashi and N. Sugimoto, *Chem. Commun.*, 2013, **49**, 8444-8446.
27. L. E. Morrison and L. M. Stols, *Biochemistry*, 1993, **32**, 3095-3104.
28. X. Chen, Y. Zhou, P. Qu and X. S. Zhao, *J. Am. Chem. Soc.*, 2008, **130**, 16947-16952.
29. M. Rajewska, K. Wegrzyn and I. Konieczny, *Fems Microbiol. Rev.*, 2012, **36**, 408-434.
30. S. P. Bell and A. Dutta, *Annu. Rev. Biochem.*, 2002, **71**, 333-374.
31. M. L. DePamphilis, in *DNA Replication in Eukaryotic Cells*, ed. M. L. DePamphilis, Cold Spring Harbor Laboratory Press, New York, 1996.
32. A. C. Leonard and M. Méchali, *Cold Spring Harb. Perspect. Biol.*, 2013, **5**, a010116.
33. M. Méchali, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 728-738.
34. J. A. Bryant and S. J. Aves, *Ann. Bot.*, 2011, **107**, 1119-1126.

35. A. C. Leonard and J. E. Grimwade, in *Annu. Rev. Microbiol., Vol 65*, eds. S. Gottesman and C. S. Harwood, Annual Reviews, Palo Alto, 2011, vol. 65, pp. 19-35.
36. H. C. M. Nelson, J. T. Finch, B. F. Luisi and A. Klug, *Nature*, 1987, **330**, 221-226.
37. M. Coll, C. A. Frederick, A. H. J. Wang and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 8385-8389.
38. C. Yoon, G. G. Prive, D. S. Goodsell and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 6332-6336.
39. A. Ghosh and M. Bansal, *J. Mol. Biol.*, 1999, **294**, 1149-1158.
40. A. Ghosh and M. Bansal, *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, 1999, **55**, 2005-2012.
41. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179-5197.
42. N. R. Markham and M. Zuker, in *Bioinformatics*, ed. J. M. Keith, Humana Press, 2008, vol. 453, pp. 3-31.
43. D. M. Gray, K. H. Johnson, M. R. Vaughan, P. A. Morris, J. C. Sutherland and R. L. Ratliff, *Biopolymers*, 1990, **29**, 317-323.
44. J. H. Riazance, W. C. Johnson, L. P. McIntosh and T. M. Jovin, *Nucleic Acids Res.*, 1987, **15**, 7627-7636.
45. J. H. Riazance, W. A. Baase, W. C. Johnson, K. Hall, P. Cruz and I. Tinoco, *Nucleic Acids Res.*, 1985, **13**, 4983-4989.
46. L. M. Nielsen, S. V. Hoffmann and S. B. Nielsen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 15054-15059.
47. C. R. Cantor, M. M. Warshaw and H. Shapiro, *Biopolymers*, 1970, **9**, 1059-1077.
48. G. D. Fasman, Handbook of Biochemistry and Molecular Biology, Volume 1: Nucleic Acids, 3rd edition, 1975, pp. 589, CRC Press.
49. A. J. Miles, S. V. Hoffmann, Y. Tao, R. W. Janes and B. A. Wallace, *Spectrosc. Int. J.*, 2007, **21**, 245-255.
50. A. J. Miles, R. W. Janes, A. Brown, D. T. Clarke, J. C. Sutherland, Y. Tao, B. A. Wallace and S. V. Hoffmann, *J. Synchrotron Radiat.*, 2008, **15**, 420-422.
51. J. H. Bredehöft, N. C. Jones, C. Meinert, A. C. Evans, S. V. Hoffmann and U. J. Meierhenrich, *Chirality*, 2014, DOI: 10.1002/chir.22329.
52. C. Meinert, J. H. Bredehoeft, J. J. Filippi, Y. Baraud, L. Nahon, F. Wien, N. C. Jones, S. V. Hoffmann and U. J. Meierhenrich, *Angew. Chem. Int. Ed.*, 2012, **51**, 4484-4487.
53. R. Owczarzy, Y. You, B. G. Moreira, J. A. Manthey, L. Huang, M. A. Behlke and J. A. Walder, *Biochemistry*, 2004, **43**, 3537-3554.
54. C. Chen, W. Wang, Z. Wang, F. Wei and X. S. Zhao, *Nucleic Acids Res.*, 2007, **35**, 2875-2884.
55. A. M. Paiva and R. D. Sheardy, *J. Am. Chem. Soc.*, 2005, **127**, 5581-5585.
56. P. W. Atkins, in *Physical Chemistry*, ed. P. W. Atkins, Oxford University Press, Oxford, 1990.
57. J. W. Nelson and I. Tinoco, *Biochemistry*, 1982, **21**, 5289-5295.
58. T. G. Dewey and D. H. Turner, *Biochemistry*, 1979, **18**, 5757-5762.
59. B. Jolles, A. Laigle, L. Chinsky and P. Y. Turpin, *Nucleic Acids Res.*, 1985, **13**, 2075-2085.
60. H. Simpkins and E. G. Richards, *Biochemistry*, 1967, **6**, 2513-2520.
61. C. S. M. Olsthoorn, L. J. Bostelaar, J. F. M. Derooij, J. H. Vanboom and C. Altona, *Eur. J. Biochem.*, 1981, **115**, 309-321.



A-tracts (AAAA..:TTTT…) form much faster than AT-tracks (ATAT…:TATA…).