EDGE ARTICLE

Building blocks for recognition-encoded oligoesters that form H-bonded duplexes†

Filip T. Szczypiński [ID] and Christopher A. Hunter [ID] *

Competition from intramolecular folding is a major challenge in the design of synthetic oligomers that form intermolecular duplexes in a sequence-selective manner. One strategy is to use very rigid backbones that prevent folding, but this design can prejudice duplex formation if the geometry is not exactly right. The alternative approach found in nucleic acids is to use bases (or recognition units) that have different dimensions. A long-short base-pairing scheme makes folding geometrically difficult and is compatible with the flexible backbones that are required to guarantee duplex formation. A monomer building block equipped with a long hydrogen bond donor (phenol, **D**) recognition unit and a monomer building block equipped with a short hydrogen bond acceptor (phosphine oxide, **A**) recognition unit were prepared with differentially protected alcohol and carboxylic acid groups. These compounds were used to synthesise the homo and hetero-sequence 2-mers **AA**, **DD** and **AD**. ^{19}F and ^{31}P NMR experiments were used to characterize the assembly properties of these compounds in toluene solution. **AA** and **DD** form a stable doubly-hydrogen-bonded duplex with an effective molarity of 20 mM for formation of the second intramolecular hydrogen bond. **AD** forms a duplex of similar stability. There is no evidence of intramolecular folding in the monomeric state of this compound, which shows that the long-short base-pairing scheme is effective. The ester coupling chemistry used here is an attractive method for the synthesis of long oligomers, and the properties of the 2-mers indicate that this molecular architecture should give longer mixed sequence oligomers that show high fidelity sequence-selective duplex formation.

1 Introduction

1.1 Background

Two sequence-complementary strands of nucleic acid will form a stable duplex due to hydrogen bonding interactions between the bases. This supramolecular structure was immediately recognised to provide a plausible mechanism for information transfer between a template strand and a copy in the key biological processes of replication, translation and transcription, where the sequence of the copy is organised by the same base-pairing interactions that lead to duplex formation.[1,2] These copying processes are currently unique to nucleic acids and represent the molecular basis for the evolution of life on this planet. Synthetic systems that form duplexes in the same way are therefore likely to provide a platform for template-directed synthesis of mixed sequence oligomers, and ultimately to the application of directed evolution for the discovery of new functional non-biological molecules.[3–8]

It is clear that duplex formation is not restricted to the precise molecular structure found in DNA and RNA. A range of nucleic acid analogues have been prepared in which the phosphate diester,[9–11] the bases,[7,12–15] and the sugar have been replaced,[16–22] and all of these oligomers form stable duplexes. Synthetic oligomers that bear no relation to nucleic acids have also been shown to form duplexes through various non-covalent interactions: metal–ligand coordination,[23,24] salt bridges,[25,26] aromatic interactions,[27] and hydrogen bonding.[28–30] By using two different complementary recognition sites as the equivalent of the nucleic acid bases, it is also possible to encode sequence information into synthetic oligomers, and sequence-selective duplex formation has been demonstrated for short sequences.[26,31]

We have been using a single hydrogen bond between a hydrogen bond donor (*e.g.* phenol, **D**) and a hydrogen bond acceptor (*e.g.* phosphine oxide, **A**) as the base-pairing interaction for duplex formation. This two letter alphabet allows information to be encoded in an oligomer as the sequence of **A** and **D** recognition sites. Provided the backbone does not contain any polar functional groups that could compete with the base-pairing interactions, the use of a single hydrogen bond as the base-pair removes any possibility of mismatches, because **A** cannot interact with **A** and **D** cannot interact with **D**. A number of different backbone architectures have been

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: herchelsmith.orgchem@ch.cam.ac.uk

characterized, and the nature of the backbone was found to play a crucial role in the assembly properties of these oligomers.

The different possible self-assembly channels are illustrated in Fig. 1. The key requirement for duplex formation is that the equilibrium constant for propagation of the intramolecular hydrogen bonds that zip up the duplex, K EM_p, is greater than one (K is the association constant for formation of an inter-molecular hydrogen bond, and EM_p is the effective molarity for propagation of intramolecular hydrogen bonds in the duplex).[32,33] One of the competing assembly channels is formation of multiple intermolecular interactions that lead to higher order networks, but this process can be avoided by operating at a concentration, c, which is lower than the value of EM_i, the effective molarity for formation of the first intra-molecular hydrogen bond that initiates duplex formation. The other major competing assembly channel is due to the forma-tion an intramolecular hydrogen bond within an oligomer, which leads to folding. The probability of this process is determined by the equilibrium constant K EM_f, where EM_f is the effective molarity for folding.

The values of the three effective molarity parameters depend on the conformational properties of the backbone. For the very flexible backbone shown in Fig. 2(a), the values of EM_i and EM_p are 10 mM to 30 mM, and the duplex channel dominates for length complementary homo-oligomers.[34] For the very rigid backbone shown in Fig. 2(b), similar results were obtained with EM_i and EM_p values of 40 mM to 70 mM.[35] Geometry is critical for more rigid backbones. The backbone shown in Fig. 2(b) has a well-defined geometry, which places the recognition groups in the correct orientation for duplex formation. However, for backbones of intermediate rigidity, where the conformational properties are more difficult to predict, mixed results were ob-tained. The backbone shown in Fig. 2(c) formed duplexes with $EM_i = EM_p = 10$ mM,[36] but the backbones shown in Fig. 2(d) and (e) did not lead to extended duplexes. For these two systems, EM_i was similar to the values found for the other
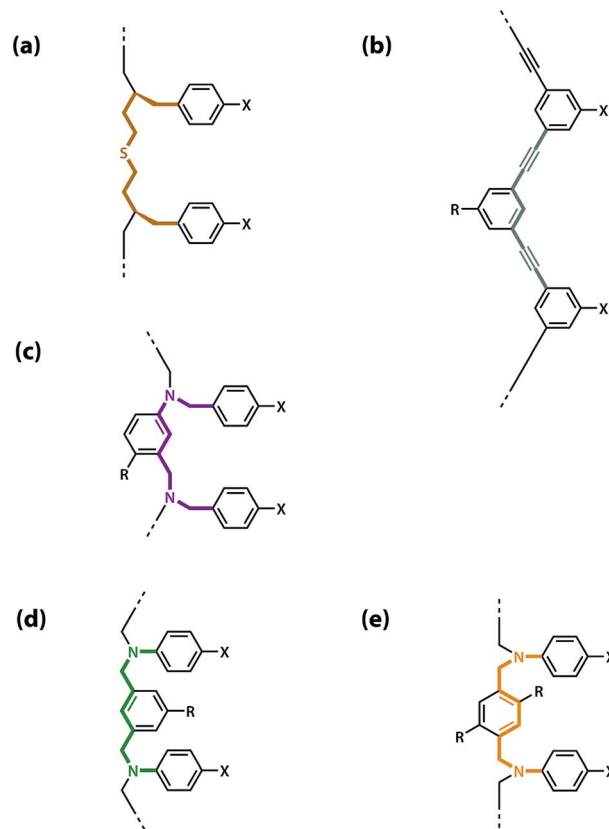

Fig. 2 Backbones (a)–(e) of the previously reported synthetic infor-mation molecules.[34–36,38]

backbones (10 mM to 20 mM), but the geometry was not compatible with duplex propagation, and EM_p was too low to measure.[37]

The results obtained for homo-oligomers suggest that highly flexible backbones should provide a reliable platform for the design of duplex-forming oligomers. Conformational flexibility ensures that the backbone will always be able to adapt to a geometry compatible with base-pair formation in an extended duplex. More rigid backbones are difficult to design with the degree of accuracy required to guarantee the geometric complementarity needed for formation of an extended duplex.[37] The values of effective molarity measured for the very flexible backbone and the very rigid backbone shown in Fig. 2 are similar, so it appears that effective molarities associated with duplex formation are not adversely affected by conformational flexibility. Very flexible backbones are easily accessed, so this approach would make backbone design straightforward.

However, the effective molarity for intramolecular folding, EM_f, also depends on the conformational properties of the backbone. As shown in Fig. 3(a), a long flexible backbone promotes 1,2-folding between **A** and **D** recognition units that are adjacent in sequence. The value of EM_f for this system is about 10 mM, which is comparable to the values of effective molarity for zipping up the duplex, so the folding channel will dominate for mixed sequence oligomers of this architecture.[39] Of course, longer mixed sequence oligomers will always be able to fold, no matter what backbone is used, and indeed sequence-
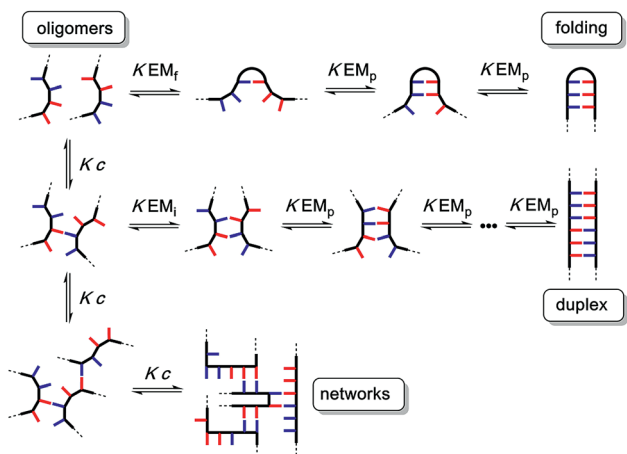

Fig. 1 Possible channels for supramolecular assembly of recognition-encoded oligomers. The outcome depends on the concentration, c, the association constant for the intermolecular base-pairing interac-tion, K, and the effective molarities for folding, EM_f, duplex initiation, EM_i and duplex propagation, EM_p.
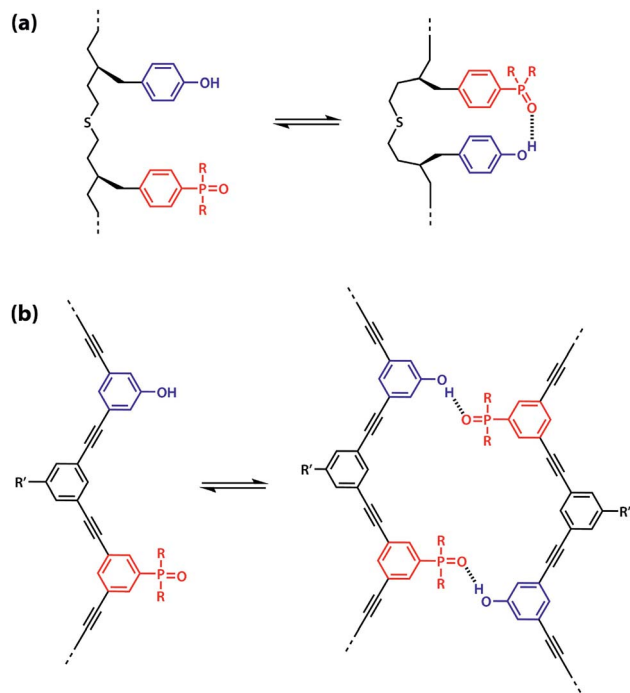
Fig. 3 (a) Intramolecular 1,2-folding in information molecules with flexible backbones. (b) Duplex formation in information molecules with rigid backbones.
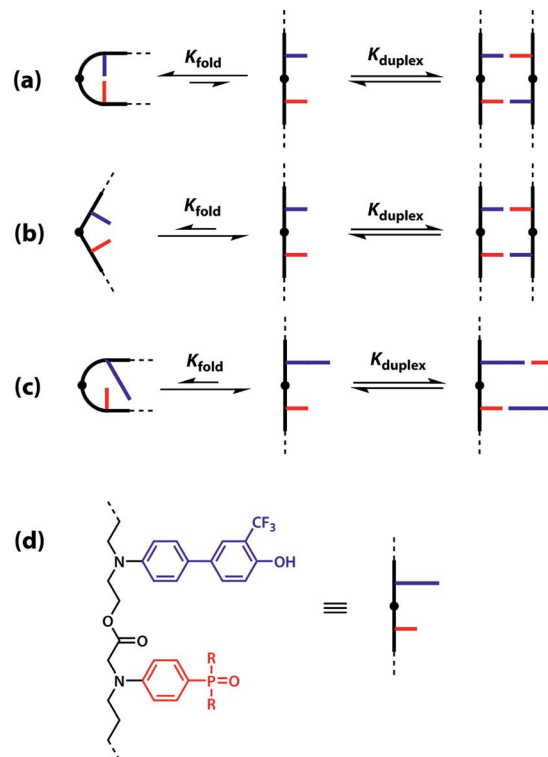

Fig. 4 (a) Flexible backbones lead to 1,2-folding; (b) rigid backbones prevent 1,2-folding; (c) recognition units of different dimensions prevent 1,2-folding; (d) molecular design corresponding to the short-long base-pairing scheme.

encoded folding of single-stranded RNA is key to the biological properties.[2] Folded nucleic acid structures involve looped out bases, so if a single-stranded nucleic acid is annealed with a sequence-complementary strand, duplex formation will dominate, because additional base-pairing interactions are made in the duplex. However, Fig. 1 shows that if 1,2-folding is possible, the number of base-pairs formed in the folding and duplex channels can be identical, so the folding channel will dominate. Minimising 1,2-folding is therefore critical to the design of recognition-encoded oligomers that form sequence-selective duplexes with high fidelity.

One strategy for avoiding 1,2-folding is to reduce the value of EM_f by increasing the rigidity of the backbone. As shown in Fig. 3(b), the very rigid backbone that we studied previously does not fold, so duplex formation is the dominant assembly channel for mixed sequence oligomers of this architecture. However, it would be preferable to work with more flexible backbones to guarantee duplex formation, as explained above. Here, we explore an alternative strategy for preventing 1,2-folding in oligomers with a very flexible backbone. If two short bases are attached to a long flexible backbone, 1,2-folding is favoured (Fig. 4(a)). Fig. 4(b) illustrates how folding can be prevented by attaching the two short bases to a rigid backbone. Fig. 4(c) shows how changing the dimensions of the bases can be used to prevent folding. By making one of the bases longer than the other, the probability of finding a backbone conformation compatible with folding is significantly reduced, and the duplex assembly channel should dominate. Fig. 4(d) shows the corresponding molecular design that we validate in this paper. It is worth noting that this short-long base-pairing

scheme has similar geometrical properties to the purine–pyrimidine base-pairing system found in nucleic acids.

The backbone proposed in Fig. 4(d) uses ester linkages as the coupling chemistry for the synthesis of oligomers. Esters are sufficiently weak hydrogen bond acceptors ($\beta \approx 5.5$) not to compete significantly with the phosphine oxide recognition units ($\beta \approx 10.5$).[40] Ester coupling is sufficiently high-yielding to be used for the synthesis of polymers, and iterative coupling could be automated in a peptide synthesiser.[41–43] Orthogonal protecting groups have been developed for the preparation of oligoesters with sequences of different building blocks.[44–51] Here, we describe synthesis of the required monomer building blocks, demonstrate their use in the synthesis of different 2-mer sequences, and show that the long-short base-pairing scheme successfully prevents 1,2-folding for this oligomer architecture.

2 Results and discussion

2.1 Synthesis

A divergent approach to the synthesis of the monomer building blocks was employed, in which a common aromatic bromide intermediate was coupled with the hydrogen bond donor and acceptor recognition units, as shown in Scheme 1. Commercially available 2-bromoethanol **5** was protected as the silyl ether **6**, which was then used for alkylation of 4-bromoaniline to yield **7**. Aniline **7** was alkylated with benzyl bromoacetate to give the key intermediate **8**. Commercially available phenol **1** was
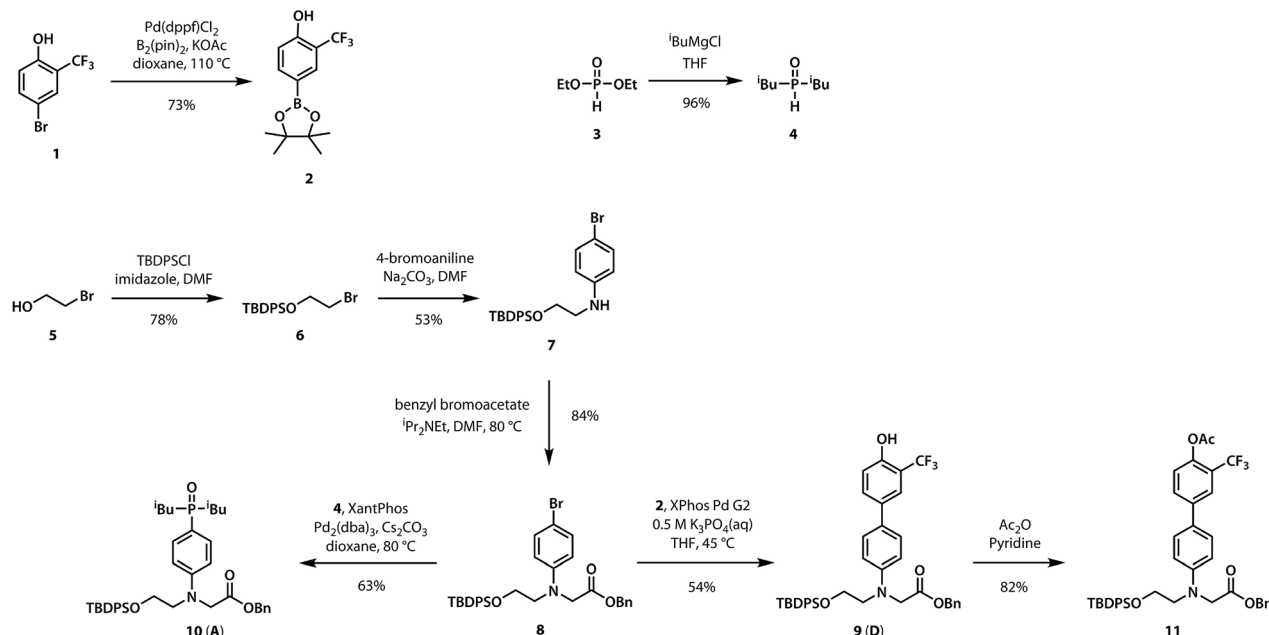
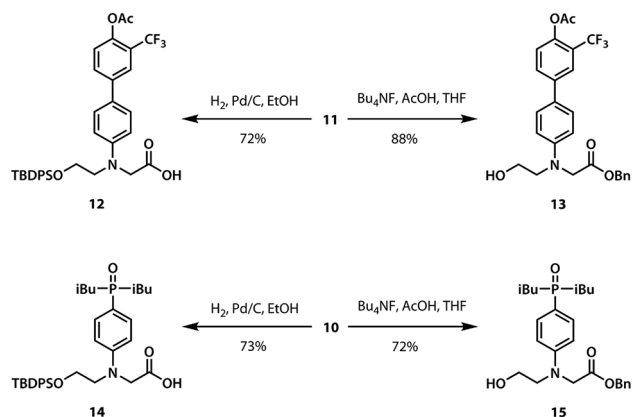**Scheme 1** Synthesis of the orthogonally-protected monomers for information oligoesters.

converted to the boronic ester **2**, which was coupled with **8** under Suzuki–Miyaura conditions to give the hydrogen bond donor monomer **9** (**D**). Treatment of commercially available diethyl phosphite **3** with iso-butylmagnesium chloride gave **4**, which was coupled with **8** using palladium(0) and XantPhos to yield the hydrogen bond acceptor monomer **10** (**A**).

For the ester coupling reactions, the potentially reactive phenol moiety in **9** was first protected as the acetyl ester **11** (Scheme 2). The benzyl and TBDPS protecting groups in **10** and **11** were removed orthogonally to give the four precursors **12–15** required for ester coupling reactions. Treatment with hydrogen gas over palladium on charcoal gave the monoprotected carboxylic acids **12** and **14**. Alternatively, reaction with *n*-tetrabutylammonium fluoride buffered with acetic acid gave the monoprotected alcohols **13** and **15**. These monoprotected hydroxyacid monomers were used to synthesise three different 2-mer sequences by EDC coupling with a catalytic amount of

N,N-dimethylaminopyridine (Scheme 3). Coupling **14** with **15** gave **AA** directly. **AD** and **DD** were obtained with the phenol groups protected as acetate esters, but these groups were removed quantitatively by stirring in a solution of ammonium acetate in water and methanol.

2.2 NMR binding studies

Duplex formation and folding were investigated using ^{19}F and ^{31}P NMR titrations and dilutions in toluene-d$_8$ at 298 K. The

Scheme 3 Synthesis of the **AA**, **DD**, and **AD** 2-mers.

Scheme 2 Synthesis of the monoprotected hydroxyacid monomers.

association constant for formation of the $\mathbf{A \cdot D}$ complex, which makes a single intermolecular hydrogen bond, was measured by titrating $\mathbf{A}$ into $\mathbf{D}$. A large upfield change in the ^{19}F NMR chemical shift of $\mathbf{D}$ was observed, and the data fit well to a 1 : 1 binding isotherm to give an association constant of $K_{\mathbf{A \cdot D}} = 3.8 \times 10^3$ M^{-1} (Table 1). The association constant for the $\mathbf{AA \cdot DD}$ complex was similarly measured by titrating $\mathbf{AA}$ into $\mathbf{DD}$, and the association constant for dimerization of $\mathbf{AD}$ was determined by a ^{19}F NMR dilution experiment in toluene-d$_8$ at 298 K. The association constants for the $\mathbf{AA \cdot DD}$ and $\mathbf{AD \cdot AD}$ complexes are both two orders of magnitude higher than that for $\mathbf{A \cdot D}$, which indicates that there are two cooperative hydrogen bonding interactions in the complexes formed by the sequence complementary 2-mers (Table 1). The limiting ^{19}F and ^{31}P NMR chemical shifts of the free species (δ_{free}) and fully bound complexes (δ_{bound}) were determined by extrapolation of the binding isotherms (Table 1). The values are similar for all three complexes. The large upfield limiting complexation-induced changes in ^{19}F NMR chemical shift (0.4 ppm) indicate that all of the phenol groups form hydrogen bonds in all of the complexes. The large downfield limiting complexation-induced changes in ^{31}P NMR chemical shift (5–7 ppm) indicate that all of the phosphine oxide groups form hydrogen bonds in all of the complexes. Comparison of the values of the free ^{19}F and ^{31}P NMR chemical shifts of $\mathbf{AA}$, $\mathbf{DD}$, and $\mathbf{AD}$ show that there is no significant intramolecular hydrogen bonding in $\mathbf{AD}$, *i.e.* there is no folding in the monomeric state. The results indicate that both the $\mathbf{AA \cdot DD}$ and $\mathbf{AD \cdot AD}$ duplexes are fully assembled through the intended base-pairing interactions at mM concentrations in toluene solution at room temperature as shown in Fig. 5.

A schematic representation of the equilibria involved in duplex assembly is shown in Fig. 6. For $\mathbf{AA \cdot DD}$, formation of the first intermolecular hydrogen bond gives an open complex, and formation of the second intramolecular hydrogen bond gives the closed duplex. Assuming that all of the hydrogen bonds in the systems described here are of similar strength, it is possible to describe the association constant for formation of the closed $\mathbf{c \cdot AA \cdot DD}$ duplex in terms of the association constant for formation of a single intermolecular hydrogen bond $K_{\mathbf{A \cdot D}}$ and the effective molarity for the intramolecular interaction EM_i. The backbone in these systems has a direction, because the hydroxyl and acid ends are different, so parallel and anti-parallel orientations of the duplex are possible. As the end groups are spatially separated from the recognition sites, we assume that the two possible $\mathbf{c \cdot AA \cdot DD}$ have similar stability.

Therefore, the open complex $\mathbf{o \cdot AA \cdot DD}$ has four equally populated states and the closed duplex $\mathbf{c \cdot AA \cdot DD}$ has two.

It is possible to express the association constants for duplex formation in terms of $K_{\mathbf{A \cdot D}}$ and EM_i:

$$K_{\mathbf{AA \cdot DD}} = 4K_{\mathbf{A \cdot D}} \times \frac{1}{2}K_{\mathbf{A \cdot D}}\text{EM}_i = 2K_{\mathbf{A \cdot D}}^2\text{EM}_i \qquad (1)$$

Hence the effective molarity for duplex formation can be determined as:

$$\text{EM}_i = \frac{K_{\mathbf{AA \cdot DD}}}{2K_{\mathbf{A \cdot D}}^2} \qquad (2)$$

The association constants in Table 1 were used to calculate EM_i for this system as 19 ± 3 mM, which is consistent with values of supramolecular effective molarities we have measured for other hydrogen bonded duplexes.[31,34–36,38,39] The equilibrium constant for closing the duplex is given by $\frac{1}{2}K_{\mathbf{A \cdot D}}\text{EM}_i$ and is 40 for this system, which implies that the duplex is fully closed and only 2% of the species populate the partially-bound open state $\mathbf{o \cdot AD \cdot AD}$.

For the closed hetero-2-mer duplex $\mathbf{c \cdot AD \cdot AD}$, there is no degeneracy associated with the backbone directionality, because the anti-parallel orientation is determined by the sequence. However, there is the possibility of intramolecular 1,2-folding in the monomeric state, which is governed by the corresponding effective molarity EM_f. Hence, the observed dimerisation constant $K_{\mathbf{AD \cdot AD}}$ depends on the concentrations of the folded ($\mathbf{AD}_{\text{folded}}$) and open ($\mathbf{AD}_{\text{open}}$) species that are populated in the monomeric state:

$$[\mathbf{AD}] = [\mathbf{AD}_{\text{open}}] + [\mathbf{AD}_{\text{fold}}] = [\mathbf{AD}_{\text{open}}](1 + K_{\mathbf{A \cdot D}}\text{EM}_f) \qquad (3)$$

$$K_{\mathbf{AD \cdot AD}} = \frac{[\mathbf{c \cdot AD \cdot AD}]}{[\mathbf{AD}]^2} = \frac{[\mathbf{c \cdot AD \cdot AD}]}{(1 + K_{\mathbf{A \cdot D}}\text{EM}_f)^2 [\mathbf{AD}_{\text{open}}]^2}$$

$$= \frac{K_{\mathbf{A \cdot D}}^2\text{EM}_i}{2(1 + K_{\mathbf{A \cdot D}}\text{EM}_f)^2} \qquad (4)$$

Assuming that the effective molarity for duplex formation, EM_i, is the same for $\mathbf{AA \cdot DD}$ and $\mathbf{AD \cdot AD}$, it is possible to combine eqn (2) and (4) to determine $(K_{\mathbf{A \cdot D}}\text{EM}_f + 1)$, which is the factor that describes the fraction of monomeric $\mathbf{AD}$ that exists in the folded state:

$$K_{\mathbf{A \cdot D}}\text{EM}_f + 1 = \sqrt{\frac{K_{\mathbf{AA \cdot DD}}}{4K_{\mathbf{AD \cdot AD}}}} \qquad (5)$$

Table 1 Association constants (K), effective molarities (EM), limiting NMR chemical shifts (δ_{free} and δ_{bound}), and complexation-induced changes in the chemical shifts ($\Delta\delta$) for the formation of duplexes in toluene at 298 K

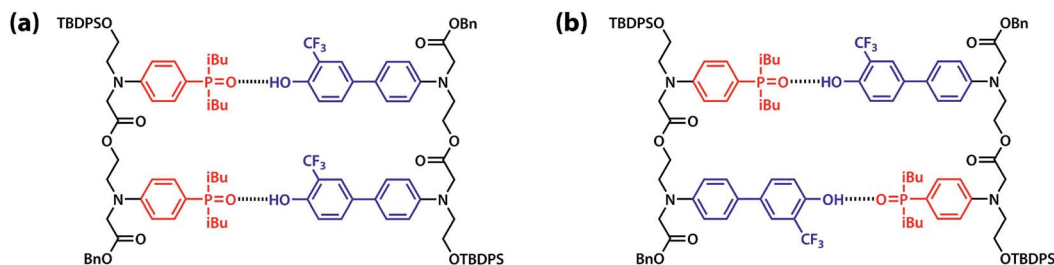| Complex | log K/M^{-1} | ^{19}F NMR | | | ^{31}P NMR | | |
		δ_{free}/ppm	δ_{bound}/ppm	$\Delta\delta$/ppm	δ_{free}/ppm	δ_{bound}/ppm	$\Delta\delta$/ppm
$\mathbf{A \cdot D}$	3.6 ± 0.1	-61.2	-61.6	-0.4	34.2	41.0	6.8
$\mathbf{AA \cdot DD}$	5.8 ± 0.1	-61.1	-61.5	-0.4	34.3	39.3	5.0
$\mathbf{AD \cdot AD}$	5.2 ± 0.1	-61.1	-61.5	-0.4	35.8	40.9	5.1

Fig. 5 Hydrogen bonded duplexes formed by (a) the **AA** and **DD** 2-mers, and (b) the self-complementary **AD** 2-mer.
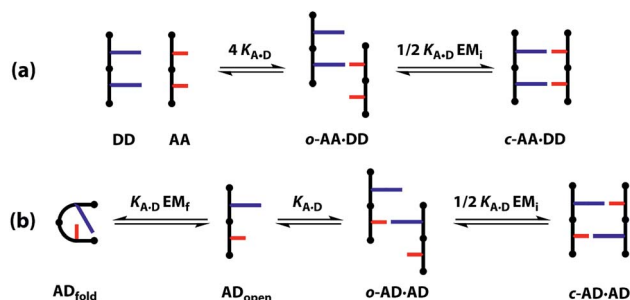

Fig. 6 Competing equilibria in the assembly of (a) the **AA·DD** duplex and (b) the **AD·AD** duplex.

Substituting the values from Table 1 into eqn (5) gives a value of 1.0 for $(K_{\mathbf{A \cdot D}} EM_f + 1)$, which is consistent with the NMR chemical shift data. These results indicate that virtually all monomeric **AD** exists in the open state and the 1,2-folding does not compete with duplex formation in this system.

If the two arrangements of the *c*-**AA·DD** were not degenerate, the statistical factor in eqn (1) would be equal to one, giving $(K_{\mathbf{A \cdot D}} EM_f + 1) \approx 1.4$. This value would require that 30% of monomeric **AD** exists in the folded state, which is not consistent with the NMR chemical shift data, suggesting that assumption that the parallel and antiparallel backbone arrangements are equally populated in the *c*-**AA·DD** duplex is reasonable.

2.3 Molecular mechanics calculations

The competition between duplex formation and intramolecular 1,2-folding in **AD** were further investigated using molecular mechanics calculations. The OPLS3 force field with implicit chloroform solvation model was employed, as implemented in the MacroModel software (the experiments were carried out in toluene, but chloroform is the only non-polar implicit solvent model implemented).[52] A conformational search was performed on the **AD** monomer and the lowest energy structure, shown in Fig. 7(a), is a folded species. The calculation is clearly inconsistent with the experimental results, reinforcing our previous findings that computational methods do not provide a reliable method for predicting the self-assembly properties of synthetic molecules of this complexity.[39] To investigate whether this folded structure is strongly preferred over duplex formation by the force-field, two molecules of **AD** were constrained to have one intermolecular hydrogen bond, and a conformational search

gave the closed *c*-**AD·AD** duplex shown in Fig. 7(b) as the lowest energy structure. No open *o*-**AD·AD** structures were found within 5 kJ mol⁻¹ of the minimum. The calculated energy of the duplex is 87 kJ mol⁻¹ lower than the energy of two folded monomers, which suggests that there is considerable strain associated with folding in this system.

2.4 Double hydrogen bonding

Oxygen hydrogen bond acceptors can interact with more than one hydrogen bond donor, which can degrade the fidelity of sequence-selective duplex formation.[31] In order to investigate whether the base-pair recognition system used here would suffer from this problem, **A** was titrated into **DD**. The changes in ¹⁹F NMR chemical shift of the **DD** did not fit to a 1 : 1 isotherm (see ESI†), so a 1 : 2 binding model was investigated:

$$\mathbf{DD} + \mathbf{A} \underset{}{\overset{K_1}{\rightleftarrows}} \mathbf{DD \cdot A} \tag{6}$$

$$\mathbf{DD \cdot A} + \mathbf{A} \underset{}{\overset{K_2}{\rightleftarrows}} \mathbf{DD \cdot A_2} \tag{7}$$

The two donor binding sites were assumed to be independent and identical, hence $K_1 K_2 = K_{\mathbf{A \cdot D}}{}^2$ could be fixed in the least squares regression analysis. The association constant for the **DD·A** was determined to be $K_1 = (15\,000 \pm 2000)$ M⁻¹, which is four times greater than the single hydrogen bond association constant $K_{\mathbf{A \cdot D}}$ and suggests additional stabilisation
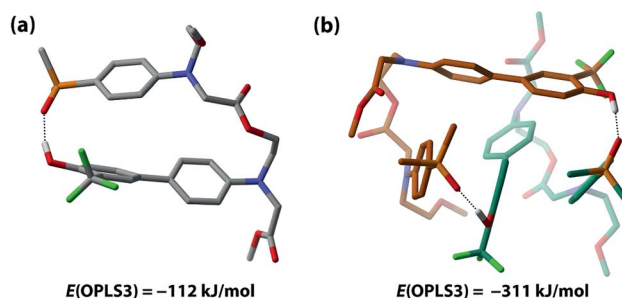

Fig. 7 Lowest-energy structures of (a) AD and (b) AD·AD. The OPLS3 force field with implicit chloroform solvation model was employed. Protecting groups and alkyl groups on phosphorus were replaced with methyl groups. Carbon atoms are colour coded by molecule, and hydrogen atoms are not shown for clarity. Hydrogen bonds are shown as dotted lines.

due to a hydrogen bond between the second phenol and the phosphine oxide. We can represent the equilibria leading to the doubly bonded complex as in Fig. 8.

Noting that both 1 : 1 complexes give rise to the observed association constant, K_1 can be expressed as:

$$K_1 = 2K_{\mathbf{A}\cdot\mathbf{D}} + 2K_{\mathbf{A}\cdot\mathbf{D}}K' \qquad (8)$$

The association constant for the formation of the second hydrogen bond is, therefore:

$$K' = \frac{K_1 - 2K_{\mathbf{A}\cdot\mathbf{D}}}{2K_{\mathbf{A}\cdot\mathbf{D}}} = \frac{K_1}{2K_{\mathbf{A}\cdot\mathbf{D}}} - 1 \qquad (9)$$

Using eqn (9) and the measured value for K_1, the association constant for the interaction of the second phenol donor with the acceptor is K' (1.0 ± 0.2), which means that the double-bonded complex represents 50% of the 1 : 1 complex. The ratio of $K_{\mathbf{A}\cdot\mathbf{D}}EM_i$ and K' describes the competition between a correctly recognised duplex and a doubly hydrogen-bonded mismatched complex. This ratio is 80 for this system, therefore sequence selectivity should be achieved for longer information oligoesters with fidelity of 99%. For comparison, the previously reported sequence-containing information oligomer shows $K' = 1.6$ and $K_{\mathbf{A}\cdot\mathbf{D}}EM_i = 9.9$, hence exhibits sequence fidelity of 86%.[31,39] While the value of K' for the system described here is comparable with that reported earlier, the exceptionally strong hydrogen-bonding interaction between the

recognition units should lead to superior performance the formation of closed duplexes with high sequence fidelity.

3 Conclusions

In conclusion, candidates for new information molecules were synthesised and their behaviour in toluene was studied through ^{19}F and ^{31}P NMR spectroscopy. The monomeric building blocks are readily accessible and 2-mers were easily synthesised through efficient ester coupling reactions, with scope for the synthesis of longer oligomers using the same methodology. A long-short base-pairing scheme akin to purines and pyrimidines in natural nucleic acids was employed in order to reduce intramolecular folding and a flexible backbone was used to ensure the geometric complementarity required for duplex formation. Homo- and hetero-2-mers were observed to form stable duplexes in toluene at 298 K with effective molarities for duplex formation of 20 mM and without any substantial 1,2-folding. The observed trends were consistent with those previously reported using ^{31}P NMR, thus providing a convenient handle for studying supramolecular association. Formation of double hydrogen bonds to the oxygen-based acceptor was found to be much less favoured than the desired base-pairing interactions. This system appears to be ideally suited to the synthesis of longer oligomers which are expected to show the possibility of high-fidelity sequence-specific information recognition *via* hydrogen bonding in organic solvents.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.

2 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.

3 J. W. Szostak, D. P. Bartel and P. L. Luisi, *Nature*, 2001, **409**, 387–390.

4 S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco and J. W. Szostak, *Nature*, 2008, **454**, 122–125.

5 V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn and P. Holliger, *Science*, 2012, **336**, 341–344.

6 A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morgunov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn and P. Holliger, *Nature*, 2014, **518**, 427–430.

7 L. Zhang, Z. Yang, K. Sefah, K. M. Bradley, S. Hoshika, M.-J. Kim, H.-J. Kim, G. Zhu, E. Jiménez, S. Cansiz, I.-T. Teng, C. Champanhac, C. McLendon, C. Liu,
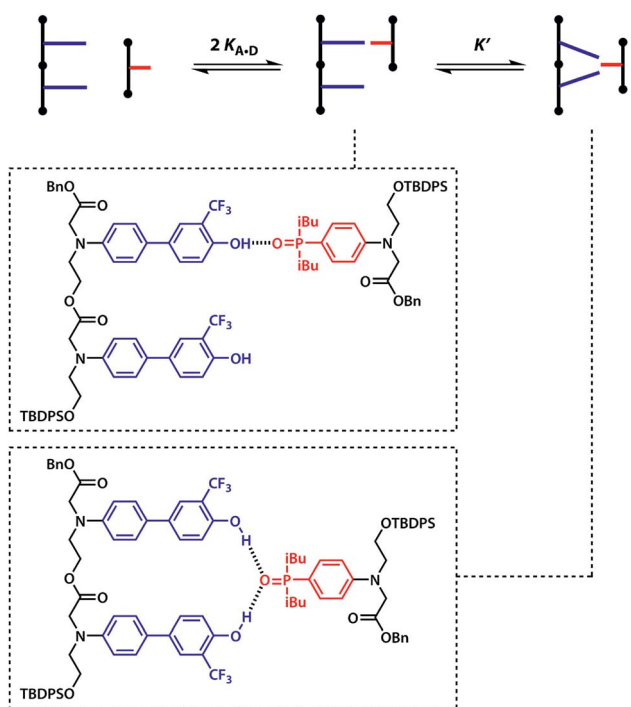
Fig. 8 Pathway towards formation of a double hydrogen bond between **DD** 2-mer and **A**. $K_{\mathbf{A}\cdot\mathbf{D}}$ is the intermolecular association constant for formation a single **A**·**D** hydrogen bond, K' is the association constant for the interaction of the second phenol with the same phosphine oxide. Statistical factors represent the degeneracy of the structures involved.

W. Zhang, D. L. Gerloff, Z. Huang, W. Tan and S. A. Benner, *J. Am. Chem. Soc.*, 2015, **137**, 6734–6737.

8 G. F. Joyce and J. W. Szostak, *Cold Spring Harbor Perspect. Biol.*, 2018, **10**, a034801.

9 D. H. Appella, *Curr. Opin. Chem. Biol.*, 2009, **13**, 687–696.

10 H. Isobe, T. Fujino, N. Yamazaki, M. Guillot-Nieckowski and E. Nakamura, *Org. Lett.*, 2008, **10**, 3729–3732.

11 K. Burgess, R. A. Gibbs, M. L. Metzker and R. Raghavachari, *J. Chem. Soc., Chem. Commun.*, 1994, 915.

12 J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, *Nature*, 1990, **343**, 33–37.

13 B. A. Schweitzer and E. T. Kool, *J. Org. Chem.*, 1994, **59**, 7238–7242.

14 H. Liu, J. Gao, S. R. Lynch, Y. D. Saito, L. Maynard and E. T. Kool, *Science*, 2003, **302**, 868–871.

15 J. C. Delaney, J. Gao, H. Liu, N. Shrivastav, J. M. Essigmann and E. T. Kool, *Angew. Chem., Int. Ed.*, 2009, **48**, 4524–4527.

16 A. Eschenmoser, *Origins Life Evol. Biospheres*, 1997, **27**, 535–553.

17 S. Obika, D. Nanbu, Y. Hari, K.-i. Morio, Y. In, T. Ishida and T. Imanishi, *Tetrahedron Lett.*, 1997, **38**, 8735–8738.

18 A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, *Tetrahedron*, 1998, **54**, 3607–3630.

19 A. Aerschot Van, I. Verheggen, C. Hendrix and P. Herdewijn, *Angew. Chem., Int. Ed.*, 1995, **34**, 1338–1339.

20 K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy and A. Eschenmoser, *Science*, 2000, **290**, 1347–1351.

21 L. Zhang, A. Peritz and E. Meggers, *J. Am. Chem. Soc.*, 2005, **127**, 4174–4175.

22 A. B. Gerber and C. J. Leumann, *Chem.–Eur. J.*, 2013, **19**, 6990–7006.

23 R. Kramer, J. M. Lehn and A. Marquis-Rigault, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5394–5398.

24 P. N. Taylor and H. L. Anderson, *J. Am. Chem. Soc.*, 1999, **121**, 11538–11545.

25 Y. Tanaka, H. Katagiri, Y. Furusho and E. Yashima, *Angew. Chem., Int. Ed.*, 2005, **44**, 3867–3870.

26 H. Ito, Y. Furusho, T. Hasegawa and E. Yashima, *J. Am. Chem. Soc.*, 2008, **130**, 14008–14015.

27 R. Amemiya, N. Saito and M. Yamaguchi, *J. Org. Chem.*, 2008, **73**, 7137–7144.

28 A. P. Bisson, F. J. Carver, D. S. Eggleston, R. C. Haltiwanger, C. A. Hunter, D. L. Livingstone, J. F. McCabe, C. Rotger and A. E. Rowan, *J. Am. Chem. Soc.*, 2000, **122**, 8856–8868.

29 H. Gong and M. J. Krische, *J. Am. Chem. Soc.*, 2005, **127**, 1719–1725.

30 Y. Yang, Z. Y. Yang, Y. P. Yi, J. F. Xiang, C. F. Chen, L. J. Wan and Z. G. Shuai, *J. Org. Chem.*, 2007, **72**, 4936–4946.

31 A. E. Stross, G. Iadevaia, D. Núñez-Villanueva and C. A. Hunter, *J. Am. Chem. Soc.*, 2017, **139**, 12655–12663.

32 C. A. Hunter and H. L. Anderson, *Angew. Chem., Int. Ed.*, 2009, **48**, 7488–7499.

33 P. Motloch and C. A. Hunter, *Adv. Phys. Org. Chem.*, 2016, **50**, 77–118.

34 D. Núñez-Villanueva and C. A. Hunter, *Chem. Sci.*, 2017, **8**, 206–213.

35 J. A. Swain, G. Iadevaia and C. A. Hunter, *J. Am. Chem. Soc.*, 2018, **140**, 11526–11536.

36 A. E. Stross, G. Iadevaia and C. A. Hunter, *Chem. Sci.*, 2016, **7**, 94–101.

37 G. Iadevaia, D. Núñez-Villanueva, A. E. Stross and C. A. Hunter, *Org. Biomol. Chem.*, 2018, 4183–4190.

38 G. Iadevaia, A. E. Stross, A. Neumann and C. A. Hunter, *Chem. Sci.*, 2016, **7**, 1760–1767.

39 D. Núñez-Villanueva, G. Iadevaia, A. E. Stross, M. A. Jinks, J. A. Swain and C. A. Hunter, *J. Am. Chem. Soc.*, 2017, **139**, 6654–6662.

40 C. S. Calero, J. Farwer, E. J. Gardiner, C. A. Hunter, M. Mackey, S. Scuderi, S. Thompson and J. G. Vinter, *Phys. Chem. Chem. Phys.*, 2013, **15**, 18262.

41 J. S. Moore and R. B. Prince, in *Materials Science and Technology: A Comprehensive Treatment*, ed. R. W. Cahn, P. Haasen and E. J. Kramer, Wiley-VCH Verlag GmbH, Weinheim, Germany, 1999, pp. 11–36.

42 O. Kuisle, E. Quiñoá and R. Riguera, *J. Org. Chem.*, 1999, **64**, 8063–8075.

43 M. Tsakos, E. S. Schaffert, L. L. Clement, N. L. Villadsen and T. B. Poulsen, *Nat. Prod. Rep.*, 2015, **32**, 605–632.

44 B. Huang and M. E. Hermes, *J. Polym. Sci., Part A: Polym. Chem.*, 1995, **33**, 1419–1429.

45 U. D. Lengweiler, M. G. Fritz and D. Seebach, *Helv. Chim. Acta*, 1996, **79**, 670–701.

46 C. M. Krell and D. Seebach, *Eur. J. Org. Chem.*, 2000, **2000**, 1207–1218.

47 J. B. Williams, T. M. Chapman and D. M. Hercules, *Macromolecules*, 2003, **36**, 3898–3908.

48 R. M. Weiss, E. M. Jones, D. E. Shafer, R. M. Stayshich and T. Y. Meyer, *J. Polym. Sci., Part A: Polym. Chem.*, 2011, **49**, 1847–1855.

49 J. Li, R. M. Stayshich and T. Y. Meyer, *J. Am. Chem. Soc.*, 2011, **133**, 6910–6913.

50 R. M. Stayshich and T. Y. Meyer, *J. Polym. Sci., Part A: Polym. Chem.*, 2008, **46**, 4704–4711.

51 M. A. Washington, D. J. Swiner, K. R. Bell, M. V. Fedorchak, S. R. Little and T. Y. Meyer, *Biomaterials*, 2017, **117**, 66–76.

52 *Schrödinger Release 2016-4: MacroModel*, 2016.