

Cite this: *Chem. Sci.*, 2016, 7, 94

Cooperative duplex formation by synthetic H-bonding oligomers†

Alexander E. Stross, Giulia Iadevaia and Christopher A. Hunter*

A series of flexible oligomers equipped with phenol H-bond donors and phosphine oxide H-bond acceptors have been synthesised using reductive amination chemistry. H-bonding interactions between complementary oligomers leads to the formation of double-stranded complexes which were characterised using NMR titrations and thermal denaturation experiments. The stability of the duplex increases by one order of magnitude for every H-bonding group added to the chain. Similarly, the enthalpy change for duplex assembly and the melting temperature for duplex denaturation both increase with increasing chain length. These observations indicate that H-bond formation along the oligomers is cooperative despite the flexible backbone, and the effective molarity for intramolecular H-bond formation (14 mM) is sufficient to propagate the formation of longer duplexes using this approach. The product *K* EM, which is used to quantify chelate cooperativity is 5, which means that each H-bond is more than 80% populated in the assembled duplex. The modular design of these oligomers represents a general strategy for the design of synthetic information molecules that could potentially encode and replicate chemical information in the same way as nucleic acids.

Received 10th September 2015
Accepted 15th October 2015

DOI: 10.1039/c5sc03414k

www.rsc.org/chemicalscience

Introduction

Formation of intermolecular complexes between two linear polymers equipped with complementary recognition sites is the molecular basis for life on Earth. As pointed out by Watson and Crick in 1953, this supramolecular architecture provides a robust mechanism for encoding molecular information and for replication of this information through template-directed synthesis.¹ Although synthetic polymers bearing side-chain recognition sites have been reported, these systems are generally of ill-defined chemical and supramolecular structure and lack the control over length and sequence found in biological polymers.² Stepwise synthesis of short oligomers provides access to more well-defined systems, and a number of synthetic supramolecular systems that form duplex structures have been reported.³ However, the recognition sites in these compounds are usually integrated into the backbone of the oligomer, which limits their versatility because a rather precise matching of molecular geometries is required. In contrast, the modular architecture of nucleic acids appears to be remarkably robust with respect to chemical manipulation.⁴ Re-engineering experiments on nucleic acids show that it is possible to substitute the sugar,⁵ the phosphate,^{6,7} or the bases,⁸ for very different

chemical components and still retain the duplex forming properties of the system. These results suggest that attempts to develop new synthetic information molecules would benefit from a modular strategy that would allow independent optimisation of the different components of the monomer units.⁹ This report describes the first steps towards such a system with the synthesis of complementary oligomers that form stable duplexes in organic solvents.

Approach

Fig. 1 illustrates the DNA molecule stripped down to the basic constituents required for duplex assembly and template-directed synthesis: a recognition-based pairing system (blue), chemistry for the synthesis of oligomers (red), and a linker that allows the recognition sites on the two polymeric backbones to reach one another (black). This basic blueprint provides guidelines for the construction of a wide range of different types of molecule that could function in the same way as nucleic acids under appropriate conditions. Nucleic acids operate in water, where stacking of the hydrophobic bases is the main driving force for duplex assembly.¹⁰ This supramolecular architecture requires mutual complementarity of the chemistry, linker and recognition elements, so that the duplex is tightly packed. Here, we target synthetic duplexes held together by H-bonding interactions in organic solvents, so that a much looser structure is possible with fewer constraints on the components. The modular design in Fig. 1 will allow for the future optimization of the properties of these components independently.

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

† Electronic supplementary information (ESI) available: Detailed experimental procedures with spectroscopic characterization data, ³¹P and ¹H NMR titration spectra, binding isotherms, limiting chemical shifts for free and bound states, and thermal denaturation spectra. See DOI: 10.1039/c5sc03414k



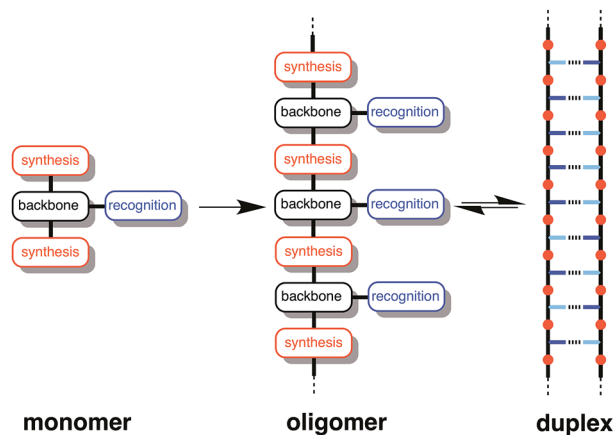


Fig. 1 Blueprint for a synthetic information molecule. The key design components are the covalent chemistry used for synthesis (red), the non-covalent chemistry used for recognition (blue), and the backbone linker that determines the geometric complementarity of the two chains (black).

The two key parameters that determine the efficiency of duplex formation between two oligomers are the association constant for the intermolecular interaction between two recognition sites (K) and the effective molarity for formation of the corresponding intramolecular interactions as the duplex zips up (EM).¹¹ Fig. 2 shows the steps involved in the assembly of a duplex from two complementary oligomers. The first step is an intermolecular interaction with an association constant K . Formation of subsequent interactions on the intramolecular pathway leading to duplex formation requires that the product KEM is greater than unity. There is a competing intermolecular assembly channel, and the ratio of EM to the operating concentration (c) determines whether uncontrolled interactions

between multiple partners leads to aggregation and formation of cross-linked supramolecular networks rather than a well-defined duplex. Thus the criteria for high fidelity duplex formation are $KEM \gg 1$ and $EM \gg c$.

Qualitatively, EM increases with geometric complementarity and decreases with conformational flexibility, but it is still difficult to make quantitative predictions of the value of EM as a function of chemical and supramolecular structure. We have shown that for supramolecular architectures of varying complementarity and flexibility, the values of EM for intramolecular interactions fall in a surprisingly narrow window (10–1000 mM)¹² compared with covalent EM s, where changes in chemical structure can lead to variations of several orders magnitude.¹³ Although there is a trade-off between conformational flexibility and geometric complementarity, the margin for error in supramolecular design is much higher if more flexible molecules are used, and the associated decrease in EM is not dramatic.¹⁴ The consequence is that if $K \gg 100 \text{ M}^{-1}$ for the recognition modules in Fig. 1, it should be possible to assemble duplex structures without the need for careful design of the linker module, provided the backbones are sufficiently flexible to ensure geometric complementarity.

This K criterion dictates the choice of recognition modules. The simplest possible recognition system is a single H-bond between a H-bond donor (D) and a H-bond acceptor (A), *i.e.* a two-letter alphabet rather than the four-letter alphabet used in nucleic acids. Provided the backbone does not have any polar functional groups that could compete for the H-bonding interactions, use of simple H-bonding modules will ensure selectivity of the recognition-based pairing system: D will only pair with A; D will not pair with D, and A will not pair with A. Phosphine oxides and phenols form exceptionally stable H-bonds in toluene (10^2 to 10^3 M^{-1}),^{14g} and so these functional

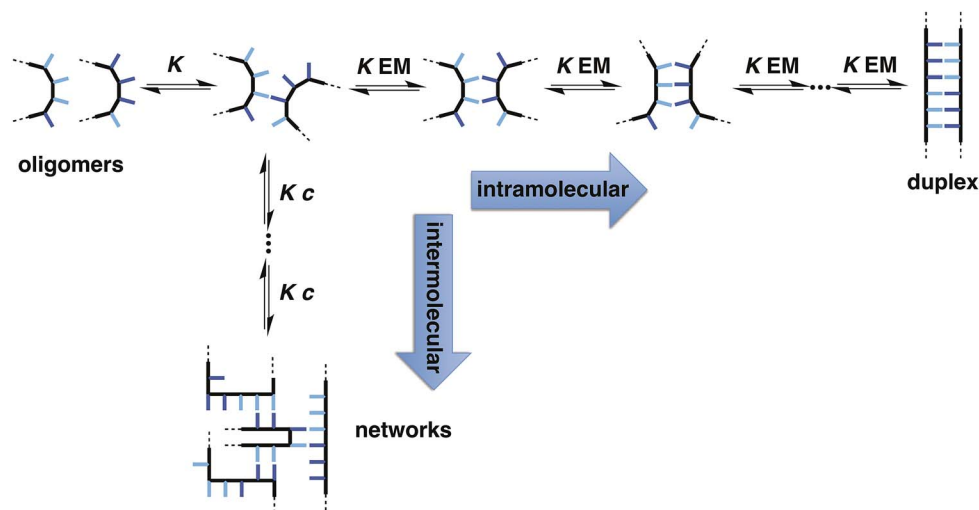


Fig. 2 Recognition-directed assembly of a duplex between two complementary oligomers (intramolecular channel) competes with uncontrolled assembly of supramolecular networks (intermolecular channel). K is the association constant for formation of an intermolecular interaction between two recognition elements (blue bars), EM is the effective molarity for formation of an intramolecular interaction in the duplex structure, and c is the operating concentration. For specific oligomers, the equilibrium constants have additional statistical factors reflecting the differences in the degeneracy of the complexes.



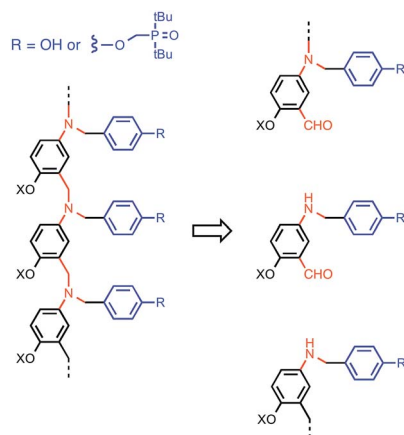


Fig. 3 Design of a synthetic information molecule (cf. Fig. 1), which can be synthesised by reductive amination of aminoaldehyde monomers (red) equipped with a phenol or a phosphine oxide recognition module (blue). X is a site for attachment of solubilising groups.

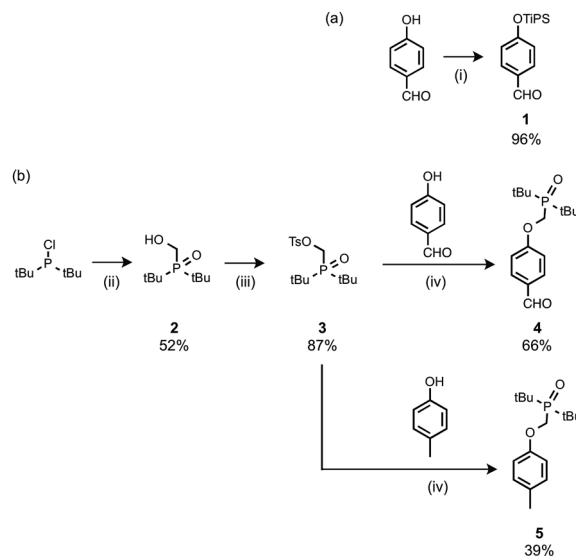
groups were chosen for the recognition modules in the oligomer design shown in Fig. 3. The backbone of the oligomer in Fig. 3 contains no H-bond donors, and the backbone aniline nitrogen and aromatic ether oxygen atoms are both weak H-bond acceptors (the H-bond acceptor parameter $\beta \approx 4$ and 3 respectively) compared with phosphine oxide ($\beta \approx 10$).¹⁵ Competition of the backbone with the recognition sites will therefore be negligible. The ether oxygen provides a convenient point for addition of side-chains to control the solubility of the oligomers (X in Fig. 3). Here we report the synthesis of phenol and phosphine oxide oligomers from the corresponding aminoaldehyde monomers and NMR experiments to characterise assembly of the corresponding duplexes.

Results and discussion

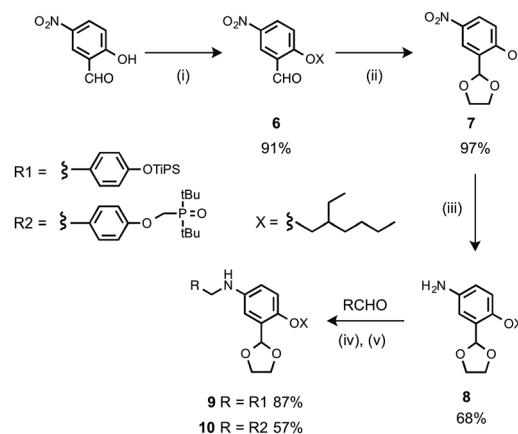
Synthesis

The recognition modules were derived from *p*-hydroxybenzaldehyde (Scheme 1). The phenol hydroxyl group was protected as the silyl ether **1** for the synthesis of oligomers.¹⁶ Oxidation of di-*t*-butylchlorophosphine in the presence of formaldehyde gave **2**, which was treated with *p*-toluenesulfonyl chloride to give **3**.^{14e} Alkylation of *p*-hydroxybenzaldehyde with **3** gave the phosphine oxide H-bond acceptor unit, **4**. Phosphine oxide **5** was prepared by alkylation of *p*-cresol with **3** (Scheme 1b). **5** is a single H-bond acceptor compound (A), which was used with *p*-cresol, the corresponding single H-bond donor compound (D), for determining the strength of an intermolecular phenol–phosphine oxide H-bond.

The monomer units for oligomer synthesis were prepared according to Scheme 2. 2-Hydroxy-4-nitrobenzaldehyde was alkylated with 2-ethylhexyl bromide to give **6**, which was protected as the corresponding acetal, **7**, by condensation with ethylene glycol. Reduction of **7** gave the primary aniline **8**. Formation of the corresponding imines using **1** or **4** and reduction with sodium borohydride gave the required monomers **9** and **10**.



Scheme 1 (i) Tri-*i*-propylsilyl chloride, imidazole; (ii) formaldehyde, aq. HCl; (iii) *p*-toluenesulfonyl chloride, NEt₃; (iv) Cs₂CO₃.



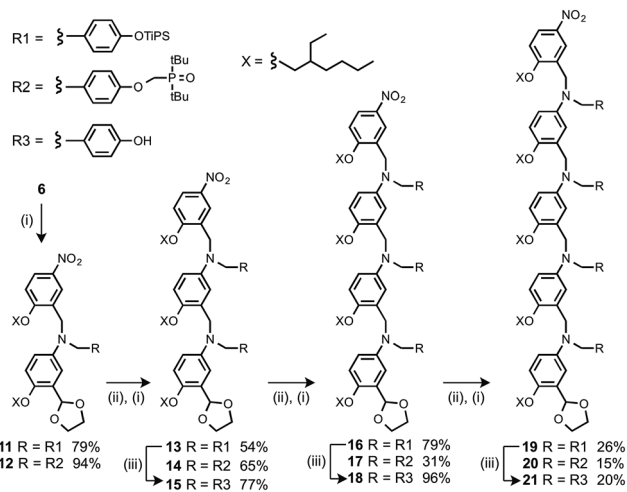
Scheme 2 (i) 2-Ethylhexyl bromide, K₂CO₃; (ii) *p*-toluenesulfonic acid, ethylene glycol; (iii) H₂, Pd/C; (iv) heat; (v) NaBH₄.

Oligomers were prepared by building a chain starting from **6** (Scheme 3). Reductive amination of **6** with the monomer units (**9** or **10**) using NaBH(OAc)₃ gave the 1-mer chains, **11** and **12**.¹⁷ The acetal groups were then deprotected using aqueous HCl, and the resulting aldehydes were coupled with a further monomer unit to give the 2-mers, **13** and **14**. **14** is the double H-bond acceptor AA, and deprotection of **13** using tetra-*n*-butylammonium fluoride (TBAF) gave **15**, the corresponding double H-bond donor DD. The same deprotection-coupling sequence shown in Scheme 3 was used to prepare the 3-mers, **17** (AAA) and **18** (DDD), and the 4-mers, **20** (AAAA) and **21** (DDDD).

Binding studies

Complexation of length-complementary oligomers was studied using ¹H and ³¹P NMR titration experiments in toluene. The chemical shifts of the ³¹P and ¹H NMR signals observed for the



Scheme 3 (i) 9 or 10, NaBH(OAc)₃; (ii) aq. HCl; (iii) TBAF.

oligomers are similar to the values observed for the monomer units, which indicates that there is no significant folding of the free oligomers and that the H-bonding sites are displayed in an accessible environment on the flexible backbone. The ¹H NMR spectra of the oligomers are complicated with many overlapping signals, due to the large number of chemically similar protons. However, the ³¹P NMR spectra are much simpler, and large changes in chemical shift were observed on titration of the donor oligomers (D_N, N = 1–4) into the acceptor oligomers (A_N). All titration data fit well to 1 : 1 binding isotherms, and the association constants and limiting changes in ³¹P NMR chemical shift are reported in Table 1 (see ESI† for details). The complexation-induced changes in ³¹P NMR chemical shift are similar for all of the signals in all of the complexes (+4–5 ppm). A large positive change in ³¹P NMR chemical shift is indicative of H-bond formation¹⁸ and the results in Table 1 imply that all of the phosphine oxide groups are fully bound in all of the complexes. In other words, fully H-bonded duplex structures are formed (Fig. 4). Although the ¹H NMR data are more difficult to interpret, the patterns of complexation-induced changes in chemical shift are also similar for all of the oligomers, which suggests that the duplexes have similar structures (see ESI†). The oligomer chains have directionality, so there are two possible structures for the duplexes: parallel and antiparallel

Table 1 Association constants (K_N), effective molarities (EM) and limiting complexation-induced changes in ³¹P NMR chemical shift for formation of duplexes in toluene at 298 K^a

Complex	$\log K_N/M^{-1}$	EM/mM	K EM	$\Delta\delta/\text{ppm}$
A·D	2.5 ± 0.1	—	—	4.9
AA·DD	3.3 ± 0.1	8 ± 3	3 ± 1	5.3, 5.3
AAA·DDD	4.2 ± 0.1	14 ± 1	5 ± 1	4.8, 4.9, 5.1
AAAA·DDDD	5.4 ± 0.5	21 ± 8	7 ± 1	3.8, 5.1, 5.1, 5.1

^a Each titration was repeated at least twice, and the average values are reported with errors (in brackets) at the 95% confidence limit.

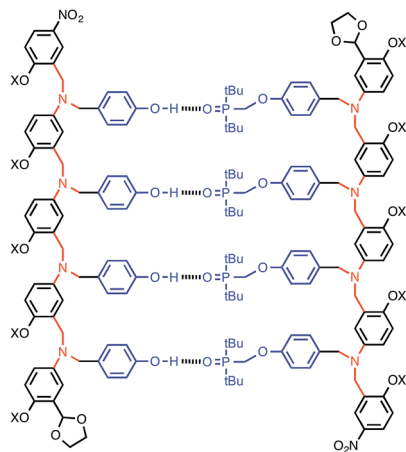
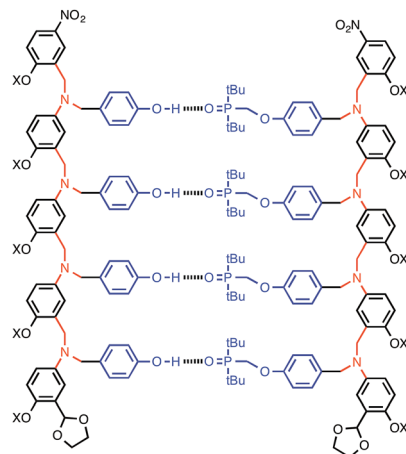


Fig. 4 The AAAA·DDDD duplex structure: parallel or antiparallel structures are possible.

(Fig. 4). The NMR data do not allow us to determine whether one of these is preferred over the other.

Fig. 5 shows the association constants for duplex formation (K_N) plotted as a function of the number of H-bonding sites N : the data fit well to a straight line, and there is a uniform increase in association constant of an order of magnitude for each additional H-bond formed (eqn (1)).

$$\log K_N = 1.0N + 1.5 \quad (1)$$

This relationship implies that the flexible backbone is able to adopt an ideal geometry for duplex assembly, so that the effective molarity (EM) for sequential H-bond formation in zipping up the duplex is constant. The association constant for duplex formation between two oligomers with N interaction sites (K_N) can therefore be expressed as the product of stepwise equilibrium constants (*cf.* Fig. 2) for assembly of the duplex using a single value of EM (eqn (2)).

$$K_N = 2K^N EM^{N-1} \quad (2)$$



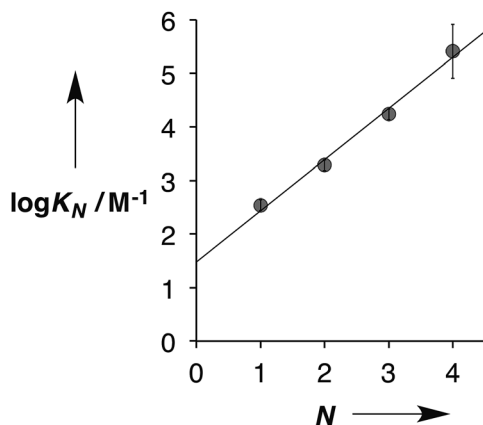


Fig. 5 Relationship between the association constants for duplex formation in toluene at 298 K (K_N) and the number of recognition modules in the oligomer (N). The best fit straight line is shown ($r^2 = 0.99$).

where the statistical factor of 2 assumes that the parallel and antiparallel structures in Fig. 4 are equally populated.

This relationship was used to calculate the values of EM in Table 1. Using the A·D association constant as K in eqn (2), gives an average value of 14 mM for the EM for duplex assembly. The chelate cooperativity associated with duplex assembly is quantified by the product K EM, and the data in Table 1 give an average value of 5. This value implies that duplex assembly is cooperative and that it should be possible to propagate the assembly of longer structures using this architecture. However, each H-bond is about 80% bound on average, so the duplex structures are dynamic with some fraying of the interaction sites.

Thermal denaturation experiments

Thermal denaturation data were measured by making 1 : 1 solutions of length-complementary oligomers at 1 mM concentrations in toluene and measuring ^{31}P NMR spectra at temperatures from 228 to 373 K. At low temperatures, the ^{31}P NMR signals moved to higher chemical shift, indicative of an increase in the extent of H-bond formation (Fig. 6).¹⁸ At higher temperatures, a decrease in ^{31}P NMR chemical shift was observed, indicative of disruption of the H-bonding interactions. When a sample of phosphine oxide **5** was monitored over the same temperature range in the absence of a H-bond donor, the variation in chemical shift (≈ 1 ppm) was small relative to the changes observed for the H-bonded complexes (>4 ppm), confirming that the temperature dependence of the ^{31}P spectra in Fig. 6 is related to duplex assembly and denaturation. The signals in the ^{31}P NMR spectra of the AAAA·DDDD duplex become very broad and overlapped with changes in temperature, so it was not possible to extract an accurate melting profile for this system. However, the data for AAAA·DDDD are qualitatively consistent with a more stable duplex that melts at higher temperatures than the other three complexes (see ESI†).

The thermal denaturation experiment can be used to extract thermodynamic parameters for duplex formation by fitting the

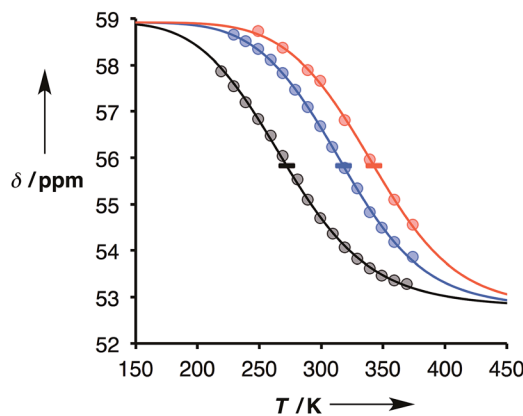


Fig. 6 Experimental ^{31}P NMR chemical shift plotted as a function of temperature for 1 : 1 mixtures (1 mM) of A·D (black), AA·DD (blue), and AAA·DDD (red) in toluene. The lines are the best fit to eqn (9) (total rmsd < 0.2 ppm), and the horizontal bars show the transition melting temperatures, $T_{m,N}$.

data to a two-state model,¹⁹ which assumes only duplex and single strands are present. At a given temperature T , the equilibrium constant for formation of a duplex between two oligomers with N binding sites, $K_N(T)$, is given by eqn (3).

$$K_N(T) = \frac{[D_N \cdot A_N]}{[D_N][A_N]} \quad (3)$$

where $[D_N]$ and $[A_N]$ are the concentrations of single strand oligomers bearing donor groups and acceptor H-bonding groups respectively, and $[D_N \cdot A_N]$ is concentration of the duplex.

If the two oligomers are present in equal concentrations, the total concentration of all free strands can be written as c and the total fraction of all bound species as α (eqn (4)).

$$K_N(T) = \frac{2\alpha}{c(1-\alpha)^2} \quad (4)$$

Rearranging eqn (4) gives an expression for α in terms of $K_N(T)$ and c (eqn (5)).

$$\alpha = \frac{K_N(T)c + 1 - \sqrt{2K_N(T)c + 1}}{K_N(T)c} \quad (5)$$

The observed chemical shift for a two state system in fast exchange, δ , can be expressed in terms of α by eqn (6).

$$\delta = \delta_f + (\delta_b - \delta_f)\alpha \quad (6)$$

where δ_b and δ_f are the chemical shifts of the duplex and single strand states.

Assuming that the enthalpy and entropy of duplex formation are temperature independent, and the change in heat capacity between free and bound states is zero, the temperature dependence of $K_N(T)$ can be analysed using the integrated form of the van't Hoff equation (eqn (7)).¹⁹



$$\ln \left\{ \frac{K_N(T)}{K_{m,N}(T)} \right\} = -\frac{\Delta H_N^\circ}{R} \left(\frac{1}{T} - \frac{1}{T_{m,N}} \right) \quad (7)$$

where $K_N(T_{m,N})$ is the association constant for formation of the $D_N \cdot A_N$ duplex at the transition melting temperature, $T_{m,N}$, ΔH_N° is the enthalpy change for formation of the $D_N \cdot A_N$ duplex, and R is the gas constant.

At the transition melting temperature $\alpha = 0.5$, and using this value in eqn (4) gives $K_N(T_{m,N}) = 4/c$. Substituting for $K_N(T_{m,N})$ in eqn (7) and rearranging gives eqn (8), the temperature dependence of the association constant $K_N(T)$.

$$K_N(T) = \frac{4}{c} e^{-\left\{ \frac{\Delta H_N^\circ}{R} \left(\frac{1}{T} - \frac{1}{T_{m,N}} \right) \right\}} \quad (8)$$

Combining eqn (5), (6) and (8), allows us to express the observed chemical shift in terms of T .

$$\delta = \delta_f + (\delta_b - \delta_f) \times \left(\frac{1 + 4e^{-\left\{ \frac{\Delta H_N^\circ}{R} \left(\frac{1}{T} - \frac{1}{T_{m,N}} \right) \right\}} - \sqrt{1 + 8e^{-\left\{ \frac{\Delta H_N^\circ}{R} \left(\frac{1}{T} - \frac{1}{T_{m,N}} \right) \right\}}}}{4e^{-\left\{ \frac{\Delta H_N^\circ}{R} \left(\frac{1}{T} - \frac{1}{T_{m,N}} \right) \right\}}} \right) \quad (9)$$

The thermal denaturation data for duplexes A·D, AA·DD and AAA·DDD were analysed by optimising the values of δ_b , δ_f , ΔH_N° and $T_{m,N}$ to minimise the difference between the experimentally measured values of δ and the values calculated using eqn (9). The values of δ_b and δ_f could depend on N , but the limiting chemical shifts determined by fitting the titration data recorded at 298 K are similar for all four duplexes ($\delta_f = 53.6$ – 53.9 ppm, and $\delta_b = 57.9$ – 59.0 ppm, see ESI†). We therefore fit the thermal denaturation data using the same value of δ_b and δ_f for all of the duplexes, and the results ($\delta_f = 52.8$ ppm, and $\delta_b = 58.9$ ppm) are consistent with the titration data. Fig. 6 shows the lines of best fit calculated using eqn (9) (solid lines) for each thermal denaturation data set, and Table 2 shows the values of the fitted parameters along with the values of ΔS_N° and ΔG_N° calculated from $\log K_N$ at 298 K.

The values of $\log K_N$ at 298 K in Table 2 are similar to the corresponding values from the titration data (Table 1), which suggests that the assumptions used in fitting the thermal

denaturation data are reasonable. The experimental data in Fig. 6 clearly follow the theoretical sigmoidal curve corresponding to the melting transition. The free-bound transitions occur over narrower temperature ranges with increasing N , and the point of inflection occurs at higher temperatures. These visual observations are confirmed by the calculated parameters shown in Table 2, which show that the increased binding constants observed for higher values of N are associated with an increase in the enthalpy change on duplex formation and with an increase in the transition melting temperature. These features are characteristic of cooperative interactions between the H-bonding sites along the duplex.¹⁰

Conclusions

We present a general strategy for the design of synthetic information molecules that could potentially encode and replicate chemical information in the same way as nucleic acids. A series of H-bond donor (phenol) and H-bond acceptor (phosphine oxide) oligomers (D_N and A_N , $N = 1$ – 4) have been synthesised using reductive amination chemistry. The assembly of duplexes, $A_N \cdot D_N$, was characterised using NMR titrations and thermal denaturation experiments in toluene. The stability of the duplex increases by one order of magnitude for every H-bonding group added to the chain. Similarly, the enthalpy change for duplex assembly and the melting temperature for duplex denaturation both increase with increasing chain length. Although the oligomers have a relatively flexible backbone, this lack of preorganisation does not significantly impede assembly of the duplex, and H-bond formation along the oligomers is cooperative. The effective molarity for intramolecular H-bond formation ($EM = 14$ mM) appears to be sufficient to propagate the formation of long duplexes using this approach. The product $K EM$, which is used to quantify chelate cooperativity is 5, which means that each H-bond is more than 80% populated in the assembled duplex. The modular design of these compounds should allow us to explore variations in the recognition and backbone linker modules to optimise the properties of the system for selective recognition of mixed sequence oligomers and template-directed synthesis.

Acknowledgements

We thank the EPSRC and ERC for funding.

Notes and references

- 1 J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 964.
- 2 (a) S. Połowiński, *Prog. Polym. Sci.*, 2002, **27**, 537; (b) P. K. Lo and H. F. Sleiman, *J. Am. Chem. Soc.*, 2009, **131**, 4182; (c) T. Terashima, T. Mes, T. F. A. De Greef, M. A. J. Gillissen, P. Besenius, A. R. A. Palmans and E. W. Meijer, *J. Am. Chem. Soc.*, 2011, **133**, 4742; (d) C. R. South and M. Weck, *Macromolecules*, 2007, **40**, 1386; (e) M. Szwarc, *J. Polym. Sci.*, 1954, **13**, 317; (f) R. McHale, J. P. Patterson, P. B. Zetterlund and R. K. O'Reilly, *Nat. Chem.*, 2012, **4**, 491; (g) Y. Kang, A. Lu, A. Ellington, M. C. Jewett and

Table 2 Thermodynamic parameters for formation of H-bonded duplexes in toluene determined using ³¹P NMR thermal denaturation data

N	$T_{m,N}/K$	$\Delta H_N^\circ/kJ mol^{-1}$	$T\Delta S_N^\circ (298)/kJ mol^{-1}$	$\Delta G_N^\circ (298)/kJ mol^{-1}$	$\log K_N (298)/M^{-1}$
1	272	−26	−9	−16	2.9
2	318	−37	−17	−21	3.7
3	342	−42	−18	−24	4.3



- R. K. O'Reilly, *ACS Macro Lett.*, 2013, **2**, 581; (h) J. Ferguson and S. A. O. Shah, *Eur. Polym. J.*, 1968, **4**, 343; (i) J. Smid, Y. Y. Tan and G. Challa, *Eur. Polym. J.*, 1983, **19**, 853; (j) J. Smid, Y. Y. Tan and G. Challa, *Eur. Polym. J.*, 1984, **20**, 887; (k) J. Smid, Y. Y. Tan and G. Challa, *Eur. Polym. J.*, 1984, **20**, 1095; (l) J. Smid, J. C. Speelman, Y. Y. Tan and G. Challa, *Eur. Polym. J.*, 1985, **21**, 141; (m) J. Smid, Y. Y. Tan, G. Challa and W. R. Hagen, *Eur. Polym. J.*, 1985, **21**, 757.
- 3 (a) R. Kramer, J.-M. Lehn and A. Marquis-Rigault, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5394; (b) A. Marquis, V. Smith, J. Harrowfield, J.-M. Lehn, H. Herschbach, R. Sanvito, E. Leize-Wagner and A. van Dorsselaer, *Chem.–Eur. J.*, 2006, **12**, 5632; (c) V. Berl, I. Huc, R. G. Khoury, M. J. Krische and J.-M. Lehn, *Nature*, 2000, **407**, 720; (d) V. Berl, I. Huc, R. G. Khoury and J.-M. Lehn, *Chem.–Eur. J.*, 2001, **7**, 2810; (e) H. L. Anderson, *Inorg. Chem.*, 1994, **33**, 972; (f) P. N. Taylor and H. L. Anderson, *J. Am. Chem. Soc.*, 1999, **121**, 11538; (g) J. Sánchez-Quesada, C. Seel, P. Prados, J. de Mendoza, I. Dalcol and E. Giralt, *J. Am. Chem. Soc.*, 1996, **118**, 277; (h) 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; (i) A. P. Bisson and C. A. Hunter, *Chem. Commun.*, 1996, **15**, 1723; (j) B. Gong, Y. Yan, H. Zeng, E. Skrzypczak-Jankun, Y. W. Kim, J. Zhu and H. Ickes, *J. Am. Chem. Soc.*, 1999, **121**, 5607; (k) B. Gong, *Synlett*, 2001, **5**, 582; (l) H. Zeng, R. S. Miller, R. A. Flowers and B. Gong, *J. Am. Chem. Soc.*, 2000, **122**, 2635; (m) H. Zeng, H. Ickes, R. A. Flowers and B. Gong, *J. Org. Chem.*, 2001, **66**, 3574; (n) B. Gong, *Polym. Int.*, 2007, **56**, 436; (o) B. Gong, *Acc. Chem. Res.*, 2012, **45**, 2077; (p) 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; (q) W.-J. Chu, Y. Yang and C.-F. Chen, *Org. Lett.*, 2010, **12**, 3156; (r) W.-J. Chu, J. Chen, C.-F. Chen, Y. Yang and Z. Shuai, *J. Org. Chem.*, 2012, **77**, 7815; (s) E. A. Archer and M. J. Krische, *J. Am. Chem. Soc.*, 2002, **124**, 5074; (t) H. Gong and M. J. Krische, *J. Am. Chem. Soc.*, 2005, **127**, 1719; (u) C. A. Hunter, P. S. Jones, P. M. N. Tiger and S. Tomas, *Chem. Commun.*, 2003, **14**, 1642; (v) G. K. Mittapalli, Y. M. Osornio, M. A. Guerrero, K. R. Reddy, R. Krishnamurthy and A. Eschenmoser, *Angew. Chem., Int. Ed.*, 2007, **46**, 2478.
- 4 (a) A. Eschenmoser, *Science*, 1999, **284**, 2118; (b) S. A. Benner, *Acc. Chem. Res.*, 2004, **37**, 784; (c) S. A. Benner, F. Chen and Z. Yang, *Synthetic Biology, Tinkering Biology, and Artificial Biology: A Perspective from Chemistry*, in *Chemical Synthetic Biology*, ed. P. L. Luisi and C. Chiarabelli, John Wiley & Sons, Ltd, Chichester, UK, 2011, pp. 69–106; (d) S. A. Benner, *Biological Theory*, 2013, **8**, 357; (e) C. Wilson and A. D. Keefe, *Curr. Opin. Chem. Biol.*, 2006, **10**, 607; (f) D. H. Appella, *Curr. Opin. Chem. Biol.*, 2009, **13**, 687; (g) E. T. Kool, *Curr. Opin. Chem. Biol.*, 2000, **4**, 602.
- 5 (a) K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy and A. Eschenmoser, *Science*, 2000, **290**, 1347; (b) A. Aerschot van, I. Verheggen, C. Hendrix and P. Herdewijn, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 1338; (c) D. Renneberg and C. J. Leumann, *J. Am. Chem. Soc.*, 2002, **124**, 5993; (d) D. A. Braasch and D. R. Corey, *Chem. Biol.*, 2001, **8**, 1; (e) S. K. Singh, A. A. Koshkin and J. Wengel, *Chem. Commun.*, 1998, 455; (f) H. V. Nguyen, Z.-Y. Zhao, A. Sallustrau, S. L. Horswell, L. Male, A. Mulas and J. H. R. Tucker, *Chem. Commun.*, 2012, **48**, 12165; (g) L. Zhang, A. Peritz and E. Meggers, *J. Am. Chem. Soc.*, 2005, **127**, 4174; (h) M. K. Schlegel, A. E. Peritz, K. Kittigowittana, L. Zhang and E. Meggers, *ChemBioChem*, 2007, **8**, 927; (i) P. Karri, V. Punna, K. Kim and R. Krishnamurthy, *Angew. Chem., Int. Ed.*, 2013, **52**, 5840.
- 6 (a) S. A. Benner and D. Hutter, *Bioorg. Chem.*, 2002, **30**, 62; (b) Z. Huang, K. C. Schneider and S. A. Benner, *J. Org. Chem.*, 1991, **56**, 3869; (c) Z. Huang and S. A. Benner, *J. Org. Chem.*, 2002, **67**, 3996; (d) C. Richert, A. L. Roughton and S. A. Benner, *J. Am. Chem. Soc.*, 1996, **118**, 4518; (e) B. R. Shaw, M. Dobrikov, X. Wang, J. Wan, K. He, J.-L. Lin, P. Li, V. Rait, Z. A. Sergueeva and D. Sergueev, *Ann. N. Y. Acad. Sci.*, 2003, **1002**, 12; (f) P. Li, Z. A. Sergueeva, M. Dobrikov and B. R. Shaw, *Chem. Rev.*, 2007, **107**, 4746; (g) H. Isobe, T. Fujino, N. Yamazaki, M. Guillot-Nieckowski and E. Nakamura, *Org. Lett.*, 2008, **10**, 3729.
- 7 (a) M. Eriksson and P. E. Q. Nielsen, *Annu. Rev. Biophys.*, 1996, **29**, 369; (b) P. E. Nielsen, *Chem. Biodiversity*, 2010, **7**, 786; (c) P. E. Nielsen and M. Egholm, *Curr. Issues Mol. Biol.*, 1999, **1**, 89; (d) P. E. Nielsen and G. Haaima, *Chem. Soc. Rev.*, 1997, **26**, 73; (e) Y. Ura, J. M. Beierle, L. J. Leman, L. E. Orgel and M. R. Ghadiri, *Science*, 2009, **325**, 73.
- 8 (a) J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, *Nature*, 1990, **343**, 33; (b) Z. Yang, D. Hutter, P. Sheng, A. M. Sismour and S. A. Benner, *Nucleic Acids Res.*, 2006, **34**, 6095; (c) F. Wojciechowski and C. J. Leumann, *Chem. Soc. Rev.*, 2011, **40**, 5669; (d) S. A. Benner, *Curr. Opin. Chem. Biol.*, 2012, **16**, 581; (e) H. Liu, J. Gao, S. R. Lynch, Y. D. Saito, L. Maynard and E. T. Kool, *Science*, 2003, **302**, 868; (f) E. T. Kool, *Acc. Chem. Res.*, 2002, **35**, 936; (g) E. T. Kool, H. Lu, S. J. Kim, S. Tan, J. N. Wilson, J. Gao and H. Liu, *Nucleic Acids Symp. Ser.*, 2006, **50**, 15; (h) J. N. Wilson and E. T. Kool, *Org. Biomol. Chem.*, 2006, **4**, 4265.
- 9 (a) Y. Tanaka, H. Katagiri, Y. Furusho and E. Yashima, *Angew. Chem.*, 2005, **117**, 3935; (b) H. Ito, Y. Furusho, T. Hasegawa and E. Yashima, *J. Am. Chem. Soc.*, 2008, **130**, 14008; (c) Y. Furusho and E. Yashima, *Macromol. Rapid Commun.*, 2010, **32**, 136; (d) H. Yamada, Y. Furusho, H. Ito and E. Yashima, *Chem. Commun.*, 2010, **46**, 3487.
- 10 (a) S. I. Chan, M. P. Schweizer, P. O. P. Ts'o and G. K. Helmkamp, *J. Am. Chem. Soc.*, 1964, **86**, 4182; (b) D. Pörschke, *Biopolymers*, 1971, **10**, 1989; (c) L. A. Marky and K. J. Breslauer, *Biopolymers*, 1982, **21**, 2185.
- 11 C. A. Hunter and H. L. Anderson, *Angew. Chem., Int. Ed. Engl.*, 2009, **48**, 7488.
- 12 (a) C. A. Hunter, M. C. Misuraca and S. M. Turega, *J. Am. Chem. Soc.*, 2011, **133**, 582; (b) C. A. Hunter, M. C. Misuraca and S. M. Turega, *J. Am. Chem. Soc.*, 2011, **133**, 20416; (c) E. Chekmeneva, C. A. Hunter, M. C. Misuraca and S. M. Turega, *Org. Biomol. Chem.*, 2012, **10**, 6022; (d) C. A. Hunter, M. C. Misuraca and



- S. M. Turega, *Chem. Sci.*, 2012, **3**, 2462; (e) C. A. Hunter, M. C. Misuraca and S. M. Turega, *Chem. Sci.*, 2012, **3**, 589; (f) H. Adams, E. Chekmeneva, C. A. Hunter, M. C. Misuraca, C. Navarro and S. M. Turega, *J. Am. Chem. Soc.*, 2013, **135**, 1853; (g) E. Chekmeneva, C. A. Hunter, M. J. Packer and S. M. Turega, *J. Am. Chem. Soc.*, 2008, **130**, 17718; (h) M. A. Jinks, H. Sun and C. A. Hunter, *Org. Biomol. Chem.*, 2014, **12**, 1440; (i) H. Sun, C. A. Hunter and E. M. Llamas, *Chem. Sci.*, 2015, **6**, 1444; (j) H. Sun, C. Navarro and C. A. Hunter, *Org. Biomol. Chem.*, 2015, **13**, 4981.
- 13 (a) C. Galli and L. Mandolini, *Eur. J. Org. Chem.*, 2000, **18**, 3117; (b) A. J. Kirby, *Adv. Phys. Org. Chem.*, 1981, **17**, 183.
- 14 (a) V. M. Krishnamurthy, V. Semetey, P. J. Bracher, N. Shen and G. M. Whitesides, *J. Am. Chem. Soc.*, 2007, **129**, 1312; (b) Z. Zhong, X. Li and Y. Zhao, *J. Am. Chem. Soc.*, 2011, **133**, 8862; (c) M. Mammen, S.-K. Choi and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 1998, **37**, 2754; (d) A. Mulder, J. Huskens and D. N. Reinhoudt, *Org. Biomol. Chem.*, 2004, **2**, 3409; (e) W. Jiang, K. Nowosinski, N. L. Löw, E. V. Dzyuba, F. Klautzsch, A. Schäfer, J. Huuskonen, K. Rissanen and C. A. Schalley, *J. Am. Chem. Soc.*, 2012, **134**, 1860; (f) E. T. Mack, P. W. Snyder, R. Perez-Castillejos, B. Bilgiçer, D. T. Moustakas, M. J. Butte and G. M. Whitesides, *J. Am. Chem. Soc.*, 2012, **134**, 333; (g) H. Sun, C. A. Hunter, C. Navarro and S. Turega, *J. Am. Chem. Soc.*, 2013, **135**, 13129; (h) M. C. Misuraca, T. Grecu, Z. Freixa, V. Garavini, C. A. Hunter, P. van Leeuwen, M. D. Segarra-Maset and S. M. Turega, *J. Org. Chem.*, 2011, **76**, 2723; (i) J.-M. Lehn, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 89; (j) D. J. Cram, G. M. Lein, T. Kaneda, R. C. Helgeson, C. B. Knobler, E. Maverick and K. N. Trueblood, *J. Am. Chem. Soc.*, 1981, **103**, 6822; (k) D. J. Cram, *Angew. Chem., Int. Ed. Engl.*, 1986, **25**, 1039.
- 15 (a) 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; (b) C. A. Hunter, *Angew. Chem., Int. Ed.*, 2004, **43**, 5310.
- 16 M. A. Ischay, Z. Lu and T. P. Yoon, *J. Am. Chem. Soc.*, 2010, **132**, 8572.
- 17 (a) A. F. Abdel-Magid and S. J. Mehrman, *Org. Process Res. Dev.*, 2006, **10**, 971; (b) A. F. Abdel-Magid, K. G. Carson, B. D. Harris, C. A. Maryanoff and R. D. Shah, *J. Org. Chem.*, 1996, **61**, 3849.
- 18 (a) U. Mayer, V. Gutmann and W. Gerger, *Monatsh. Chem.*, 1975, **106**, 1235; (b) U. Mayer, W. Gerger and V. Gutmann, *Monatsh. Chem.*, 1977, **108**, 489; (c) V. Gutmann, *Pure Appl. Chem.*, 1979, **51**, 2197.
- 19 (a) J. Applequist and V. Damle, *J. Am. Chem. Soc.*, 1965, **87**, 1450; (b) K. J. Breslauer, *Methods Enzymol.*, 1995, **259**, 221; (c) T. Ratilainen, A. Holmén, E. Tuite, P. E. Nielsen and B. Nordén, *Biochemistry*, 2000, **39**, 7781.

