Article

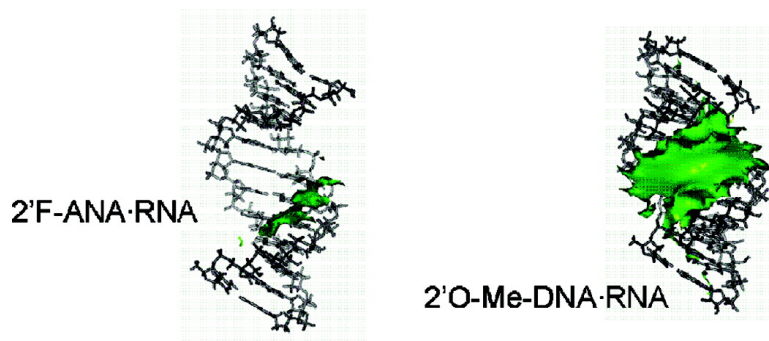# Theoretical Analysis of Antisense Duplexes: Determinants of the RNase H Susceptibility

Agnes Noy, F. Javier Luque, and Modesto Orozco

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Theoretical Analysis of Antisense Duplexes: Determinants of the RNase H Susceptibility

Agnes Noy,[†,‡] F. Javier Luque,[§] and Modesto Orozco*,[†,‡,||]

Joint IRB-BSC Research Program in Computational Biology, Institut de Recerca Biomèdica Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028 and Barcelona Supercomputer Center, Jordi Girona 31, Edifici Torre Girona, Barcelona 08028, Spain, National Institute of Bioinformatics, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, Departament de Fisicoquímica and Institut de Biomedicina, Facultat de Farmácia, Universitat de Barcelona, Avgda Diagonal 643, Barcelona 08028, Spain, and Departament de Bioquímica, Facultat de Biologia, Universitat de Barcelona, Avgda Diagonal 647, Barcelona 08028, Spain

Received September 6, 2007; E-mail: modesto@mmb.pcb.ub.es

Abstract: The structure and dynamic properties of different antisense related duplexes (DNA·RNA, 2′O−Me-DNA·RNA, 2′F-ANA·RNA, C5(Y)-propynyl-DNA·RNA, ANA·RNA, and control duplexes DNA·DNA and RNA·RNA) have been determined by means of long molecular dynamics simulations (covering more than 0.5 μs of fully solvated unrestrained MD simulation). The massive analysis presented here allows us to determine the subtle differences between the different duplexes, which in all cases pertain to the same structural family. This analysis provides information on the molecular determinants that allow RNase H to recognize and degrade some of these duplexes, whereas others with apparently similar conformations are not affected. Subtle structural and deformability features define the key properties used by RNase H to discriminate between duplexes.

Introduction

Antiparallel DNA·RNA hybrids are stable structures formed when complementary strands of DNA and RNA bind following recognition rules analogous to those of homopolymeric duplexes (i.e., those formed by pure DNA (DNA$_2$) and RNA (RNA$_2$) strands). A large series of crystallographic studies[1−6] suggested that in the crystal phase the DNA·RNA hybrid adopts a pure A-type structure, which should in practice be close to the homopolymeric RNA duplex. However, low-resolution CD and NMR data[7−9] and more recently high-resolution NMR experiments[10−16] showed that the structure of the DNA·RNA hybrid in solution is more complex, because it appears to have many characteristics of a canonical A-duplex, but some structural features resemble those of a B-type duplex. This view has received support by nanosecond-scale molecular dynamics (MD) simulations,[17] which suggest that the hybrid is closer to the A-type form than to the B-type one, but the structure does not strictly fit the canonical A conformation (we named this conformation the A/B form), showing a unique DNA versus RNA strand asymmetry.

The cell makes use of DNA·RNA hybrids at specific physical locations and at well-defined times of its life cycle (for example, in the nuclei during replication) but in general the formation of the hybrid duplexes is associated with potential damages for the cell-like viral infection,[18] triggering then survival responses that are mostly mediated by the enzyme RNase H.[19−21] This enzyme is a nuclease that degrades the RNA strand of the hybrid

---

† Joint IRB-BSC Research Program in Computational Biology.
‡ National Institute of Bioinformatics.
§ Departament de Fisicoquímica and Institut de Biomedicina. Universitat de Barcelona.
|| Departament de Bioquímica. Universitat de Barcelona.

(1) Wang, A. H.; Fujii, S.; van Boom, J. H.; van der Marel, G. A.; van Boeckel, S. A.; Rich, A. *Nature* **1982**, *299*, 601.
(2) Katahira, M.; Lee, S. J.; Kobayashi, Y.; Sujeta, H.; Kyogoku, Y.; Iwai, S.; Ohtsuka, E.; Benevides, J. M.; Thomas, G. J. *J. Am. Chem. Soc.* **1990**, *112*, 4508.
(3) Egli, M.; Usman, N.; Zhang, S. G.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 534. (b) Egli, M.; Usman, N.; Rich, A. *Biochemistry* **1993**, *32*, 3221.
(4) Horton, N. C.; Finzel, B. C. *J. Mol. Biol.* **1996**, *264*, 521.
(5) Conn, G. L.; Brown, T.; Leonard, G. A. *Nucleic Acids Res.* **1999**, *27*, 555.
(6) Xiong, Y.; Sundaralingam, M. *Nucleic Acids Res.* **2000**, *28*, 2171.
(7) (a) Chou, S. H.; Flynn, P.; Reid, B. *Biochemistry* **1989**, *28*, 2422. (b) Chou, S. H.; Flynn, P.; Reid, B. *Biochemistry* **1989**, *28*, 2435.
(8) Hall, K. B.; McLaughlin, L. W. *Biochemistry* **1991**, *30*, 10606.
(9) Fedoroff, O. Y.; Salazar, M.; Reid, B. R. *J. Mol. Biol.* **1993**, *233*, 509.
(10) (a) Lane, A. N.; Ebel, S.; Brown, T. *Eur. J. Biochem.* **1993**, *215*, 297. (b) Gyi, J. I.; Conn, G. L.; Lane, A. N.; Brown, T. *Biochemistry* **1996**, *35*, 12538. (c) Gyi, J. I.; Lane, A. N.; Conn, G. L.; Brown, T. *Biochemistry* **1998**, *37*, 73. (d) Gyi, J. I.; Gao, D.; Conn, G. L.; Trent, J. O.; Brown, T.; Lane, A. N. *Nucleic Acids Res.* **2003**, *31*, 2683.

(11) (a) Salazar, M.; Fedoroff, O. Y.; Miller, J. M.; Ribeiro, N. S.; Reid, B. R. *Biochemistry* **1993**, *32*, 4297. (b) Fedoroff, O. Y.; Ge, Y.; Reid, B. R. *J. Mol. Biol.* **1997**, *269*, 225.
(12) (a) Gao, X.; Jeffs, P. W. *J. Biomol. NMR.* **1994**, *4*, 367. (b) Cross, C. W.; Rice, J. S.; Gao, X. *Biochemistry* **1997**, *36*, 4096.
(13) (a) Gonzalez, C.; Stec, W.; Kobylanska, A.; Hogrefe, R. I.; Reynolds, M.; James, T. L. *Biochemistry* **1994**, *33*, 1062. (b) Gonzalez, C.; Stec, W.; Reynolds, M.; James, T. L. *Biochemistry* **1995**, *34*, 4969.
(14) Nishizaki, T.; Iwai, S.; Ohkubo, T.; Kojima, C.; Nakamura, H.; Kyogoku, Y.; Ohtsuka, E. *Biochemistry* **1996**, *35*, 4016.
(15) Szyperski, T.; Gotte, M.; Billeter, M.; Perola, E.; Cellai, L.; Heumann, H.; Wuthrich, K. *J. Biomol. NMR.* **1999**, *13*, 343.
(16) Hantz, E.; Larue, V.; Ladam, P.; Le Moyec, L.; Gouyette, C.; Huynh Dinh, T. *Int. J. Biol. Macromol.* **2001**, *28*, 273.
(17) Noy, A.; Pérez, A.; Márquez, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2005**, *127*, 4910.
(18) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 3rd Edition; Garland Publishing Inc.: New York, 1994.
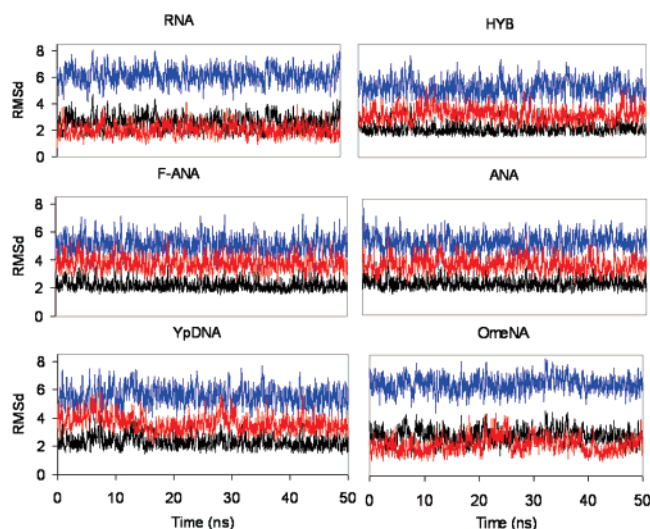(19) Peliska, J. A.; Benkovic, S. J. *Science* **1992**, *258*, 1112.

Figure 1. RMSd (for the central 10-mer equivalent (Methods) backbone atoms, in Å) with respect to DD sequence canonical A (red), B (blue), and NMR-based (1EFS; black) structures of the different duplexes with the DD sequence containing at least 1 RNA strand. Equivalent plots for the other two sequences (Figures S4 and S5), as well as for the DNA duplexes (Figure S6) are displayed in Supporting Information.

Table 1. Total Simulation Time (in ns) of the Different Trajectories Considered in This Study for Different Duplexes Starting from A- and B-forms and from NMR-Based Structure (1EFS)[a,b]

Seq 1	DNA$_2$	RNA$_2$	HYB	F-ANA	ANA	YpDNA	OmeNA
NMR-form			50	50	50	50	not stable
A form		50	3[c]	not stable	not stable	3[c]	50
B form	50						

Seq 2	DNA	RNA	HYB	F-ANA	ANA	YpDNA	OmeNA
equil(1)[d]	20	20	20	20	20	20	20

Seq 3	DNA	RNA	HYB	F-ANA	ANA	YpDNA	OmeNA
equil(1)[d]	20	20	20	20	20	20	20

[a] Only long simulations (i.e., 50 and 20 ns) were analyzed. [b] 1, 2, and 3 stand for the three sequences considered here: d/r(CGCGAATTCGCG)$_2$, d/r(CATAGGCCTATG)$_2$, and d/r(CACAGATCTGTG)$_2$, respectively. [c] After 3 ns, converged to the NMR-form. [d] Trajectories starting from the snapshots at the 20th ns for sequence 1.

without affecting the complementary DNA strand,[22,23] which can thus bind another complementary RNA strand. RNase H does not have a clear sequence specificity[6,9,24–26] but exhibits an astonishing ability to discriminate between DNA·RNA hybrids and other related structures such as DNA$_2$, RNA$_2$, or single-stranded DNAs or RNAs.[22,23,25,27–30]

Its ability to degrade DNA·RNA has been exploited for biomedical and biotechnological purposes in the so-called antisense therapy, where a single-stranded DNA (i.e., antisense)

is introduced in the cell to hybridize with a desired (i.e., sense) messenger RNA, which (if expressed) could trigger a given pathology. The first antisense drugs are already on the market,[31] and many other are under clinical trials to treat a wide range of pathologies, such as, cancer, inflammation, and viral infections.[29,32,33] Unfortunately, the accumulated experience demonstrates that the first-generation antisense drugs face major practical problems, which has stimulated an intense research effort to design modified nucleic acids (XNA): (i) resistant to interfering nucleases and (ii) able to bind with high stability and specificity to the target sense RNA.[34,35] There are many examples of XNAs, which, as single strands are resistant to nucleases, can enter efficiently into the cell and form stable and specific XNA·RNA duplexes, but their usefulness is limited by the fact that the corresponding hybrid is not degraded by RNase H.[28] Keeping in mind the chemical difficulties of developing modified nucleic acids, it would be worth identifying the molecular determinants that an XNA should satisfy to be recognized as a substrate by RNase H prior to its synthesis. Unfortunately, despite the large amount of structural data on XNA·RNA hybrids (17 structures in PDB), the vast biochemical knowledge on the reaction,[4,9,11,14,15,22–30] and the availability of few structures of the enzyme,[36,37] the structural reasons allowing the enzyme to distinguish between different hybrids remain to be elucidated.

To identify the molecular basis of the RNase H susceptibility, we report here a massive MD study for a variety of XNA·RNA duplexes, some of them degraded by the enzyme whereas others not. The analysis reveals the existence of clear differences in the structure and deformability pattern between duplexes susceptible or resistant to the enzyme. On the basis of these findings, we outline a clear and simple protocol to be used prior to the chemical synthesis of the XNA to evaluate its susceptibility to RNase H.

Methods

Sequence Selection. As is common in the field,[38,39] Dickerson's dodecamer (DD[40]) was selected as a typical sequence (d/r(CGCGAATTCGCG)$_2$) to study different types of XNA·RNA duplexes. However, to make our conclusions more general and to escape from potential sequence-specific artifacts, we also analyzed two additional dodecamers designed by permutations of DD nucleobases subject to the following restrictions: (i) all unique dinucleotide steps should be represented, (ii) a similar Pyr/Pur ratio should exist in each strand to avoid the adoption of unusual structures, and (iii) only palindromes were considered to avoid sequence-induced asymmetries between sense and antisense strands. With these premises, the additional chosen sequences were d/r(CATAGGCCTATG)$_2$ and d/r(CACAGATCTGTG)$_2$.

(20) Rumbaugh, J. A.; Murante, R. S.; Shi, S.; Bambara, R. A. *J. Biol. Chem.* **1997**, *272*, 22591.
(21) Crouch, R. J.; Toulmé, J. J., Eds.; *Ribonucleases H.* John Libbey: Paris, 1998.
(22) Han, G. W.; Kopka, M. L.; Cascio, D.; Grzeskowiak, K.; Dickerson, R. E. *J. Mol. Biol.* **1997**, *269*, 811.
(23) Lima, W. F.; Crooke, S. T. *Biochemistry* **1997**, *36*, 390.
(24) Roberts, W. R.; Crothers, D. M. *Science* **1992**, *258*, 1463.
(25) Oda, Y.; Iwai, S.; Ohtsuka, E.; Ishikawa, M.; Ikehara, M.; Nakamura, H. *Nucleic Acids Res.* **1993**, *21*, 4690.
(26) Ho, S. P.; Britton, D. H. O.; Stone, B. A.; Behrens, D. L.; Leffet, L. M.; Hobbs, F. W.; Miller, J. A.; Trainor G. L. *Nucleic Acids Res.* **1993**, *21*, 4690.
(27) Stein, H.; Hausen, P. *Science* **1969**, *166*, 393.
(28) Altmann, K. H.; Fabbrot, D.; Dean, N. M.; Geiger, T.; Monia, B. P.; Muller, M.; Nicklin, P. *Biochem. Soc. Trans.* **1996**, *24*, 630.
(29) Crooke, S. T. *Annu. Rev. Med.* **2004**, *55*, 61.
(30) Agrawal, S.; Iyer, R. P. *Pharmacol. Ther.* **1997**, *76*, 151.
(31) The Vitravene Study Group. *Am. J. Ophtalmol* **2002**, *133*, 467.
(32) Kurreck, J. *Eur. J. Biochem.* **2003**, *270*, 1628.
(33) Gleave, M. E.; Monia, B. P. *Nat. Rev. Cancer* **2005**, *5*, 468.
(34) Mangos, M. M.; Damha, M. J. *Curr. Top. Med. Chem.* **2002**, *2*, 1147.
(35) Chan, J. H. P.; Lim, S.; Wong, W. S. F. *Clin. Exp. Pharmacol. Physiol.* **2006**, *33*, 533.
(36) Sarafianos, S. G.; Das, K.; Tantillo, C.; Clark, A. D.; Ding, J.; Whitcomb, J. M.; Boyer, P. L.; Hughes, S. H.; Arnold, E. *EMBO J.* **2001**, *20*, 1449.
(37) Nowotny, M.; Gaidamakov, S. A.; Crouch, J.; Yang, W. *Cell* **2005**, *121*, 1005.
(38) Orozco, M.; Perez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350.
(39) Cheatham, T. E., 3rd. *Curr. Opin. Struct. Biol.* **2004**, *14*, 360.
(40) Pérez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E. 3rd; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.

Table 2. Rotational[a] and Translational[b] Helical Parameters for the Ten Unique Dinucleotide Steps for the Different Duplexes Studied Here[c,d]

step	DNA$_2$	RNA$_2$	DNA	F-ANA	ANA	YpDNA	OmeNA
GC·GC	34.7 ± 9.6	30.2 ± 6.0	31.8 ± 7.1	29.6 ± 6.6	30.8 ± 7.6	33.7 ± 6.3	30.0 ± 5.8
	−0.5 ± 9.6	4.7 ± 8.5	3.4 ± 8.7	1.8 ± 8.1	3.2 ± 8.7	0.9 ± 7.7	5.4 ± 8.5
	0.4 ± 7.5	−0.1 ± 7.1	0.9 ± 7.2	2.7 ± 7.0	3.8 ± 7.4	2.3 ± 6.8	0.4 ± 7.0
	3.3 ± 0.5	*3.3 ± 0.4*	*3.3 ± 0.5*	*3.3 ± 0.4*	*3.4 ± 0.5*	*3.4 ± 0.4*	*3.3 ± 0.4*
	−0.3 ± 1.0	*−1.6 ± 0.9*	*−1.1 ± 0.9*	*−1.3 ± 0.9*	*−1.3 ± 1.0*	*−1.6 ± 0.9*	*−1.8 ± 0.8*
	−0.1 ± 1.1	*0.0 ± 0.9*	*0.1 ± 1.0*	*−0.2 ± 0.9*	*0.2 ± 1.0*	*0.2 ± 0.8*	*0.1 ± 0.8*
GG·CC	32.0 ± 7.5	29.6 ± 4.8	29.2 ± 5.3	29.6 ± 4.9	27.8 ± 5.5	28.9 ± 4.8	29.4 ± 4.5
	4.0 ± 7.4	8.3 ± 7.1	6.6 ± 7.1	5.4 ± 7.3	6.8 ± 7.1	5.1 ± 6.6	8.8 ± 7.0
	0.1 ± 6.2	0.0 ± 5.8	2.9 ± 5.9	2.9 ± 5.7	3.3 ± 5.8	2.5 ± 5.5	0.3 ± 5.7
	3.5 ± 0.5	*3.3 ± 0.4*	*3.4 ± 0.4*	*3.4 ± 0.4*	*3.4 ± 0.4*	*3.3 ± 0.4*	*3.3 ± 0.4*
	−1.1 ± 1.0	*−1.9 ± 0.5*	*−1.8 ± 0.6*	*−1.8 ± 0.5*	*−1.7 ± 0.6*	*−1.8 ± 0.5*	*−2.1 ± 0.4*
	0.0 ± 1.0	*0.0 ± 0.8*	*0.1 ± 0.8*	*0.3 ± 0.8*	*0.2 ± 0.7*	*0.1 ± 0.7*	*0.0 ± 0.7*
GT·AC	30.6 ± 8.8	29.0 ± 5.1	29.9 ± 5.7	29.4 ± 6.7	29.4 ± 5.4	30.2 ± 5.5	29.8 ± 4.6
	0.6 ± 8.1	8.7 ± 8.4	5.1 ± 7.9	2.3 ± 8.1	3.8 ± 7.4	4.5 ± 7.8	7.2 ± 7.0
	0.2 ± 6.0	0.1 ± 6.5	0.0 ± 6.0	4.0 ± 9.9	2.8 ± 6.2	1.5 ± 6.1	−0.7 ± 5.8
	3.3 ± 0.4	*3.3 ± 0.4*	*3.2 ± 0.4*	*3.3 ± 0.5*	*3.4 ± 0.4*	*3.2 ± 0.4*	*3.3 ± 0.4*
	−0.5 ± 0.7	*−1.3 ± 0.7*	*−1.0 ± 0.7*	*−1.2 ± 1.0*	*−1.4 ± 0.8*	*−1.1 ± 0.7*	*−1.5 ± 0.7*
	−0.1 ± 1.1	*0.0 ± 1.0*	*0.0 ± 1.0*	*−0.3 ± 1.2*	*−0.2 ± 1.0*	*0.2 ± 1.1*	*0.2 ± 0.8*
GA·TC	36.9 ± 11.7	30.3 ± 6.7	29.7 ± 8.1	30.9 ± 7.3	30.0 ± 9.1	30.3 ± 7.2	29.8 ± 6.4
	1.8 ± 11.2	8.3 ± 12.8	7.9 ± 12.7	5.8 ± 12.2	5.9 ± 12.0	5.2 ± 11.2	9.2 ± 12.0
	−0.3 ± 9.0	0.0 ± 9.6	2.8 ± 9.7	2.6 ± 9.3	2.6 ± 9.5	1.7 ± 8.8	1.4 ± 8.8
	3.5 ± 0.6	*3.2 ± 0.7*	*3.4 ± 0.7*	*3.4 ± 0.7*	*3.4 ± 0.7*	*3.3 ± 0.7*	*3.3 ± 0.6*
	−0.3 ± 1.4	*−1.6 ± 0.9*	*−1.3 ± 1.1*	*−1.5 ± 1.0*	*−1.6 ± 1.0*	*−1.5 ± 0.9*	*−1.9 ± 0.8*
	0.0 ± 1.4	*0.0 ± 1.4*	*0.1 ± 1.3*	*0.2 ± 1.3*	*0.2 ± 1.4*	*0.1 ± 1.2*	*−0.1 ± 1.2*
AA·TT	35.0 ± 6.4	28.3 ± 5.1	28.7 ± 5.3	28.9 ± 5.1	28.2 ± 4.9	28.0 ± 5.1	28.6 ± 4.4
	2.1 ± 7.0	9.0 ± 9.1	7.1 ± 8.5	5.0 ± 8.4	4.8 ± 8.4	6.0 ± 8.6	8.0 ± 8.3
	−0.1 ± 5.7	0.2 ± 7.6	1.0 ± 7.3	1.5 ± 6.9	1.8 ± 7.1	2.9 ± 6.7	−0.3 ± 6.8
	3.4 ± 0.4	*3.3 ± 0.5*	*3.3 ± 0.5*	*3.3 ± 0.5*	*3.3 ± 0.5*	*3.3 ± 0.5*	*3.2 ± 0.5*
	−0.6 ± 0.8	*−1.6 ± 0.6*	*−1.3 ± 0.7*	*−1.4 ± 0.7*	*−1.6 ± 0.7*	*−1.6 ± 0.6*	*−1.8 ± 0.6*
	−0.1 ± 0.8	*0.0 ± 1.0*	*0.1 ± 0.9*	*0.2 ± 0.9*	*0.1 ± 0.8*	*0.1 ± 0.9*	*−0.1 ± 0.8*
AG·CT	30.2 ± 12.7	28.3 ± 6.7	27.5 ± 7.8	27.7 ± 8.5	28.1 ± 7.6	27.5 ± 7.2	27.8 ± 6.6
	2.7 ± 11.6	10.6 ± 12.0	6.9 ± 11.6	4.8 ± 11.6	2.0 ± 11.4	6.8 ± 11.0	9.0 ± 10.8
	0.3 ± 8.7	0.1 ± 9.2	1.6 ± 9.3	3.2 ± 9.6	0.5 ± 9.7	1.8 ± 9.0	−1.2 ± 8.7
	3.3 ± 0.7	*3.4 ± 0.6*	*3.4 ± 0.7*	*3.4 ± 0.7*	*3.5 ± 0.7*	*3.4 ± 0.6*	*3.3 ± 0.6*
	−0.7 ± 1.4	*−1.7 ± 0.8*	*−1.6 ± 1.0*	*−1.6 ± 0.9*	*−1.8 ± 0.9*	*−1.6 ± 0.8*	*−1.9 ± 0.7*
	0.2 ± 1.5	*0.0 ± 1.4*	*−0.1 ± 1.4*	*0.0 ± 1.6*	*−0.2 ± 1.3*	*0.1 ± 1.4*	*−0.1 ± 1.4*
AT·AT	30.9 ± 7.8	27.6 ± 7.1	28.3 ± 7.5	26.8 ± 7.0	27.5 ± 6.4	27.5 ± 6.5	27.8 ± 6.8
	−0.1 ± 9.6	8.3 ± 12.6	4.3 ± 11.4	1.6 ± 10.3	2.3 ± 8.7	2.3 ± 9.5	5.8 ± 10.2
	0.1 ± 7.7	0.0 ± 10.3	0.7 ± 9.4	3.1 ± 9.1	2.2 ± 7.8	2.2 ± 7.8	−0.4 ± 8.8
	3.3 ± 0.6	*3.2 ± 0.6*	*3.2 ± 0.6*	*3.3 ± 0.6*	*3.3 ± 0.5*	*3.3 ± 0.5*	*3.3 ± 0.5*
	−0.9 ± 0.9	*−1.4 ± 1.0*	*−1.2 ± 1.0*	*−1.4 ± 1.0*	*−1.7 ± 1.0*	*−1.7 ± 0.9*	*−1.9 ± 1.0*
	0.1 ± 1.3	*0.0 ± 1.6*	*−0.1 ± 1.5*	*−0.4 ± 1.4*	*−0.3 ± 1.3*	*−0.3 ± 1.3*	*0.2 ± 1.3*
CG·CG	26.0 ± 12.6	31.0 ± 4.8	31.8 ± 6.9	31.0 ± 5.4	30.1 ± 6.5	26.3 ± 8.0	31.0 ± 4.7
	7.3 ± 8.4	14.9 ± 9.7	7.9 ± 9.8	9.8 ± 9.3	8.7 ± 9.6	11.2 ± 10.2	15.4 ± 9.2
	0.7 ± 7.4	−0.1 ± 6.8	−1.1 ± 8.0	2.9 ± 6.8	2.7 ± 7.1	4.4 ± 7.3	−1.1 ± 6.7
	3.1 ± 0.5	*3.6 ± 0.6*	*3.8 ± 0.6*	*3.5 ± 0.6*	*3.4 ± 0.6*	*3.8 ± 0.7*	*3.6 ± 0.6*
	−0.2 ± 0.8	*−1.6 ± 0.6*	*−1.4 ± 0.8*	*−1.4 ± 0.7*	*−1.2 ± 0.8*	*−1.8 ± 1.0*	*−1.8 ± 0.6*
	0.1 ± 1.1	*0.0 ± 1.0*	*−0.3 ± 1.0*	*0.4 ± 0.9*	*0.1 ± 1.1*	*0.1 ± 1.2*	*−0.1 ± 1.0*
CA·TG	30.1 ± 12.6	31.5 ± 4.7	29.6 ± 6.4	30.6 ± 4.9	30.5 ± 5.4	29.5 ± 5.3	31.4 ± 4.4
	8.5 ± 9.8	16.7 ± 9.9	12.9 ± 10.5	10.6 ± 9.0	10.2 ± 9.0	11.5 ± 9.5	16.1 ± 9.1
	−0.4 ± 7.3	0.1 ± 6.5	3.0 ± 7.4	1.2 ± 6.5	1.3 ± 6.8	2.4 ± 6.5	−0.3 ± 6.3
	3.3 ± 0.5	*3.6 ± 0.5*	*3.4 ± 0.7*	*3.4 ± 0.5*	*3.4 ± 0.5*	*3.5 ± 0.6*	*3.5 ± 0.5*
	−0.4 ± 1.0	*−1.5 ± 0.5*	*−1.0 ± 0.9*	*−1.2 ± 0.6*	*−1.3 ± 0.6*	*−1.2 ± 0.6*	*−1.6 ± 0.5*
	−0.1 ± 1.1	*0.0 ± 0.9*	*0.2 ± 1.0*	*0.4 ± 0.9*	*0.3 ± 0.9*	*−0.1 ± 1.0*	*−0.1 ± 0.8*
TA·TA	32.1 ± 10.7	31.3 ± 4.8	29.4 ± 5.1	31.3 ± 5.1	30.4 ± 5.3	29.8 ± 5.4	31.8 ± 4.3
	8.1 ± 11.6	17.5 ± 10.2	16.5 ± 9.8	11.9 ± 9.1	10.5 ± 9.9	14.3 ± 10.1	16.4 ± 9.8
	−0.5 ± 7.3	−0.1 ± 7.1	1.6 ± 7.3	−0.5 ± 7.5	0.2 ± 7.4	2.0 ± 7.8	0.1 ± 6.6
	3.3 ± 0.5	*3.4 ± 0.5*	*3.4 ± 0.5*	*3.3 ± 0.5*	*3.2 ± 0.5*	*3.5 ± 0.7*	*3.4 ± 0.5*
	−0.3 ± 1.1	*−1.4 ± 0.5*	*−1.2 ± 0.5*	*−1.2 ± 0.6*	*−1.2 ± 0.6*	*−1.2 ± 0.6*	*−1.5 ± 0.5*
	−0.2 ± 1.3	*0.0 ± 0.8*	*0.3 ± 0.8*	*0.4 ± 0.7*	*0.3 ± 0.8*	*0.2 ± 0.9*	*0.0 ± 0.7*
all	32.1 ± 7.1	29.5 ± 3.7	29.4 ± 4.4	29.3 ± 4.3	29.2 ± 4.3	29.6 ± 5.1	29.5 ± 3.5
	2.9 ± 6.6	10.1 ± 7.3	7.4 ± 7.4	5.3 ± 7.0	5.3 ± 6.7	6.6 ± 7.2	9.7 ± 7.0
	0.0 ± 4.7	0.0 ± 4.8	1.5 ± 5.2	2.5 ± 5.1	2.1 ± 5.0	2.3 ± 4.7	−0.3 ± 4.5
	3.3 ± 0.3	*3.3 ± 0.4*	*3.4 ± 0.4*	*3.3 ± 0.3*	*3.4 ± 0.4*	*3.4 ± 0.4*	*3.3 ± 0.3*
	−0.5 ± 0.7	*−1.6 ± 0.5*	*−1.3 ± 0.6*	*−1.4 ± 0.6*	*−1.5 ± 0.6*	*−1.5 ± 0.5*	*−1.8 ± 0.4*
	0.0 ± 0.8	*0.0 ± 0.7*	*0.0 ± 0.7*	*0.1 ± 0.7*	*0.1 ± 0.7*	*0.1 ± 0.7*	*0.0 ± 0.6*

[a] Twist, roll, and tilt in roman and degrees. [b] Rise, slide, and shift in italics and Å. [c] For non-homopolymeric duplexes, it is labeled by the antisense strand. [d] Standard deviations were obtained by linear propagation of deviations in samplings and structures (Methods).
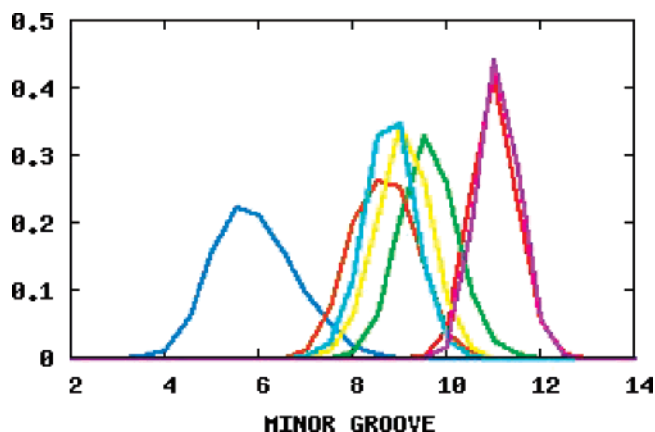
Figure 2. Population (in fractions of 1) of minor-groove widths (in Å) for the different duplexes considered here. Dark blue, DNA$_2$; red, RNA$_2$; green, DNA·RNA; brown, F-ANA·RNA; light blue, ANA·RNA; yellow, YpDNA· RNA; magenta, OmeNA·RNA. Values were derived by pooling the results obtained for the three sequences.

Hybrids Considered. Besides the standard DNA·RNA hybrid, a variety of duplexes formed upon binding the sense RNA strand to different complementary XNAs acting as antisense candidates were considered: (i) arabinonucleic acids (ANA·RNA duplex), (ii) 2′-F-arabinonucleic acids (F-ANA· RNA duplex), (iii) 2′-O−Me-DNA (OmeNA·RNA duplex), and (iv) C5-propynyl-DNA (YpDNA·RNA duplex). With these XNAs, we include modifications in the sugar and in the nucleobase, which give rise to hybrids of different stability and RNase H susceptibility.[41−44] The corresponding homopolymers (DNA$_2$ and RNA$_2$) were also included as a reference. Accordingly, a total of 7 structures were analyzed for 3 different sequences, leading to 21 different dodecamers.

Structure Setup. The generation of the starting structures was first carried out for the DD sequences, which were then used as templates for the other sequences (below). For homopolymeric duplexes (DNA$_2$ and RNA$_2$), the corresponding X-ray structures were chosen.[45] For the hybrids, two starting structures were generated: (i) one corresponding to the pure A form and (ii) the other to the NMR-based structure of the DNA· RNA hybrid (1EFS[16]) manipulated as described elsewhere.[17] All of the structures were neutralized by Na$^+$ ions and immersed in rectangular (around $60 \times 60 \times 70$ Å^3) boxes of TIP3P waters (from 4385 to 6661 molecules), so that there is at least 12 Å from any atom of the DNA duplex to the edges of the box. The systems were minimized, thermalized ($T = 298$ K), and equilibrated using our standard equilibration protocol,[46] doubling the length of the individual periods to ensure the lack of equilibration artifacts.

For DNA·RNA and YpDNA·RNA, both structures converge quickly (in less than 700 ps) to samplings close to the NMR-

based structure (Figure S1 in the Supporting Information and Figure 1). Despite all of our efforts, the arabinonucleic acids (ANA·RNA and F-ANA·RNA) failed to yield stable trajectories starting from the A form due to the very bad interactions around the sugar when the pure A form was imposed. Accordingly, only trajectories starting from the NMR-based structure were followed. The reverse problem was found for OmeNA·RNA duplexes, where the only stable trajectories were those starting from the pure A form.

The starting structures for the simulations with the other two sequences (d/r(CATAGGCCTATG)$_2$ and d/r(CACAGATCT-GTG)$_2$) were created by sequence substitution of the snapshots at the 20th ns of the trajectories obtained for the DD sequence. Because the differences in sequence only involved changes of A to G and T to C, common atoms were retained and the rest of them were added by using the LEAP module of the *AMBER8.0* program[47] (those changes affected C6 and C2 positions for purines, and C5, C4, and C3 ones in pyrimidines). The set of simulations performed in this study is summarized in Table 1.

Force Field. Canonical strands (DNA or RNA) were represented using the parmbsc0[40] revision of the parm99 force field,[48] which yields reliable trajectories for a large variety of nucleic acids up to (at least) the microsecond time scale.[49] Parametrization of the C5-propynil derivative was done using RESP/6-31G(d) charges[50] and the gaff force-field[51] parameters for the propynyl group. Parameters for the 2′-O−Me, arabino, and 2′-F-arabino derivatives were taken from Venkateswarlu's study.[52] Charges for arabino and 2′-F-arabino derivatives were obtained from RESP/6-31G(d) calculations.[50] To derive suitable torsional parameters for the modified sugars, the pseudorotation profile of thymidine nucleoside containing either arabino or 2′-F-arabino sugars was computed at the MP2/6-31G(d) level and fitted using a Monte Carlo procedure as described elsewhere.[40] Geometries were fully optimized except for the backbone dihedral angles (β, γ, ϵ, and χ), which were fixed at standard hybrid values (i.e., values from NMR structure (1EFS); 177, 80, −170, and −140 respectively), and the two internal dihedrals of the sugars required fixing the phase angle at a given value. The final parameters reproduce quite well the QM profiles (Figure S2 in the Supporting Information), and their goodness is further supported by the results obtained in test calculations (20 ns trajectories in aqueous solution) of small arabino and 2′-F-arabino hairpins,[53] which were chosen because of the availability of NMR data (Table S1 in the Supporting Information). The optimized parameters (downloadable at http://mmb.pcb.ub.es/antisense) lead to pseudo-

(41) Wagner, R. W.; Matteucci, M. D.; Grant, D.; Huang, T.; Froehler, B. C. *Nat. Biotechnol.* **1996**, *14*, 840.
(42) Nishizaki, T.; Iwai, S.; Ohtsuka, E.; Nakamura, H. *Biochemistry* **1997**, *36*, 2577.
(43) (a) Wilds, C. J.; Damha, M. J. *Nucleic Acids Res.* **2000**, *28*, 3625. (b) Noronha, A. M.; Wilds, C. J.; Lok, C. N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J. *Biochemistry* **2000**, *39*, 7050. (c) Damha, M. J.; Noronha, A. M.; Wilds, C. J.; Trempe, J. F.; Denisov, A.; Pon, R. T.; Gehring, K. *Nucleosides Nucleotides Nucleic Acids* **2001**, *20*, 429.
(44) (a) Barnes, T. W.; Turner, D. H. *Biochemistry* **2001**, *40*, 12738. (b) Barnes, T. W.; Turner, D. H. *J. Am. Chem. Soc.* **2001**, *123*, 4107.
(45) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V. *Science* **1982**, *216*, 475.
(46) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1999**, *119*, 7463.

(47) Case, D. A. et al. *AMBER8*; University of California: San Francisco, 2004.
(48) (a) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179. (b) Cheatham, T. E. 3rd; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
(49) Pérez, A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2007**, *129*, 14739−45.
(50) Bayly, C. E.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(51) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
(52) (a) Venkateswarlu, D.; Lind, K. E.; Mohan, V.; Manoharan, M.; Ferguson, D. M. *Nucleic Acids Res.* **1999**, *27*, 2189. (b) Venkateswarlu, D.; Ferguson, D. M. *J. Am. Chem. Soc.* **1999**, *121*, 5509.
(53) (a) Trempe, J.; Wilds, C. J.; Denisov, A. Y.; Pon, R. T.; Damha, M. J.; Gehring, K. *J. Am. Chem. Soc.* **2001**, *123*, 4896. (b) Denisov, A. Y.; Noronha, A. M.; Wilds, C. J.; Trempe, J.; Pon, R. T.; Gehring, K.; Damha, M. J. *Nucleic Acids Res.* **2001**, *29*, 4284.

rotational profiles significantly different from those obtained for 2′-deoxyribose (Figure S3 in the Supporting Information).

Simulation Details. The equilibrated structures of DD were subject to 50 ns of unconstrained MD simulation (100 ns for the DNA·DNA duplex[40]) at constant temperature (298 K) and pressure (1 atm) using periodic boundary conditions and particle mesh Ewald.[54] *SHAKE*[55] was used to constrain all bonds involving hydrogen atoms, which allowed us to use an integration step of 2 fs. Default *AMBER8.0* parameters were used for the remaining simulation conditions.[47] On the basis of the strong stability of the trajectories observed for DD sequences, only 20 ns simulations were performed for the other two sequences (which in fact started from 20 ns pre-equilibrated structures; above and Table 1). All of the MD simulations were performed using the PMEMD module of the *AMBER8.0* computer program.[47]

Analysis. Standard geometrical analysis was performed to follow the main structural features of the duplexes. The recognition properties were examined from classical molecular interaction potentials (CMIP[56]) using Na^+ as a probe. The essential dynamics of the duplexes was derived by diagonalization of the covariance matrix,[38,57] which yields a set of eigenvectors $\{\nu_i\}$ describing the nature of the essential movements and the associated eigenvalues $\{\lambda_i\}$, which indicate the magnitude of the displacement expected along each mode. Note that in the harmonic limit the stiffness constants associated to each deformation mode can be derived from λ_i as shown in eq 1.

$$K_i = k_B T / \lambda_i \tag{1}$$

where k_B is Boltzman's constant and T is the absolute temperature.

To compare the similarity in the set of essential movements in two trajectories, we used absolute (eqs 2 and 4) and relative (eqs 3 and 5) similarity indexes previously described.[38,58] Note that index ξ_{AB} (eq 4) is equivalent to γ_{AB}(eq 2) but considers explicitly the similarity between all of the associated eigenvalues included in the set of important eigenvectors (i.e., those needed to explain a given percentage of the total structural variance). Furthermore, note that the relative indexes eliminate the noise in the absolute index arising from the limited length of simulations.

$$\gamma_{AB} = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} (\nu_i^A \nu_j^B)^2 \tag{2}$$

where n is the number of important eigenvectors (as usual we consider 10 eigenvectors, which explains on average around 75% of variance), and ν_i^X stands for the *i*-unitary eigenvector of trajectory X (note that γ_{AB} takes values from 0

to 1, which means null similarity and identical essential dynamics).

$$\kappa_{AB} = 2 \frac{\gamma_{AB}}{(\gamma_{AA}^T + \gamma_{BB}^T)} \tag{3}$$

where γ_{XX}^T is the absolute self-similarity index (eq 2) for trajectory X obtained by comparing the first and second halves of the trajectory.

$$\xi_{AB} = \frac{2 \sum_{i=1}^{i=z} \sum_{j=1}^{j=z} (\nu_i^A \nu_j^B) \left[\dfrac{\exp\left\{ -\dfrac{(\Delta x)^2}{\lambda_i^A} - \dfrac{(\Delta x)^2}{\lambda_j^B} \right\}}{\sum_{i=1}^{i=z} \exp\left\{ -\dfrac{(\Delta x)^2}{\lambda_i^A} \right\} \sum_{j=1}^{j=z} \exp\left\{ -\dfrac{(\Delta x)^2}{\lambda_j^B} \right\}} \right]^2}{\sum_{i=1}^{i=z} \left(\dfrac{\exp\left\{ -2\dfrac{(\Delta x)^2}{\lambda_i^A} \right\}}{\left(\sum_{i=1}^{i=z} \exp\left\{ -\dfrac{(\Delta x)^2}{\lambda_i^A} \right\} \right)^2} \right)^2 + \sum_{j=1}^{j=z} \left(\dfrac{\exp\left\{ -2\dfrac{(\Delta x)^2}{\lambda_j^B} \right\}}{\left(\sum_{j=1}^{j=z} \exp\left\{ -\dfrac{(\Delta x)^2}{\lambda_j^B} \right\} \right)^2} \right)^2} \tag{4}$$

where λ_i is the eigenvalue (in Å^2) associated to eigenvector ν_i, Δx is set to standard values for duplexes,[58] and the sum is extended to the same important space used in eq 2 ($z = 10$).

$$\delta_{AB} = 2 \frac{\xi_{AB}}{(\xi_{AA}^T + \xi_{BB}^T)} \tag{5}$$

Similarity indexes between several trajectories were clustered to summarize information. To this end, we used agglomeration hierarchical clustering, which consists of building up the tree from elements by progressively merging clusters. Thus, a first cluster is defined by the two elements with shorter distance; then, values from this cluster to all of the elements are recalculated, assuming an average and shorter distance is chosen again to make a new cluster or to increase one. The clustering procedure was performed using disimilarity distances defined as $1-\delta$, an intuitive measure and fulfill statistics requirements for a distance ($d_{AB} = d_{BA}$; $d_{AA} = 0$; $d_{AB} \geq 0$). When duplexes with the same sequence were compared, all of the common atoms including those at the nucleobases were considered (i.e., excluding 5-H/methyl/propynyl of pyrimidines and 2′-H/OH/F/O-methyl). However, when comparison involved different sequences only the backbone ending at C1′ was included. In all of the cases, the analysis was performed for the duplex as well as for the separated strands.

The ability of a given duplex to adapt its structure to that required for the productive binding to RNase H was analyzed by computing the distance in essential space needed to convert the unbound form to the bound conformation defined by the structure of a RNAse H-bound DNA·RNA hybrid (1ZBI structure[37]) and the associated energy. This distance in essential space was determined using Mahalanobis metrics,[59] which consists in defining Euclidean distances

(54) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(55) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327.

(56) Gelpi, J. L.; Kalko, S. G.; Barril, X.; Cirera, J.; de La Cruz, X.; Luque, F. J.; Orozco, M. *Proteins* **2000**, *45*, 428.

(57) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412.

(58) (a) Cubero, E.; Abrescia, N. G. A.; Subirana, J. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 14603. (b) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007. (c) Pérez, A.; Blas, J. R.; Rueda, M.; Lopez-Bes, J. M.; de la Cruz, X.; Orozco, M. *J. Chem. Theory Comput.* **2005**, *1*, (5), 790−800.

(59) (a) Noy, A.; Pérez, A.; Laughton, C. A.; Orozco, M. *Nucleic Acids Res.* **2007**, *35*, 3330. (b) Mahalanobis, P. C. *Proc. Nat. Inst. Sci. India* **1936**, *2*, 49.
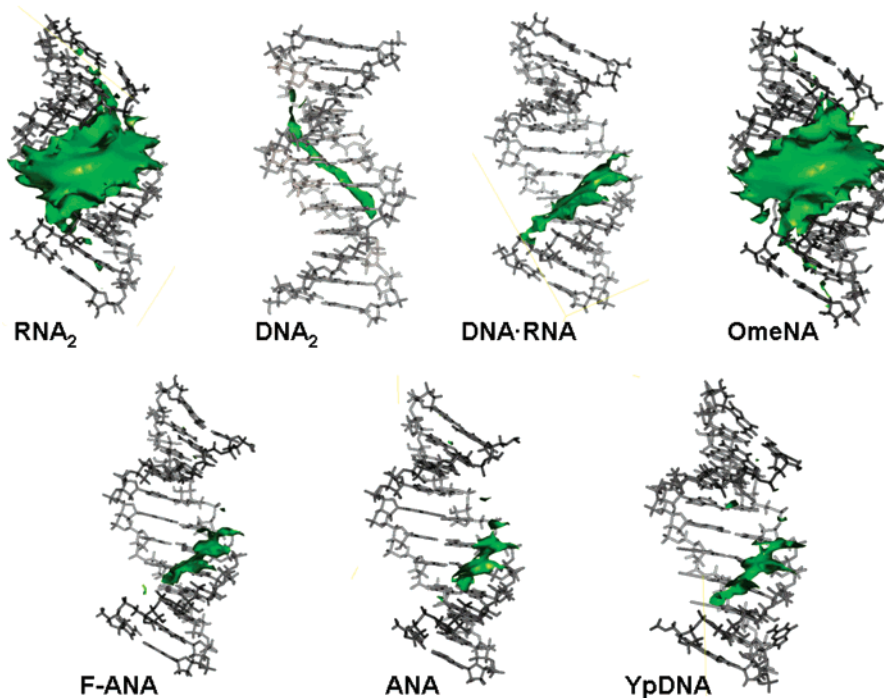
Figure 3. cMIP isocontours (-3 kcal/mol) for the interaction with a Na$^+$ probe for the different average duplexes with the DD sequence.
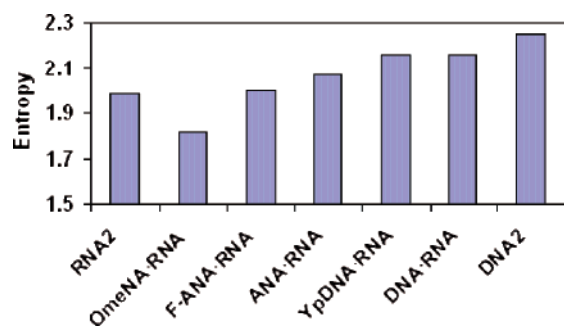

Figure 4. Entropies (in kcal/mol K) extrapolated at infinite simulation time for trajectories performed for duplexes with the DD sequence (Methods for details).

Figure 5. Force constants (in cal/mol Å^2) assigned to the first 10 essential movements (5 in the insert) of the duplexes with the DD sequence (very similar plots were obtained for the other 2 sequences). Dark blue, DNA$_2$; red, RNA$_2$; green, DNA·RNA; brown, F-ANA·RNA; light blue, ANA·RNA; yellow, YpDNA·RNA; magenta, OmeNA·RNA.

weighted by the variance of every degree of freedom (eq 6), which when computed using orthogonal vectors (as those derived from essential mode analysis) can be written as shown in eq 7.

$$d_{\mathrm{M}} = [x^{\mathrm{T}} C^{-1} x]^{1/2} \qquad (6)$$

where x is the Euclidean distance vector and C is the covariance matrix.

$$d_{\mathrm{M}} = \left[\sum_{i=1}^{n} \left(\frac{x_i}{\lambda_i^{1/2}} \right)^2 \right]^{1/2} \qquad (7)$$

where x_i is the displacement along individual eigenvectors, λ_i stands for the corresponding eigenvalue (in units of distance^2), and the sum extends to the space of important essential movements ($n = 10$ in this article).

The minimum Mahalanobis distance between two structures was computing by selecting an iteratively small displace-
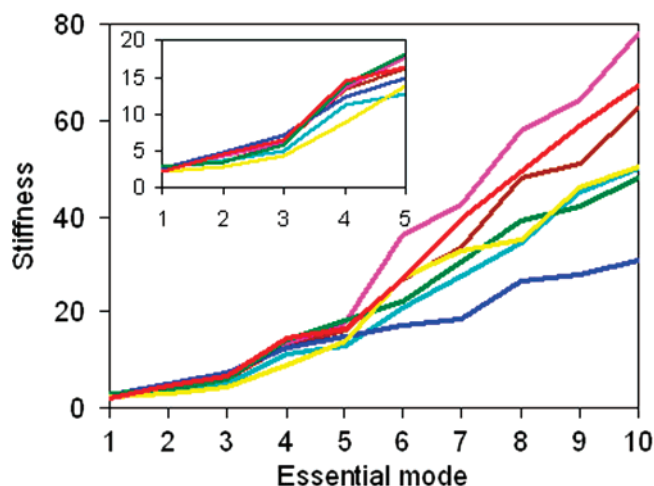
ment along the different eigenvectors that better reduces the RMSd from the original to the target structure while keeping eq 7 to a minimum. In practice, this distance can be assimilated as the easiest deformation pattern to drive a transition assuming a harmonic relationship between displacement from minimum and energy. Note that in the harmonic limit the energy associated to a displacement along normal modes can be easily determined from Mahalanobis metrics as shown in eq 8.

$$E = \frac{k_{\mathrm{B}} T}{2} d_{\mathrm{M}}^2 \qquad (8)$$

In both cases, the transition was considered completed when the perturbed unbound reached a cutoff defined by eq 9

(around 0.8 and 1.3 Å for antisense and sense strands respectively) that takes in to account RMSd value of DNA·RNA trajectory with a bound conformation caused by thermal fluctuations.

$$RMSd(XNA,HYB_{1ZBI}) \leq RMSd_{avg}(HYB_{MD},HYB_{1ZBI}) - 2RMSd_{sd}(HYB_{MD},HYB_{1ZBI}) \quad (9)$$

where indexes avg and sd stand for average and standard deviation, respectively.

Entropies were determined by diagonalization of the mass-weighted covariance matrix obtained during the trajectory using the Schlitter model[60] for all common atoms (eq 10). Entropy estimates at infinite simulation time were obtained using Harris's extrapolation technique[61] (eq 11),

$$S \approx 0.5k \sum_i \ln\left(1 + \frac{e^2}{\alpha_i^2}\right) \quad (10)$$

where $\alpha_i = \hbar\omega_i/k_BT$, ω denotes the eigenvalues obtained by diagonalization of the mass-weighted covariance matrix, and the sum extends to all of the nontrivial vibrations (all of the other symbols have the standard physical meaning);

$$S(t) = S_\infty - \frac{\alpha}{t^\beta} \quad (11)$$

where α and β are fitted parameters and t is the simulation time.

Finally, the elastic force-constants associated with helical deformation at the base-pair step level were determined by inversion of the covariance matrix in helical space, which yields a stiffness matrix (eq 12) whose diagonal elements provide the stiffness constants associated to pure rotational (twist, roll, and tilt) and translational (rise, slide, and shift) deformation of the 10 unique base-pair steps. These constants were obtained for all the sequences and averaged then to provide more reliable, environment-independent, stiffness parameters for the different steps. A global view of helical deformability can be obtained by further averaging the different steps and by defining global rotational (K_{rot}) and translational (K_{trans}) deformability indexes (eqs 13 and 14)

$$\Xi = k_BTC_h^{-1}\begin{bmatrix} k_{twist} & k_{tw-ro} & k_{tw-ti} & k_{tw-ri} & k_{tw-sl} & k_{tw-sh} \\ k_{tw-ro} & k_{roll} & k_{ro-ti} & k_{ro-ri} & k_{ro-sl} & k_{ro-sh} \\ k_{tw-ti} & k_{ro-ti} & k_{tilt} & k_{ti-ri} & k_{ti-sl} & k_{ti-sh} \\ k_{tw-ri} & k_{ro-ri} & k_{ti-ri} & k_{rise} & k_{ri-sl} & k_{ri-sh} \\ k_{tw-sl} & k_{ro-sl} & k_{ti-sl} & k_{ri-sl} & k_{slide} & k_{sl-sh} \\ k_{tw-sh} & k_{ro-sh} & k_{ti-sh} & k_{ri-sh} & k_{sl-sh} & k_{shift} \end{bmatrix} \quad (12)$$

where C_h is the covariance matrix in helical space and Ξ is the stiffness matrix whose elements in the diagonal correspond to stiffness constants for rotations (twist, roll, and tilt) and translations (rise, slide, and shift) of base pairs and those out of the diagonal to coupling between helical parameters.
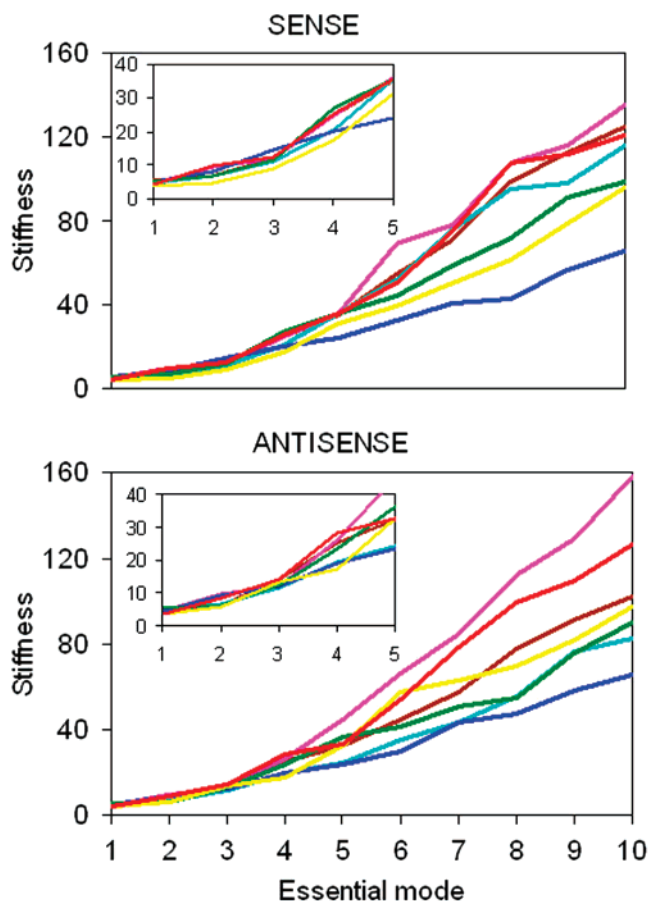
(60) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617−21.
(61) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658.

Figure 6. Force constants (in cal/mol Å²) of the first 10 essential movements (5 in the insert) of the sense (RNA) and antisense (XNA) strands of duplexes with the DD sequence (very similar plots were obtained for the other 2 sequences). Dark blue, DNA₂; red, RNA₂; green, DNA·RNA; brown, F-ANA·RNA; light blue, ANA·RNA; yellow, YpDNA·RNA; magenta, OmeNA·RNA.

$$K_{tor} = k_{twist}k_{rol}k_{tilt} \quad (13)$$

$$K_{trans} = k_{rise}k_{shift}k_{slide} \quad (14)$$

The average helical parameters were obtained for all of the steps and were then averaged to provide more reliable, environment-independent values. Standard deviations associated with the averages were obtained by the linear propagation of errors (eq 15),

$$sd_{step} = \sqrt{\sum_i sd_i^2} \quad (15)$$

where i stands for all of the occurrences of a step in the three different sequences simulated.

Standard geometrical and energetic analysis was done using *X3DNA*,[62] *PTRAJ*,[47] and in house programs. The essential dynamics was performed with the *PCAZIP* program[63] (http://mmb.pcb.ub.es/software/pcasuite.html & http://www.ccpb.ac.uk/events/workshops/previous/ analysis/) and other local programs. Clustering of duplexes according to their essential deformation pattern was done using the *R* statistical program.[64] The different

(62) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300*, 819.
(63) Meyer, T.; Ferrer-Costa, C.; Pérez, A.; Rueda, M.; Bidon-Chanal, A.; Luque, F. J; Laughton, C. A.; Orozco M. *J. Chem. Theory Comput.* **2006**, *2*, 251.
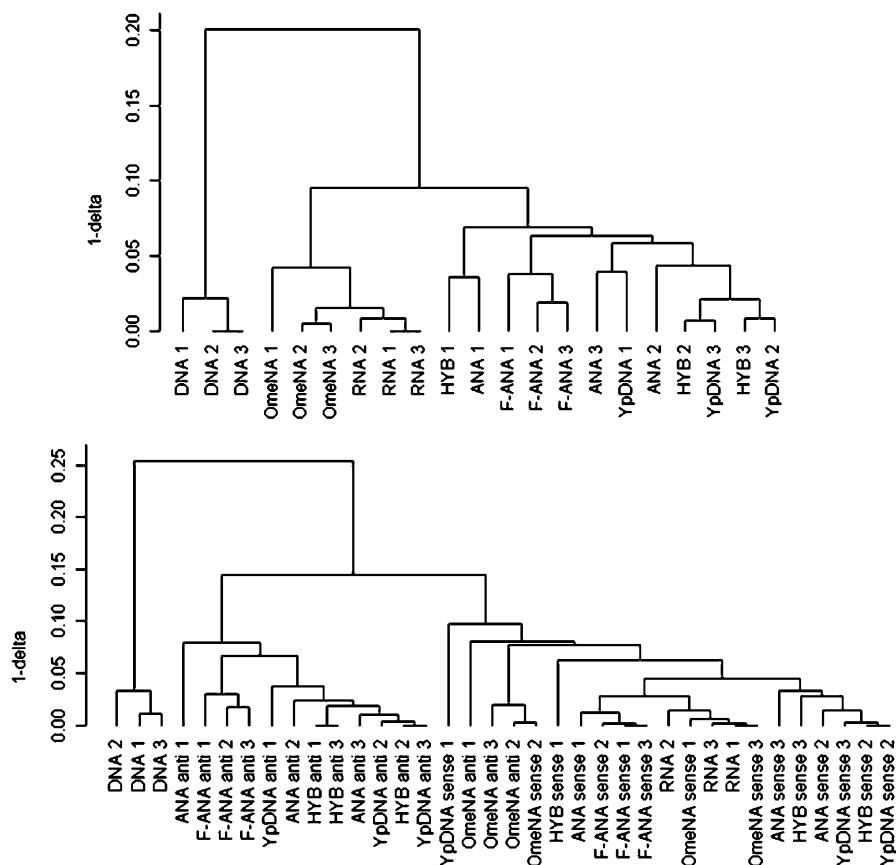(64) Ihaka, R.; Gentleman, R. *J. Comp. Graph. Stat.* **1996**, *5*, 299.

Figure 7. Clustering of the different molecules (and sequences) based on similarity measures (results correspond to $1-\delta$ distances (eq 5), though similar values were obtained using the κ index). Top, considering the simultaneous dynamics of both strands; bottom, considering the dynamics of each strand separately. Similarities were determined using the 10 central steps. 1, 2, and 3 stand for the three sequences considered here: $d/r(CGCGAATTCGCG)_2$, $d/r(CATAGGCCTATG)_2$ and $d/r(CACAGATCTGTG)_2$.

trajectories are available in compressed format (95% variance threshold) at http://mmb.pcb.ub.es/antisense/ and can be decompressed with the *PCAZIP* program (http://mmb.pcb.ub.es/software/pcasuite.html).[63]

Results and Discussion

General Structural Properties. MD simulations lead to very stable trajectories (Figure 1), where the helical structure and the pattern of Watson−Crick hydrogen bonding are well preserved. Comparison of the DNA DD trajectory presented here with a reference 1.2 μs simulation performed in our laboratory[49] shows that current simulation times are long enough to capture most of the key structural and flexibility features of the equilibrium structure of 12-mer duplexes like those considered here.

Inspection of the different trajectories (Figure 1 and Figures S4−S6 in the Supporting Information) suggest the existence of three structural groups: (i) the DNA duplex, which samples B-like conformations close to those experimentally characterized in crystal and solution phases, (ii) the DNA·RNA, YpDNA·RNA, ANA·RNA, and F-ANA·RNA hybrids, whose conformation resembles the NMR-based structure for DNA·RNA duplexes (i.e., it pertains to the A family, but with some B-like features; A/B conformation), and (iii) RNA$_2$ and OmeNA·RNA duplexes, which display a pure A conformation. The analysis of the helical parameters (Table 2) shows that all of the duplexes except DNA$_2$ pertain to the A family with average twist angles of about 30 degrees but also reveals differences in the roll and

rise between pure A duplexes and A/B structures. The sequence-dependent variability is very strong in DNA$_2$, small for pure A forms, and moderate for A/B structures. It is worth noting that sequence-dependence rules valid for one type of duplexes are not for the others, showing that not only nucleobase−nucleobase interactions (mostly stacking) but also the nature of the backbone modulate the local geometry of the different sequences.

The width of the minor groove reflects very clearly the different nature of the three conformational groups mentioned above. The DNA duplex shows a spread distribution of minor-groove widths reflecting the intrinsic flexibility of the duplex and the dependence of conformation on sequence (Figure 2). Averaging data for all of the steps and sequences yields a minor-groove width equal to 6.2 Å. The pure A-form duplexes (RNA$_2$ and OmeNA·RNA) display wide (11.2 Å), sequence-independent rigid minor grooves. Finally, the A/B hybrids show intermediate minor grooves ranging from 9.0 (ANA·RNA) to 9.9 Å (DNA·RNA). The rigidity of these minor grooves is also intermediate compared to those of pure A and B forms.

As expected, the different geometry of the grooves changes drastically the interaction profile of the duplexes (Figure 3, for the shake of simplicity this analysis is shown only for the DD sequence). Thus, pure A-form duplexes mainly interact with cationic groups along the major groove, whereas DNA$_2$ displays the classical profile with strong propensity for binding cations along the bottom of the narrow minor groove. All of the A/B duplexes display a common pattern with well-defined

Table 3. Rotational[a] and Translational[b] Helical Force Constants of the 10 Unique Dinucleotide Steps for the Different Duplexes[c]

step	DNA$_2$	RNA$_2$	DNA	F-ANA	ANA	YpNA	OmeNA
GC·GC	31.1	59.3	47.3	55.1	47.1	57.5	66.4
	27.1	27.1	26.8	29.2	26.0	36.0	27.0
	37.9	36.7	36.0	38.8	35.5	43.6	39.3
	9.7	*12.7*	*11.1*	*11.5*	*10.0*	*11.6*	*13.0*
	3.1	*3.2*	*3.0*	*3.4*	*3.0*	*3.6*	*3.8*
	1.7	*2.3*	*2.1*	*2.3*	*2.0*	*3.0*	*2.8*
GG·CC	32.3	61.8	58.0	61.8	52.3	62.8	70.8
	22.8	25.4	24.8	23.7	25.2	30.9	25.3
	43.4	49.7	45.1	48.0	46.9	51.5	50.7
	7.8	*9.5*	*8.9*	*8.5*	*9.0*	*8.8*	*9.9*
	1.7	*5.4*	*4.6*	*5.3*	*4.9*	*5.2*	*7.0*
	1.5	*2.3*	*2.2*	*2.3*	*2.3*	*2.8*	*2.8*
GT·AC	21.4	50.9	41.2	63.2	53.0	44.9	64.5
	23.1	20.0	24.6	28.5	24.3	22.9	27.8
	36.6	29.2	36.9	41.4	34.5	34.5	42.2
	8.1	*10.4*	*8.4*	*9.6*	*9.5*	*10.6*	*10.0*
	2.5	*3.1*	*3.3*	*3.7*	*2.8*	*3.4*	*3.5*
	1.1	*1.1*	*1.2*	*1.5*	*1.2*	*1.0*	*1.9*
GA·TC	33.8	63.9	50.9	57.7	47.3	59.7	67.2
	21.8	19.8	19.4	20.6	20.8	24.9	19.9
	41.9	40.2	37.0	39.9	39.1	42.2	43.3
	8.6	*8.9*	*8.1*	*8.0*	*8.0*	*8.1*	*9.2*
	2.1	*4.0*	*2.9*	*3.6*	*3.5*	*4.0*	*5.1*
	1.6	*1.5*	*1.5*	*1.8*	*1.6*	*1.9*	*1.8*
AA·TT	49.6	57.4	54.0	61.1	65.2	60.2	70.0
	25.4	19.2	21.8	22.6	21.5	22.1	21.4
	43.5	30.1	32.9	35.7	33.1	36.1	38.4
	8.1	*8.4*	*7.9*	*7.8*	*7.0*	*8.1*	*8.8*
	3.2	*3.8*	*2.9*	*3.6*	*3.6*	*3.9*	*4.6*
	2.2	*1.3*	*1.7*	*1.8*	*2.0*	*1.9*	*2.0*
AG·CT	24.9	62.9	56.3	63.3	61.6	62.1	69.7
	20.2	19.6	21.1	21.7	21.6	23.5	22.8
	39.9	34.5	36.6	37.9	35.7	38.1	41.6
	7.4	*7.8*	*6.9*	*6.9*	*6.5*	*7.4*	*8.1*
	1.7	*4.5*	*3.7*	*4.5*	*4.8*	*4.6*	*6.0*
	1.5	*1.3*	*1.5*	*1.7*	*1.8*	*1.6*	*1.6*
AT·AT	47.8	54.9	51.8	61.6	58.4	52.6	62.3
	29.2	19.5	22.8	25.8	26.5	24.9	26.5
	40.4	22.9	31.7	34.3	33.8	34.2	34.8
	8.5	*10.3*	*9.5*	*10.0*	*9.1*	*9.3*	*10.8*
	3.9	*3.3*	*3.4*	*3.9*	*3.0*	*3.4*	*3.5*
	1.6	*1.0*	*1.1*	*1.3*	*1.2*	*1.2*	*1.5*
CG·CG	11.8	59.1	45.7	55.9	42.4	38.4	62.4
	19.5	14.5	14.6	14.7	15.1	15.3	16.1
	27.7	29.3	25.6	29.8	26.4	29.5	31.0
	7.1	*3.9*	*3.7*	*3.7*	*3.9*	*3.4*	*4.2*
	2.4	*3.6*	*3.3*	*3.6*	*2.9*	*2.8*	*4.2*
	1.2	*1.4*	*1.5*	*1.7*	*1.4*	*1.4*	*1.4*
CA·TG	13.5	59.0	39.8	56.0	51.2	48.3	66.7
	15.9	15.3	14.4	17.5	16.8	16.8	17.4
	25.1	28.9	24.7	29.0	27.1	29.3	30.8
	5.5	*4.7*	*3.7*	*4.8*	*4.6*	*4.5*	*5.2*
	1.3	*4.7*	*2.3*	*3.3*	*3.7*	*3.8*	*5.1*
	1.1	*1.6*	*1.4*	*1.9*	*1.8*	*1.6*	*2.0*
TA·TA	18.9	55.2	50.5	53.8	50.1	46.2	67.0
	15.2	13.6	15.5	15.9	14.8	14.2	15.3
	23.3	24.4	24.3	25.7	24.0	22.6	28.6
	6.4	*5.1*	*5.3*	*5.2*	*5.2*	*3.7*	*5.5*
	1.5	*4.9*	*4.3*	*4.5*	*4.1*	*4.0*	*5.7*
	0.9	*2.1*	*2.3*	*2.7*	*2.3*	*1.8*	*2.7*
average	28.5	58.4	49.6	59.0	52.9	52.3	66.7
	22.0	19.4	20.6	22.0	21.3	23.2	22.0
	36.0	32.6	33.1	36.1	33.6	36.2	38.1
	7.7	*8.2*	*7.4*	*7.6*	*7.3*	*7.6*	*8.5*
	2.3	*4.1*	*3.4*	*3.9*	*3.6*	*3.9*	*4.9*
	1.4	*1.6*	*1.7*	*1.9*	*1.8*	*1.8*	*2.1*
$\Xi_{rot}(av)$[d]	23 10^3	37 10^3	34 10^3	46 10^3	38 10^3	45 10^3	58 10^3
$\Xi_{trans}(av)$	26	53	41	57	47	53	84

[a] Twist, roll, and tilt in roman and cal/mol·deg^2. [b] Rise, slide, and shift in italics and kcal/mol·Å^2. [c] For non-homopolymeric duplexes it is labeled by the antisense strand. [d] Values in the last rows correspond to average global translational and rotational constants (in kcal^3/mol^3·Å^6 and cal^3/mol^3·deg^6).
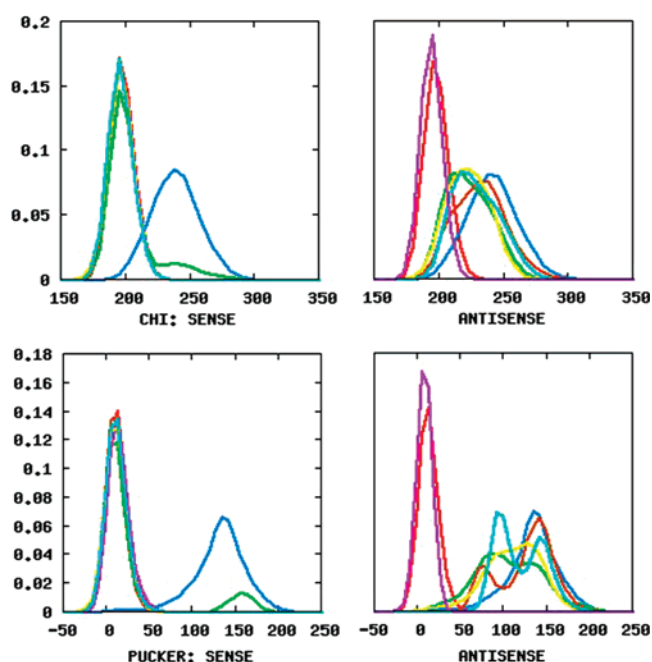
Figure 8. Population (in fractions of 1) of glycosidic torsion (χ in degrees) and phase angle (pucker in degrees) for the sense and antisense strands of the different duplexes considered here. Blue, DNA2; red, RNA2; green, DNA·RNA; brown, F-ANA·RNA; light blue, ANA·RNA; yellow, YpDNA·RNA; magenta, OmeNA·RNA. Values shown are obtained by pooling results from the three sequences.

Table 4. Percentage of Noncanonical α/γ and ϵ/ζ Torsions in the Different Duplexes[a]

α/γ	DNA$_2$	RNA$_2$	HYB	F-ANA	ANA	YpDNA	OmeNA
duplex	0.61	0.06	2.14	7.26	1.31	6.24	0.02
sense			0.25	0.64	0.94	12.39	0.04
antisense			4.04	13.87	1.67	0.1	0.01

ϵ/ζ	DNA	RNA	HYB	F-ANA	ANA	YpDNA	OmeNA
duplex	20.14	0.01	2.27	0.07	4.33	0.98	0.00
sense			1.72	0.04	0.03	0.88	0.00
antisense			2.82	0.10	8.63	1.08	0.00

[a] Because of the slow convergence of α/γ torsion, values are shown only for the duplexes with the DD sequence.

regions of favorable interaction located asymmetrically in the major groove around the phosphate groups of the sense strand. Note then that, despite the greater similarity with the A form, the A/B-like hybrids show a quite distinct interaction pattern in the grooves, which also differs from that of pure B-DNA duplexes.

Global Flexibility and Strand Asymmetry. Intramolecular entropy analysis for the duplexes with the DD sequence (those for which longer trajectories are available) demonstrates that DNA$_2$ is more flexible than any of the other duplexes studied here, whereas pure A forms, particularly the OmeNA·RNA duplex, are the most rigid ones (Figure 4). All of the A/B hybrids display entropy values intermediate between those of DNA$_2$ and RNA$_2$, with the arabino derivatives closer to RNA$_2$ and the DNA·RNA and YpDNA·RNA duplexes closer to DNA$_2$.

The higher global flexibility of DNA$_2$ relative to hybrids and these relative to pure A duplexes is not homogeneous for all of the essential deformations, as noted by the force constants associated with the essential deformation modes. Thus, for the very first modes, pure A duplexes can be very soft, but the situation is reversed after ca. the fifth mode (Figure 5), thus

Table 5. Minimum Mahalanobis Distances in Essential Deformation Space between the Structure of the Antisense (Roman), Sense (Italics) Strands in Relaxed Forms and in the RNase H-bound Conformations (Methods)[a,b,c]

sequence	Seq. 1	Seq. 2	Seq. 3	av distance	energy
DNA(a-sense)	3.41	1.96	2.54	2.69	2.17
DNA(sense)	*7.59*	*6.04*	*7.26*	*6.96*	*14.53*
RNA(a-sense)	7.82	8.60	8.34	8.25	20.3
RNA(sense)	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
HYB(a-sense)	1.92	1.35	1.49	1.59	0.75
HYB(sense)	*2.20*	*1.83*	*1.12*	*1.72*	*0.89*
F-ANA(a-sense)	1.68	2.16	1.96	1.90	1.1
F-ANA(sense)	*2.09*	*2.25*	*1.93*	*2.09*	*1.31*
ANA(a-sense)	1.80	1.34	1.62	1.59	0.75
ANA(sense)	*2.53*	*3.80*	*2.20*	*2.84*	*2.42*
YpDNA(a-sense)	1.55	1.85	1.62	1.67	0.83
YpDNA(sense)	*2.03*	*2.17*	*1.75*	*1.98*	*1.18*
OmeNA(a-sense)	8.74	9.84	7.23	8.60	22.1
OmeNA(sense)	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

[a] The last two columns correspond to the average distance values and the associated harmonic deformation energy (in kcal/mol). [b] All of the values were determined considering the central 6-mer portion (i.e., the region in contact with the protein). [c] 1, 2, and 3 stand for the three sequences considered here: $d/r(CGCGAATTCGCG)_2$, $d/r(CATAGGCCTATG)_2$, and $d/r(CACAGATCTGTG)_2$, respectively.

confirming[65] that A duplexes are guided by simple dynamics where only a very small number of modes contribute to the flexibility, whereas A/B hybrids and specially DNA_2 show a more complex deformability pattern, which involves a larger number of essential movements. Interestingly, the two strands of A/B hybrids show a quite remarkable difference in the deformability scheme, as can be seen in the stiffness constants associated with the first deformation modes of the two independent strands (Figure 6). Thus, sense strands behave close to those of pure A forms (RNA_2 and OmeNA·RNA), whereas the antisense strands (of A/B hybrids) are closer to those of DNA_2 (Figure 6). This makes clear that the duplex structure does not make the two strands uniform in terms of flexibility.

Comparison of the nature of the essential deformation movements (Methods) shows an overall good similarity between the different duplexes (Table S2 in the Supporting Information). However, a careful clustering of similarity data shows the existence of two main groups: one corresponds to the three DNA_2's considered here, whereas the other contains all of the other duplexes (Figure 7). This latter cluster contains two other well differentiated families: (i) the pure A duplexes and (ii) the A/B hybrids. Note that the differences between the essential dynamics related to sequences might be sizable but do not alter the assignment of a duplex into a given cluster, thus confirming the consistency of the deformation behavior of the duplexes and the possibility to classify duplexes on the basis of the deformability pattern. When the similarity analysis is performed for the individual strands, the asymmetry between them becomes clear (Table S2 in the Supporting Information and Figure 7). Three clusters appear: (i) DNA, (ii) all of the antisense strands of A/B hybrids, and (ii) RNA_2, OmeNA·RNA, and all of the sense strands. This demonstrates that irrespective of the nature of the hybrid, all of the sense strands maintain an internal deformability pattern that resembles that of a RNA strand in an A duplex, which differs from the deformability pattern of the antisense strands.
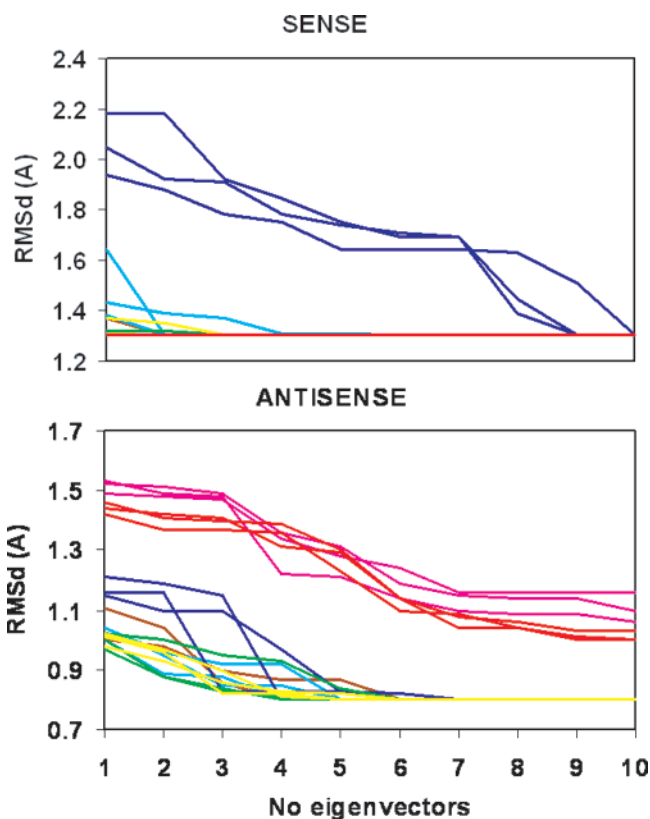
(65) Noy, A.; Perez, A.; Lankas, F.; Luque, F. J.; Orozco, M. *J. Mol. Biol.* **2004**, *343*, 627.

Figure 9. RMSd (in Å) of the sense and antisense strands to the RNase H-bound structure reached when movements are allowed along an increasing (from 1 to 10) number of eigenvectors (i.e., selecting iteratively small distances over eigenvectors to reduce as much as posible RMSd while keeping the Mahalanobis distance and associated energy to a minimum; eqs 7 and 8; Methods). Using values of RMSd produced by thermal fluctuations for DNA·RNA trajectory as reference, a transition is considered done when RMSd(sense) < 1.3 Å and RMSd(antisense) < 0.8 Å (Methods). The different lines with the same color correspond to the three sequences considered. Blue, DNA_2; red, RNA_2; green, DNA·RNA; brown, F-ANA·RNA; light blue, ANA·RNA; yellow, YpDNA·RNA; magenta, OmeNA·RNA.

Stiffness analysis associated to helical deformations provides information on the flexibility of the different duplexes at the base-pair step level. The results show that DNA_2 is more flexible than all of the other duplexes both in terms of rotations and in terms of translations of base-pair steps (Table 3), which is mainly due to twist and shift deformations because no relevant stiffness differences are found for the other helical deformations. The OmeNA·RNA duplex is the stiffest one at the base-pair step level, followed by RNA_2, whose rigidity at this level is rather similar to that obtained for A/B hybrids. DNA_2 has the largest sequence-dependent variability of helical stiffness, and pure A duplexes (RNA_2 and OmeNA·RNA) display the smallest, whereas the A/B hybrids are placed in between but generally closer to the pure A family than to the DNA_2 (Table 3).

Interestingly, steps that are very stiff for one type of helical deformation might be soft for others duplexes (i.e., the CG step, which is the easiest point for unwinding in DNA_2, is quite stiff for the same deformation in RNA_2), suggesting that concepts such as step deformability or rigidity might be used with caution, and that, in contrast to with general belief, alterations in the backbone can affect in different ways distinct base steps, changing the sequence rules for stiffness (Table 3). These findings, in conjunction with the geometrical data for steps (above and Table 2), demonstrate that there is a subtle and

complex coupling between sequence-dependent helical geometry/deformability and the backbone, which must not be ignored.

The flexibility of the backbone is mostly dominated at the microscopic level by: (i) rotations around glycosidic angle (χ), (ii) sugar puckering, (iii) the concerted α/γ rotation, and (iv) ϵ/ζ coupled rotations. DNA_2 shows the largest flexibility in terms of χ torsions (Figure 8), which correlates with the largest flexibility in terms of twist. The pure A duplexes show a very sharp χ distribution (Figure 8), whereas the situation for the A/B hybrids depends on the strand (below) with the sense strand identical to pure RNA and the antisense one closer to DNA. Sugar puckering widely oscillates in the South to Southeast region for DNA_2, whereas it is fixed in the North conformation for A duplexes. Again, the two strands of A/B-duplexes show clear distinct trends: sense sugars fixed in the North conformation (a residual population of South puckering is found in DNA·RNA hybrids), whereas wide distributions of phase angles in the South−Southeast regions occur in the antisense strand.

The use of the new parmbsc0 force field allowed us to evaluate the concerted α/γ and ϵ/ζ coupled rotations, which are crucial to understand backbone flexibility. As previously reported,[40] DNA_2 displays reversible α/γ rotations in the multi-nanosecond time scale, with a minor population (<1%) of all of the α/γ pairs in noncanonical regions (Table 4). The population of noncanonical conformations becomes nearly zero for pure A duplexes but increases significantly for all of the A/B hybrids because of the tendency to display unusual α/γ conformers in the antisense strand, which seems to agree with the fact that unusual α/γ conformers are found in the DNA·RNA−RNase H complex in hybrid regions close to the enzyme active site.[37] The largest flexibility with respect to ϵ/ζ rotation is found for DNA_2 as a result of the well-known (B_I/B_{II}) transition (around 20% of all ϵ/ζ torsions are in the noncanonical conformation). Such transitions do not exist for pure A forms and are less prevalent for the hybrids, where the transitions are located in the antisense strand.

In summary, our analysis shows that the flexibility of A/B hybrids is larger than that of pure A-form duplexes and smaller than that of DNA_2. The backbone of all A/B hybrids has a unique deformation pattern, with a strong asymmetry between sense and antisense strands and a good memory of each strand of its behavior in the corresponding homopolymer.

Biologically Relevant Deformations. One of the most distinctive features of the A/B hybrids is the unique structure of the grooves. However, considering the flexibility of nucleic acids, this structural feature alone is not expected to be enough as to discriminate between RNase H substrates and nonsub-strates.[14,15] Therefore, the enzyme should exploit additional discriminative properties. Our results suggest that deformability can be a key differential feature because hybrids susceptible to RNase H have a deformability pattern clearly different to that of nonsubstrate hybrids. Thus, the higher flexibility of A/B hybrids relative to A forms should facilitate the binding to the enzyme in a suitable conformation, and the strong strand asymmetry should help the enzyme recognize and cleave the sense strand, keeping intact the antisense one.

The deformation of pure A type hybrids or RNA_2 implies a pathway in the essential space associated with high-energy requirements (Table 5), due mostly to the difficulty in deforming the very rigid antisense (and sense) strand (Figure 9). As a result, a productive binding of A-type duplexes to the enzyme is energetically very demanding. On the contrary, deformation of A/B hybrids is achieved without a significant energy cost. Finally, it is worth noting that the productive binding of DNA_2 is mostly handicapped by the energy cost of deforming the sense strand from a pure B to a pure A conformation, in other words, the cost of changing its general shape to the general A form expected by the enzyme. It is worth noting that these findings are not dependent on the sequence and should be considered universal for a given duplex type.

In summary, our results point toward a double sieve: the general shape, which precludes the binding of a B-like structure, and the flexibility pattern, which avoids the productive binding of rigid pure A forms, whose general structure is not so far from that expected by the enzyme. Our results not only rationalize the apparently paradoxical behavior of the enzyme but also outline a protocol to recognize RNase H substrates. This implies 20−50 ns MD simulations from which geometry and flexibility will be determined. On the basis of this, any hybrid can be grouped into those susceptible to the enzyme and those that cannot be recognized in a productive way, opening then the possibility for more efficient design of antisense drugs.

Supporting Information Available: Complete ref 47, Figures S1−S6, and Table S1, S2. This material is available free of charge via the Internet at http://pubs.acs.org.

JA076734U