The Origins and the Biological Consequences of the Pur/Pyr DNA·RNA Asymmetry

Montserrat Terrazas,[1,10] Vito Genna,[1,10] Guillem Portella,[2] Núria Villegas,[1] Dani Sánchez,[1] Carme Arnan,[3,4,5] Carlos Pulido-Quetglas,[3,4,5] Rory Johnson,[3,4,5,6,7] Roderic Guigó,[3,4,5] Isabelle Brun-Heath,[1] Anna Aviñó,[8] Ramon Eritja,[8] and Modesto Orozco[1,9,11,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain.

[2]Department of Chemistry, University of Cambridge, Cambridge, UK.

[3]Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, Barcelona, Spain.

[4] Universitat Pompeu Fabra (UPF), Barcelona, Spain.

[5]Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain.

[6]Department of Clinical Research, University of Bern, Bern, Switzerland.

[7]Department of Medical Oncology, Inselspital, University Hospital and University of Bern, Bern, Switzerland.

[8]Institute for Advanced Chemistry of Catalonia (IQAC), CSIC, Networking Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Jordi Girona 18-26, E-08034 Barcelona, Spain

[9]Department of Biochemistry and Biomedicine, University of Barcelona, 08028 Barcelona, Spain.

[10]These authors contributed equally.

[11]Lead contact.

*Correspondence: modesto.orozco@irbbarcelona.org

SUMMARY

We analyze the physical origin and the chemical and biological consequences of the asymmetry in DNA·RNA hybrids occurring when the ratio purine/pyrimidine (Pu/Py) is different in the DNA and RNA strands. When the DNA strand of the hybrid is Py-rich the duplex is much more stable, rigid and A-like than when the DNA strand is Pu-rich. The origins of this dramatic asymmetry are double: first the apparently innocuous substitution dT→rU produces a significant decrease in stacking and second backbone distortions are larger for DNA(Pu)·RNA(Py) hybrids than for the mirror RNA(Pu)·DNA(Py) ones. The functional impact of the structural/dynamical asymmetry in the biological activities of hybrids is dramatic and can be used to improve the efficiency of antisense-type strategies based on the degradation of hybrids by RNase H or gene-editing using CRISPR-Cas9 technology.

INTRODUCTION

RNA·DNA hybrids are believed to be a direct consequence of the replication of RNA virus and, consequently, pathological entities that need to be quickly eliminated. However these duplexes are present in uninfected eukaryotic cells, where they are found as transient structures in the transcription forks, Okazaki's fragments, and R-loops[1-3] as well as inserted in eukaryotic chromosomes.[4] Additionally, hybrids have a wide range of biomedical applications related to the knock-down or knock-out of targeted genes by RNase H antisense[5] technology, or to gene editing by CRISPR-Cas9 approaches.[6] Hence, RNA·DNA hybrids are not exotic structures created in the laboratory, but molecules with a clear biological role whose presence in the cell is controlled by a myriad of enzymatic systems,[7] and which might be the ultimate tool to fight genetic-based diseases.

From a structural point of view, RNA·DNA hybrids form regular right-handed duplexes that, based on fiber diffraction data, were first believed to be in canonical A conformation.[1,8-10] Nonetheless, recent theoretical and experimental studies have shown that RNA·DNA hybrids not always adopt a standard A-form helix, mostly due to an intrinsic asymmetry between DNA and RNA strands which maintain memory of their preferences in the respective homopolymers.[11-21] Very puzzling, and yet not fully explored is another source of asymmetry impacting stability which is originated when the ratio Purine/Pyrimidine (Pu/Py) is unbalanced between the DNA and RNA strands.[22-27] By combining spectroscopic studies, melting experiments and molecular dynamics (MD) simulations we characterize the nature of the dPu·rPy/rPu·dPy asymmetry. We also describe the changes in structural and dynamic properties originated from such asymmetry and how this affects the biological activity of the hybrid. Analysis of the data allowed us to derive very simple rules to improve the efficiency of hybrid-based therapies without modifying the targeted sequence.

RESULTS AND DISCUSSION

Stability of homo-polymers and hybrids

Melting experiments were first performed on a variety of homopurine·homopyrimidine hairpins having 100%, 70% and 50% A·T/U content in the stem and an inter $(EG)_6$-loop (Figure 1a). These hairpins are ideal model systems to avoid the structural polymorphism expected for polyG and polyA strands, as for low/medium oligonucleotide concentration, only intramolecular Watson-Crick duplexes can be formed. For each sequence, four classes of hairpins were prepared: (i) deoxypurine ·deoxypyrimidine (dPu·dPy; **1**, **5** and **9**), (ii) ribopurine ·ribopyrimidine (rPu·rPy; **4**, **8** and **12**), (iii) ribopurine ·deoxypyrimidine (rPu·dPy; **2**, **6** and **10**), and (iv) deoxypurine·ribopyrimidine (dPu·rPy; **3**, **7** and **11**). Melting temperatures change dramatically depending on the sequence and the nature of the nucleic acid. The general believe that the RNA·RNA duplexes are always more stable than the DNA·DNA ones is incorrect, as for the A-rich strands the condition is the opposite (Figure 1a and Figure S1a). Another common believe: the stability of DNA·RNA hybrids is in between that of pure DNA and pure RNA duplexes is also false, as the relative stabilities of hybrids vs homopolymers are strongly sequence-dependent (Figure 1a and Figure S1a).[28] Melting experiments show also a large difference in the stability of hybrids depending on whether the purines are concentrated in the DNA or in the RNA strand. Such a difference is strongly dependent on the A/G ratio, as for pure adenosine-duplexes rPu·dPy duplex is more than 30 ºC more stable than the dPu·rPy one, while for duplexes with only 50% of adenosines the difference is reduced to 18 ºC. The same conclusions can be reached by looking at intermolecular duplexes (Figure 1b and Figure S1b), discarding then the hypothesis that we are facing a hairpin artifact. Thus, the rPu(rich)·dPy(rich) duplex **14** is nearly 30 ºC more stable than the dPu(rich)·rPy(rich) **15** one, while control hybrids **18** and **19** display the same melting temperature. In summary, an uneven distribution of Pu and Py between the DNA and RNA strands lead to a surprisingly large difference in hybrid stability.

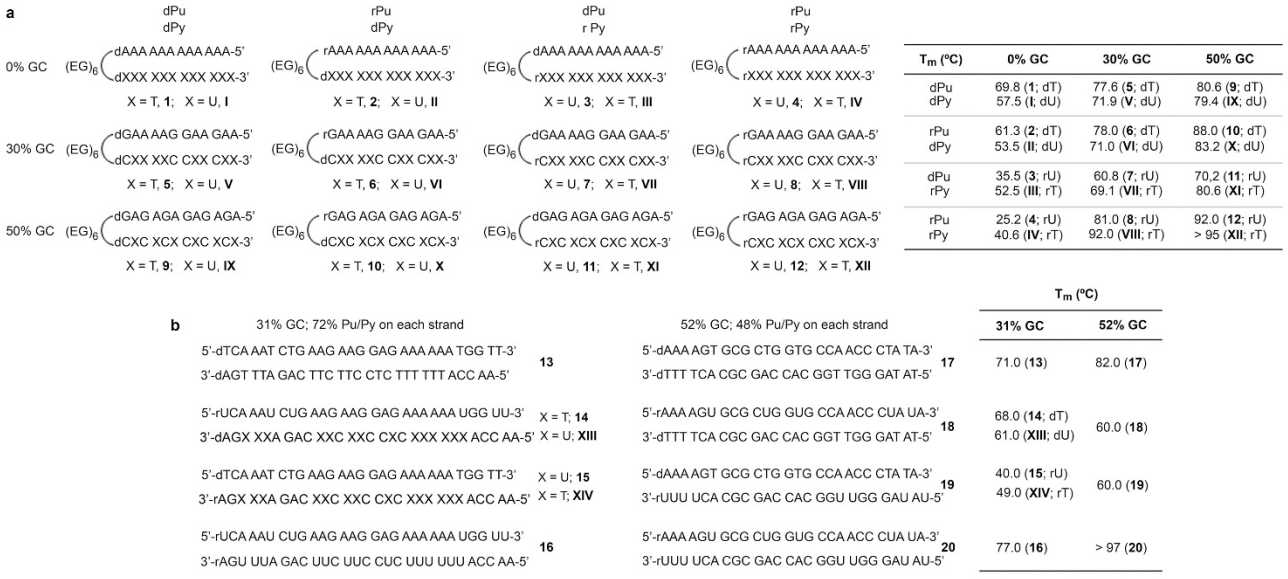

Figure 1. (a, b) Thermal stability of 12mer hairpins **1-12** and **I-XII** and large intermolecular duplexes **13-16**, **17-20, XIII** and **XIV**, in 10 mM sodium cacodylate buffer (pH 8.0) containing 100 mM NaCl and 10 mM $MgCl_2$ (see Materials and Methods for details). Duplexes numbered using Arabic notation (**1-20**) are formed by natural nucleotides, whereas the ones numbered using Roman notation (**I-XIV**) are formed by a combination of natural nucleotides and the non-natural nucleotides dU or rT.

To understand the connection between sequence asymmetry and stability, we first computed the stacking and H-bond population of the dA·rU and rA·dT hybrids, as well as of the non-natural hybrids dA·rT and rA·dU from MD ensembles. Interestingly, while the interbase H-bond population is in all the cases similar (see Figure S2 and Experimental Procedures for details of the calculations), major differences appear in the stacking energy between dA·rU and rA·dT hybrids, with the latter being favored by ± 3.75 kcal/mol per base pair, the difference mainly being assigned to Coulomb interactions of the 5-methyl group. To validate this hypothesis, we synthesize hybrids substituting d(T) by d(U) and r(U) by r(T). Results in Figure 1 show

that the change of d(T) to d(U) decreases by 8 ºC the melting temperature of hairpin **2** (a rPu·dPy hybrid), while the reverse change r(U) to r(T) stabilizes 17 ºC hairpin **3** (a dPu·rPy hybrid). Similar results are found in large intermolecular duplexes (see for example the relative stability of **14** and **XIII** or **15** and **XIV** in Figure 1b) thus confirming that better stacking of T in the pyrimidine strand is a major determinant of the greatest stability of rPu·dPy duplexes. However, detailed analysis of experimental results in Figure 1 demonstrates also that additional terms must contribute to the huge difference of stability between rPu·dPy and dPu·rPy hybrids, as dA·rT (**III**) melts at a temperature 9 ºC lower than rA·dT (**2**), and substitution of rU by rT in intermolecular duplex **15** (i.e. the change from **15** to **XIV**) is not sufficient to recover the stability of the natural hybrid **14** (see Figure 1; similar conclusions can be reached by looking at duplexes **6**, **7** and **10**, **11**). The discussion of the origin of this hidden term requires a detailed structural and physicochemical analysis of both rPu·dPy and dPu·rPy hybrids.

Structural studies

The CD spectra of the different intermolecular duplexes **13-20** and **XIII-XIV** confirm that the overall structure is closer to the A- than to the B-form,[29] but also that the hybrid duplexes are not canonical A-type duplexes (Figure S3). The differences between dPu(rich)·rPy(rich) and rPu(rich)·dPy(rich) spectra are modest, but not negligible indicating subtle conformational differences between both families of hybrids. In any case, structural interpretation of CD data is difficult due to the convoluted connection between helical geometry and the spectra. Thus, to gain detailed structural and mechanistic insights, we performed all-atom molecular dynamics (MD) simulations of hybrid duplexes dPu·rPy and rPu·dPy along with the control homo duplexes (DNA or RNA duplexes) (see Experimental Procedures and Supplemental Information). Ensembles are consistent with an overall A-like structure for both hybrids (Figure 2 and Figure S4), as suggested by CD spectra and all previous knowledge (see Introduction). However, differences from

standard A- conformation are clear, with the rPu·dPy being closer to the A-form than the dPu·rPy one, which is more flexible and has structural characteristics approaching to the B-form (Figure 2). The structural differences between dPu·rPy and rPu·dPy duplexes are clear in the distribution of many geometrical parameters, like twist, roll, sugar pucker and groove dimensions which are always much more A-like for the most stable and rigid rPu·dPy hybrid than for the dPu·rPy one (Figure 2). Experimental data available is too scarce (see Suppl. Table S1) to obtain statistically significant results, however the available one (suppl. Table S1) strongly suggest that rPu·dPy hybrids show larger roll and shallower minor grooves than dPu·rPy duplexes, in perfect agreement with our simulations (see Figure 2 and Suppl. Table S1).

Finally, as previously suggested[17] DNA and RNA strands maintain, in the hybrid, memory of the conformation adopted in the homopolymer, which is specially visible in the population of South puckerings in the DNA strand (sugar phase around 140°), something that is not expected in an A-like duplex. As discussed above this B-memory is much stronger for dPu·rPy than for the rPu·dPy ones (Figure 2), something that fully agrees with the scarce available experimental data available (see puckering in Supp. Table S1). Clearly, backbone geometries are quite different in both types of hybrids, which might be the reason for the stacking-independent greater stability of rPu·dPy compared to dPu·rPy. To evaluate this hypothesis we compared the backbone intramolecular energy of both duplexes using the respective MD ensembles. Results summarized in Figure S5a indicate that indeed a more relaxed geometry of the backbone stabilizes rPu·dPy compared to dPu·rPy (see Figure S5b). Breaking the analysis to strands and using the homoduplexes as reference (Figure S5a), we can conclude that rigid RNA strand is reluctant to modify its geometry to adopt that required for the hybrid duplex, irrespectively of whether it is rPu·dPy or dPu·rPy. On the contrary, flexible DNA can easily adapt its geometry to that required in the hybrid, but energy cost is significantly larger for dPu·rPy than for rPu·dPy, which explains the (stacking independent) greater stability of rPu·dPy hybrids (see Supplementary Movie).

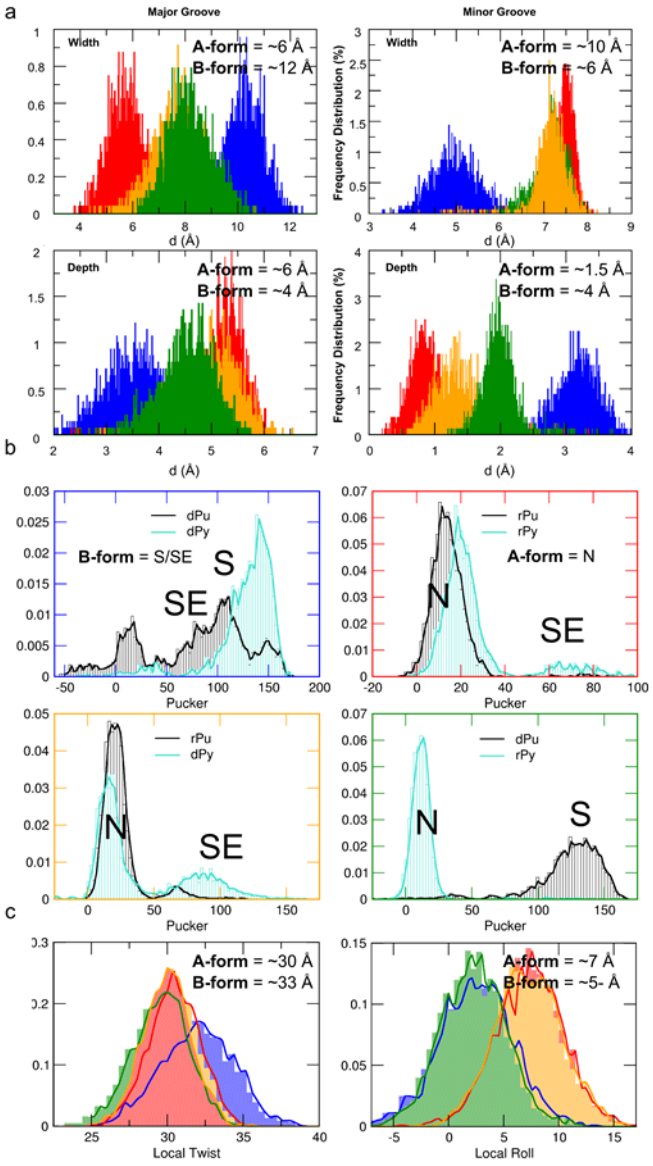

Figure 2. Structural characterization of hybrid duplexes. Color code: dPu:dPy (blue), rPu:rPy (red), dPu:rPy (green) and rPu:dPy (orange). Distances are expressed in Å while angle are measured in degrees. (a) Frequency distribution of major and minor groove structural properties. (b) Frequency distribution of sugar

pucker (phase) in each strand of all the investigated duplexes. (c) Local helical twist and roll. Canonical A- and B-from double-helix refer to references 30 and 31.

Implications of hybrid structure and stability on RNase H activity

We especulated that changes in physical properties of the hybrid implicit to the rPu·dPy/dPu·rPy asymmetry should impact biological properties of the hybrids. We explored first the impact of the Pur/Pyr DNA/RNA asymmetry in RNase H activity (Figure 3) using hybrids **14**, **15, 18** and **19**, using as negative controls the corresponding DNA·DNA (**13** and **17**; D·D) and RNA·RNA (**16** and **20**; R·R). As observed in Figure 3b, hybrids possessing 52% GC content and the same amount of purine and pyrimidine bases on each strand (**18** and **19**) are good substrates of RNAse H irrespective of the rPu·dPy or dPu·rPy configuration. On the contrary, significant differences are found for duplexes over-enriched in A·T/U (possessing 31% GC content) and with a strong Pu/Py strand asymmetry (with 72% Pu/Py on each strand) (Figure 3a). Thus, hybrid **15** (dPu-rich·rPy-rich) is as better substrate of RNase H than its rPu-rich·dPy-rich analogue (**14**). The integrated intensities of the gel bands showed that only 35% of the original hybrid (**15**) population remained intact after 5 minutes of incubation with the enzyme, while 95% of its rPu-rich·dPy-rich analogue (**14**) remained intact after the same period of time. In summary, the less stable and more flexible dPu·rPy hybrids are better substrates of the enzyme than the rPu·dPy ones. The explanation is clear: dPu·rPy structural and dynamics properties fits much better (Figure 1 and Figure S4) the two known requirements for subtrate recognition by RNase H: large flexibility[32] (average RMSD = 2.24 Å higher if compared with that of rPu·dPy) and shallow minor groove[27,32,33] (± 1.2 Å less deeper with respect to rPu·dPy duplex), as shown in Figure 2.

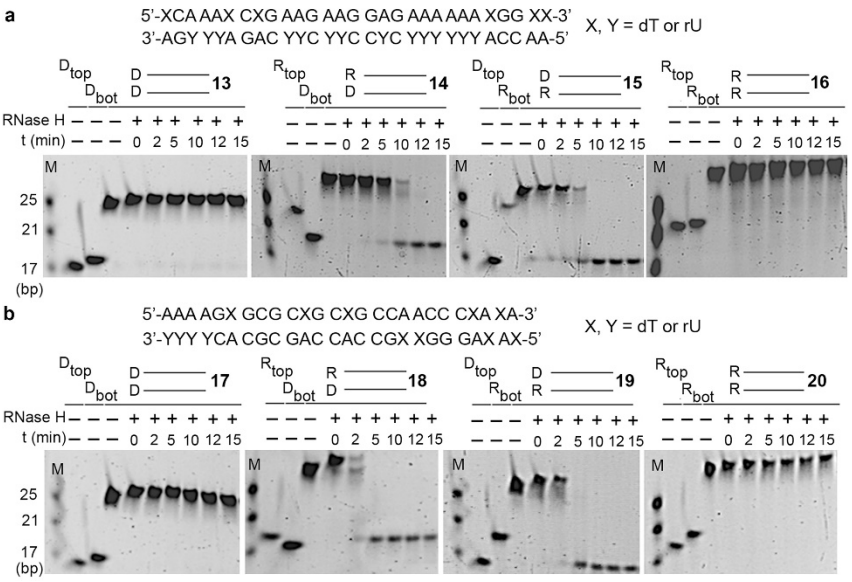

Figure 3. Analysis of the RNase H cleavage of hybrids **14**, **18** (R·D), **15** and **19** (D·R). D·D (**13** and **17**) and R·R (**16** and **20**) natural duplexes were used as controls. Annealed duplexes were incubated with RNase H at 37 ºC, and aliquots were withdrawn at the indicated time points. The reaction mixtures were analyzed by native PAGE. D: DNA strand; R: RNA strand. Top: top strand; Bot: bottom strand.

Modulation of sgRNA-CRISPR-Cas9-mediated gene deletion by modifying sgRNA base composition

We then analyzed the change in efficiency of CRISPR-Cas9-mediated cleavage of the highly expressed MALAT1[34] oncogene by changing the nature of the nucleobases in complexes formed by sgRNAs and genomic DNA (rPu-rich·dPy-rich versus dPu-rich·rPy-rich, see Experimental Procedures). As seen in Figure 4b much more intense Knock Out PCR product bands appear in the cases of cells treated with the

purine-rich sgRNAs **rAg** and **rGg** (making rPu-rich·dPy-rich hybrids) with respect to cells treated with their pyrimidine-rich **rUg** and **rCg** sgRNA counterparts (making dPu-rich·rPy-rich hybrids). Quantitative CRISPR PCR (QC-PCR)[35] of independent biological replicates (Figure 4c) confirmed higher cleavage efficiency for rPu-rich sgRNAs **rAg** and **rGg** with respect to their their rPy-rich sgRNA counterparts **rUg** and **rCg** (which form dPu-rich genomic·rPy-rich sgRNA complexes dA·rU and dG·rC, respectively). Very interestingly, the difference in efficiency of sgRNA (Pu-rich and Py-rich) is more pronounced for the A/U-rich sgRNA pair than for the G/C-rich sgRNA one (above 40% when comparing A-rich and U-rich sgRNAs **rAg** and **rUg**, $P < 0.0001$, versus around 20% when comparing the G-rich and C-rich sgRNAs **rGg** and **rCg**, $P < 0.001$). Thus, results highlight a strong asymmetry in CRISPR -Cas9 efficiency that correlates with the expected stability of the DNA·RNA hybrids. These results agree with a model in which the CRISPR-Cas9 activity is related to the ability of the incoming RNA to disrupt the duplex structure. This suggests that in general targeting Py-rich rather than the Pu-rich genomic sequences would increase CRISPR- Cas9 gene editing capabilities.

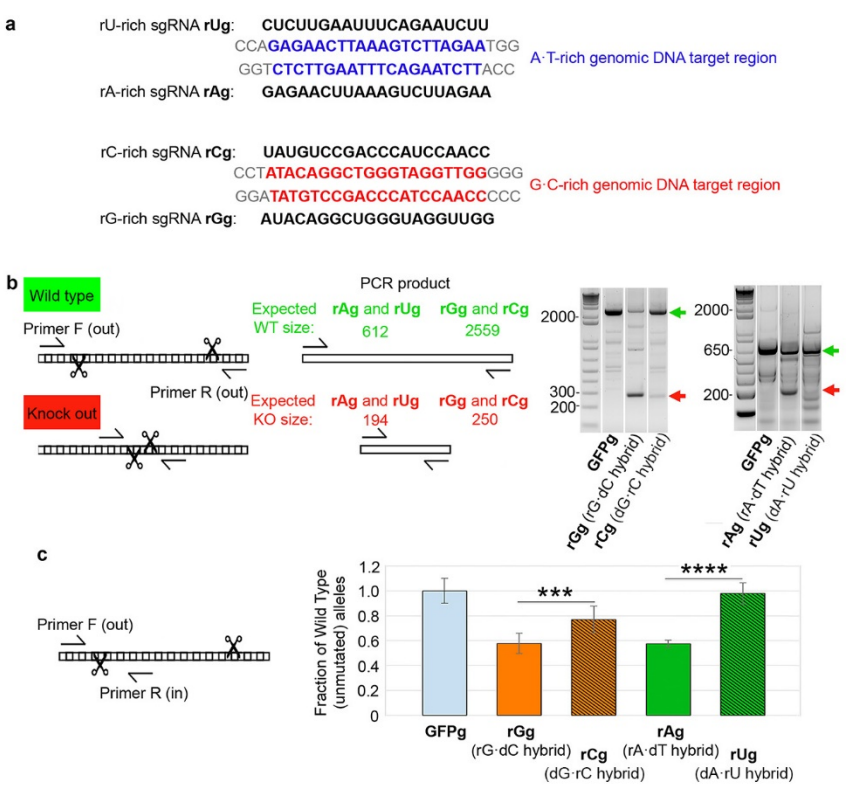

Figure 4. (a) rU-rich, rA-rich, rG-rich and rC-rich sgRNAs **rUg**, **rAg**, **rGg** and **rCg** and the corresponding A·T-rich/G·C-rich target regions of genomic MALAT1. (b) Results for genomic PCR. Left: outline of the genomic PCR method for assessing deletion of the target regions. The expected sizes of the Wild Type (WT) and the Knock Out (KO) PCR products is shown. Right: Agarose gels showing genotyping results from bulk (unsorted cells, with primers flanking the deleted region). sgRNA targeting GFP (**GFPg**) is used as negative control. (c) QC PCR results from four independent biological replicates using primers in-out. Y-axis shows the normalized fraction of the unmutated, wild type alleles, using primers amplifying the targeted region. Data were normalized to control cells transfected vector expressing sgRNA GFPg. Asterisks indicate significant differences between indicated samples (dPu·rPy and rPu·dPy hybrids): *** $P<0.001$, **** $P<0.0001$.

CONCLUSIONS

In conclusion, DNA·RNA hybrids show a double asymmetry with complex origins and dramatic biological impact. The first one is related to the fact that DNA and RNA strands behave differently, keeping structural and dynamical properties resembling those in the homopolymer. A second source of asymmetry

emerges when Pur/Pyr is uneven in the DNA and RNA strands. Hybrids where the RNA strand is Py- rich are less stable, more flexible and less A-like than those where the Pyr are concentrated in the DNA strand. The better stacking of thymidine with respect to uracil explains a significant part of the differential stability; the rest being explained by the reduced strain in the oligonucleotide backbone in dPu·rPy hybrids compared with the rPu·dPy ones. The different structural and dynamic characteristics of the two types of hybrids impact their biological properties in a quite complex manner. As a result, it is better to target a Py-rich RNA sequence than the equivalent Pu-rich one if we are pursuing RNAse H mediated degradation of a pathological RNA. On the contrary, it is better to target the Py-rich DNA strand of a gene if we want to use the hybrid to trigger CRISPR-Cas9 editing. These simple rules emerging from this study can lead to more efficient hybrid-based therapies.

EXPERIMENTAL PROCEDURES

Theoretical structural calculations. Theoretical calculations were performed on a variety of hybrids, DNA and RNA containing 50% of CG base pair. All of them were generated with Curves+ webserver[36] taking full advantage from the Arnott A-conformation as reference point.[37] The all-atom AMBER/parmbsc1[38] force-field was adopted for DNA strands while Shaw's force-field[39] was used to describe RNA strands; for discussion on the ability of these force-fields to represent both type of nucleic acids see Methods and references 38, 39, 40-43, where we demonstrate the good quality of parmbsc1(DNA) and Shaw's force-field(RNA) to reproduce regular helical structures of both DNA and RNA polymers. The length of all covalent bonds, including hydrogen atoms, was set using the LINCS algorithm,[44] allowing a time-integration step of 2 fs. All simulations were performed using GROningen MAchine for Chemical

Simulations (Gromacs) 2016 code.[45] Long-range electrostatic interactions were calculated with the particle mesh Ewald method (PME) with a real space cut-off of 12 Å and periodic boundary conditions in the three directions of Cartesian space were used. Constant temperature (310 K) was imposed using Langevin dynamics[46] with a damping coefficient of 1 ps. A constant pressure of 1 atm was maintained with Langevin-Piston dynamics[47] with a 200 fs decay period and a 50 fs time constant. All the simulated systems were hydrated using TIP3P water molecules.[48] Na^+ and Cl^- ions were added to neutralize the total charge of the system. The size of the final box was approximately 60 Å × 70 Å × 70 Å, with ~9 200 water molecules, resulting in a total number of ~28 500 atoms. The simulation protocol we have adopted is the following: all the systems were initially minimized using a steepest-descent algorithm and then were slowly heated up to 310 K in 10 ns. The first 50 ns of production run are considered as equilibration phase. Approximately ~500 ns of MD simulations were collected in the NPT ensemble for each of the five systems, resulting in a total of ~2 μs of dynamics. Coordinates of the systems were collected every 5 ps and statistics were collected considering the equilibrated trajectories only (thus discarding the first ~50 ns of simulation for each systems). Analysis of trajectory were performed using Curves+[36] and our NAFLEX server. Trajectories will be stored in our BigNAsim database (https://mmb.irbbarcelona.org/BIGNASim/index.php).

Oligonucleotide synthesis, deprotection and purification. All hairpins were synthesized via solid phase synthesis using standard phosphoramidite methods.[49] Commercially available 5'-*O*-DMT-U-, 5'-*O*-DMT-T-3'-succinyl-LCAA-CPG (Link Technologies) and Glen UnySupport 1000 CPG (Glen Research) were used as the solid supports. Phosphoramidite monomers of dA^{Bz}, dC^{Ac}, dG^{dmf}, T and dU, 2'-*O*-TBDMS-protected phosphoramidite monomers of A^{Bz}, C^{Ac}, G^{dmf}, U, and 5-Me-U, spacer CE phosphoramidite 18 (HEG), deblocking solution (3% TCA in CH_2Cl_2), activator solution (0.3 M 5-benzylthio-1-H-tetrazole in CH_3CN), CAP A solution (acetic anhydride/pyridine/THF), CAP B solution (THF/*N*-methylimidazole 84/16) and oxidizing solution (0.02 M iodine in THF/pyridine/water (7:2:1)) were obtained from commercial sources.

The coupling time was 15 min. All oligonucleotides were synthesized in DMT-ON mode. After the solid-phase synthesis, the solid support was incubated at 55 ºC for 2 h with 1.5 mL of NH_3 solution (33%) and 0.5 mL of ethanol. The supernatant was evaporated to dryness and the residue was treated with triethylamine (75 μL) and triethylamine trihydrofloride (60 μL) in DMSO (115 μL) at 65 ºC for 2.5 h. The oligonucleotides were purified using Glen-Pack Cartridges (Glen Research) following manufacturer's instructions. The purified oligonucleotides were quantified by absorption at 260 nm and confirmed by MALDI mass spectrometry (see Table S2).

UV-monitored thermal denaturation studies. Absorbance versus temperature curves of oligonucleotide duplexes were measured at 3 μM oligonucleotide concentration in 10 mM sodium cacodylate buffer (pH 8.0) containing 100 mM NaCl and 10 mM $MgCl_2$. Experiments were performed in 1 cm path length quartz cells on a Varian-Cary-100 spectrophotometer equipped with thermoprogrammer. The samples were heated to 95 ºC, allowed to slowly cool to 10 ºC, and then warmed during the denaturation experiments at a rate of 0.5 ºC/min to 100 ºC, monitoring absorbance at 260 nm. Melting temperatures (T_m) were determined by fitting of the first derivative of absorbance with respect to 1/T.

CD measurements. CD spectra (200-320 nm with a 100 nm min^{-1} scan rate) were recorded at room temperature on a Jasco J-810 spectropolarimeter under the same buffer conditions as for UV melting curves and at 15 μM oligonucleotide concentration.

RNase H cleavage reaction of natural and hybrid duplexes. Oligonucleotide duplexes (0.4 μM) were mixed with RNase H enzyme (0.02 U/μL; Thermo Scientific) in 20 mM Tris·HCl, 40 mM KCl, 8 mM $MgCl_2$, 1 mM DTT, 0.03 mg/mL BSA, pH 8.3 buffer. The mixtures were incubated at 37 °C and aliquots (5 μL) were taken from the mixture after 0, 2, 5, 10, 12, 15 and 30 minutes for their analysis by 15% non-denaturing PAGE at 4 ºC. The gels were visualized with SYBR Gold.

CRISPR-Cas9 efficiency experiments. We evaluate the efficiency of CRISPR-Cas9 editing by tacking either a Pu or a Py region as targets in a gene (MALAT1) and the sgRNA technology.[50]

sgRNA design and preparation. Nucleotide sequence of MALAT1 was obtained from GENCODE genome version hg19 and searched for pairs of target sequences meeting the following requirements: (i) a 20mer (neutral sequence, with similar A·T and G·C content) flanked by a NGG PAM sequence (target 1), and (ii) an A·T rich or G·C-rich 20mer flanked by PAM sequences in both strands {CCN(20N)NGG} (target 2). Design of the sgRNA variable sequences using the software: CRISPETa[35] (http://crispeta.crg.eu).

For each sequence, off-targets were obtained using a pre-computed database[35] and efficiency score was calculated using a previously described scoring algorithm.[51] sgRNA pairs targeting (i) either the Pur-rich or the complementary Pyr-rich strand of these A·T-rich or G·C-rich regions and (ii) the corresponding neutral region (Figure S6) were generated using DECKO (Double Excision CRISPR knockout) system, which applies a two-step cloning to generate lentiviral vectors expressing two guide RNAs (gRNAs) simultaneously.[50] This vector system uses a single 165 bp starting oligonucleotide carrying the variable sequences of both sgRNAs and expresses the two sgRNAs in tandem from U6 and H1 promoters (Figure S6). Following this approach, vectors expressing neutral sgRNA 1 (**N1g**) : rU-rich gRNA (**rUg**) pair (vector **prUg**), **N1g** : rA-rich gRNA (**rAg**) pair (vector **prAg**), neutral sgRNA 2 (N2g) : rG-rich gRNA (**rGg**) pair (vector **prGg**), and N2g : rC-rich gRNA (**rCg**) (vector **prCg**) were generated (Figure S6).

Design and cloning of plasmids. The sequences of insert-1A, insert-1U, insert-1G and insert-1C (inserts 1 corresponding to **prAg**, **prUg**, **prGg** and **prCg**, respectively) were designed by combining the four pairs of designed target sequences, as shown in Figures S7A,B. For each of the four cases, the corresponding set of oligo1_R, oligo2_F, oligo3_R, oligo4_F, oligo5_R and oligo6_F (Table S3; at 20 ng/µl concentration) were cloned using Gibson assembly method[52] into pDECKO mCherry vector (Addgene ref. 78534)[35] digested with BsmBI (Thermo Fisher). We mixed 20 ng of insert 1A, 1U, 1G or 1C with 100–150 ng of BsmBI-

digested plasmid in 10 µL volume, with 10 µL of 2x Gibson mix. We incubated the mixture at 50 °C for 1 h, and fast transformed 2 µL of this into 50 µl of z-Stbl3 competent cells (prepared with Mix and Go *E. coli* Transformation Kit, Zymo Research). The resulting intermediate plasmid, that contained additional internal BsmBI sites, was digested with BsmBI and dephosphorylated with alkaline phosphatase (Thermo Fisher EF0654). The Insert-2 sequence was previously amplified by PCR (Figure S7C) and cloned by Gibson assembly as described in Pulido-Quetglas et al.[35] 2 µL of the Gibson product were transformed into 50uL of z-Stbl3 competent cells. Clones were tested by colony PCR and by Sanger sequencing using primer sequences found in Table S4.

CRISPR-Cas9-induced gene knock-out experiments. Cas9 stable expressing HEK-293T cell line was created as decribed,[50] by transfection with Cas9 plasmid with blasticidin resistance. Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Cells were transfected with **prAg**, **prUg**, **prGg** or **prCg** plasmids, which contain puromycin resistance. 48 h later, cells were selected for at least 5 days with blasticidin (10 µg/mL) and puromycin (2 µg/mL). gDNA was extracted with NucleoSpin Tissue purification kit (Macherey-Nagel) and PCR was done from 500 ng of purified genomic DNA using primer sequences found in Table S5. Products from PCR were loaded in 2% agarose gels. Quantitative real time PCR (qPCR) from 50 ng of purified gDNA was performed using Lightcycler 480 SYBR Green master kit (Roche) on a LightCycler 480 Real-Time PCR System (Roche). Primer sequences can be found in Table S6. Target sequence primers were normalized to primers GAPDH F/R amplifying a distal, non-targeted region, using the ΔΔCt method,[53] incorporating primer efficiencies. Data were analyzed by using the GraphPad Prism 5 program (GraphPad Software). Where appropriate, the results are expressed as mean ± standard deviation (SD). *P*-values of 0.05 or less were accepted as indicators of statistically significant data. Significant differences were assessed by Student's *t*-tests. Each experiment was performed in triplicate.

SUPPLEMENTAL INFORMATION

Supplemental information includes 7 Figures and 6 Tables and can be found with this article online at .

ACKNOWLEDGMENTS

Research supported by the Spanish MINECO (BIO2015-64802-R; BFU2015-61670-EXP to M.O.) and the ERC Council (SimDNA, grant 291433, to M.O.). IRB is the recipient of a Severo Ochoa Award and Excellence from MINECO. M.T. thanks the ISCIII for Miguel Servet grants (CP13/00211 and CPII18/00032). V.G. thanks the European Molecular Biology Organization (EMBO) for financial support (ALTF 103-2018).

AUTHOR CONTRIBUTIONS

M.O. conceived and supervised the project and wrote the manuscript. R.E., R.G., R.J. and I.B.-H. supervised the project. M.T., A.A. and D.S. performed synthetic experiments. V.G. and G.P. performed the computational studies. M.T., N.V. and C.A. performed the biological work. C. P.-Q. designed the sgRNA sequences.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES AND NOTES

1. Egli, M., Usman, N., Zhang, S., and Rich,A. (1992) Crystal structure of an Okazaki fragment at 2-A resolution. Proc. Natl. Acad. Sci. USA *89*, 534-538.
2. Zheng, L., and Shen, B. (2011) Okazaki fragment maturation: nucleases take centre stage. J. Mol. Cell. Biol. *3*, 23-30.
3. Constantino, L., and Koshland, D. (2015) The Yin and Yang of R-loop biology. Curr. Opin. Cell Biol. *34*, 39-45.
4. Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., Broin, P.Ó., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A., Seoighe, C., and Greally, J.M. (2015) RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenetics & Chromatin *8*, 46.
5. Sahu, N.K., Shilakari, G., Nayak, A., and Kohli, D.V. (2007) Antisense Technology: A Selective Tool for Gene Expression Regulation and Gene Targeting. Curr. Pharm. Biotechnol. *8*, 291-304.
6. Doudna, J.A., and Carpentier, E. (2014) The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1077.
7. Wang, I.X., Grunseich, C., Fox, J., Burdick, J., Zhu, Z., Ravazian, N., Hafner, M., and Cheung, V.G. (2018) Human proteins that interact with RNA/DNA hybrids. Genome Res. *28*, 1405-1414.
8. Egli, M., Usman, N., and Rich, A. (1993) Conformational influence of the ribose 2'-hydroxyl group: Crystal structures of DNA-RNA chimeric duplexes. Biochemistry *32*, 3221-3237.
9. Conn, G.L., Brown, T., and Leonard, G.A. (1999) The crystal structure of the RNA/DNA hybrid r(GAAGAGAAGC)·d(GCTTCTCTTC) shows significant differences to that found in solution. Nucleic Acids Res. *27*, 555-561.

10. Xiong, Y., and Sundaralingam, M. (2000) Crystal structure of a DNA·RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). Nucleic Acids Res. *28*, 2171-2176.

11. Fedoroff, O.Y., Salazar, M., and Reid, B.R. (1993) Structure of a DNA : RNA Hybrid Duplex: Why RNase H Does Not Cleave Pure RNA. J. Mol. Biol. *233*, 509-523.

12. Lane, A.N., Ebel, S., and Brown, T. (1993) NMR assignments and solution conformation of the DNA·RNA hybrid duplex d(GTGAACTT) · r(AAGUUCAC). Eur. J. Biochem. *215*, 297-306.

13. Gao, X., and Jeffs, P.W. (1994) Sequence-dependent conformational heterogeneity of a hybrid DNA·RNA dodecamer duplex. J. Biomol. NMR *4*, 367-384.

14. Gonzalez, C., Stec, W., Reynolds, M.A., and James, T.L. (1995) Structure and Dynamics of a DNA·RNA Hybrid Duplex with a Chiral Phosphorothioate Moiety: NMR and Molecular Dynamics with Conventional and Time-Averaged Restraints. Biochemistry *34*, 4969-4982.

15. Gyi, J.I., Conn, G.L., Lane, A.N., and Brown, T. (1996) Comparison of the Thermodynamic Stabilities and Solution Conformations of DNA·RNA Hybrids Containing Purine-Rich and Pyrimidine-Rich Strands with DNA and RNA Duplexes. Biochemistry *35*, 12538-12548.

16. Fedoroff, O.Y., Ge, Y., and Reid, B.R. (1997) Solution structure of r(gaggacug):d(CAGTCCTC) hybrid: implications for the initiation of HIV-1(+)-strand synthesis. J. Mol. Biol. *269*, 225-239.

17. Gyi, J.I., Lane, A.N., Conn, G.L., and Brown, T. (1998) Solution Structures of DNA·RNA Hybrids with Purine-Rich and Pyrimidine-Rich Strands: Comparison with the Homologous DNA and RNA Duplexes. Biochemistry *37*, 73-80.

18. Szyperski, T., Götte, M., Billeter, M., Perola, E., Cellai, L., Heumann, H., and Wüthrich, K. (1999) NMR structure of the chimeric hybrid duplex r(gcaguggc)·r(gcca)d(CTGC) comprising the tRNA-DNA junction formed during initiation of HIV-1 reverse transcription. J. Biomol. NMR *13*, 343-355.

19. Hantz, E., Larue, V., Ladam, P., Le Moyec, L., Gouyette, C., and Dinh, T.H. (2001) Solution conformation of an RNA–DNA hybrid duplex containing a pyrimidine RNA strand and a purine DNA strand. Int. J. Biomol. Macromol. *28*, 273-284.

20. Gyi, J.I., Gao, D., Conn, G.L., Trent, J.O., Brown, T., and Lane, A.N. (2003) The solution structure of a DNA·RNA duplex containing 5-propynyl U and C; comparison with 5-Me modifications. Nucleic Acids Res. *31*, 2683-2693.

21. Noy, A., Luque, F.J., and Orozco, M. (2008) Theoretical Analysis of Antisense Duplexes: Determinants of the RNase H Susceptibility. J. Am. Chem. Soc. *130*, 3486-3496.

22. Hall, K.B., and McLaughlin, L.W. (1991) Thermodynamic and structural properties of pentamer DNA·DNA, RNA·RNA and DNA·RNA duplexes of identical sequence. Biochemistry *30*, 10606-10613.

23. Roberts, R.W., and Crothers, D.M. (1992) Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science *258*, 1463-1466.

24. Hung, S.-H., Yu, Q., Gray, D.M., and Ratliff, R.L. (1994) Evidence from CD spectra that d(purine)-r(pyrimidine) and r(purine)-d(pyrimidine) hybrids are in different structural classes. Nucleic Acids Res. *22*, 4326-4334.

25. Ratmeyer, L., Vinayak, R., Zhong, Y.Y., Zon, G., and Wilson, W.D. (1994) Sequence Specific Thermodynamic and Structural Properties for DNA·RNA Duplexes. Biochemistry *33*, 5298-5304.

26. Lesnik, E.A., and Freier, S.M. (1995) Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure. Biochemistry *34*, 10807-10815.

27. Suresh, G., and Priyakumar, U.D. (2014) DNA–RNA hybrid duplexes with decreasing pyrimidine content in the DNA strand provide structural snapshots for the A- to B-form conformational transition of nucleic acids. Phys. Chem. Chem. Phys. *16*, 18148-18155.

28. Gavette, J.V., Stoop, M., Hud, N.V., and Krishnamurthy, R. (2016) RNA-DNA chimeras in the context of an RNA world transition to an RNA/DNA world. Angew. Chem. Int. Ed. *55*, 13204-13209.

29. Kypr, J., Kejnovská, I., Bednářová, K., and Vorlíčková, M. (2012) Circular dichroism spectroscopy of nucleic acids, Comprehensive chiroptical spectroscopy. In Applications in stereochemical analysis of synthetic compounds, natural products, and biomolecules, N. Berova, P.L. Polavarapu, K. Nakanishi and R.W. Woody, eds. (John Wiley & Sons), pp. 573-584.

30. Noy, A., Pérez, A., Márquez, M., Luque, F.J., and Orozco, M. (2005) Structure, recognition properties, and flexibility of the DNA·RNA hybrid. J. Am. Chem. Soc. *127*, 4910-4920.

31. Hantz, E., Larue, V., Ladam, P., Le Moyec, L., Gouyette, C., and Huynh-Dinh, T. (2001) Solution conformation of an RNA-DNA hybrid duplex containing a pyrimidine RNA strand and a purine DNA strand. Int. J. Biol. Macromol. *28*, 273-284.

32. Sarafianos, S.G., Das, K., Tantillo, C., Clark,Jr, A.D., Ding, J., Whitcomb, J.M., Boyer, P.L., Hughes, S.H., and Arnold, E. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J. *20*, 1449-1461.

33. Nowotny, M., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2005) Crystal Structures of RNase H Bound to an RNA/DNA Hybrid: Substrate Specificity and Metal-Dependent Catalysis. Cell *121*, 1005-1016.

34. Gutschner, T., Baas, M., and Diederich, S. (2011) Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. Genome Res. *21*, 1944-1954.

35. Pulido-Quetglas, C., Aparicio-Prat, E., Arnan, C., Polidori, T., Hermoso, T., Palumbo, E., Ponomarenko, J., Guigó, R., and Johnson, R. (2017) Scalable Design of Paired CRISPR Guide RNAs for Genomic Deletion. PLOS Comput. Biol. *13*: e1005341.

36. Blanchet, C., Pasi, M., Zakrzewska, K., and Lavery, R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. Nucleic Acids Res. *39*, W68-73.

37. Selsing, E., Arnott, S., and Ratliff, R.L. (1975) Conformations of poly[d(A-T-T)]·poly[d(A-A-T)] J. Mol. Biol. *98*, 243-244.

38. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., Portella, G., Battistini, F., Gelpi, J.L., González, C., Vendruscolo, M., Laughton, C.A., Harris, S.A., Case, D.A., and Orozco, M. (2016) Parmbsc1: a refined force field for DNA simulations. Nat. Methods *13*, 55-58.

39. Tan, D., Piana, S., Dirks, R., and Shaw, D.E. (2018) RNA force field with accuracy comparable to state-of-the-art protein force fields. Proc. Natl. Acad. Sci. USA *115*, E1346-E1355.

40. Dans, P.D., Gallego, D., Balaceanu, A., Darré, L., Gómez, H., and Orozco, M. (2019) Modeling, simulations and bioinformatics at the service of RNA structure. Chem *5*, 51-73.

41. Dans, P.D., Ivani, I., Hospital, A., Portella, G., González, C., and Orozco, M. (2017) How accurate are accurate force-fields for DNA. Nucleic Acids Res. *45*, 4217-4230.

42. Kuzmanic, A., Dans, P.D., and Orozco, M. (2019) An in-depth look at DNA crystals through the prism of molecular dynamics simulations. Chem *5*, 649-663.

43. Dans, P.D, Walther, J., Gómez, H., and Orozco, M. (2016) Multiscale simulation of DNA. Current Opin. Struct. Biol. *37*, 29-45.

44. Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. *18*, 1463-1472.

45. Berendsen, H.J.C., Vanderspoel, D., and Vandrunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. *91*, 43-56.

46. Grest, G.S., and Kremer, K. (1986) Molecular dynamics simulation for polymers in the presence of a heat bath. Phys. Rev. A *33*, 3628-3631.

47. Feller, S.E., Zhang, Y.H., Pastor, R.W., and Brooks, B.R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. *103*, 4613-4621.

48. Jorgensen, W.L., Chandrasekhar, J., and Madura, J.D. (1983) Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926-935.

49. Beaucage, S.L., and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. *22*, 1859-1862.

50. Aparicio-Prat, E., Arnan, C., Sala, I., Bosch, N., Guigó, R., and Johnson, R. (2015) DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. BMC Genomics *16*: 846.

51. Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hedge, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014) Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nature Biotechnol. *32*, 1262-1267.

52. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison 3rd ,C.A., and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343-345.

53. Schmittgen, T.D., and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C_T method. Nat. Protoc. *3*, 1101-1108.