

# Crystal structure, stability and *in vitro* RNAi activity of oligoribonucleotides containing the ribo-difluorotoluy nucleotide: insights into substrate requirements by the human RISC Ago2 enzyme

Feng Li<sup>1</sup>, Pradeep S. Pallan<sup>1</sup>, Martin A. Maier<sup>2</sup>, Kallanthottathil G. Rajeev<sup>2</sup>, Steven L. Mathieu<sup>3</sup>, Christoph Kreuz<sup>4</sup>, Yupeng Fan<sup>2</sup>, Jayodita Sanghvi<sup>2</sup>, Ronald Micura<sup>4</sup>, Eriks Rozners<sup>3</sup>, Muthiah Manoharan<sup>2</sup> and Martin Egli<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry, School of Medicine, Vanderbilt University, Nashville, TN 37232, <sup>2</sup>Department of Drug Discovery, Alnylam Pharmaceuticals, Inc., 300 Third Street, Cambridge, MA 02142, <sup>3</sup>Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA and <sup>4</sup>Institute of Organic Chemistry, Center for Molecular Biosciences (CMBI), Leopold-Franzens University, 6020 Innsbruck, Austria

Received July 20, 2007; Revised and Accepted August 10, 2007

## ABSTRACT

Short interfering RNA (siRNA) duplexes are currently being evaluated as antisense agents for gene silencing. Chemical modification of siRNAs is widely expected to be required for therapeutic applications in order to improve delivery, biostability and pharmacokinetic properties. Beyond potential improvements in the efficacy of oligoribonucleotides, chemical modification may also provide insight into the mechanism of mRNA downregulation mediated by the RNA–protein effector complexes (RNA-induced silencing complex or RISC). We have studied the *in vitro* activity in HeLa cells of siRNA duplexes against firefly luciferase with substitutions in the guide strand of U for the apolar ribo-2,4-difluorotoluy nucleotide (rF) [Xia, J. *et al.* (2006) *ACS Chem. Biol.*, 1, 176–183] as well as of C for rF. Whereas an internal rF:A pair adjacent to the Ago2 ('slicer' enzyme) cleavage site did not affect silencing relative to the native siRNA duplex, the rF:G pair and other mismatches such as A:G or A:A were not tolerated. The crystal structure at atomic resolution determined for an RNA dodecamer duplex with rF opposite G manifests only minor deviations between the geometries of rF:G and the native U:G wobble pair. This is in contrast to the previously found, significant deviations between the geometries of rF:A and U:A pairs. Comparison between the structures of the RNA duplex containing rF:G and a new structure of an RNA with A:G mismatches with the structures of

standard Watson–Crick pairs in canonical duplex RNA leads to the conclusion that local widening of the duplex formed by the siRNA guide strand and the targeted region of mRNA is the most likely reason for the intolerance of human Ago2 (hAgo2), the RISC endonuclease, toward internal mismatch pairs involving native or chemically modified RNA. Contrary to the influence of shape, the thermodynamic stabilities of siRNA duplexes with single rF:A, A:A, G:A or C:A (instead of U:A) or rF:G pairs (instead of C:G) show no obvious correlation with their activities. However, incorporation of three rF:A pairs into an siRNA duplex leads to loss of activity. Our structural and stability data also shed light on the role of organic fluorine as a hydrogen bond acceptor. Accordingly, UV melting ( $T_M$ ) data, osmotic stress measurements, X-ray crystallography at atomic resolution and the results of semi-empirical calculations are all consistent with the existence of weak hydrogen bonds between fluorine and the H-N1(G) amino group in rF:G pairs of the investigated RNA dodecamers.

## INTRODUCTION

Messenger RNAs (mRNAs) represent a necessary step in gene expression, being the intermediary between gene and protein. Short interfering RNAs (siRNAs) are duplexes of RNA, 21–23 nt long, and consist of the sense and the guide strand. The latter becomes incorporated into a complex of proteins and serves to

\*To whom correspondence should be addressed. Tel: +1 615 343 8070; Fax: +1 615 322 7122; Email: martin.egli@vanderbilt.edu

identify a complementary sequence in an mRNA. This either causes the mRNA to be cleaved by the protein complex or prevents it from being translated into protein (1–8), thus effectively silencing the encoding gene. Therefore, RNA interference (RNAi) is a powerful biological process for specific silencing of gene expression in diversified eukaryotic cells and has tremendous potential for functional genomics, drug discovery through *in vivo* target validation, and development of novel gene-specific medicine. It is hoped that this approach might be used to shut down disease-related genes in humans.

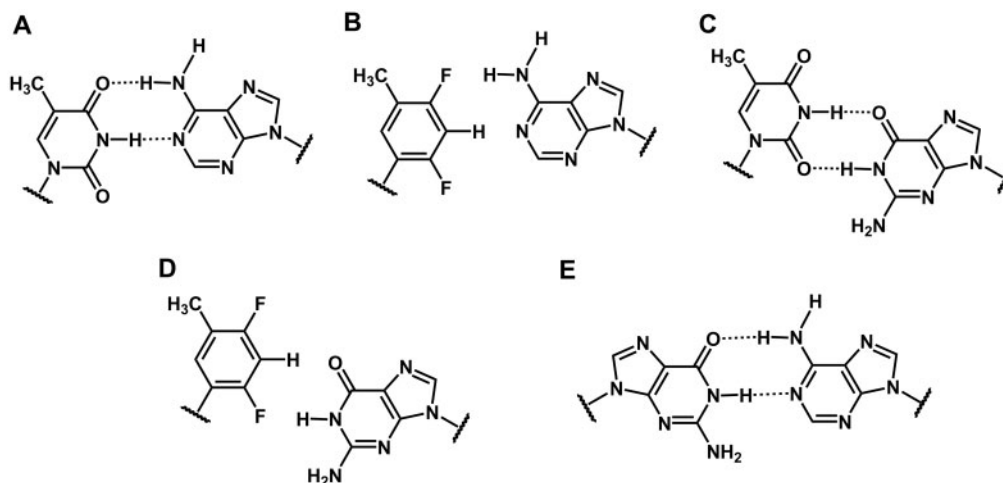
The RNAi process within cells comprises at least four sequential steps (9–15): (i) ATP-dependent processing of double-stranded RNA into siRNAs by ‘Dicer’, an RNase III-family enzyme; (ii) incorporation of these naturally derived siRNAs, or synthetic siRNAs introduced into the cell, into an inactive ribonucleoprotein complex; (iii) ATP-dependent unwinding of the siRNA duplex to generate an active RNA-induced silencing complex (RISC) whereby ATP is used to maintain 5'-phosphates on siRNAs and (iv) cleavage of the RNA target in an ATP-independent reaction. RISC is a sequence-specific endonuclease complex that contains Argonaute (Ago) proteins and the single-stranded guide siRNA strand. RISC mediates cleavage of target RNAs with multiple turnovers in the absence of additional ATP. The high specificity of the process is based on Watson–Crick pairing between the siRNA antisense (guide) oligonucleotide and the target mRNA. Within the RISC complex, the Ago2 enzyme is exclusively responsible for cleavage of the mRNA (16,17). Structural data indicate that the PIWI domain of Ago2 that adopts an RNase H-like fold is responsible for the ‘Slicer’ activity (18).

The cleavage position on the mRNA in a 21-mer siRNA duplex with mRNA is located between nucleotides 10 and 11 upstream from the target nucleotide paired to the 5'-terminus of the antisense strand (12). The sites of active duplexes are not predictable, but numerous siRNAs targeted against a particular message have been studied in attempts to correlate duplex features with function. An evaluation of the activity of 25 duplexes targeting the open reading frame and 3'-untranslated region of lamin A/C found no bias based on position within the message (19). Variations in siRNA efficiency as a function of small positional shifts on the target (19,20) could be due to competition for binding with protein factors, different thermodynamic stabilities, or sequence dependence of the RISC complex. In addition, the structure and accessibility of the target message will also play an important role (21) and prediction algorithms may offer some guidance for siRNA sequence selection (22). The most effective siRNAs have a tendency to pair more weakly at the 5'-antisense side of the duplex relative to the 3'-antisense side (22,23). Several criteria including base composition, strength of base pairing at particular positions in the duplex, overall thermodynamic stability and base identities were used to select siRNA duplexes against five genes (24). Duplexes that were selected based on these criteria resulted in a mean silencing of 76% compared to 39% mean silencing for those that were randomly selected. Introduction of mismatches in the central region of the antisense strand

was found to abolish silencing activity (12,25), but the sense strand can tolerate limited mismatches (25,26). Interestingly, when mismatches were placed in the terminal four-base region between the 5'-end of the antisense strand and the 3'-end of the sense strand improvements in activity were observed (27). Chemically modified siRNAs linked by a hairpin loop can still trigger target RNA cleavage through the RNAi pathway (19,28–30). However, it appears that recruitment by RISC requires siRNA duplexes because most antisense-only, single-stranded siRNAs show much less activity than the corresponding duplexes (10). Approaches to the successful identification of effective and specific siRNAs have recently been reviewed (25).

Chemical modification of siRNA duplexes on one or both strands will likely be required to improve their efficacy for *in vivo* applications despite the fact that short double-stranded RNAs show relatively high stability and activity in cell culture (31–38). Chemical modification is desirable to enhance: (i) the siRNA-binding affinity by, for example, conformationally preorganizing the guide strand for the mRNA; (ii) the nuclease stability and therefore the half-life of the siRNA duplexes in circulation *in vivo* and (iii) the biodistribution and pharmacokinetic properties of siRNAs and to allow targeting of siRNA to certain cell types. Incorporation of phosphorothioate (PS) linkages, 2'-OMe- or 2'-fluoro-nucleotides into siRNAs was found to affect their activities dependent on the position and extent of the modifications (35–38). For example, siRNA duplexes with the sense or antisense strand fully 2'-OMe-modified showed a significantly diminished activity in HeLa cells despite efficient silencing by the corresponding native RNAs (39). And siRNAs with both strands fully 2'-OMe-modified were inactive. On the other hand siRNAs that featured fully PS- and 2'-OMe-modified sense strands displayed similar efficacy and potency as the unmodified parent duplexes in some cases (40). The limited activity seen in some cases for duplexes with completely modified antisense strands may therefore relate to altered unwinding of the modified duplex or loading of the unmodified antisense-strand into RISC rather than inaccessibility of the target.

We recently evaluated the *in vitro* activity of siRNAs containing apolar ribo-2,4-difluorotoluy nucleotides (rF) in place of U (Figure 1A and B) (41). Modification of the guide strand at the 5'-end with rF was found not to affect silencing compared with the unmodified control. Moreover, siRNAs with an internal rF modification, despite a slightly reduced affinity for their target effectively silenced gene expression and they also exhibited improved nuclease resistance in serum. A RACE analysis demonstrated that the siRNA-mediated cleavage occurred at the predicted site both for the unmodified and the rF-containing duplex. The crystal structure of an RNA duplex with rF:A pairs revealed local conformational variations relative to canonical RNA (41). rF and A paired more loosely compared with U and A and there was an absence of ordered water molecules around the difluorotoluy moiety in both grooves. Whereas substitution of rF for U essentially had no effect on



**Figure 1.** Chemical structures of base pairs. (A) T:A; (B) F:A; (C) U:G; (D) F:G; (E) G:A. Hydrogen bonds are indicated by dashed lines.

silencing, base mismatches at the corresponding positions led to a loss of (G:A, A:A) or reduced (C:A) activity (41).

Here, we report *in vitro* activity data of siRNAs with U:G or rF:G pairs (Figure 1C and D) and crystal structures near 1Å-resolution of RNA dodecamer duplexes with either rF:G (FG 12-mer) or A:G pairs (GA 12-mer) (Figure 1E). To further investigate the relative stability of T:A, rF:A, T:G and rF:G pairs we have carried out semi-empirical calculations at the MP2/6-31G\* and MP2/6-31+G\*\* levels. The combined data allow insights into substrate specificity by hAgo2 and indicate that RNA duplex shape at or near the cleavage site is a more important determinant of RISC-mediated activity of siRNAs than thermodynamic stability.

## MATERIALS AND METHODS

### Oligoribonucleotide synthesis and purification

The rF phosphoramidite building block for incorporation into RNAs was prepared as described in ref. (41). All siRNA molecules and the rF-containing dodecamers for stability and crystallographic studies were synthesized on a 394 ABI synthesizer using previously described protocols (41). The oligonucleotides were cleaved from the CPG, with simultaneous deprotection of base and phosphate groups, with 1.0 ml of a mixture of ethanolic ammonia (ammonia-ethanol, 3:1) for 16 h at 55°C. The solution was decanted, lyophilized and re-suspended in 0.2 ml of triethylamine trihydrofluoride (TEA.3HF, Aldrich) and was incubated at 65°C for 90 min to remove the t-butyldimethylsilyl (TBDMS) groups. The completely deprotected oligonucleotides were then precipitated from anhydrous methanol (MeOH, 0.4 ml). The oligonucleotides were purified by PAGE (41) and the desired bands were excised and shaken overnight in 5 ml of 100 mM sodium acetate. The extracted oligonucleotides were desalted using C-18 Sep-Pak cartridges (Waters). The oligonucleotides were characterized by electrospray mass spectroscopy (ES/MS) and analytical

anion-exchange HPLC and/or capillary gel electrophoresis (CGE).

The GA dodecamer r(CGCGAAUUAGCG) was synthesized on a Pharmacia Gene Assembler Plus instrument using 2'-O-TOM nucleoside phosphoramidites (42) and previously described protocols (43). Deprotection and cleavage from the solid support was achieved in a mixture of MeNH<sub>2</sub> in EtOH (8 M, 0.6 ml) and MeNH<sub>2</sub> in H<sub>2</sub>O (40%, 0.6 ml) for 5 h. After the solution was completely evaporated, tetrabutylammonium fluoride trihydrate (TBAF.3H<sub>2</sub>O) in THF (1 M, 0.95 ml) was added. The reaction mixture was slowly shaken for 14 h at room temperature to remove the 2'-O-silyl ethers. The reaction was quenched by the addition of triethylammonium acetate (TEAA) (1 M, pH 7.4, 0.95 ml). The volume of the solution was reduced to 1 ml and the solution was loaded on a Amersham HiPrep 26/10 Desalting column (2.6 × 10 cm; Sephadex G25). The crude RNA was eluted with H<sub>2</sub>O and dried. Analysis of crude RNA products after deprotection was performed by anion-exchange chromatography on a Dionex DNAPac100 column (4 × 250 mm) at 80°C. Flow rate: 1 ml/min; eluant A: 25 mM Tris-HCl (pH 8.0), 6 M urea; eluant B: 25 mM Tris-HCl (pH 8.0), 0.5 M NaClO<sub>4</sub>, 6 M urea; gradient: 0–60% B in A within 45 min; UV-detection at 265 nm. The crude RNA (trityl-off) was purified on a semi-preparative Dionex DNAPac100 column (9 × 250 mm). Flow rate: 2 ml/min; gradient: 5–10% B in A within 20 min. Fractions containing RNA were loaded on a C18 SepPak cartridge (Waters/Millipore), washed with 0.1–0.2 M (Et<sub>3</sub>NH)HCO<sub>3</sub>, H<sub>2</sub>O, and eluted with H<sub>2</sub>O/CH<sub>3</sub>CN (6/4). RNA fractions were lyophilized.

### *In vitro* analysis of luciferase expression

HeLa cells that stably express both firefly and *Renilla* luciferase (HeLa Dual-luc cells) were grown as previously described (41). *In vitro* activity of siRNAs was determined using a high-throughput 96-well plate format assay for luciferase activity. HeLa Dual-luc cells were transfected with firefly luciferase targeting siRNAs at concentrations



ranging from 0.04 to 30 nM using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. After 24 h, cells were analyzed for both firefly and *Renilla* luciferase expression using a plate luminometer (VICTOR<sub>2</sub>, PerkinElmer) and the Dual-Glo Luciferase Assay kit (Promega). Firefly/*Renilla* luciferase expression ratios were used to determine the percentage of gene silencing relative to untreated controls.

### Thermal denaturation ( $T_M$ ) and osmotic stress studies

Melting of each oligonucleotide (4  $\mu$ M) was done in 10 mM sodium cacodylate, 0.1 mM EDTA and 300 mM NaCl in presence of 0, 5, 10, 15 and 20% weight/volume of each of the three organic co-solutes in Table 1. Oligonucleotide concentrations were calculated from the nearest-neighbor approximation (44) using  $\epsilon = 1000 \text{ M}^{-1} \text{ cm}^{-1}$  ( $\lambda = 260 \text{ nm}$ ) as the extinction coefficient for rF nucleoside (45). Absorbance versus temperature profiles were measured at 280 nm on a Varian Bio 100 spectrometer equipped with a six-position Peltier temperature controller. The temperature was increased at 0.5°C per minute. Five samples were measured concurrently in the double-beam mode. The melting temperatures and thermodynamic parameters were obtained using Varian Cary software (Version 02.00). The experimental absorbance versus temperature curves were converted into a fraction of strands remaining hybridized ( $\alpha$ ) versus temperature curves by fitting the melting profile to a two-state transition model, with linearly sloping lower and upper base lines. The melting temperatures ( $T_M$ ) were obtained directly from the temperature at  $\alpha = 0.5$ . The final  $T_M$  was an approximation of usually five to eight measurements.

The thermodynamic parameters were determined from the width at the half-height of the differentiated melting curve. The fraction of strands remaining hybridized ( $\alpha$ ) versus temperature curves were converted into differentiated melting curves [ $\delta\alpha/\delta(T_M^{-1})$  versus  $T_M$ ] using Varian Cary software (Version 02.00). The width of the differentiated melting curve at the half-height is inversely proportional to the van't Hoff transition enthalpy; for a bimolecular transition  $\Delta H = 10.14/(T_1^{-1} - T_2^{-1})$  where  $T_1$  is the lower temperature and  $T_2$  is the upper temperature (both in K) at one-half of [ $\delta\alpha/\delta(T_M^{-1})$ ] (46).

The changes in the number of water molecules associated with the melting process  $\Delta n_w$  were determined as described by Spink and Chaires (47)

$\Delta n_w = (-\Delta H/R)[d(T_M^{-1})/d(\ln a_w)]$ , where  $-\Delta H$  is the enthalpy determined from the width at the half-height of differentiated melting curve in pure buffer and  $R$  is the universal gas constant ( $1.986 \text{ cal mol}^{-1} \text{ K}^{-1}$ ). The experimentally determined values of water activity ( $\ln a_w$ ) at given co-solute concentrations were provided by Professors Spink and Chaires. The slope of the plot of reciprocal temperature (in K) of melting versus the logarithm of water activity ( $\ln a_w$ ) at different concentrations (0, 5, 10, 15 and 20%) of small co-solutes gave the value of  $d(T_M^{-1})/d(\ln a_w)$ . The final  $\Delta n_w$  were obtained by linear fitting using KaleidaGraph software (Version 3.51) with a confidence level usually better than 98%. The experimental uncertainties were obtained as previously reported (48).

### Crystallization of RNA dodecamers containing rF:G and A:G pairs and X-ray diffraction data collection

Crystals of the FG dodecamer r(CGCFAAUUGGCG) were grown at 18°C by the sitting-drop vapor diffusion method, using the Grid Screen Ammonium Sulfate (Hampton Research Inc., Aliso Viejo, CA, USA). Droplets containing 0.5 mM oligonucleotide, 0.8 M ammonium sulfate and 0.05 M MES, pH 6.0, were equilibrated against a reservoir of 1.6 M ammonium sulfate, 0.1 M MES, pH 6.0. Crystals appeared after 3 days. Crystals of the AG dodecamer r(CGCGA AUUAGCG) were grown at 18°C by the hanging-drop vapor diffusion method, using the Nucleic Acid Miniscreen (Hampton) (49). Both conditions 14 and 16 resulted in crystals. Crystals for diffraction data collection were grown from droplets containing 0.5 mM oligonucleotide, 5% 2-methyl-2,4-pentanediol (MPD), 20 mM sodium cacodylate, pH 7.0, 6 mM spermine-4HCl, 40 mM sodium chloride and 10 mM magnesium chloride that were equilibrated against a reservoir of 35% MPD. Crystals appeared after 2 days. In both cases, crystals were mounted in nylon loops and frozen in liquid nitrogen. Diffraction data were collected at 120 K on the insertion device beamlines of the Southeast Regional Collaborative Access Team (SER-CAT, 22-ID, FG dodecamer) and the DuPont-Northwestern-Dow Collaborative Access Team (DND-CAT, 5-ID, GA dodecamer) at the Advanced Photon Source (APS), Argonne, IL. The wavelengths for the data collections varied between *ca.* 0.9 and 1.0 Å. In both cases high- and low-resolution data sets were

**Table 1.** Thermodynamic stability ( $T_M$ ) and osmotic stress data for self-complementary RNA 12-mers with U:G, rF:A, rF:G and A:G pairs

Sequence	Name	$T_M^a$ (°C)	$\Delta n_w$ Ethylene glycol	$\Delta n_w$ Glycerol	$\Delta n_w$ Acetamide
CGCUAAUUGGCG	UG	57.7 ± 0.2	18.5 ± 3.0	20.3 ± 3.4	33.1 ± 3.0
CGCFAAUUGGCG	FG	54.9 ± 0.2	14.6 ± 2.1	26.0 ± 3.3	29.2 ± 2.4
CGCGAAUUGGCG	GF	53.9 ± 0.3	13.5 ± 2.4	24.7 ± 3.7	29.6 ± 3.1
CGCFAAUUAGCG <sup>b</sup>	FA	53.2 ± 0.3	13.2 ± 3.1	12.7 ± 3.3	23.6 ± 2.9
CGCGAAUUAGCG	GA	58.2 ± 0.2	21.2 ± 2.6	23.8 ± 3.1	Not measured
CGCAAUUGGCG <sup>c</sup>	AU	58.7 ± 0.6	22.8 <sup>d</sup>	18 <sup>d</sup>	43.2 <sup>d</sup>

<sup>a</sup>4  $\mu$ M RNA, 300 mM NaCl, 10 mM Na cacodylate, pH 6.5.

<sup>b</sup>The crystal structure of this duplex is described in ref. (41).

<sup>c</sup>Taken from ref. (48).

<sup>d</sup>Experimental uncertainties were not determined.

collected separately. Diffraction data were processed with the programs HKL2000 (50) (FG) and XDS (51) (GA).

### Structure determination and refinement

Both structures were determined by the molecular replacement method using the program EPMR (52). For the FG RNA structure, a canonical A-RNA dodecamer constructed with the program TURBO-FRODO (53) was used as the search model. A previously published structure at low resolution (54) served as the search model for structure determination of the GA RNA duplex. In both cases, the structures were initially refined with CNS (55) and at a later stage with the program SHELX-97 (56). In the subsequent refinement cycles, solvent water molecules were added, and anisotropic temperature factors refinement was carried out. The program TURBO-FRODO was used to display electron density maps for manual rebuilding of the models. To calculate the R-free, 5% randomly chosen reflections were set aside in both cases.

### Computational methods

*Ab initio* calculations were carried out using the GAUSSIAN 03 code (57) on a 64-bit Linux machine. Optimizations were initiated at the HF-3-21G level, followed by calculations at the MP2/6-31G\* and finally at the MP2/6-31+G\*\* levels. No constraints, i.e. to maintain planarity of the base pairs were used. Interaction energies were computed as the difference in energy between the optimized base pair on one hand, and the sum of energies of optimized components on the other. To calculate the point energies of base pairs based on the crystallographic conformations, hydrogen atoms were added by the program. All quantities were corrected for basis set superposition error by the standard counterpoise procedure (58).

### Coordinates

Final coordinates and structure factors for the FG and GA RNAs have been deposited in the Protein Data Bank (<http://www.rcsb.org>.) The PDB ID codes are 2Q1O (FG duplex) and 2Q1R (GA duplex).

## RESULTS

### Gene silencing activity of siRNAs with U or C replaced by rF

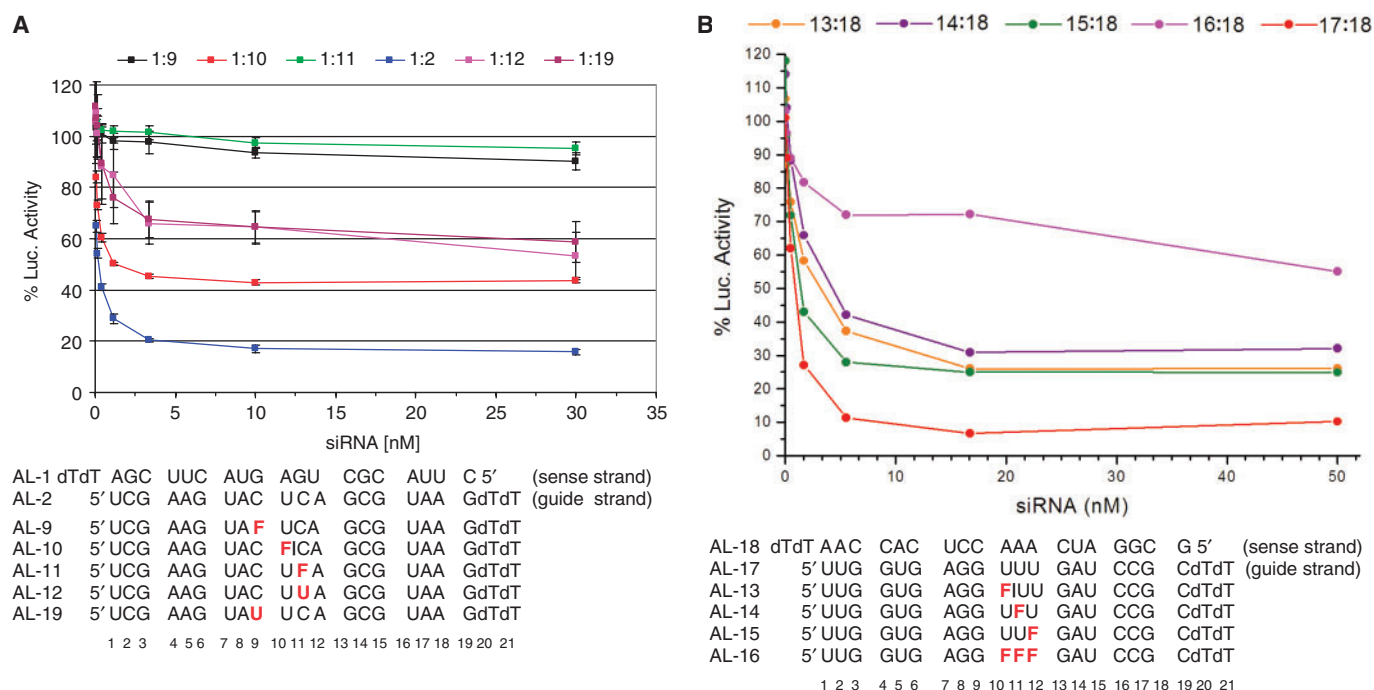
To examine the effects on gene silencing of rF incorporation into siRNAs, we employed an *in vitro* firefly luciferase assay in HeLa cells. Two regions of the luciferase gene were targeted separately with RNA duplexes whose guide strands contained single or multiple rF residues (Figure 2) (41). The silencing activities of modified duplexes applied in concentrations ranging between 0.04 and 30 nM were measured in the form of a decreased luciferase luminescence and compared with the signals resulting from treatment with native control siRNA duplexes. A single uridine positioned three or more nucleotides upstream or downstream from the cleavage site in the guide strand could be replaced by rF without a significant change in the

50% inhibitory concentration ( $IC_{50}$ ) (41). Moreover, replacement of the U adjacent to the cleavage site by rF only increased the  $IC_{50}$  from 0.21 to 1.2 nM (Figure 2A). However, substitution of either C<sub>9</sub> or C<sub>11</sub> for U or rF, resulting in the formation of U:G and rF:G pairs, respectively, significantly diminished or even abolished the silencing activity altogether ( $IC_{50} > 30$  nM; Figure 2A). Similarly, replacement of U<sub>10</sub> by A or G resulted in a loss of activity, whereas a U<sub>10</sub>→C substitution (leading to a C:A mismatch) yielded an  $IC_{50}$  of 0.92 nM, comparable to the activity of an siRNA duplex with the U<sub>10</sub>→rF substitution (41). Therefore, it appears that a single purine–purine mismatch pair or U:G/rF:G wobble pairs (assuming the rF:G pair adopts a wobble geometry) at or adjacent to the cleavage site are not tolerated by the hAgo2 endonuclease. The C:A mismatch and rF:A constitute noteworthy exceptions. A RACE analysis of the cleavage products had established that the position of the cleavage site remained unchanged as a result of using siRNA duplexes with single rF:A pairs compared with native siRNAs (41).

Whereas a single rF residue in place of U was tolerated in the guide strand, consecutive replacement of three Us also led to a loss of activity (Figure 2B). The thermodynamic stabilities (UV melting  $T_M$ 's) of siRNA 21-mer duplexes with single rF:A, A:G, A:A or C:A pairs are comparable (between 65 and 67°C relative to the native construct = 73°C) (41). Therefore, the observed difference in silencing activity between siRNA duplexes containing rF:A or any of the above mismatch pairs is unlikely to be a consequence of reduced stability. Conversely, three consecutive rF:A pairs may lead to local melting of the RNA duplex and considerably distort or destroy the secondary and tertiary structure of the hAgo2 substrate.

### Thermodynamic stability of and osmotic stress measurements for oligoribonucleotides with U→rF or C→rF substitutions

To assess the effect of rF incorporation opposite A or G on the thermodynamic stability of an RNA dodecamer duplex we carried out UV melting experiments. Initial runs at a wavelength of 260 nm showed nonlinear upper or lower base lines, a problem that was eliminated by measuring optical densities as a function of temperature at 280 nm. The resulting  $T_M$ 's are listed in Table 1. For the comparison here between the six sequences we have listed  $T_M$  differences per modified base pair (note that there are two substitutions per self-complementary dodecamer duplex). Remarkably, substituting U for rF opposite G resulted in a  $T_M$  reduction of just 1.4°C per modification (UG versus FG duplex). In the case of the dodecamer with G and rF switched (GF) the reduction in  $T_M$  relative to the UG duplex amounts to 1.9°C per modification. Switching a U:G (UG) to an rF:A pair (FA) lowers the  $T_M$  by 2.3°C. Therefore, the rF:G is more stable than the rF:A pair ( $\Delta T_M = 0.9^\circ\text{C}$ ; FG versus FA duplex). The difference in  $T_M$  between the FA and A:U pairs (AU) is 2.8°C. For completeness, we also measured the stability of the GA duplex with two G:A mismatch pairs. The GA duplex was slightly less stable than the AU duplex ( $\Delta T_M = -0.3^\circ\text{C}$ ; GA versus AU duplex) but slightly



**Figure 2.** Concentration-dependent *in vitro* silencing activity of siRNA duplexes directed against firefly luciferase relative to *Renilla* luciferase. (A) Incorporation of a single rF residue opposite A in the sense [data not shown, see ref. (41)] and guide strands (1:10) is tolerated albeit with a decrease in activity for the latter (reference duplex 1:2). On the other hand, incorporation of rF or U opposite G is not or significantly less tolerated (1:9, 1:11 and 1:12, 1:19, respectively). (B) Contrary to the presence of a single rF residue in the guide strand opposite A (13:18, 14:18, 15:18; reference duplex 17:18), multiple rF:A pairs in an siRNA duplex prevent silencing (16:18). In the AL-10, AL-11, AL-13 and AL-14 oligonucleotides, the rF substitution is adjacent to the cleavage site (marked by a vertical line).

more stable than the UG duplex ( $\Delta T_M = 0.3^\circ\text{C}$ ; GA versus UG duplex). Overall the observed order of stability of the pairs at pH 6.5 is: A:U > G:A > U:G > rF:G > G:rF > rF:A.

In order to gain insight into changes in RNA hydration as a result of rF incorporation, we employed the osmotic stress method (47,59). The reductions in the RNA melting temperatures upon addition of ethylene glycol, glycerol or acetamide were measured at different concentrations of these small co-solutes (48,60). The changes in the number of water molecules  $\Delta n_w$  associated with the melting process were determined as described by Spink and Chaires (47):  $\Delta n_w = (-\Delta H/R)[d(T_M^{-1})/d(\ln a_w)]$ , where  $\Delta H$  is the measured enthalpy,  $R$  is the universal gas constant ( $1.986 \text{ cal mol}^{-1} \text{ K}^{-1}$ ) and  $\ln a_w$  are experimentally determined values of water activity at given co-solute concentrations. This approach to establish the number of water molecules bound to a nucleic acid duplex is based on the assumption that the co-solutes (ethylene glycol, glycerol, acetamide) only change the water activity and do not directly interact with the RNA.

Rather than interpreting the absolute values obtained for  $\Delta n_w$  we believe it may be more instructive to consider the relative differences in amounts of water released when comparing the duplexes (Table 1). Interestingly, in the ethylene glycol and acetamide series the general trend regarding hydration followed the above order of thermodynamic stabilities ( $T_M$ ): A:U > G:A > U:G > rF:G > G:rF > rF:A. Thus, introduction of the G:A and U:G mismatches gradually decreased the hydration.

Introduction of rF as a base-pairing partner decreased  $\Delta n_w$  even more, although the effect may be considered a relatively minor one, especially if one considers the highly hydrophobic nature of the rF modification. In all three series, hydration of the duplexes with either rF:G or G:rF pairs was more favorable than the hydration of the duplex with rF:A pairs, which is in agreement with the observations made in crystal structures (*vide infra*). Overall, the trends were less clear in the glycerol series. As previously observed (48,60), the  $\Delta n_w$  values were somewhat higher in the acetamide than in the ethylene glycol or glycerol series. A possible explanation for such discrepancies might be the fact that the organic co-solutes display some interaction with the RNA (61,62). Record and coworkers (62) showed that most small organic co-solutes (including glycerol) were not completely excluded from the surface of bovine serum albumin. When used in osmotic stress experiments, such co-solutes are expected to underestimate the magnitude in change of hydration (62). Whereas the inertness of the co-solutes and reliability of the absolute numbers of  $\Delta n_w$  is therefore still under debate (61,62), we (48,60) and others (47) have used the relative trends to gain insight into the changes in hydration in native and modified nucleic acids.

#### X-ray crystallography of RNA dodecamers with rF:G and A:G mismatches

In order to gain a better understanding of the *in vitro* siRNA activity and thermodynamic stability/hydration data of RNAs containing rF residues and to compare the



geometry of the rF:G pair with that of the previously visualized rF:A pair (41), we determined the crystal structure at 1.10 Å resolution of the RNA dodecamer duplex [r(CGCFAAUUGGCG)]<sub>2</sub> with rF:G pairs (FG duplex, Table 1). The structure of the RNA dodecamer duplex [r(CGCGAAUUAGCG)]<sub>2</sub> with A:G mismatch pairs (GA duplex, Table 1) at 1.12 Å was determined to potentially provide a structural rationale for the inactivity of siRNA duplexes with an incorporated mismatch pair at the cleavage site in *in vitro* assays. A summary of crystal data and refinement parameters for both structures is provided in Table 2. Sum (2F<sub>o</sub>-F<sub>c</sub>) electron density maps based on the final models of the FG and GA duplexes around selected regions are depicted in Figure 3.

The geometries of the FG and GA duplexes (all riboses adopt a C3'-endo pucker) deviate from a more or less canonical A-form only in the vicinity of the rF:G and A:G pairs (Figure 4). rF:G pairs exhibit the shear typical for the U:G wobble pair whereby the guanine and difluorotoluene moieties are pushed into the minor and major grooves, respectively (Figures 1C and 3C). In the GA duplex, A and G are both in the standard 'anti' conformation and pair in a pseudo Watson-Crick fashion (Figures 1E and 3D). As a result, the minor groove of the GA duplex is widened by ~1 Å with a concomitant narrowing of the major groove. The FG structure features two independent duplexes per crystallographic asymmetric unit, leading to four observations of the rF:G pair.

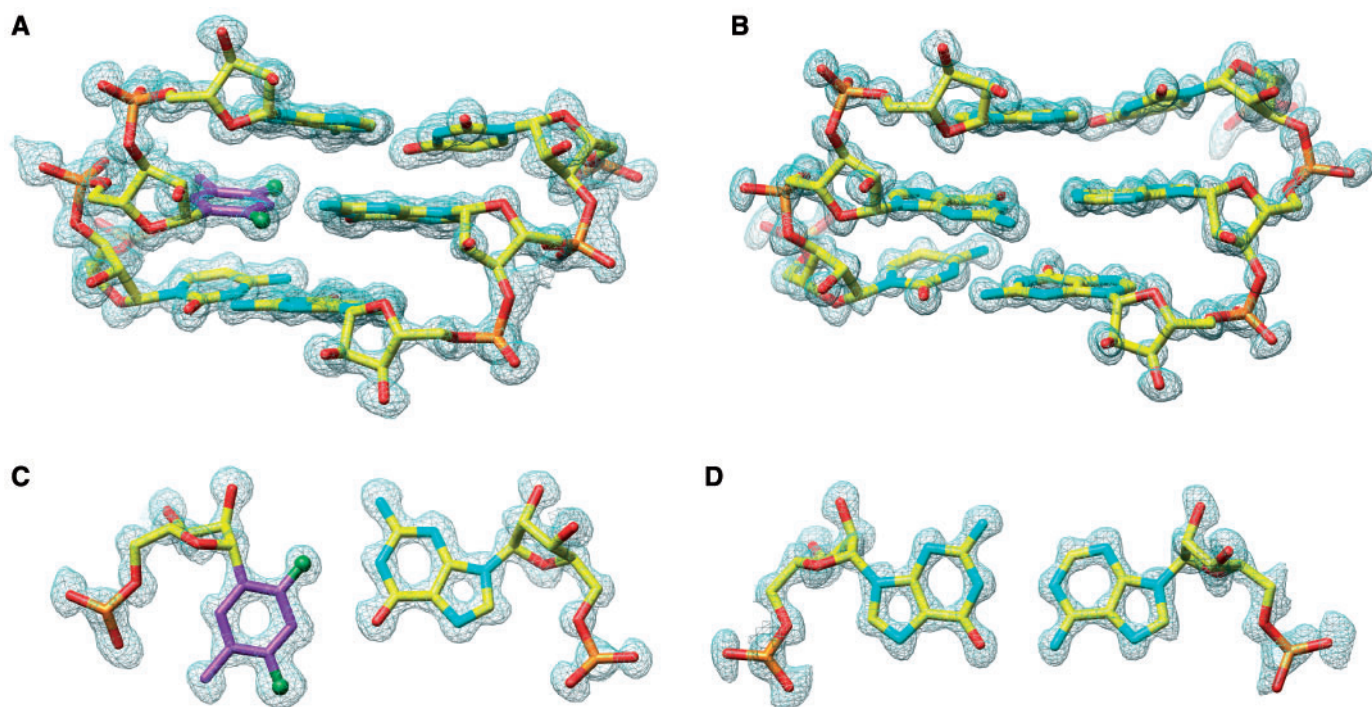
**Table 2.** Crystal data, data collection parameters and structure refinement statistics

Structure	5'-CGCFAAUUGG CG-3' (FG)	5'-CGCGAAUUAG CG-3' (GA)
Crystal data		
Space group	P1	C2
Number of strands per asym. unit	4	1
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	22.90, 32.06, 42.93	41.34, 34.89, 32.12
$\alpha$ , $\beta$ , $\gamma$ (°)	96.2, 83.9, 93.9	90.0, 129.3, 90.0
Data collection		
Number of reflections	45 752	13 514
Resolution (last shell)	1.10 (1.14–1.10)	1.12 (1.16–1.12)
Completeness (%; last shell)	93.9 (90.7)	96.4 (89.3)
<i>R</i> <sub>merge</sub> (last shell)	0.059 (0.239)	0.035 (0.277)
Refinement		
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.112/0.142	0.147/0.187
Number of atoms		
RNA	1016	255
Water	324	76
Ions	–	2 Mg <sup>2+</sup>
B-factors		
Nucleic acid (Å <sup>2</sup> )	9.9	13.4
Water (Å <sup>2</sup> )	29.0	29.7
r.m.s.d.		
Bond lengths (Å)	0.014	0.014
Bond angles (1...3; Å)	0.029	0.031

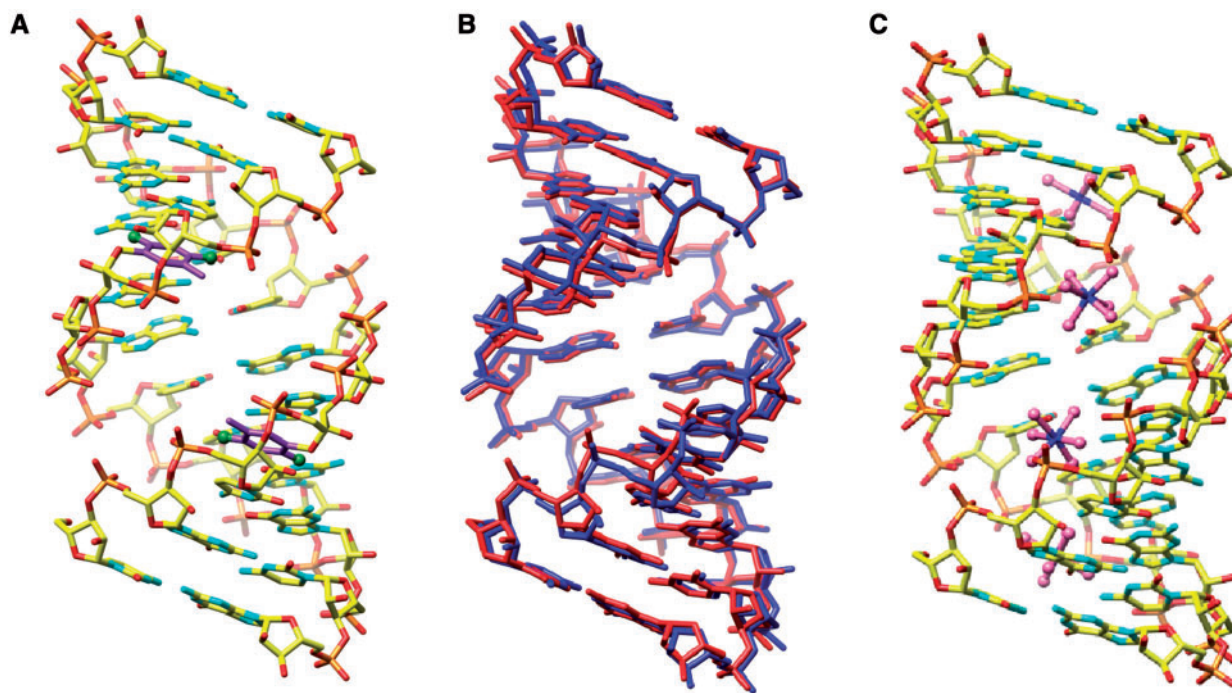
Both duplexes have very similar geometries as revealed by a superimposition (Figure 4B). The structure of the GA duplex was previously reported at a lower resolution by others (1.9 Å) (54). However, in that structure with space group *P*2<sub>1</sub> the asymmetric unit consisted of a single duplex in a general position (the presence of a strong non-crystallographic 2-fold symmetry was noted by the authors at the time). Conversely, the space group of the GA RNA crystals analyzed here is *C*2, but the unit cell constants are virtually identical to those reported earlier. The conversion from pseudo-2-fold to crystallographic 2-fold symmetry is most likely a consequence of different crystallization conditions (see Materials and Methods section). In the *C*2 structure, the duplex is located on a crystallographic dyad and the asymmetric unit consists of a single dodecamer strand (Table 2). Summaries of the helical parameters calculated with the program CURVES (63) for the two FG duplexes and the GA duplex are provided in the Supplementary Data. Residues in the GA duplex are numbered C101 to G112 and residues in the four independent strands of the FG duplexes are numbered C101 to G112 (strand 1), C201 to G212 (strand 2), C301 to G312 (strand 3) and C401 to G412 (strand 4) (Figures 3 and 5).

The four rF:G pairs exhibit similar geometries (Figure 5); the (rF)C<sub>3</sub>-H ... O<sub>6</sub>(G) hydrogen bond distance varies between 3.18 and 3.31 Å. In three of the pairs the (rF)F<sub>2</sub> ... N<sub>1</sub>(G) distance is below 3.1 Å and in the rF304:G409 pair this distance is somewhat longer (3.45 Å, Figure 5D). Therefore, the fluorine-amine interaction here is considerably shorter than the sum of the van der Waals radii. If we assume the radii of fluorine and hydrogen to be 1.35 Å and 1.2 Å, respectively (64), and a N-H distance of 1 Å, the F ... N distances in three of the four rF:G pairs are *ca* 0.5 Å below the sum (3.55 Å). The wisdom of assigning a larger radius to fluorine than to hydrogen has been questioned recently (65). Still, even we assume the same radius for fluorine and hydrogen (i.e. 1.2 Å), the observed distances between F<sub>2</sub> and N<sub>1</sub> in the rF:G pairs are indicative of an attractive interaction. The shortest distance (3.03 Å) between F<sub>2</sub>(rF) and N<sub>1</sub>(G) seen here is actually more than 0.8 Å shorter than the average distance between F<sub>4</sub>(rF) and N<sub>6</sub>(A) in rF:A pairs (41)! This difference is noteworthy because the distances of the (rF)C<sub>3</sub>-H ... O<sub>6</sub>(G)/N<sub>1</sub>(A) hydrogen bonds in the rF:G and rF:A pairs are very similar. The close resemblance of the rF:G and U:G pairs (Figure 5) can itself be taken as evidence for an attractive interaction between fluorine and N<sub>1</sub>-H(G). If the (rF)C<sub>3</sub>-H ... O<sub>6</sub>(G) hydrogen bond provided the only or the main contribution to the stability of the rF:G pair, the shear seen between F and G would seem quite unnecessary. Likewise, shearing does not benefit stacking interactions; if the geometry of rF:G and rF:A pairs were chiefly determined by stacking it would be difficult to explain their different geometries.

The similarity between rF:G and U:G is not limited to pairing geometry but extends to nearest neighbor interactions and hydration. Thus, superimposition of (CpU):(GpG) and (CprF):(GpG) base-pair steps reveals virtually identical conformations (Figure 6). The four rF:G pairs in the FG structure display a conserved hydration pattern

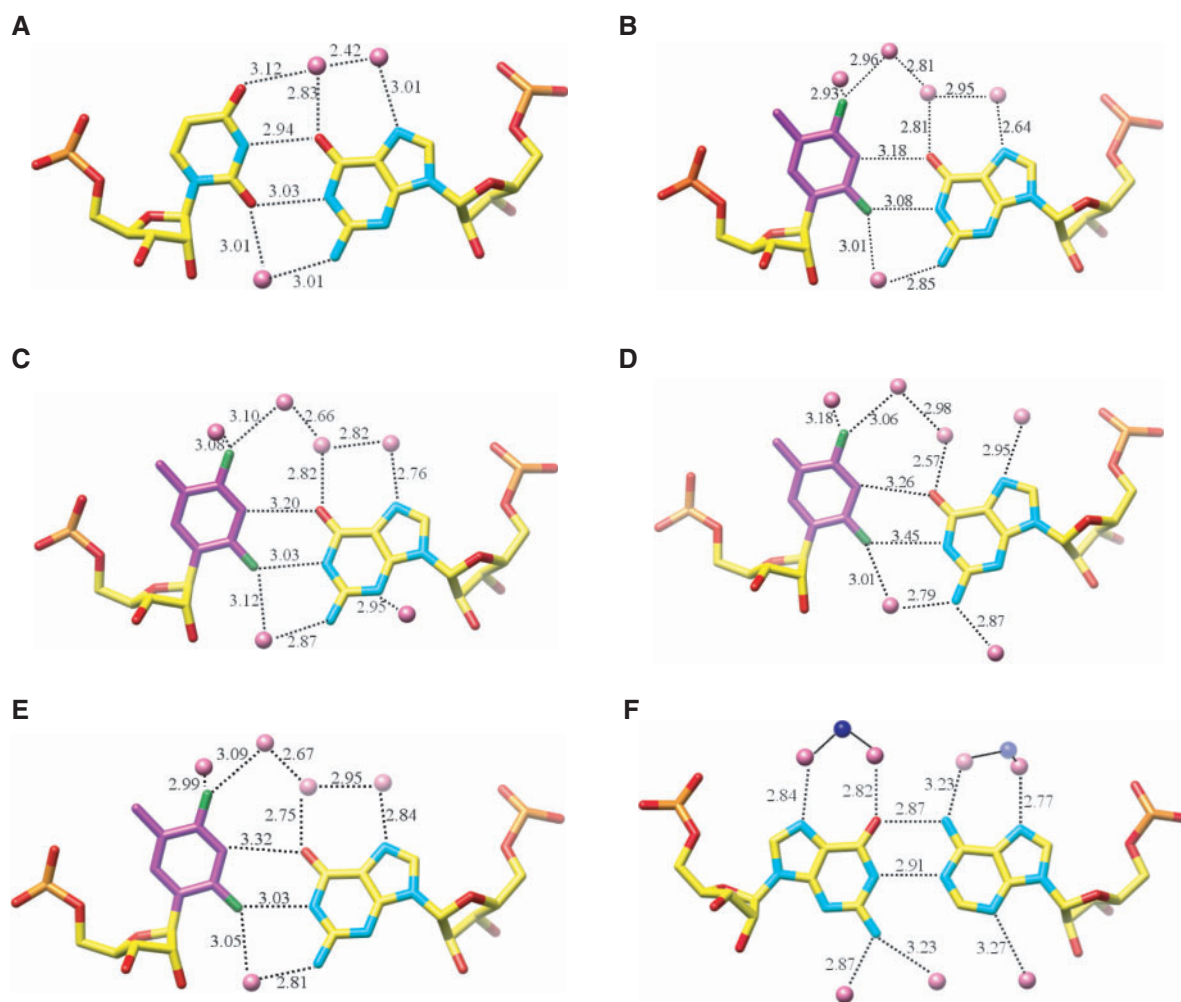


**Figure 3.** Quality of the final Fourier ( $2F_o - F_c$ ) sum electron density around the RNA duplexes in the region of the chemically modified or mismatched base pairs. (A) Portion of the FG duplex viewed into the minor groove; (B) portion of the GA duplex viewed into the minor groove; (C) rF104:G209 in the FG duplex and (D) G104:A209 in the GA duplex. Atoms are colored yellow, red, cyan and orange for carbon, oxygen, nitrogen and phosphorus, respectively. Carbon atoms of the difluorotoluy moiety are highlighted in magenta and fluorine atoms are shown as green spheres.

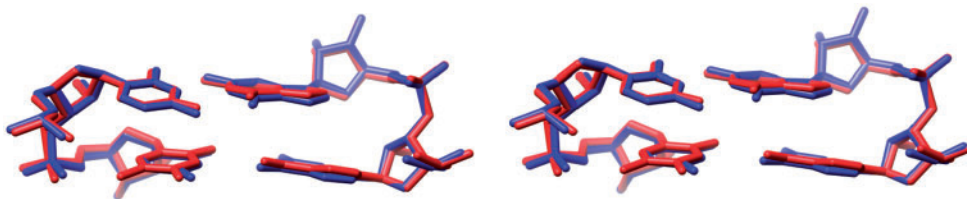


**Figure 4.** Overall conformation of RNA dodecamer duplexes with rF:G pairs. (A) One of the two RNA duplexes (duplex 1) per crystallographic asymmetric unit viewed into the major groove. Carbon atoms of the F moiety are highlighted in magenta and fluorine atoms are shown as green spheres. (B) Superimposition of the two duplexes; duplex 1 is red and duplex 2 is blue. (C) Overall conformation of the RNA dodecamer duplex with A:G pairs; the view is into the major groove that contains four Mg<sup>2+</sup> hexahydrate complexes.





**Figure 5.** Geometry and hydration of rF:G, U:G and G:A base pairs. (A) U:G pair in structure NDB AR0009 (66); (B) rF104:G209; (C) rF204:G109; (D) rF304:G409; (E) rF404:G309; (F) G104:A209<sup>#</sup> (# designates a symmetry-related residue). Hydrogen bonds in the U:G and G:A base pairs and the corresponding interactions in the rF:G pairs are shown as dashed lines with distances in Å. Water molecules and Mg<sup>2+</sup> ions are pink and blue spheres, respectively.



**Figure 6.** Stereo illustration of the superimposition of CpU-GpG [blue, NDB AR0009 (66)] and CprF-GpG (red) dinucleotide steps. The rF residue is located on the lower left.

(Figure 5B–E) that is similar to the water structure around the U:G pair (Figure 5A) (66–69). In both cases the water structure extends to the periphery of the grooves (data not shown), with ribose 2'-hydroxyl groups serving as bridge-heads in the minor groove (70). The extensive hydration of the difluorotoluy moiety observed in the FG structure further distinguishes the rF:G from the rF:A pair and is in good agreement with the osmotic stress data discussed above. No water molecules were observed at hydrogen-bonding distance from difluorotoluyls in rF:A pairs,

although we need to keep in mind that the resolutions of the FG and FA (41) structures are different (1.1 versus 1.6 Å). Conversely, water molecules are located as close as 2.93 Å from fluorine in the FG structure (Figure 5B). If we assume that hydrogen atoms from water molecules are pointing more or less in the direction of fluorine (the resolution of the structure is not sufficient to allow visualization of H atoms), the observed O<sub>W</sub>-fluorine distances are inconsistent with mere van der Waals interactions.

The GA structure also revealed extensive hydration of A:G pairs in the major and minor grooves (Figure 5F) which again is in good agreement with high  $\Delta n_w$  obtained using osmotic stress. Our structure at higher resolution revealed the coordination of four  $Mg^{2+}$ -hexahydrate ions inside the narrow and deep major groove. No metal cations were observed in the previously reported crystal structure of the GA dodecamer (54). Interestingly,  $Mg^{2+}$  ions engage in water-mediated contacts to the major groove edges of both G and A (Figure 5F).

### Semi-empirical calculations of F:G and F:A pairs and their native counterparts

All calculations were carried out with the program GAUSSIAN (57) and a counterpoise correction (58) was applied in each case. We ran equilibrium geometry optimizations as well as point-energy calculations based on the crystallographically observed geometries. Energies were computed both at the MP2/6-31G\* and MP2/6-31+G\*\* levels of theory. These are compared in Table 3 for F:G and F:A pairs along with their native counterparts T:G and T:A. Note that we used T as the reference nucleobase due to the presence of a methyl group in 2,4-difluorotoluene. Equilibrium geometries for the various pairs obtained at both levels of theory are depicted in Figure 7. The relative ranking of energies is T:G > T:A  $\gg$  F:G > F:A with energy differences between the T:G and T:A and the F:G and F:A pairs of around 3 kcal mol<sup>-1</sup>. The T:G pair is  $\sim$ 8 kcal mol<sup>-1</sup> more stable than the F:G pair (MP2/6-31+G\*\* level). Interestingly, the order of stabilities for isolated pairs obtained from semi-empirical calculations is similar to the one established experimentally based on UV melting data for RNA duplexes containing these pairs (U instead of T; in UV melting UA is slightly more stable than UG). The geometry calculated for the F:G pair (Figure 7A) shows close resemblance to the geometries of rF:G pairs in the crystal structure of the FG duplex (Figure 5B–E). On the other hand, the calculated and experimentally observed geometries of the F:A pair differ drastically; the F4 ... N6 distance in the calculated pair is *ca.* 0.7 Å shorter than the corresponding distance in the crystal structure (41). Despite this divergence between the computed and experimental pairing geometries for F:G and F:A, the differences between the average point energy based on crystallographic data and the equilibrium geometry energy for the F:G (–5.1 versus –7 kcal mol<sup>-1</sup>, respectively) and F:A pairs (–2.9 versus –4.1 kcal mol<sup>-1</sup>, respectively) are quite similar.

## DISCUSSION

### Insights into Ago2 substrate recognition

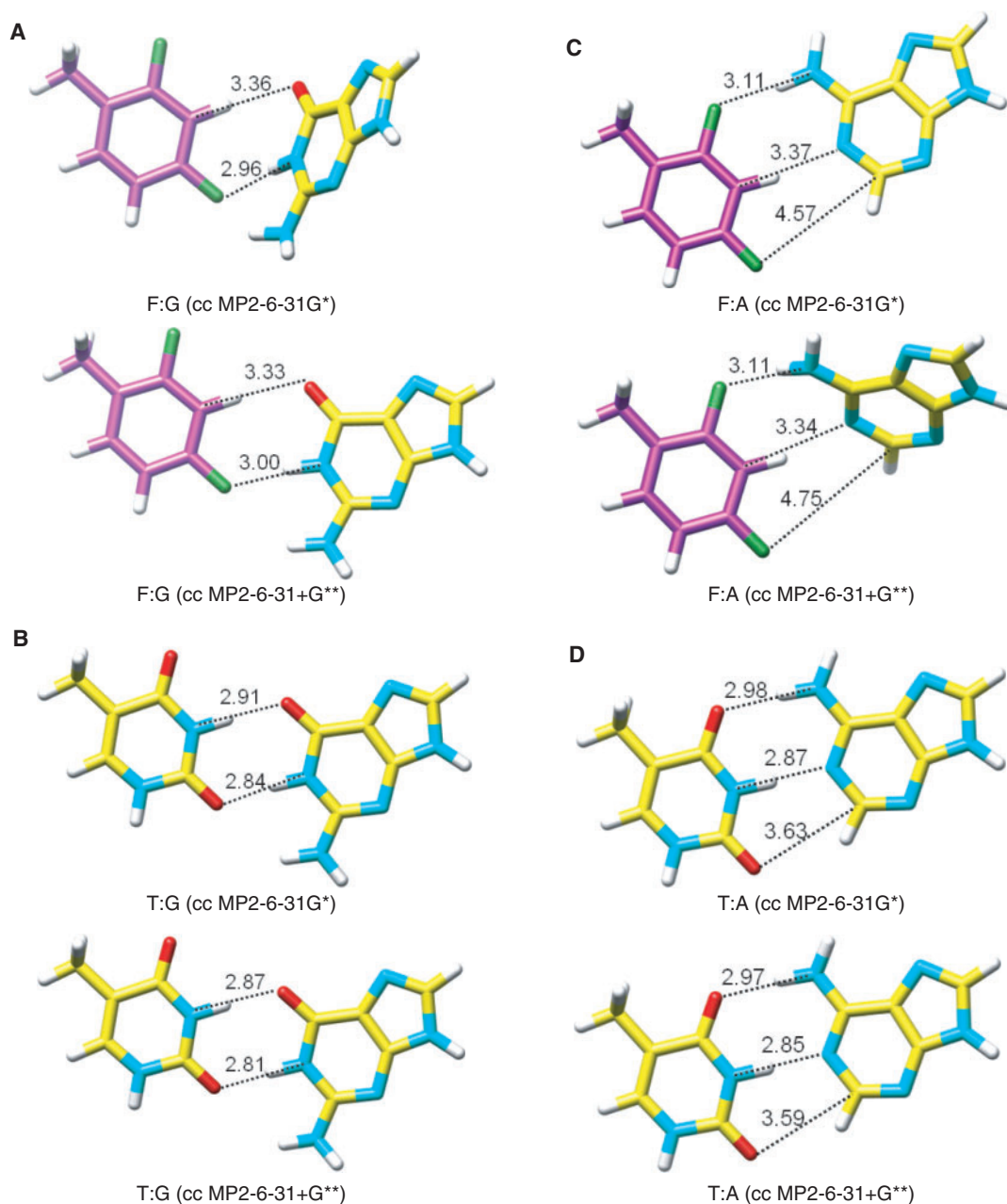
*In vitro* siRNA experiments in HeLa cells using a dual luciferase assay demonstrated that unlike a single rF:A pair, an rF:G pair at or adjacent to the cleavage site is not tolerated by the Ago2 slicer enzyme. Similarly, a single U:G mismatch drastically reduces activity and the only mismatch pair exhibiting activity (somewhat reduced relative to a canonical siRNA duplex) is the A:C pair (41).

**Table 3.** Energies ( $-E$  in kcal mol<sup>-1</sup>) obtained from semi-empirical calculations (CC = Counterpoise Correction)

Basis set level/ Base pair	MP2		MP2	
	6-31G*	6-31+G**	MP2 6-31G* (CC)	MP2 6-31+G** (CC)
Point energy calculation based on crystallographic data				
F104:G209	8.3	6.9	4.4	4.8
F204:G109	9.2	7.7	5.1	5.6
F304:G409	7.2	6.4	4.2	4.7
F404:G309	8.4	7.1	4.7	5.1
F4:A21	5.4	4.9	2.6	2.8
F16: A9	5.6	4.9	2.6	2.9
average F:G	8.3	7	4.6	5.1
average F:A	5.5	4.9	2.6	2.9
Energy calculation from base-pair geometry optimization				
T:A	18.2	16.5	12.3	12.7
F:A	7.6	7.2	4	4.3
T:G	19.6	19.3	14.1	15.5
F:G	11.5	8.4	6.4	7

This is consistent with a requirement for a Watson–Crick geometry of base pairs flanking the cleavage site. It is unlikely that the observed differences in activity between siRNAs containing mismatches or the rF residue at various locations are related to other factors, *i.e.* efficiencies in siRNA strand dissociation and biases in strand loading. Importantly, there is no correlation between the thermodynamic stability and silencing activity of the investigated siRNAs (Figure 2A, Table 1). Moreover, detailed positional analyses demonstrated that a siRNA with the rF residue in the guide strand, adjacent to the Ago2 cleavage site led to reduced silencing activity relative to positions further removed from the cleavage site or at the 5'-terminus (41,71). rF in the sense strand of the siRNA had no effect on the activity (71). However, an earlier RACE (rapid amplification of cDNA ends) analysis of the siRNA-induced cleavage of luciferase mRNA provides the strongest support that the observed silencing activities are directly related to Ago2-induced cleavage (41). Thus, cleavage at the normal site was found with siRNA AL-10 (rF:A, Figure 2A), whereas no cleavage was observed with siRNA AL-9 (rF:G, Figure 2A).

Although rF and A are more loosely paired relative to U:A, geometric constraints of the enzyme active site may produce a geometry that is not too different from that of a standard Watson–Crick base pair. Observations in connection with *in vitro* DNA polymerization experiments involving F-containing templates or incoming dFTP are instructive in this context (72–75). Accurate incorporation of dATP opposite template F or dFTP opposite template A is obviously not due to complementary hydrogen bond formation between the two partners (76). Rather, the active sites of high-fidelity DNA polymerases act as a mold that provides tight geometric constraints for a replicating base pair, virtually forcing it to adopt to a Watson–Crick-like arrangement (77). By contrast, the



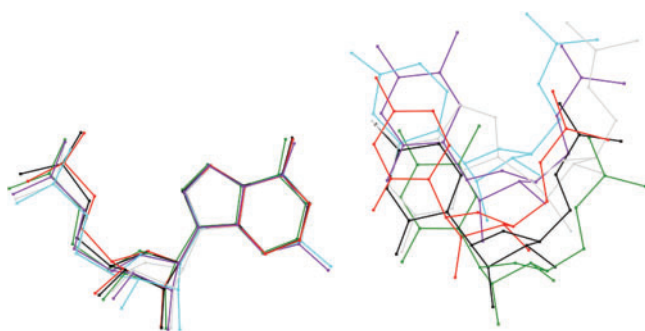
**Figure 7.** Equilibrium geometries of the (A) F:G, (B) T:G, (C) F:A and (D) T:A pairs based on semi-empirical calculations [cc MP2/6-31G\* and cc MP2/6-31+G\*\* levels; cc = counterpoise correction (58)]. Distances are in Ångstrom.

A:G, U:G and rF:G mismatches cannot be accommodated in this fashion. A superimposition of these mismatch pairs demonstrates that they result in larger deviations from a canonical backbone geometry than the rF:A and A:C pairs (Figure 8). In fact the observed siRNA activities are inversely correlated with the backbone distortions present in the RNA duplexes as a result of mismatch pairs. Although the A<sup>+</sup>:C pair with its two hydrogen bonds is expected to be isosteric with the U:G wobble pair (78), the particular example of the former used in the superimposition exhibits a smaller backbone distortion than both U:G and rF:G. Moreover, all measurements of siRNA activity were carried out at pH 7 and it is unlikely

that either A or C is protonated under these conditions. A neutral A:C pair with a single hydrogen bond between N6(A) and N1(C) or N1(A) and N4(C) may in fact display a geometry that is even more similar to that of a standard Watson-Crick base pair.

Contrary to the correlation between local distortions in the backbone geometry and *in vitro* siRNA activities, there is no obvious relationship between activity and thermodynamic stability. Thus, both the rF:A pair and the A:C pair (41) are more destabilizing than the U:G or rF:G pairs (Table 1). But the latter two lead to a loss of siRNA activity whereas the former result in a somewhat diminished activity compared to a canonical (and





*in vitro* siRNA activity: A:U > A:C = A:rF >> G:rF ≈ G:U = A:G  
 local backbone distortion: A:U < A:C = A:rF << G:rF < G:U = A:G

**Figure 8.** Comparison between the geometry of various base pairs. The diagram was generated by superimposing the purine bases (A or G, left) of different pairs. Distances between phosphates of residues on the right relative to the phosphate group of U are as follows: A:U (black, base pair A106:U207, FG structure—this article; reference structure), A:rF [green, base pair A21:rF4, 0.79 Å, (41)], G:U [cyan, AR0009, base pair G10:U5, 3.32 Å, (66)], G:rF (purple, base pair G209:rF104, 2.07 Å), A:G (gray, base pair A209:G104, GA structure—this article, 2.97 Å), and A<sup>+</sup>:C [red, AR0039, base pair A2:C17, 0.79 Å, (78)].

thermodynamically more stable) U:A pair at the Ago2 cleavage site (Figure 2). The stability data for the dodecamer duplexes with mismatches provided here and those based on longer siRNA duplexes published previously (41) are consistent with the thermodynamics at pH 7 of RNA hairpins containing single internal mismatches: G:C ≫ A:U ≫ U:G > G:A > A:C (79). Unlike siRNAs with single rF residues in place of U in the guide strand that retain activity [ref. (41) and this article], and appear to enhance sequence selectivity beyond that of the natural base in some cases (71), an siRNA with three consecutive rF:A pairs spanning the cleavage site did not display any activity (Figure 2B). This observation could indicate that multiple rF:A pairs result in more severe helical distortions (80) that are not tolerated by the Ago2 enzyme. Alternatively, the loss of activity may be related to a significant weakening of the hybridization between guide strand and mRNA target, or a combination of changes in stability and conformation.

### rF:G and rF:A pairs display considerable differences

The main surprise regarding the geometry of the rF:G pair is how closely it resembles the native U:G pair, including the shear that is a hallmark of the wobble pair. Based on the significantly larger separation between rF and A relative to U and A (>0.8 Å on average based on the various atom pairs), one would perhaps have expected a looser association of rF and G as well, weakly stabilized by a C-H ⋯ N hydrogen bond. Instead, the average distance between atom pairs in the four independent rF:G pairs (Figure 5B–E) present in the crystal structure of the FG duplex is just 0.3 Å longer than a typical hydrogen bonding distance of *ca* 2.9 Å in the U:G pair (Figure 5A). Not only is the pairing tighter than expected, but the rF:G pairs display a conserved hydration pattern that matches the arrangement of water molecules around the native U:G wobble pair. By comparison, the hydration of the

rF:A pair was much more limited (41). It is unlikely that this observation is simply due to the lower resolution of the earlier structure compared to that of the FG duplex (1.6 Å). Osmotic stress measurements are consistent with the structural data and support the notion that the rF:G pair is surrounded by more solvent molecules. Furthermore, the thermodynamic stability ( $T_M$ ) of an RNA dodecamer duplex containing two rF:G pairs is higher than that of the corresponding duplex with rF:A pairs (Table 1). This is noteworthy because U:G pairs are typically less stable than U:A pairs (79). By analogy, one would have predicted that rF:G pairs are less stable than rF:A pairs. Indeed, our findings with modified RNA duplexes differ from published stability data for DNA dodecamer duplexes with single F:G or F:A pairs (81). Accordingly, the relative order of stability in DNA is T:A ( $T_M = 50^\circ\text{C}$ ) > T:G (43°C) > F:A (36°C) > F:G (35°C). Although there are no crystal structures of DNA duplexes with incorporated F:A or F:G pairs it is unlikely that the differences between DNA and RNA arise from differential stacking interactions involving the difluorotoluene moiety. Incorporation of F in place of T into DNA leads to improved stacking (82). By comparison, the particular pairing mode between F/T and G pushes the former into the major groove, thus presumably reducing the overlap with the surrounding nucleobases. Overall, X-ray crystallography, osmotic stress measurements and UV melting data all clearly indicate that F pairs more favorably with G than with A in the case of RNA.

### Organic fluorine as a hydrogen-bond acceptor

The question whether fluorine can act as an acceptor of hydrogen bonds has recently been discussed in detail and with authority (65,83). In general, fluorine is a poor acceptor despite being very electronegative, and a screen of small molecule crystal structures only recovered a few examples with short X-H ← *d* → F-Y contacts (2.1 < *d* < 2.3 Å). Many of these structures do not contain a substantial amount of solvent, but their precision obviously exceeds that of a macromolecular crystal structure such as that of the FG duplex even at 1.1 Å. Hydrogen atoms can typically not be located in crystal structures of macromolecules at resolutions of around 1 Å. On the other hand bond lengths and angles in such structures are sufficiently precise (r.m.s.d. bond lengths <0.02 Å, Table 2) to allow reliable statements as to the presence or absence of a hydrogen bond. In drawing this conclusion we need to be aware that a short contact involving potential acceptor and donor functions is not necessary making a crucial contribution to stability.

The crystal structure of the FG duplex reveals distances between fluorine and N-H in three of the four rF:G pairs that are consistent with formation of a hydrogen bond: The F ⋯ N distances range between 3.03 and 3.08 Å and, therefore, the F ⋯ H distances can be expected to be around 2.1 Å (Figure 5B, C and E). Interestingly, the calculated energies (Table 3; MP2 6-31+G<sup>\*\*</sup>) based on the crystallographically observed geometries of F:G pairs correlate nicely with the F ⋯ N distances. Thus, the point energy of the F304:G409 pair with an F ⋯ N distance of

3.45 Å is  $-4.7$  kcal/mol compared to  $-5.6$  kcal/mol for the F204:G109 pair with a F  $\cdots$  N distance of 3.03 Å. When this fluorine (F2) is replaced with a hydrogen, point energy calculations at the MP2 6-31+G\*\* (CC) level using the geometries of the four experimentally determined F:G pairs (Figure 5, Table 3) for the 4-fluorotoluene:G pair indicate destabilizations ranging between 3.5 and 4.5 kcal mol<sup>-1</sup>. The notion that a short distance involving fluorine is associated with a stabilizing contribution certainly does not contradict the conclusion that fluorine is engaging in a hydrogen bond in F:G pairs. Further the calculations allow the conclusion that the pairing between F and G accounts for about half the stability resulting from pairing between T and G. It is unlikely, given the above observations, that pairing of F and G only involves the formation of a C-H  $\cdots$  N hydrogen bond. And in line with the experimental data, the semi-empirical calculations demonstrate that the F:G pair is more stable than the F:A pair.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are grateful for financial support by the US National Institutes of Health (grants R01 GM55237 to M.E. and R01 GM071461 to E.R.), the Austrian Science Fund FWF (P17864 to R.M.), and the BM-BWK (Gen-AU program, "Non-coding RNAs", No. P7260-012-011, to R.M.). S.L.M. was supported by a Matz Co-op Scholarship (Northeastern University). We would like to thank Dr Zdzislaw Wawrzak for help with data collection and processing and Mr Benjamin Diop-Frimpong with assistance in the refinement of the AG RNA dodecamer. Data were collected on the 22-ID beamline of the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at [www.ser-cat.org/members.html](http://www.ser-cat.org/members.html). The DuPont-Northwestern-Dow Collaborative Access Team (DND-CAT) Synchrotron Research Center at the Advanced Photon Source (Sector 5) is supported by E. I. DuPont de Nemours & Co., The Dow Chemical Company, the National Science Foundation and the State of Illinois. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Funding to pay the Open Access publication charges for this article was provided by grant NIH R01 GM055237.

*Conflict of interest statement.* None declared.

## REFERENCES

- Novina, C.D. and Sharp, P.A. (2004) The RNAi revolution. *Nature*, **430**, 161–164.
- Mello, C.C. and Conte, D.Jr. (2004) Revealing the world of RNA interference. *Nature*, **431**, 338–342.
- Hannon, G.J. and Rossi, J.J. (2004) Unlocking the potential of the human genome with RNA interference. *Nature*, **431**, 371–378.
- Tuschl, T. and Borkhardt, A. (2002) Small interfering RNAs: a revolutionary tool for the analysis of gene function and gene therapy. *Mol. Interv.*, **2**, 158–167.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
- Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**, 494–498.
- Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, **15**, 188–200.
- Caplen, N.J., Parrish, S., Imani, F., Fire, A. and Morgan, R.A. (2001) Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc. Natl Acad. Sci. USA*, **98**, 9742–9747.
- Nykanen, A., Haley, B. and Zamore, P.D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, **107**, 309–321.
- Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, **110**, 563–574.
- Schwarz, D.S., Hutvagner, G., Haley, B. and Zamore, P.D. (2002) Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol. Cell*, **10**, 537–548.
- Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. and Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.*, **20**, 6877–6888.
- Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature*, **431**, 343–349.
- Hutvagner, G. and Zamore, P.D. (2002) RNAi: nature abhors a double-strand. *Curr. Opin. Genet. Dev.*, **12**, 225–232.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*, **15**, 185–197.
- Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441.
- Song, J.J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, **305**, 1434–1437.
- Harborth, J., Elbashir, S.M., Vandenberg, K., Manninga, H., Scaringe, S.A., Weber, K. and Tuschl, T. (2003) Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense Nucleic Acid Drug Dev.*, **13**, 83–105.
- Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E. and Prydz, H. (2002) Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res.*, **30**, 1757–1766.
- Kretschmer-Kazemi Far, R. and Sczakiel, G. (2003) The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Res.*, **31**, 4417–4424.
- Gong, D. and Ferrell, J.E.Jr (2004) Picking a winner: new mechanistic insights into the design of effective siRNAs. *Trends Biotechnol.*, **22**, 451–454.
- Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
- Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S. and Khvorova, A. (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.*, **22**, 326–330.
- Pei, Y. and Tuschl, T. (2006) On the art of identifying effective and specific siRNAs. *Nat. Methods*, **3**, 670–676.
- Hamada, M., Ohtsuka, T., Kawaida, R., Koizumi, M., Morita, K., Furukawa, H., Imanishi, T., Miyagishi, M. and Taira, K. (2002) Effects on RNA interference in gene expression (RNAi) in cultured mammalian cells of mismatches and the introduction of chemical

- modifications at the 3'-ends of siRNAs. *Antisense Nucleic Acid Drug Dev.*, **12**, 301–309.
27. Schwarz,D.S., Hutvagner,G., Du,T., Xu,Z., Aronin,N. and Zamore,P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208.
  28. Yu,J.Y., DeRuiter,S.L. and Turner,D.L. (2002) RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *Proc. Natl Acad. Sci. USA*, **99**, 6047–6052.
  29. Paddison,P.J., Caudy,A.A., Bernstein,E., Hannon,G.J. and Conklin,D.S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.*, **16**, 948–958.
  30. Czauderna,F., Fechtner,M., Aygun,H., Arnold,W., Klippel,A., Giese,K. and Kaufmann,J. (2003) Functional studies of the PI(3)-kinase signalling pathway employing synthetic and expressed siRNA. *Nucleic Acids Res.*, **31**, 670–682.
  31. Soutschek,J., Akinc,A., Bramlage,B., Charisse,K., Constien,R., Donoghue,M., Elbashir,S., Geick,A., Hadwiger,P. *et al.* (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature*, **432**, 173–178.
  32. Kurreck,J. (2003) Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.*, **270**, 1628–1644.
  33. Manoharan,M. (2003) RNA interference and chemically modified siRNAs. *Nucleic Acids Res. (Suppl.3)*, 115–116.
  34. Dorsett,Y. and Tuschl,T. (2004) siRNAs: applications in functional genomics and potential as therapeutics. *Nat. Rev. Drug Discov.*, **3**, 318–329.
  35. Manoharan,M. (2004) RNA interference and chemically modified small interfering RNAs. *Curr. Opin. Chem. Biol.*, **8**, 570–579.
  36. Eckstein,F. (2005) Small non-coding RNAs as magic bullets. *Trends Biochem. Sci.*, **30**, 445–452.
  37. Morrissey,D.V., Zinnen,S.P., Dickinson, B.A.; Jensen,K., McSwiggen,J.A., Vargeese,C. and Polisky,B. (2005) Chemical modifications of synthetic siRNAs. *Pharm. Discovery Dev.*
  38. Bumcrot,D., Manoharan,M., Koteliansky,V. and Sah,D.W. (2006) RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nat. Chem. Biol.*, **2**, 711–719.
  39. Czauderna,F., Fechtner,M., Dames,S., Aygun,H., Klippel,A., Pronk,G.J., Giese,K. and Kaufmann,J. (2003) Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res.*, **31**, 2705–2716.
  40. Kraynack,B.A. and Baker,B.F. (2006) Small interfering RNAs containing full 2'-O-methylribose nucleotide-modified sense strands display Argonaute2/eIF2C2-dependent activity. *RNA*, **12**, 163–176.
  41. Xia,J., Noronha,A., Toudjarska,I., Li,F., Akinc,A., Braich,R., Frank-Kamenetsky,M., Rajeev,K.G. *et al.* (2006) Gene silencing activity of siRNAs with a ribo-difluorotoluy nucleotide. *ACS Chem. Biol.*, **1**, 176–183.
  42. Pitsch,S., Weiss,P.A., Jenny,L., Stutz,A. and Wu,X. (2001) Reliable chemical synthesis of oligoribonucleotides (RNA) with 2'-O-[(triiisopropylsilyl)oxylmethyl(2'-O-tom)-protected phosphoramidites. *Helv. Chim. Acta*, **84**, 3773–3795.
  43. Hobartner,C. and Micura,R. (2003) Bistable secondary structures of small RNAs and their structural probing by comparative imino proton NMR spectroscopy. *J. Mol. Biol.*, **325**, 421–431.
  44. Puglisi,J.D., Tinoco, and I. Jr. (1989) Absorbance melting curves of RNA. *Meth. Enzymol.*, **180**, 304–325.
  45. Kim,T.W. and Kool,E.T. (2004) A set of nonpolar thymidine nucleoside analogues with gradually increasing size. *Org. Lett.*, **6**, 3949–3952.
  46. Breslauer,K.J. (1995) Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Meth. Enzymol.*, **259**, 221–242.
  47. Spink,C.H. and Chaires,J.B. (1999) Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA. *Biochemistry*, **38**, 496–508.
  48. Rozners,E. and Moulder,J. (2004) Hydration of short DNA, RNA and 2'-OME oligonucleotides determined by osmotic stressing. *Nucleic Acids Res.*, **32**, 248–254.
  49. Berger,I., Kang,C.H., Sinha,N., Wolters,M. and Rich,A. (1996) A highly efficient 24-condition matrix for the crystallization of nucleic acid fragments. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 465–468.
  50. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307–326.
  51. Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, **26**, 795–800.
  52. Kissinger,C.R., Gehlhaar,D.K. and Fogel,D.B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 484–491.
  53. Cambillau,C. and Roussel,A. (1997) Turbo Frodo, Version OpenGL.1.
  54. Leonard,G.A., McAuley-Hecht,K.E., Ebel,S., Lough,D.M., Brown,T. and Hunter,W.N. (1994) Crystal and molecular structure of r(CGCGAAUUAGCG): an RNA duplex containing two G(anti).A(anti) base pairs. *Structure*, **2**, 483–494.
  55. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
  56. Sheldrick,G.M. and Schneider,T.R. (1997) SHELXL: high-resolution refinement. *Meth. Enzymol.*, **277**, 319–343.
  57. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Montgomery,J.A.J., Vreven,T., Kudin,K.N. *et al.* (2004) edn. *Revision C02 edn.* Gaussian, Inc., Wallingford, CT.
  58. Boys,S.F. and Bernardi,F. (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.*, **19**, 553–566.
  59. Parsegian,V.A., Rand,R.P. and Rau,D.C. (1995) Macromolecules and water: probing with osmotic stress. *Meth. Enzymol.*, **259**, 43–94.
  60. Rozners,E., Smicius,R. and Uchiyama,C. (2005) Expanding functionality of RNA: synthesis and properties of RNA containing imidazole modified tandem G-U wobble base pairs. *Chem. Commun. (Camb.)* 5778–5780.
  61. Timasheff,S.N. (1998) In disperse solution, “osmotic stress” is a restricted case of preferential interactions. *Proc. Natl Acad. Sci. USA*, **95**, 7363–7367.
  62. Courtenay,E.S., Capp,M.W., Anderson,C.F. and Record,M.T.Jr (2000) Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants *in vivo* and for the interpretation of “osmotic stress” experiments *in vitro*. *Biochemistry*, **39**, 4455–4471.
  63. Lavery,R. and Sklenar,H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.
  64. Pauling,L. (1939) *The Nature of the Chemical Bond*, 2nd edn. Cornell University Press, Ithaca, New York.
  65. Dunitz,J.D. and Schweizer,W.B. (2006) Molecular pair analysis: C-H...F interactions in the crystal structure of fluorobenzene? And related matters. *Chem. Eur. J.*, **12**, 6804–6815.
  66. Mueller,U., Schubel,H., Sprinzl,M. and Heinemann,U. (1999) Crystal structure of acceptor stem of tRNA(Ala) from *Escherichia coli* shows unique G.U wobble base pair at 1.16 Å resolution. *RNA*, **5**, 670–677.
  67. Auffinger,P. and Westhof,E. (1998) Hydration of RNA base pairs. *J. Biomol. Struct. Dyn.*, **16**, 693–707.
  68. Masquida,B. and Westhof,E. (2000) On the wobble GoU and related pairs. *RNA*, **6**, 9–15.
  69. Westhof,E. and Fritsch,V. (2000) RNA folding: beyond Watson-Crick pairs. *Structure*, **8**, R55–R65.
  70. Egli,M., Portmann,S. and Usman,N. (1996) RNA hydration: a detailed look. *Biochemistry*, **35**, 8489–8494.
  71. Somoza,A., Chelliserrykattil,J. and Kool,E.T. (2006) The roles of hydrogen bonding and sterics in RNA interference. *Angew. Chem. Int. Ed. Engl.*, **45**, 4994–4997.
  72. Moran,S., Ren,R.X. and Kool,E.T. (1997) A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl Acad. Sci. USA*, **94**, 10506–10511.
  73. Moran,S., Ren,R.X.-F., Rumney,S. and Kool,E.T. (1997) Difluorotoluene, a nonpolar isostere of thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.*, **119**, 2056–2057.



74. Morales, J.C. and Kool, E.T. (1998) Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nat. Struct. Biol.*, **5**, 950–954.
75. Delaney, J.C., Henderson, P.T., Helquist, S.A., Morales, J.C., Essigmann, J.M. and Kool, E.T. (2003) High-fidelity *in vivo* replication of DNA base shape mimics without Watson-Crick hydrogen bonds. *Proc. Natl Acad. Sci. USA*, **100**, 4469–4473.
76. Wang, X. and Houk, K.N. (1998) Difluorotoluene, a thymine isostere, does not hydrogen bond after all. *Chem. Comm. (Camb.)* 2631–2632.
77. Kim, T.W., Brieba, L.G., Ellenberger, T. and Kool, E.T. (2006) Functional evidence for a small and rigid active site in a high fidelity DNA polymerase: probing T7 DNA polymerase with variably sized base pairs. *J. Biol. Chem.*, **281**, 2289–2295.
78. Lima, S., Hildenbrand, J., Korostelev, A., Hattman, S. and Li, H. (2002) Crystal structure of an RNA helix recognized by a zinc-finger protein: an 18-bp duplex at 1.6 Å resolution. *RNA*, **8**, 924–932.
79. Meroueh, M. and Chow, C.S. (1999) Thermodynamics of RNA hairpins containing single internal mismatches. *Nucleic Acids Res.*, **27**, 1118–1125.
80. Zacharias, M. and Engels, J.W. (2004) Influence of a fluorobenzene nucleobase analogue on the conformational flexibility of RNA studied by molecular dynamics simulations. *Nucleic Acids Res.*, **32**, 6304–6311.
81. Guckian, K.M., Krugh, T.R. and Kool, E.T. (1998) Solution structure of a DNA duplex containing a replicable difluorotoluene-adenine pair. *Nat. Struct. Biol.*, **5**, 954–959.
82. Guckian, K.M., Schweitzer, B.A., Ren, R.X., Sheils, C., Paris, P.L., Tahmassebi, D.C. and Kool, E.T. (1996) Experimental measurement of aromatic stacking affinities in the context of duplex DNA. *J. Am. Chem. Soc.*, **118**, 8182–8183.
83. Dunitz, J.D. (2004) Organic fluorine: odd man out. *ChemBiochem*, **5**, 614–621.