Organic & Biomolecular Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Organic & Biomolecular Chemistry

ARTICLE

NMR study on the interaction of the conserved CREX 'stem-loop' in the Hepatitis E virus genome with a naphthyridine-based ligand

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

N. Dyubankova, M. Froeyen, M. Abramov, H.P. Mattelaer, P. Herdewijn, E. Lescrinier*

A 2-amino-1,8-napthyridine derivative that is described to bind single guanine bulges in RNA-DNA and RNA-RNA duplexes,[1] was synthesized and its interaction with the single G bulge in the conserved CREX of the Hepatitis E Virus (HEV) genome was explored by NMR and molecular modeling. Results indicate that the ligand intercalates in the internal loop, though none of the expected hydrogen bonds with the single G in the bulge could be demonstrated.

Introduction

The Hepatitis E Virus is an example of RNA viruses that produce subgenomic negative-sense RNA as an intermediate template for the synthesis of a capped subgenomic positive-sense RNA. This strategy enables translation initiation at a start-codon not located at the 5'-end of the viral genome. Its genome[2] contains three partially overlapping 'open reading frames' (ORFs) flanked by 'non-coding regions' (NCR, Figure 1A). ORF1 is located at the 5' end and codes for a non-structural polyprotein that is processed after translation into individual functional proteins involved in RNA synthesis. ORF2 is located at the 3' end and encodes capsid proteins. The 5' end of ORF2 overlaps partially with ORF3 which could encode a protein of maximum 123 amino acids to which some regulatory functions are attributed.[3] ORF1 is considered to be translated immediately after cell entry, yielding the enzymes that are required for RNA-synthesis while ORF2 and ORF3 are translated from a single subgenomic RNA. To unravel the molecular basis of HEV translation, mutants have been constructed that revealed the presence of a highly conserved sequence in the junction region between ORF1 and ORF2 is indispensable for the synthesis of both proteins encoded by the subgenomic RNAs. Genomic analysis revealed that this highly conserved cis-reacting element (CRE) and its flanking nucleotides fold into a stable stem-loop structure (CREX).[4] In silico and biochemical studies demonstrated that mutations in the lower stem of the CREX are tolerated if they conserve CREX at its structural level.[5] Huang et al.[4] predict a significantly different folding for the genotype 4 sequence, though the sequence of the junction region in genotype 4 could also fold in an alternative way including the typical CREX that is also present in other genotypes (Figure 1B).

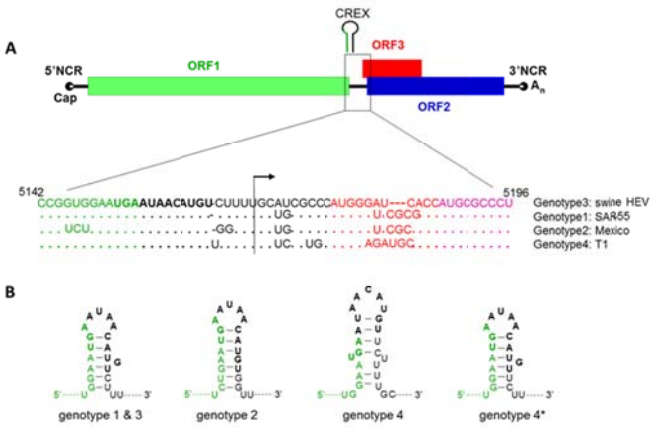

Figure 1: (A) Schematic diagram for the HEV genome representing open reading frames (ORF1, ORF2, ORF3), junction region and non-coding regions (NCR). Nucleotide sequence alignment of the junction region is shown for the different HEV genotypes. Dots denote sequence identity, and dashes denote deletions. Green and blue nucleosides belong to ORF1 and ORF3 respectively. Nucleosides that are part of overlapping ORF2 and ORF3 are depicted in purple. The arrow indicates the 5'-end of the small subgenomic RNA. The location of the CREX is indicated in the schematic diagram at the top. (B) Secondary structure of CREX sequences in genotypes 1-4 as predicted by Huang et al.[4] Genotype 4* denotes an alternative secondary structure we propose for an identical sequence. Nucleotides in bold belong to the cis-reactive element.

Experience with for example HIV TAR has demonstrated that a bulge region in the viral genome can be exploited as binding site for small molecule ligands that interfere with the viral life cycle.[6] Similarly, the single G bulge in the conserved CREX of the RNA genome of HEV can be targeted: stabilizing the structure of the CREX would prevent production of the negative viral genome strand and consequently block the viral reproduction while destabilizing the structure of the regulatory element would hamper production of the subgenomic RNA that is required for the viability of the virus in vivo. In this work, a solution NMR structure of CREX was

[a] Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium
* Corresponding author: Tel. +32 16 337391, Fax +32 16 337340, E-mail: Eveline.Lescrinier@rega.kuleuven.be
Electronic Supplementary Information (ESI) available: NMR and Molecular modeling & simulation . See DOI: 10.1039/x0xx00000x

determined that can be used as starting point for rational ligand design.

The predicted CREX in genotypes 1 and 3, as well as in the alternatively folded genotype 4* contain a single guanosine bulge. In literature naphthyridine analogues are described to specifically recognize single-G bulges in dsRNA and RNA-DNA duplexes.[7] Therefore, the 2-amino-1,8-naphthyridine derivative was selected as a starting point in a ligand optimization strategy to target the single G bulge in the conserved junction region in the RNA genome of HEV. In this work the first NMR study on a naphthyridine analogue binding at a single G bulge in an RNA duplex is described. High resolution NMR structures are available for dsDNA in complex with dimeric 2-amino-1,8-naphthyridine compounds[8] and some of their naphthyridine-azaquinolone analogues[9] that have two heterocycles interconnected via a linker. In both of these structures ligand binding disrupts a G-C Watson-Crick base pair to form a G-naphthyridine pair where naphthyridine displaces the 'widowed' C out of the duplex.

Experimental Part

Sample preparation

3-amino-N-(7-methyl-1,8-naphthyridine-2-yl)propanamide (compound **1**) was prepared as described in literature.[10] An unlabeled 17-mer RNA was obtained from Eurogentec. Samples were annealed prior to NMR experiments by briefly heating at 80°C and snap cooling on ice to promote hairpin formation. Samples were lyophilized and dissolved in 100% D_2O or in a 90%/10% mixture H_2O/D_2O. The pH was adjusted to 6.8 by adding small amounts of HCl 0.1M.

NMR experiments

Spectra involving phosphorus nuclei were obtained on a Bruker Avance 500 spectrometer equipped with a TXI Z gradient probe. Other spectra were recorded on a Bruker Avance II 600MHz spectrometer equipped with a 5 mm TXI HCN Z gradient cryoprobe.

The water signal in samples with 90% H_2O was suppressed using excitation sculpting.[11] The 2D NOESY in H_2O (mixing time = 200 ms; at 5°C) was recorded with a sweep width of 14400 Hz in both dimensions, 256 scans, 2048 data points in t_2 and 256 FIDs in t_1. The data were apodized with a shifted sine-bell square function in both dimensions.

The 2D DQF-COSY,[12] TOCSY[13] and NOESY[14] spectra in D_2O were recorded with a sweep width of 5400 Hz in both dimensions. The total TOCSY mixing time was set to 64 ms. All spectra were acquired with 32 scans, 2048 data points in (t_2) and 512 FIDs in (t_1). The data were apodized with a shifted sine-bell square function in both dimensions and processed to a 4K x 1K matrix. The NOESY experiments were acquired with mixing times of 150 and 300ms.

A [^{1}H-^{31}P]-HETCOR[15] spectrum was acquired with 256 scans, 2048 data points in the proton dimension, t_2 and 512 increments in the phosphorus dimension, t_1, over sweep widths of 5400 and 2430 Hz, respectively. The data were apodized with a shifted sine-bell square function in both dimensions and processed to a 2K x 1K matrix.

Natural abundance [^{1}H,^{13}C]-HSQC[16] were recorded with sensitivity enhancement and gradient coherence selection optimized for selection of CH groups ($^1J_{CH}$ = 150Hz) using 64 scans and 256/1024 complex data points and 60/9 ppm spectral widths in t_1 and t_2, respectively to measure ^{1}H-^{13}C correlations in the sugar part of the nucleosides and H5-C5 correlations in pyrimidine nucleobases. A separate natural abundance [^{1}H,^{13}C]-HSQC was measured to determine the remaining ^{1}H-^{13}C correlations in nucleobases. It was optimized for selection of aromatic CH groups ($^1J_{CH}$ = 180Hz) using 32 scans and 128/1024 complex data points and 40/9 ppm spectral widths in t_1 and t_2, respectively.

Natural abundance [^{1}H,^{13}C]-HMBC were measured in D_2O to correlate H2 and H8 protons to C4 in adenine[17] using 128 scans and 4096/128 complex data points and 10/60 ppm spectral widths in t_2 and t_1, respectively ($^1J_{CH}$ = 200Hz and $^3J_{CH}$ = 8Hz). This spectrum was also used to determine the chemical shift of C2 and C4 in uridine by their correlation to H6.

NMR derived restraints

Distance restraints between non-exchangeable protons were derived from a NOESY spectrum 150 ms mixing times. Based on ISPA, inter-proton distances were calculated. An experimental error (±20%) was used on the calculated inter-proton distances. The calibration of NOE cross peak intensities was done against the H5–H6 cross peaks as an internal standard. The distances implying exchangeable protons were collected from the NOESY spectrum in 90% H_2O and manually set to 1.0-2.0 Å, 1.0-3.6 Å and 1.0-5.0 Å for strong medium and weak cross-peaks intensities respectively.

Sugar puckers of the riboses were inferred from the weak H1' to H2' scalar couplings. Residues that have H1' to H2' scalar couplings varying from 0 and 2 Hz were restrained to the N pucker conformation [dihedral restraints applied: H1'-C1'-C2'-H2' (99.2 ± 20°), H2'-C2'-C3'-C3' (39.4 ± 20°), H3'-C3'-C4'-H4' (–162.0 ± 20°)]. In residues that have larger H1' to H2' couplings (between 3 and 7 Hz), no restraints were applied on the sugar rings of these residues during the structure calculation.

The normal ^{31}P chemical shifts of most residues were used to restrain α and ζ torsion angles (0 ± 120°),[18] except for those in the internal loop despite the 'normal' ^{31}P chemical shifts in this part of the molecule. The α and ζ torsion angles of involving P that showed a chemical shift deviating from the typical values of A-form RNA, were not restrained. The nicely resolved H4'-P(n)-cross peaks in the 2D ^{1}H-detected [^{1}H,^{31}P]HETCOR, allowed us to restrain the ß and γ torsion angles to 180 ± 30° and 54 ± 30°, respectively.[18] The ß and γ torsion angles of

residues that contain a phosphorus with an altered pattern in the 2D ^{1}H-detected [^{1}H,^{31}P]HETCOR were not restrained.
The ε torsion angles were restrained (to 230° ± 70°) based on steric arguments.[19]

Hydrogen bond restraints in Watson-Crick base pairs of the stem with unambiguously assigned imino signals, were applied as NOE-distance restraints as well as hydrogen bonds between A4:N6 – U13:O4 and A3:N6 – U15:O4 determined from ^{13}C chemical shifts of C4 in uracil. No planarity restraints were used in the structure calculations.

NMR Structure calculation

All structure calculations were performed with X-PLOR-NIH V3.851.[20] A set of 100 structures was generated by torsion angle molecular dynamics, starting from an extended strand and using NMR derived restraints. After the torsion angle molecular dynamics round on the 5'-GGAAUGGAAACAUGUCC-3' oligo using 123 dihedral angle restraints, 109 interresidue and 107 intraresidue restraints, 48 structures had converged to very similar structures with similar total energies and no violations of the NOE and dihedral restraints. Ten lowest energy structures were used for further refinement during the 'gentle molecular dynamics' round in water.[21]

A preliminary structure of 5'-GGAAUGGAAACAUGUCC-3' in complex with compound **1** was obtained following the same procedure as for the RNA oligomer itself. However, during this calculations 117 dihedral angle restraints, 99 interresidue and 107 intraresidue distance were used in the RNA oligo supplemented with 3 intermolecule distance restraints. Only the lowest energy structure from the torsion angle dynamics round was further refined in a 'gentle molecular dynamics' round in water.
The obtained structures were analyzed with the Curves+ program (version 1.31).[22]

Molecular modeling

Molecular dynamics simulations with the AMBER14 software were performed on the average NMR structure in which molecule **1** was docked next to G14 using Autodock4.2[23] and B-type dsDNA containing a single G bulge with compound **2** having Watson-Crick-type base pairing as described in literature.[1] The parameters for the RNA structure were taken from the ff14SB force field.[24] The ligands **1** & **2** were parametrized using the general Amber force field (gaff). More details are given in the supporting information.

Finally, some visual representations of the molecules were obtained with Chimera.[25]

Results and Discussion

Resonance assignment and structure determination

Standard methods were applied to assign non-exchangeable protons starting from a classic anomeric-to-aromatic proton walk.[18] From TOCSY, DQF-COSY and NOESY spectra, assignment of remaining proton signals in sugar rings could be established. All H2 signals in adenine nucleobases were unambiguously assigned through their long range correlation to C4, the carbon that also has cross peak with purine H8 in a [^{1}H,^{13}C]-HMBC (Figure S4).[17] Heteronuclear experiments on natural abundance samples were performed for assignment of ^{13}C and ^{31}P signals. Our spectra showed the typical chemical shift signature and nuclear Overhauser interactions (NOE) of previously described GAAA tetraloops,[27] confirming the proposed fold of our oligomer into a hairpin structure.

Three the Watson-Crick base pairs of the stem showed sharp imino signals (Figure S2). Strong NOE interaction to A12:H2 was exploited to assign U5:H3. An imino-to-imino of U5:H3 to G11:H2 was used to assign the latter. The remaining strong signal at 12.4ppm was assigned to G2:H1 through its NOE interaction with A3:H1', C17:H1' and C16:H5. Several broadened imino signals were observed. None of them had any detectable NOE-interactions in our 2D NOESY spectra in 90% H_2O (mixing times 100 and 200ms). One of them was assigned to the closing G1 – C17 base pair (12.6ppm). Others could belong U15:H3 and U13:H3, though they were not used as a hydrogen bond restraint between A:N1 and U:N3 since these weak signals lacked observable NOE interactions.

We used long range correlations of C2 and C4 to H6 in a [^{1}H,^{13}C]-HMBC to determine the chemical shifts of these carbonyls (Figure S3).[26] This approach allows retrieving hydrogen bonding information on uracil residues in D_2O, without measurable H3 imino signals or isotopic enrichment. Limited dispersion of chemical shifts in both dimensions could lead to peak overlaps, though this was not the case for the studied oligomer. In U13 as well as U15, chemical shifts of carbonyl atoms in the nucleobases are comparable to those of uridine in the classical A12:U5 base pairs whose chemical shift signature corresponds to hydrogen bonding at C4 and not at C2. This information was used implemented as A3:N6 – U15:O4 and A4:N6 – U13:O4 hydrogen bonds during structure calculations.

Overall geometry

The structure of the RNA hairpin was calculated from 217 distance and 123 torsion angle restraints (see Material and Methods). Although only A:N6-U:O4 hydrogen bond restraints were introduced on bases that flank the G bulge, hydrogen bond donors and acceptors in A3-U15 are close enough to form a Watson-Crick base pair.

At neutral pH, this sequence did form a stable hairpin with a shifted signal in the ^{31}P spectrum typical for the 'turning phosphate' in a GAAA tetraloop. The stem of the obtained structure has an overall A-type structure that is disturbed at the internal loop formed by G14 (Table S2).

Preferred conformation of the single guanosine bulge

Our spectra did not show indications that the bulge region of the molecule exists in several stable conformations in solution (no alternative set(s) of resonances and no exchange signals were observed). Within the signal-to-noise ratio of the experiments we can state that at least 95 % of the bulge RNA is in a stable conformation in solution.

Sequential NOEs between U13 sugar protons (H1′/H2′/H3′) and the G14-H8 proton and the NOEs between G14 sugar protons and U15-H6 are indicative for a stacked conformation of the bulged G14. Moreover, sequential aromatic-to-aromatic cross-peaks were observed for U13-G14. However no cross-peak was observed between G14-H8 and U15-H6. Sequential H2′(n)→H1′(n+1) NOE connectivity occurs for residues U13-G14, but not between G14 and U15. These observations suggest a close stacking of U13 and G14 combined with some spatial separation of G14 and U15. In the opposed strand, sequential NOEs between A3 sugar protons and A4-H8 are present together with H2′(n)→H1′(n+1) NOEs. The H2 proton of A4 shows inter-strand NOE interaction to the H1′ sugar protons of both G14 and U15.

Phosphor signals of residues in the stem, including those in the internal loop, resonate at chemical shifts typical for A-form helices. The clustering of these phosphor signals suggests no major deviation from typical A-form backbone torsion angles.

To validate the obtained structures, NMR chemical shifts (CCS) were back-calculated with the program NUCHEMICS for each of the calculated structures, considering ring-current effects and magnetic-anisotropy terms while ignoring charge contribution.[28] The good correlation between calculated and experimental values of residues 3-4 and 13-14-15 in the internal loop is reflected by their average correlation coefficient of 0.981 (standard deviation between correlation coefficients of 10 structures: 0.005). According to the Nuchemics calculation, the observed upfield shift of U13:H5 is mainly caused by ring current effects of the flanking A12 and G14 nucleobases while those of the A4 nucleobase are the major cause for an upfield shift of U5:H5 (U5:H5 and U13:H5 are observed at 5.13ppm and 5.03ppm respectively).

The stem region in calculated structures accommodates the unpaired nucleoside G14 that is stacked below U13 and slides slightly into the major groove where its Watson Crick binding face is exposed to the solvent (figure 2D). In all calculated structures G14:HO2′ is hydrogen bonded to U15:O5′ (average distance 1.80Å, st. dev=0.33). The presence of the internal loop disturbs the overall helix of the stem. Small deviations in the sugar phosphate backbones of both strands allow for the accommodation of G14 (Table S2). Sugar phosphate backbones are locally moved away from each other to provide the required space for the unpaired nucleoside. The minor groove width measured from the Pi to Pj+3 distance is increased.[19] The longest average distance is A4:P – C17:P which amounts to 16.4Å (st. dev: 0.3), corresponding to a minor groove with of 10.6Å. Unfortunately, our stem is too short to determine its major groove distance (Pi to Pj+6).

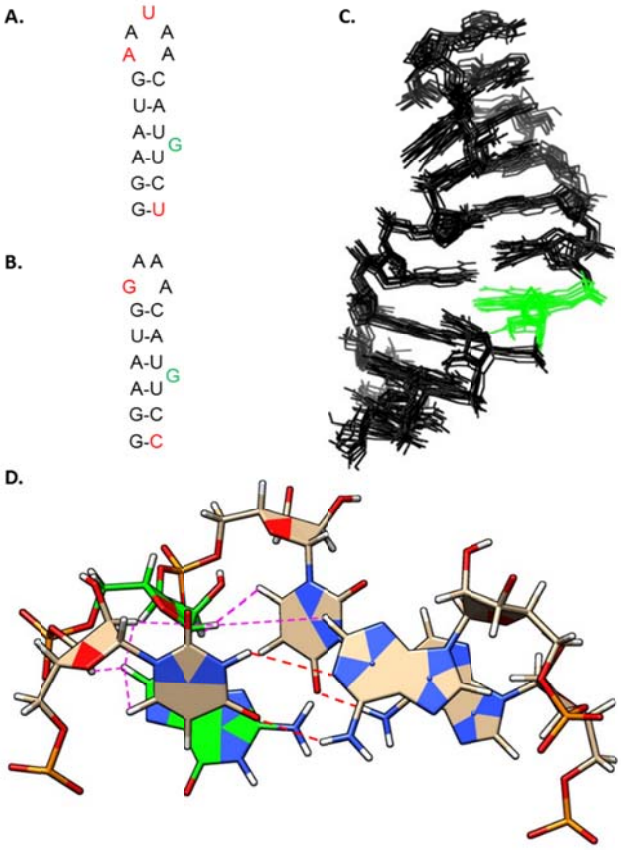

Figure 2: A: The conserved first stem-loop in of the noncoding junction region between ORF1 and ORF2 in the HEV genome. **B:** Predicted secondary structure of the stem-loop that was studied by NMR differs from the native one (A) in the loop part and closing base-pair to increase the stability in solution. **C:** overlay of the 10 calculated structures (accession number 2M4W at the PDB database) . **D:** 'close up' on top of the bulge in the lowest energy structure, flanked by A4:U13 and A3:U15 base pairs depicted on top and below G14 (green) respectively. Key inter-residue NOE contacts are indicated by purple dashed lines (A4:2-G14:1′, U13:2′-G14:1′, U13:2′/3′/6-G14:8, U13:6-G14:8 and G14:2′-U15:6)

Comparison with other structures

The Protein Data Bank[29] was searched for nucleic acid structures with single guanine bulges flanked by two Watson-Crick or GU wobble base pairs. In X-ray crystal structures containing G bulges in DNA and RNA duplexes, the guanine base is usually flipped out of the helix, while in our NMR structure it is stacked into the helix comparable to the single G flanked by AU and CG base pairs in the NMR structure of an RNA hairpin. The difference between NMR and crystal

structures can be attributed to the differences in solution and solid-state conditions. Looped-out guanosines are typically stabilized by inter-molecular contacts in crystals. Stacking of an unpaired nucleoside into a duplex region induces helix bending which is unfavored for crystallization. In the protein data bank, more examples exist on NMR structures with single-A bulges in a stacked conformation (e.g. 2JXS [35]).

Hastings et al.[30] analyzed examples of single-base bulge structures from the Protein Data Bank using molecular dynamics. They classified RNA single bulges into three different classes. Bulges in their 'first class' appear in a stacked conformation, while in other classes the bulge is predominantly or exclusively looped-out. The bulge determined in our structure can be considered as 'first class' according to Hastings' classification. As typical for this class, the overall backbone remains relatively close to that in regular A-type helices. Single-G bulges that were used in the molecular dynamics study of Hastings et al. did not belong to the 'first class' in their initial X-ray structure deposited at the protein databank. During molecular dynamics simulation, two of them switched from a 'looped-out' to the stacked conformation after approximately 1ns (Figure 3).

An extensive NMR study was performed on the base flipping in the U6 RNA internal stem loop where the unpaired U80 in a single-base bulge is stacked within the helix at neutral pH while it is flipped out into the major groove at lower pH upon protonation of the preceding A79.[31] The structure we obtained at pH=6.8 resembles the conformation observed for the RNA internal stem loop at neutral pH. In contrast to their observation of significant line broadening of the U80 H1' resonance relative to H1' resonances in nucleotides in Watson-Crick base pairs (11.6Hz versus 6.5Hz at pH=7), signals in our loop residue all have comparable line widths to those in nucleotides in Watson-Crick base pairs (figure 4). These line widths indicate that signals in our G bulge have similar T2 relaxation compared to those from Watson-Crick base paired residues in the stem and support the idea that the single G-base does not tend to undergo a conformational equilibrium as described for U80 in the U6 RNA internal stem loop but rather adopts a stable conformation in the selected conditions. Spin relaxation experiments combined with heteronuclear nuclear Overhauser effect would monitor an (absence of) interconversion between conformational states unambiguously, though it is not feasible to perform them on our unlabeled RNA sample.

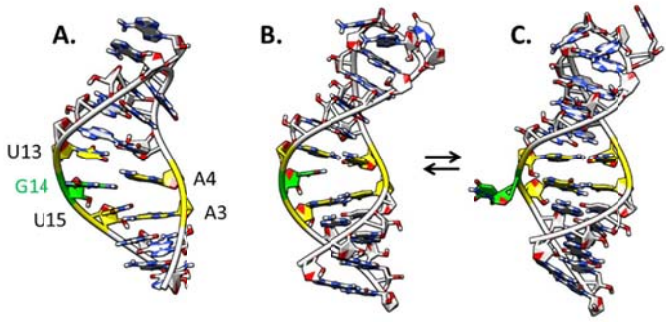

Figure 3: comparison of the determined solution structure with a single bulge G (green) flanked by two A:U base pairs (yellow) (**A**.) and the structures determined for U6 RNA internal stem loop where the unpaired U80 (green) is stacked between a G:C and A:C base pairs (yellow) at neutral pH (**B**.) while it is flipped out into the major groove at lower pH upon protonation of the preceding A79 (**C**.)

NMR study on the binding of a naphthyridine analogue

A naphthyridine analogue (Figure 5B) that was designed to specifically bind unpaired guanosine nucleobases in nucleic acids[1, 7] was synthesized according to literature. The titration of the studied RNA hairpin with compound **1** was monitored by 1D and 2D spectra (Figure 4). Several RNA signals start to broaden upon adding the ligand. Also ligand signals are strongly broadened, even at ligand/RNA ratio's that exceeded 1/1. Decreasing pH to 4.9 by adding small amounts of HCl 0.1M increased broadening at 15°C. Since basic pH is unfavorable to study RNA in solution for stability reasons (hydrolysis of phosphodiester linkages) and disappearance of imino signals in NMR spectra, a pH of 6.8 was preferred for further studies.

As expected, signals of the loop region were not influenced upon adding compound **1** up to a ratio of 1:1, while effects where observed on signals of nucleosides flanking the bulged G14. Surprisingly, signals of G14 hardly changed upon adding the naphthyridine analogue. Most affected are signals in residues A3, A4 and U15 that undergo line broadening as well as chemical shift changes for their ^{1}H and ^{13}C signals in sugars and nucleobases during the titration experiment. While uracil signals in residue U13 keep their chemical shift, changes occur in the sugar part of this residue. The opposite is observed for U5 that undergoes some chemical shift changes of nucleobase signals while signals of its sugar part are hardly affected. Line broadening of signals in the nucleobases of A3, A4, U13 and U15 hampered the detection of long range H-C correlations in an HMBC spectrum that could yield some information on changes in the hydrogen bonding of these residues.

In the [^{31}P, ^{1}H]-correlation, the only H3'-P signal with a significant change upon binding compound **1** belongs to the phosphodiester linkage between A3 and A4 (changes of 0.09 ppm and 0.33 ppm in ^{1}H and ^{31}P dimension respectively, Figure S4). Suggesting that no major structural changes are induced in the backbone of the U13-G14-U15 segment of hairpin upon ligand binding.

Due to strong broadening of signals in compound **1**, only those of its methyl and propanamide groups could be assigned. Weak NOE interactions are observed from its methyl group to

U15:1', A4:1' and A3:2 protons in the RNA hairpin. Simultaneously, sequential NOE interactions between G14 and G15 disappeared, as well as those between A3 and A4. These results indicate that the naphthyridine analogue intercalates between A4 and the A3-U15 base pair, with its methyl positioned in the minor groove of the stem. The research on sequence dependency for the binding of compound **1** to a DNA duplex already suggested that stacking interaction of the ligand to the base pair at the 3'-side of the guanine bulge is the molecular basis for its strong binding.[10]

The observed broadening of signals can be caused by a conformational or chemical exchange process at a rate that is intermediate on the chemical shift time scale. Since broadening does not improve in conditions with ligand excess, the observed line broadening cannot be fully attributed to exchange between bound and unbound state. Interchange between different binding modes of compound **1** at the bulge could also contribute to the observed effects. A similar phenomenon is reported for the binding of antitumoral bisnaphthalimides that intercalate DNA.32 Their planar rings were found to undergo 180° rotating motions in an intercalated state. Authors observed ring rotations by presence of exchange crosspeaks between naphthalimide proton resonances of both states when the exchange rate is slow on the chemical shift scale at low temperature (~2°C).

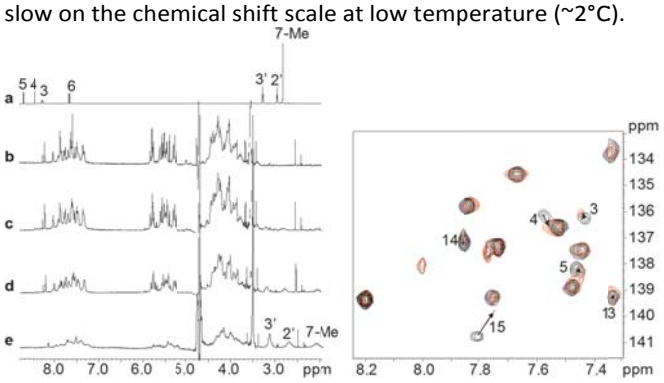

Figure 4: (left) 1D proton spectra in D_2O of compound 1 (**a**), 17mer RNA (**b**) and mixtures of both in a 1:2, 1:1 and 5:1 ratio (**c**, **d** and **e** resp). Signals of compound B that could be assigned are labeled according to numbering in figure 5. (right) Section of a natural abundance [^{1}H,^{13}C]-HSQC measured before and after adding an equimolecular amount of compound **1** to the 17mer RNA (black and red contour levels resp.). Signals are labeled with RNA residue numbers. Chemical shift changes upon adding the ligand are indicated by arrows (bold labels). Depicted spectra were recorded at 15°C.

They report selective broadening of naphthalimide ring resonances as the exchange rate becomes intermediate on the chemical shift time scale when temperature is increased while average chemical shifts were observed at higher temperatures (~40°C). Also in our sample, broadening improved upon increasing temperature to 25°C. Unfortunately, the melting point of our RNA hairpin did not allow studies at 40°C. Lowering temperature in our sample to 5°C increased line-broadening of naphthalimide and RNA signals at the binding site, however even at 5°C no separate sets of signals were observable.

The NMR behavior for the interaction of compound **1** with the studied G-bulge in the RNA duplex region are significantly different from reported data on a dimeric naphthyridine ligand and napthyridine-azaquinolone analogue binding a DNA duplex with respectively a G-G[8] and A-A mismatch flanked by GC base pairs.[9] For the latter sharp signals are observed including imino protons involved in hydrogen bonding between mismatched A residues and azaquinolones as well as hydrogen bonding between napthyridines and G residues flanking the A-A mismatch. In the high resolution NMR structure that was determined on the 11-mer duplex, 5'-d(CTAACAGAATG)-3'/5'-d(CATTCAGTTAG)-3' binding 2 napthyridine-azaquinolone ligands, the naphthyridine-guanine and 8-azaquinolone–adenine pairs are well stacked in the right-handed DNA helix while two 'widowed' cytidine nucleotides are looped out. A similar structure is proposed for the DNA with a G-G mismatch flanked by GC base pairs binding two dimeric naphthyridine ligands.[8]

Molecular Modeling

Due to strong line broadening of signals in the internal loop, no high resolution structure of naphthyridine analogue binding a G-bulge in the stem region of the RNA hairpin could be determined from NMR restraints. To obtain a model of the complex, the ligand was docked into the internal loop of the lowest energy structure of the RNA hairpin determined above. The optimal docked structure has hydrogen bonding from the N2 of G14 to compound B (Figure 5A) instead of the originally proposed pairing as shown in figure 5B and observed in the naphthyridine-guanine pairs of DNA duplex described above.

The backbone angles at the phosphodiester linkage between A3 and A4 significantly differ between the NMR structure of the free RNA and the docking model obtained for the complex (table 1). These local changes are in agreement with the observed change in chemical shift of the involved phosphorus. The overall structure of the modeled complex is not significantly different from the hairpin structure determined by NMR. This result is in agreement with similar CD spectra reported by Tok et al[7] for G-bulges in absence and presence of a naphthyridine analogue. The lack of obvious shifts in those CD spectra is indicative for similar secondary structures before and after adding the ligand.

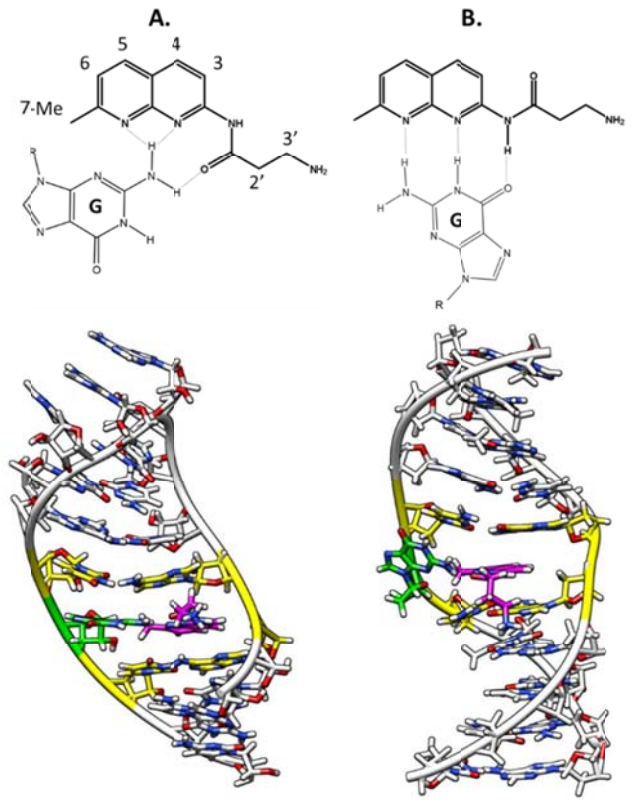

Figure 5: A. top: Chemical structure of the 3-amino-N-(7-methyl-1,8-napthyridine-2-yl)propanamide (compound **1**, in bold) used for NMR binding studies and schematic representation of hydrogen bonding between G14 and compound **1** that is present in the proposed model; bottom: End structure after 10ns molecular dynamics simulation of compound **1** (purple) docked into the internal loop of the lowest energy NMR structure of RNA hairpin. **B.** top: Chemical structure of the 3-amino-N-(7-methyl-1,8-napthyridine-2-yl)butanamide (compound **2**, in bold) with hydrogen bonding to G as proposed in literature; bottom: End structure after 10ns molecular dynamics simulation of compound **2** (purple) bound to a single G bulge in a DNA duplex starting from a Watson-Crick type hydrogen bonding between the ligand and the bulged G as proposed in literature .The single G bulge and flanking G-C base pairs are colored in green and yellow respectively.

An MD simulation on the obtained complex was stable for at least 20 ns. After clustering, the largest cluster of 10 contained 3100 structures out of 6000. The hydrogen bonding depicted figure 5A was present for 48, 33, 25% of the simulation time for G14:N2 to the ring N (at the left in Fig 5A), G14:N2 to amide CO and G14:N2 to the second ring N (at the right in Fig 5A), corresponding to the largest cluster, spread out over the full simulation time. The proposed binding mode with hydrogen bonding interactions of the intercalated compound **1** with the amino group of a bulged G in the model was unexpected. However, it does explain the stabilizing effect of compound **1** upon binding to a C-bulge observed as an increase in Tm.[1] since comparable interactions are feasible for the free amino group in a bulged C.

Nakatani et al,[1] proposed a model with a hydrogen bonded 'base pair' of G with a similar naphthyridine analogue (compound **2**, figure 5B) stacked between the flanking Watson-Crick base pairs in an B-type DNA duplex with single G bulge. To check the stability of such a model, an MD simulation of a dsDNA helical structure with compound 2 having Watson-Crick type base-pairing with a bulged G was performed. During this simulation, compound 2 remained in the cavity of the internal loop flanked by two G-C base pairs. However, the 'bulged' G14 started turning out of the helix after 3ns, making the WC base pairing with compound 2 impossible. Flipping out of G14 in the major groove of the DNA stem during molecular dynamics simulations indicates that there is not enough space in the internal loop to accommodate ligand as well as the 'single' guanine base in a stable structure. This sterical problem is not expected if the naphthyridine analogue binds to a G opposed to an abasic site, explaining the stronger increase in Tm observed by Gao et al[33] (8.6°C) compared to binding to a G-bulge (1.2-5.0°C).[1]

Conclusion

It is common knowledge that single base bulges enhance binding of intercalators and promote allosteric transitions in the targeted nucleic acids.[34] While the 2-amino-1,8-napthyridine ligand (**1**) was originally designed to form Watson-Crick type hydrogen bonds with an unpaired guanosine, no evidence for such hydrogen bonding pattern with the bulged G14 in the studied RNA stem could be observed by NMR. Based on chemical shift perturbations upon adding compound **1**, its binding site is located between the base pairs that flank G14. Observed line broadening is attributed to motions of the naphthyridine analogue intercalated in the cavity of the internal loop in the stem region of the RNA hairpin.

Molecular docking was used to obtain a model of compound **1** bound to the G-bulge in the studied RNA sequence. In this model an unexpected hydrogen bonding pattern exists between compound **1** and G14, though the proposed model is more stable in an MD simulation than the model originally proposed in literature. Our model is also in agreement with stabilizing effect of compound **1** upon binding to a C-bulge.

Acknowledgements

Authors thank C. Biernaux for the editorial help. This study made use of the BioMacs facility at KU Leuven. Equipment in the facility was purchased with funds from the Flemish government ('impuls project') and FWO-Flanders. This research was supported by a GOA grant of KU Leuven (GOA/10/13). HP Mattelaer is funded by FWO-Flanders. The refined coordinates of the ten lowest energy structures of the RNA hairpin are deposited in the PDB database (accession number: 2M4W)

Notes and references

1. K. Nakatani, S. Sando and I. Saito, *J. Am. Chem. Soc.*, 2000, **122**, 2172-2177.

2. S. K. Panda, D. Thakral and S. Rehman, *Rev. Med. Virol.*, 2007, **17**, 151-180.
3. V. Chandra, S. Taneja, M. Kalia and S. Jameel, *J. Bioscience*, 2008, **33**, 451-464.
4. Y. W. Huang, T. Opriessnig, P. G. Halbur and X. J. Meng, *J. Virol*, 2007, **81**, 3018-3026.
5. M. K. Parvez, *Gene*, 2015, **559**, 149-154.
6. M. Yang, *Curr. Drug Targets: Infect. Disord.*, 2005, **5**, 433-444.
7. J. B. Tok, L. Bi and M. Saenz, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 827-831.
8. M. Nomura, S. Hagihara, Y. Goto, K. Nakatani and C. Kojima, *Nucleic Acids Symp. Ser.*, 2005, DOI: 10.1093/nass/49.1.213, 213-214.
9. K. Nakatani, S. Hagihara, Y. Goto, A. Kobori, M. Hagihara, G. Hayashi, M. Kyo, M. Nomura, M. Mishima and C. Kojima, *Nat. Chem. Biol.*, 2005, **1**, 39-43.
10. K. Nakatani, S. Horie, T. Murase, S. Hagihara and I. Saito, *Bioorg. Med. Chem.*, 2003, **11**, 2347-2353.
11. P. Plateau and M. Gueron, *J. Am. Chem. Soc.*, 1982, **104**, 7310-7311.
12. M. Rance, O. W. Sørensen, G. Bodenhausen, G. Wagner, R. R. Ernst and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 1983, **117**, 479-485.
13. A. Bax and D. G. Davis, *J. Magn. Reson.*, 1985, **65**, 355-360.
14. J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546-4553.
15. V. Sklenar, H. Miyashiro, G. Zon, H. T. Miles and A. Bax, *FEBS Lett.*, 1986, **208**, 94-98.
16. J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S. J. Glaser, O. W. Sorensen and C. Griesinger, *J. Biomol. NMR*, 1994, **4**, 301-306.
17. M. J. P. Van Dongen, S. S. Wijmenga, R. Eritja, F. Azorin and C. W. Hilbers, *J. Biomol. NMR*, 1996, **8**, 207-212.
18. S. S. Wijmenga and B. N. M. van Buuren, *Prog. Nucl. Mag. Res. Sp.*, 1998, **32**, 287-387.
19. W. Saenger, *Principles of nucleic acid structure*, Springer-Verlag, New York, 1984.
20. C. D. Schwieters, J. J. Kuszewski, N. Tjandra and G. M. Clore, *J. Magn. Reson.*, 2003, **160**, 65-73.
21. J. P. Linge, M. A. Williams, C. A. E. M. Spronk, A. M. J. J. Bonvin and M. Nilges, *Proteins*, 2003, **50**, 496-506.
22. R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute and K. Zakrzewska, *Nucleic Acids Res.*, 2009, **37**, 5917-5929.
23. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785-2791.
24. D. A. Case, J. T. Berryman , R. M. Betz, D. S. Cerutti, I. Cheatham, T.E., T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York and P. A. Kollman, *Amber 2014*, 2014, University of California, San Francisco. .
25. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.
26. E. Lescrinier, N. Dyubankova, K. Nauwelaerts, R. Jones and P. Herdewijn, *ChemBioChem*, 2010, **11**, 1404-1412.
27. F. M. Jucker, H. A. Heus, P. F. Yip, E. H. Moors and A. Pardi, *J. Mol. Biol.*, 1996, **264**, 968-980.
28. J. A. M. T. C. Cromsigt, C. W. Hilbers and S. S. Wijmenga, *J. Biomol. NMR*, 2001, **21**, 11-29.
29. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-242.
30. W. A. Hastings, Y. G. Yingling, G. S. Chirikjian and B. A. Shapiro, *J. Comput. Theor. Nanos.*, 2006, **3**, 63-77.
31. N. J. Reiter, H. Blad, F. Abildgaard and S. E. Butcher, *Biochemistry*, 2004, **43**, 13739-13747.
32. J. Gallego, *Nucleic Acids Res.*, 2004, **32**, 3607-3614.
33. Q. Gao, H. Satake, Q. Dai, K. Ono, S. Nishizawa and N. Teramae, *Nucleic Acids Symp Ser (Oxf)*, 2005, DOI: 10.1093/nass/49.1.219, 219-220.
34. S. A. White and D. E. Draper, *Nucleic Acids Res.*, 1987, **15**, 4049-4064.
35. L. Popenda, R. W. Adamiak and Z. Gdaniec, *Biochemistry.*, 2008, **47**, 5059-5067.