

Functional binding of hexanucleotides to 3C protease of hepatitis A virus

Bärbel S. Blaum¹, Winfried Wünsche², Andrew J. Benie¹, Yuri Kusov^{3,4},
Hannelore Peters¹, Verena Gauss-Müller³, Thomas Peters^{1,*} and Georg Sczakiel^{2,*}

¹Institute of Chemistry, ²Institute of Molecular Medicine, ³Institute for Virology and Cell Biology and
⁴Institute for Biochemistry, University of Luebeck, Center for Structural and Cell Biology in Medicine (CSCM),
Ratzeburger Allee 160, D-23538 Luebeck, Germany

Received May 2, 2010; Revised November 9, 2011; Accepted November 10, 2011

ABSTRACT

Oligonucleotides as short as 6 nt in length have been shown to bind specifically and tightly to proteins and affect their biological function. Yet, sparse structural data are available for corresponding complexes. Employing a recently developed hexanucleotide array, we identified hexadeoxyribonucleotides that bind specifically to the 3C protease of hepatitis A virus (HAV 3C^{pro}). Inhibition assays *in vitro* identified the hexanucleotide 5'-GGGGGT-3' (G₅T) as a 3C^{pro} protease inhibitor. Using ¹H NMR spectroscopy, G₅T was found to form a G-quadruplex, which might be considered as a minimal aptamer. With the help of ¹H, ¹⁵N-HSQC experiments the binding site for G₅T was located to the C-terminal β-barrel of HAV 3C^{pro}. Importantly, the highly conserved KFRDI motif, which has previously been identified as putative viral RNA binding site, is not part of the G₅T-binding site, nor does G₅T interfere with the binding of viral RNA. Our findings demonstrate that sequence-specific nucleic acid–protein interactions occur with oligonucleotides as small as hexanucleotides and suggest that these compounds may be of pharmaceutical relevance.

INTRODUCTION

Specific protein–nucleic acid interactions are essential for many biological processes. The structural diversity of RNA and DNA generates a plethora of structural motifs that serve as specific recognition elements e.g. in gene

regulation and has led to the development of so called aptamer technologies that aim at nucleic acid sequences with tailored binding specificities (1–6). Based on these discoveries, the idea has been advanced that even smaller oligonucleotides may accomplish specific interactions with protein targets and first examples of such interactions were reported, in some cases with DNA oligonucleotides as small as hexamers (7–10). To further test and substantiate this concept, we investigated the ability of the viral protease of hepatitis A virus, the 3C protease (HAV 3C^{pro}, picornain 3C, EC 3.4.22.28) to bind to DNA hexanucleotides.

HAV belongs to the family of picornaviridae, all of which possess a 3C protease. This enzyme plays a central role in the picornaviridae viral life cycle and serves a dual purpose, it is the major protease that processes the viral polyprotein and it binds to regulatory structural elements of the 5'-untranslated region of the viral RNA, thereby controlling viral genome synthesis (11). The 3C^{pro} proteolytic activity has been investigated in detail in numerous studies aiming at the development of anti-viral drugs. Three-dimensional structures of 3C^{pro} are available from X-ray and NMR analyses for a number of picornaviruses including HAV (12–14), some complexed with substrate peptides or inhibitors (15–18). Binding of viral RNA to 3C^{pro}, on the other hand, is much less well understood although it is essential for viral genome replication. Mutational analyses revealed that a highly conserved KFRDI motif in 3C^{pro} is critical for RNA binding (19,20), and recent structural studies shed light on the details of the 3C^{pro}/RNA interaction at atomic resolution (21,22). The binding site for viral RNA with the KFRDI motif is located opposite the 3C^{pro} proteolytic cleft. For coxsackievirus B3, the 3C^{pro} binding site has also been mapped on the surface of the viral RNA (23).

*To whom correspondence should be addressed. Tel: +49 451 500 4241; Email: thomas.peters@chemie.uni-luebeck.de
Correspondence may also be addressed to Georg Sczakiel. Tel: +49 451 500 2730; Fax: +49 451 500 4689; Email: sczakiel@imm.uni-luebeck.de
Present addresses:

Bärbel S. Blaum, Interfaculty Institute for Biochemistry, University of Tübingen, Hoppe-Seyler-Strasse 4, D-72076 Tübingen, Germany.

Andrew J. Benie, Department of Science Formulation and Biophysics, Novo Nordisk A/S, Novo Nordisk Park, G8.1.435, DK-2760 Måløv, Denmark.

We set out to determine whether small oligonucleotides may constitute efficient ligands for nucleic acid binding 3C proteases employing HAV 3C^{pro} as an example. To avoid cysteine-mediated dimerization of the protein the C24S mutant of HAV 3C^{pro} was used. We employed a hexanucleotide chip technology (24) to search the DNA hexanucleotide sequence space for sequences that bind to HAV 3C^{pro}. We then used NMR spectroscopy and biochemical assays to locate the binding site of one of the binding hexanucleotides, G₅T, and investigated the conformation adopted by this guanine-rich oligonucleotide. A NMR assignment of the 25 kDa 3C^{pro} is available in the literature (25). This study was performed at acidic pH at which the enzyme has almost no protease activity. We therefore reassigned the protein at physiological pH employing standard triple resonance experiments and ²H/¹⁵N/¹³C labeling. We used ¹H, ¹⁵N-HSQC spectra to identify amino acids that were affected by G₅T. Surprisingly, G₅T and other hexanucleotides identified in the array did not mimic the viral RNA, nor did they interfere with the HAV 3C^{pro}-RNA complex in a gel shift assay. Instead, we found that the DNA hexanucleotides inhibit the proteolytic activity of HAV 3C^{pro}, identifying these compounds as a possible starting point for antiviral drug development.

MATERIALS AND METHODS

Oligonucleotides

All oligonucleotides were purchased from Biomers (Ulm, Germany) at HPLC grade purity. Beside the strands listed in Figure 1, the following sequences showed no signal in the hexanucleotide array with 3C^{pro} and were used as controls: 5'-TAGGAC-3', 5'-GGGTGG-3', 5'-ACTACA-3'.

Expression and purification of HAV 3C^{pro}

All experiments were conducted using the C24S mutant of HAV 3C^{pro}. The protein was expressed and an initial purification step was conducted as described previously (27). Protein samples were further purified using high-resolution cation exchange and size exclusion chromatography. Nearly, 20–30 mg of protein was applied to a 6-ml Resolve S cation exchange column at a flow rate of one column volume per minute. The protein was then eluted using a gradient of 0 to 1 M NaCl in 10 mM potassium phosphate buffer, pH 7.4. Under these conditions, the NaCl concentration at which elution of 3C^{pro} occurred was 116 mM. The main peak was further purified using a HiPrep 26/60 Sephacryl S-300 HR size exclusion column at a flow rate of 1 ml/min in 10 mM potassium phosphate, pH 7.4. Protein samples were checked for purity and activity using polyacrylamide gel electrophoresis (SDS-PAGE) and a proteolytic activity assay (vide infra).

Fluorescence labeling of HAV 3C^{pro}

Fluorescent labeling of the lysine residues HAV 3C^{pro} was performed with the Alexa Fluor 488 Monoclonal

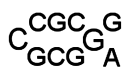
group	sequence (5' to 3')	sequence characteristics
I	CCGGAC CCGGAG* CCGGAT	CCGGAN
II	GGGGGA GGGGGC GGGGGG GGGGGT GGGCGT GGGGTG	GGGGGN
III	AGGCCA AGGCTA*	AGGCPyA
IV	CCGCGG CCC GGG CGGCCA	

Figure 1. Sequences of array-bound hexanucleotides with strongest binding to HAV 3C^{pro}. Related species are grouped (I–IV) and their major sequence characteristics are depicted on the right hand panel. Asterisks indicate sequences present within the HAV 5'-UTR: CCGGAG (pos. 24–29) and AGGCTA [pos. 90–95, numbers according to (20)].

Antibody Labeling Kit (Invitrogen, Carlsbad, CA, USA) according to the manufactures instructions. Briefly, 50 µg of PBS buffered HAV 3C^{pro} in a total volume of 90 µl was supplemented with 0.1 M sodium bicarbonate to pH ~8.3 before adding Alexa Fluor 488 reactive dye. After 60 min of incubation at room temperature the protein solution was dialyzed over night at 4°C against PBS using Spectra/Por dialysis membrane (Roth, Karlsruhe, Germany) with a cut off ~3500 Da. Subsequently, labeled protein was analyzed on 10% SDS-PAGE followed by determination of the labeling efficiency on a PhosphoImager Typhoon 8600 (Amersham Biosciences, Freiburg, Germany) with a 526-nm short pass filter.

Analysis of HAV 3C^{pro} on a hexamer array

For microarray analysis, a hexamer array representing the complete hexameric sequence space (4096 hexameric oligonucleotides) was used (24). The array was blocked with 2% (w/v) casein in buffer (50 mM Tris/HCl pH 8.0, 5 mM KCl, 5 mM MgCl₂) for 4 h at room temperature and washed five times with 25 mM HEPES pH 7.4 before starting incubation with 600 µl of 1.3 µM Alexa 488 labeled HAV 3C^{pro} for 5 min at room temperature. After washing five times for 5 s with 25 mM HEPES pH 7.4, and once for 2 s with water and rinsing for 1 s with absolute ethanol the array was scanned on PhosphoImager Typhoon 8600 with a 526 nm short pass filter. Spot intensities were quantified with the program Image Quant.

Proteolytic activity assay

The proteolytic activity of HAV 3C^{pro} was measured with the peptide substrate Ac-ELRTQ-pNA as described

previously (27). For the ranking of hexanucleotides according to their inhibitory activity $6.4\ \mu\text{M}$ 3C^{pro} and $32\ \mu\text{M}$ hexamer (monomeric concentration) in $50\ \text{mM}$ HEPES pH 7.4 (adjusted with $1\ \text{M}$ KOH, yielding a final K^+ concentration of $25\ \text{mM}$) were incubated for five minutes at room temperature in a volume of $150\ \mu\text{l}$ before addition of $1.67\ \text{mM}$ peptide substrate Ac-ELRTQ-pNA. Measurement of the kinetics was performed on a spectrophotometer (Beckman DU-600, Fullerton, CA, USA) at $405\ \text{nm}$ for 5 min taking values every 30 s. The $\Delta\text{OD}/\text{min}$ was calculated in the linear interval and the cleavage activity of 3C^{pro} in the presence of hexamer was compared to the activity in the absence of hexameric oligonucleotides which was set to 100%. For comparison of leupeptin and G_5T $6.2\ \mu\text{M}$ 3C^{pro} , $32\ \mu\text{M}$ leupeptin and $32\ \mu\text{M}$ G_5T were used. For the evaluation of ionic strength dependency on the inhibitory activity of G_5T $6.2\ \mu\text{M}$ 3C^{pro} and $32\ \mu\text{M}$ G_5T were used, and the ionic strength was adjusted using $10\ \text{mM}$ KCl, $10\ \text{mM}$ NaCl for one sample and $50\ \text{mM}$ KCl, $50\ \text{mM}$ NaCl in another sample. For the Lineweaver–Burk plot $33\ \mu\text{M}$ 3C^{pro} samples without inhibitor and G_5T concentrations of 8.5, 17.8 and $26.7\ \mu\text{M}$ were used.

HAV RNA samples for gel shift analysis

The plasmid pHAV/7 harboring the sequence of the attenuated HAV strain HM175 was used as a template of RNA transcripts (41,42). After linearization of pHAV/7 with SspI, transcription *in vitro* was performed in a $50\ \mu\text{l}$ solution containing $600\ \text{ng}$ template DNA and $200\ \text{U}$ of SP6 RNA polymerase in the presence of $50\ \mu\text{Ci}$ [α - ^{32}P] CTP, $100\ \text{nmol}$ of NTPs and $1\ \text{U}$ of RNase inhibitor for 2 h at 37°C , followed by treatment with $10\ \text{U}$ of DNaseI for 20 min at 37°C . Subsequently, the transcript was purified by phenol/chloroform extraction and gel filtration (G50 column). The quality of transcripts was checked on 6% native polyacrylamide gel.

Gel shift assay with radiolabeled HAV RNA

RNA–protein binding reactions were performed as described elsewhere (45). Briefly, a $15\ \mu\text{l}$ reaction mixture containing $0.4\ \text{nM}$ ^{32}P -labeled HAV 5'-UTR RNA ($154\ \text{nt}$), $25.6\ \mu\text{M}$ 3C^{pro} and $50\ \mu\text{M}$ hexamer was incubated in the presence of $20\ \text{U}$ of RNase inhibitor (RiboLock, Fermentas) in binding buffer [$5\ \text{mM}$ HEPES pH 7.9, $25\ \text{mM}$ KCl, $2\ \text{mM}$ MgCl_2 , $1.75\ \text{mM}$ ATP, $6\ \text{mM}$ DTT, $0.05\ \text{mM}$ PMSF, $0.167\ \text{mg/ml}$ tRNA, 5% (v/v) glycerol] for 20 min at 37°C . The reaction mixture was supplemented with $5\ \mu\text{l}$ of loading buffer [50% (v/v) glycerol, $1\ \text{mM}$ EDTA, 0.25% bromphenol blue] and analyzed on a 6% native polyacrylamide gel that had been prerun for 45 min at 4°C and $12\ \text{V/cm}$. Electrophoresis was conducted at $17\ \text{V/cm}$ at 4°C until the bromphenol blue marker had migrated to a position of 2/3 of the gel length. The gels were dried and subjected to autoradiography.

Gel assay analysis of G-quartet structures

The analysis of G-quartet structures of hexameric species was performed as described previously (28). Briefly, $500\ \text{pmol}$ of hexanucleotide incubated in the presence or absence of $10\ \text{mM}$ KCl for 30 min at 95°C and 60 min at 4°C before adding an equal amount of 25% (v/v) Ficoll 400 in TBE buffer. Samples were loaded on a 20% native polyacrylamide gel that had been prerun for 30 min and $12\ \text{V/cm}$ at 4°C . Electrophoresis was carried out at 4°C with $17\ \text{V/cm}$ for about 1.5 h. The detection of the nucleic acids was performed with Stains All (Sigma–Aldrich, Deisenhofen, Germany) according to manufactures protocol.

Assay for HAV-driven gene expression

Fifteen thousand Huh-T7 cells were seeded in 96-well plates and cultured overnight in Dulbecco modified eagle medium (DMEM) supplemented with 10% (v/v) fetal calf serum, $2\ \text{mM}$ glutamine, $100\ \text{U/ml}$ penicillin and $100\ \mu\text{g/ml}$ streptomycin. Then, a $50\ \mu\text{l}$ mixture of $50\ \text{ng}$ of a replication-competent pT7-18f-Luc containing HAV sequence in which the P1 domain had been replaced by firefly luciferase sequences (42), $0.1\ \text{ng}$ of phRL-SV40 coding for *Renilla* luciferase serving as transfection control as well as $5\ \mu\text{M}$ hexamer and $10\ \mu\text{l}$ Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) was added to the cell monolayer under serum-free conditions in Opti-MEM (Invitrogen) and transfection was continued for 4 h. The medium was replaced by serum-containing DMEM. As a negative control a transfection of $1\ \text{ng}$ firefly luciferase coding plasmid (pGL3), $0.1\ \text{ng}$ phRL-SV40 and $5\ \mu\text{M}$ hexamer was incubated in parallel. After 72 h, the cells were washed with $100\ \mu\text{l}$ PBS and lysed with $20\ \mu\text{l}$ passive lysis buffer (Promega, Madison, WI, USA). For determination of luciferase activity, samples were measured with dual luciferase assay (Promega) on an Anthos Lucy 3 luminometer (Mikrosysteme GmbH, Krefeld, Germany).

Sample preparation for NMR resonance assignment

NMR samples for backbone assignment via triple resonance experiments contained $0.2\ \text{mM}$ ^2H , ^{13}C , ^{15}N labeled 3C^{pro} in a total volume of $200\ \mu\text{l}$ $200\ \text{mM}$ deuterio-Tris pH 7.4, $10\ \text{mM}$ deuterio-DTT and 10% D_2O . The percentage of incorporated deuterium was roughly 75% for the H^α protons. TROSY variants of the HNCO, HN(CA)CO, HNCACB and HN(CO)CACB experiments were acquired at 30°C on a Bruker Avance 700 MHz spectrometer fitted with a TXI z-gradient cryoprobe. For the acquisition of paramagnetic relaxation enhancement (PRE) effects, a $0.2\ \text{mM}$ sample of ^{15}N labeled 3C^{pro} was labeled with the spin label *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) following the method described by Battiste and Wagner (46), with the modification that unreacted 3C^{pro} was not removed and a PD-10 column was used for the removal of unreacted MTSL. The spectra used for the analysis of paramagnetic relaxation enhancement (PRE) effects were acquired at 37°C on a Bruker DRX 500 MHz

spectrometer equipped with a TXI z-gradient probehead and the sample was reduced *in situ* with a 3-fold excess of ascorbic acid prior to repeating the measurement. All spectra were processed using NMRPipe (47) with linear prediction in the indirect dimensions and signal intensities were determined through the use of the nLinS module of NMRPipe. Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) was used for the backbone assignment.

DNA oligonucleotide samples for ^1H NMR

About 3.4 mg of G₅T (Biopolymers) were dissolved in 108 μl of D₂O to give a 16.7 mM stock solution and stored at -20°C . For 1D ^1H NMR measurements, 5 μl of this stock was diluted with 195 μl of 50 mM potassium phosphate pH 7.5, 100 mM sodium chloride and 10% D₂O to give a DNA concentration of 417 μM . The pH was adjusted to 7.5 using 1 M NaOH and the sample was transferred to a 3 mm (OD) NMR tube (Bruker Biospin Match system). Six 1D spectra were collected between 5 and 30°C using 5°C steps to observe sharpening of amine peaks.

^{15}N -HAV 3C^{pro} for binding study

A sample containing 100 μM uniformly ^{15}N labeled 3C^{pro} in 50 mM potassium phosphate, 150 mM sodium chloride, 2 mM deuterio-DTT, 0.25 mM deuterio-EDTA, pH 7.5 and 10% D₂O was prepared and transferred to a 3 mm (OD) NMR tube (see above). For the ^1H , ^{15}N -HSQC titration, G₅T was added to yield 3C^{pro}:G₅T ratios of 1:4, 1:20 and 1:40 (monomeric G₅T concentrations). The pH of each NMR sample was checked and, if necessary, adjusted prior to spectra collection. Spectra were collected at 30°C on a 500 MHz Bruker DRX spectrometer equipped with a cryogenic probe.

Diffusion time measurements

Diffusion ordered spectroscopy (DOSY) spectra for 3C^{pro}, G₅T and the 3C^{pro}:G₅T complex was acquired at 37°C on the above mentioned 500 MHz NMR spectrometer using an STE sequence with bipolar gradients and WATERGATE solvent suppression. For each sample, 1D ^1H NMR spectra with 1024 scans were recorded at 10 different gradient strengths reaching from 0.96 to 45.7 G/cm while the diffusion time (Δ) and gradient length ($\delta/2$) were kept constant. The gradient length was 1.8 ms for each sample and the diffusion time was varied between 100 and 170 ms for different diffusion coefficients to sample the whole decay curve for every sample. The relative signal intensities were plotted as functions of the gradient strength and the curves were fitted according to the equation:

$$I = I_0 \exp(-D\gamma^2 g^2 \delta^2 (\Delta - \delta/3 - \tau/2))$$

where I is the signal intensity, D the translational diffusion coefficient, γ the proton gyro-magnetic ratio, g is the gradient strength, δ the gradient length, Δ the diffusion time and τ the time between gradients of a gradient pair (218 μs in our measurements). For samples containing

3C^{pro}, the two most upfield methyl resonances were chosen for the analysis because they did not overlap with other signals. For free G₅T, several aromatic signals were chosen and these displayed very little variation. A 2 mM sample of hen egg lysozyme was first used to verify correct z-gradient calibration, which is critical for accurate DOSY measurements. At 20°C , a translational diffusion coefficient of $11.2 \pm 0.1 \times 10^{-7} \text{cm}^2/\text{s}$ was obtained for this sample, which is in good agreement with published values (48).

RESULTS

Screening hexanucleotide-3C^{pro} interactions

First, we systematically investigated possible interactions between HAV 3C^{pro} and DNA hexanucleotides using an array that contains the complete hexanucleotide sequence space (24). This array is designed such that hexanucleotides are attached to the chip surface via a non-nucleic acid linker attached to the 3' terminus of the DNA. In terms of specificity, interactions between proteins and surface-bound hexanucleotides have been shown to be compatible with interactions in solution (9). In the case of 3C^{pro}, array binding studies revealed a number of binding oligonucleotides which were grouped according to their primary sequence (Figure 1). The majority of these hexanucleotides exhibit high purine content. The consensus sequence of group II indicates their potential to form G-quadruplexes. Two out of three sequences shown in group IV are palindromes and could form antiparallel double strands. In the hexanucleotide array, formation of higher order structures should be sterically possible because the hexanucleotides are attached to the chip surface via a 38 atom flexible linker (24).

G₅T does not interfere with viral RNA binding

To investigate whether the hexanucleotides identified in the DNA array bind to the viral RNA binding site of 3C^{pro}, we studied binding of the *in vitro* transcribed 5'-terminal 154 nt of the 5'-UTR of HAV (26) to 3C^{pro} in the presence of hexanucleotides (Figure 2). No competitive effects were observed in this study, i.e. addition of hexanucleotides did not cause release of the RNA transcript from the complex with 3C^{pro}. In the presence of G₅T (group II), the experiment resulted in a super-shift of the complex formed between HAV RNA and 3C^{pro} indicating a ternary complex formed by RNA, 3C^{pro} and G₅T. Furthermore, no influence of G₅T on the concentration dependency of HAV RNA/3C^{pro} binding was observed, and in the absence of 3C^{pro} no complex was formed between HAV RNA and G₅T (data not shown). In summary, these data suggest binding of G₅T to 3C^{pro} at a site that is distinct from the putative binding site for viral RNA (the KFRDI motif) and does not interfere with viral RNA binding. None of the hexanucleotides representative for groups I, III and IV (CCGGAG, AGGCTA, CGGCGA) interfered with viral RNA binding and no super shift was detected for these hexanucleotides.

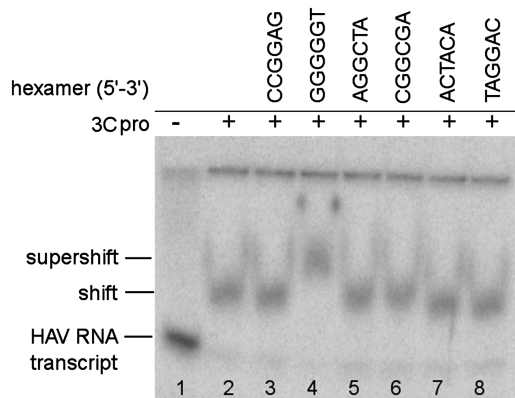


Figure 2. Supershift of the HAV 5'-UTR and 3C^{pro} in the presence of G₅T. An *in vitro* transcribed HAV 5'-UTR (154 nt) was shifted in the presence of 3C^{pro} (lanes #1 versus #2). This signal is further shifted in the presence of G₅T (lane #4) but not in the presence of other hits from the hexanucleotide array (lanes #3, #5 and #6) or control hexanucleotides (lanes #7 and #8) indicating a higher molecular weight complex, presumably a ternary complex formed by 3C^{pro}, G₅T and the viral 5'-UTR RNA.

The protease activity of 3C^{pro} is decreased in the presence of G₅T

Finally, the interaction of G₅T and other hexanucleotides with 3C^{pro} was further investigated in functional terms by testing the protease activity *in vitro* in the presence of these oligonucleotide ligands (27). This experiment showed decreased cleavage activity in the presence of G₅T, and to a smaller extent in the presence of G₅A, but almost no decrease for the closely related sequences TG₅ and AG₅ or any of the hexanucleotides representing the different groups listed in Figure 1 (Figure 3). Further, the deoxyribose backbone of G₅T seems to contribute to the inhibition of 3C as indicated by the inhibition data of chemically modified and extended derivatives (Figure 3).

Among all oligonucleotides tested, G₅T exhibited the highest inhibitory potency and underwent specific binding to 3C^{pro} at a site distinct from the putative RNA binding site. Therefore, this hexanucleotide was further investigated from a structural point of view. Figure 4 shows a comparison of G₅T's inhibitory activity with the activity of leupeptin, an established reversible inhibitor for cysteine proteases. In a low salt buffer G₅T achieved an inhibitory effect of 60% at a molar concentration at which leupeptin did only reduce the protease activity by about 10%. If the quadruplex structure of G₅T is taken into consideration (see below), its inhibitory activity is about 24 times higher than the inhibitory activity of leupeptin. However, when repeating the activity test with different salt concentrations we found a strong salt dependency of the G₅T inhibitory capacity (Figure 4). Addition of 10 mM NaCl and 10 mM KCl reduced the G₅T inhibitory activity from 60% to 47%, and addition of 50 mM NaCl and 50 mM KCl only 28% inhibitory activity was observed for G₅T. Because the protease activity of 3C^{pro} itself is not altered within this salt concentration range we conclude that the

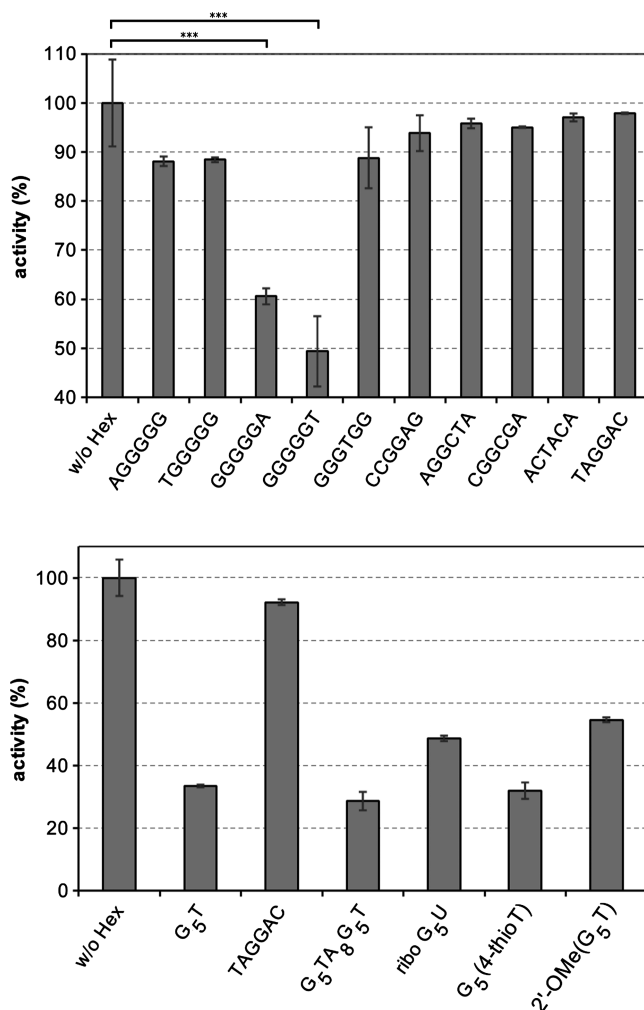


Figure 3. Inhibition of HAV 3C protease by unmodified (top) and modified (bottom) hexameric oligonucleotides. Hexamer sequences are indicated from 5' to 3' position. Oligonucleotides were used in a 5-fold molar excess over protein. The activity of the enzyme in the absence of hexamer was set to 100%. Data were compared using unpaired *t* test and the program Prims 4 (GraphPad Software). Indicated are mean values \pm SEM; ****P* < 0.001. The buffer used was 50 mM HEPES at pH 7.4 (adjusted with 1 M KOH, yielding a final K⁺ concentration of 25 mM).

complex between G₅T and 3C^{pro} must be mediated by ionic interactions to a substantial degree.

In an attempt to classify the inhibitory mechanism by which G₅T inhibits the 3C^{pro} protease activity, we measured the inhibitory activity at different G₅T concentrations and analyzed the data using a Lineweaver–Burk plot (Supplementary Figure S1). Unfortunately, these data do not allow unambiguous conclusions on the mechanism of inhibition.

Higher order structure of G₅T—quadruplex formation

For more detailed binding studies, we first focused on the solution structure of the ligands G₅T and G₅A. Because of the high guanosine content of group II sequences (Figure 1), the ability of G₅T and G₅A to form higher order structures such as G-quadruplexes

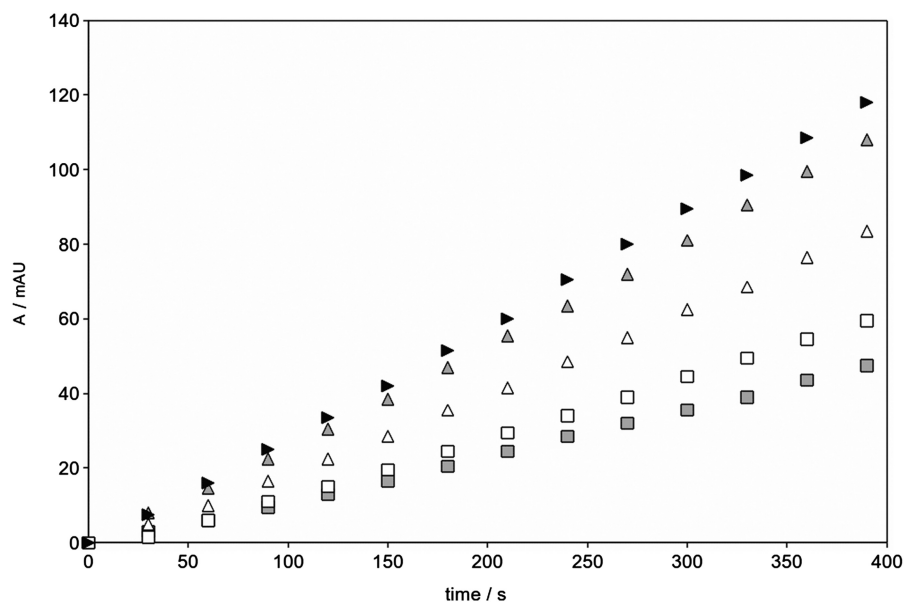


Figure 4. Comparison of inhibitory activities of leupeptin and G₅T and effect of ionic strength on the inhibitory activity of G₅T. Black triangles: 6.4 μM 3C^{pro} without inhibitor, gray triangles: with 32 μM leupeptin, gray squares: with 32 μM G₅T (concentration relating to the monomer) white squares: with 32 μM G₅T, additional 10 mM KCl and 10 mM NaCl, white triangles: with 32 μM G₅T, additional 50 mM KCl and 50 mM NaCl.

was studied. We initially noticed the ability of these oligonucleotides to oligomerize on a routine denaturing gel employed to check the purity of the hexanucleotides in the presence of 8 M urea (data not shown). Oligomerization was subsequently verified by NMR spectroscopy, which revealed characteristics typically observed for G-quadruplexes. The K⁺-dependency of quadruplex formation was tested in a gel assay following an established protocol (28), and it was found that both G₅T and G₅A form higher order structures, presumably quadruplexes, in the presence of K⁺, as indicated by slow migration (complex A in Figure 5A) that is characteristic for high molecular weight complexes and that was much less pronounced for TG₅ and AG₅ (complex B in Figure 5A). The control hexanucleotide G₃TG₂ in which the consecutive G stretch is interrupted migrated faster in native polyacrylamide gels in the presence of K⁺ as G₅T/G₅A which is consistent with a monomeric form (Figure 5A).

Additional evidence for the existence of a higher order structure of G₅T was obtained from 1D ¹H NMR spectra of the oligonucleotide obtained at different temperatures in the presence of 50 mM K⁺ or Na⁺. When compared to monomeric control hexanucleotides (CCGGAG and AGGCTA) spectra of G₅T exhibited significant peak broadening, stronger spectral overlap, and a distinctive set of resonances around 11 ppm (Figure 5B). Similar signals had been identified as imino resonances in previous studies of guanine-rich DNA sequences. These signals are characteristic for guanine quartets where extensive Hoogsteen hydrogen bonding protects guanine imino and amine protons from exchange with water protons (29,30). 1D ¹H NMR spectra were recorded between 6°C and 65°C for both samples containing either 100 mM Na⁺ or 100 mM K⁺ to monitor the stability of

the higher order structure of G₅T. No significant difference was observed between the two samples or the different temperatures (data not shown) except for a set of broad signals around 9 ppm, which became more intense at lower temperatures. The latter set of resonances is characteristic for amino protons that suffer line broadening as a consequence of exchange with water.

Although overlapping and thus difficult to quantify, the number of imino resonances in the 1D ¹H NMR spectra appeared to be larger than five (Figure 5B middle spectrum). In a parallel G-quadruplex chemically equivalent nucleotides at identical positions in the four different strands would also be magnetically equivalent and consequently only five imino resonances would be expected in total (i.e. there would be only one imino peak for all G2 residues, one for all G3 residues and so on) (29). Thus, a more complicated scenario seems to be the case for G₅T. 2D ¹H, ¹H-TOCSY and NOESY spectra were recorded at 6 and 30°C in the presence of 100 mM K⁺ to gain further insight into the preferred quadruplex conformation. Strong spectral overlap and multiple conformations complicated the analysis of 2D spectra but one prevalent DNA conformation could nevertheless be identified. For a tentative assignment of G₅T, the T6 methyl group was used as a starting point, readily allowing the assignment of T6 and G5. G4, G3 and G2 were assigned based on internal and sequential H8-H1' cross peaks (Figure 6A). The G2-H8 resonance, however, exhibited cross peaks with two sets of resonances in the H1' region (in addition to its internal H8-H1' cross peak) as well as for further deoxyribose resonances. Because none of the two H1' resonances thus associated with G2 exhibited further sequential connectivities, it was assumed that they belong to different conformations of G1 (subsequently termed G1 and G1*). The observation of two distinct cross peaks

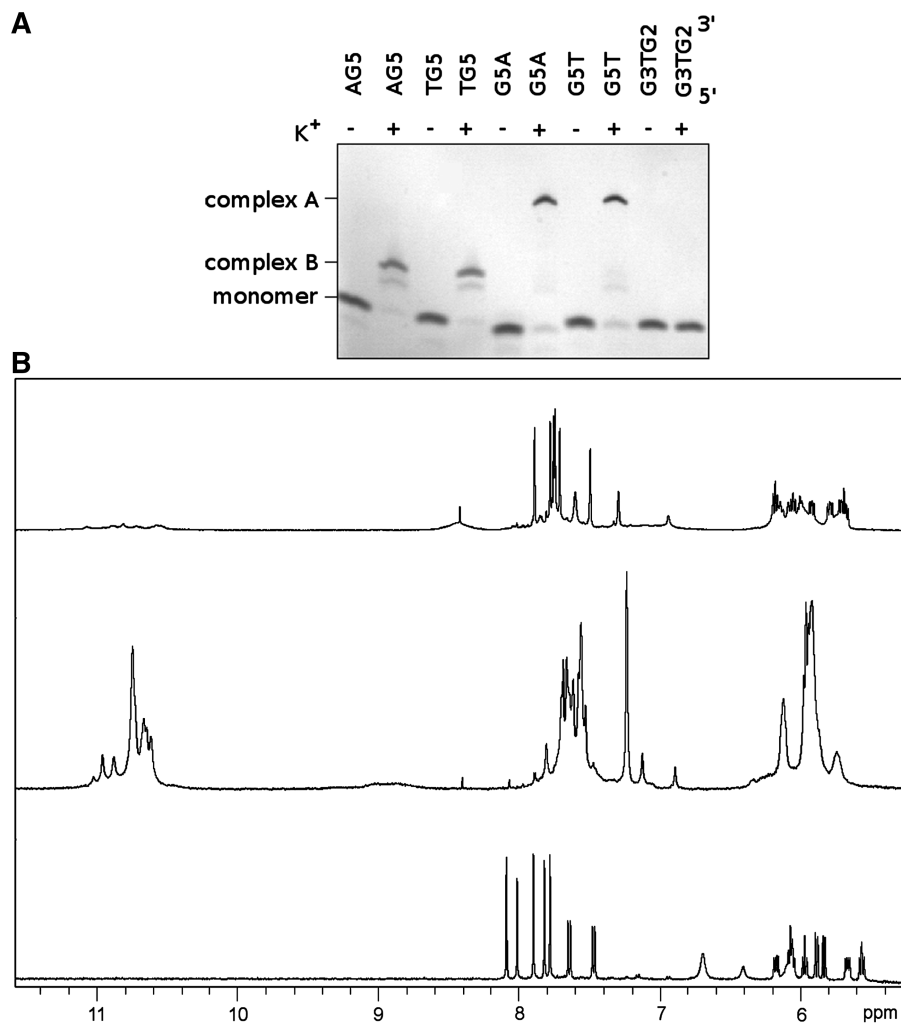


Figure 5. Formation of higher order structures/quadruplexes. **(A)** Hexanucleotides containing five consecutive guanines at their 3'-terminus (AG₅ and TG₅; lanes #1–4) show a mobility shift in native polyacrylamide gel electrophoresis in the presence of 10 mM K⁺ (complex B) which is even more dramatic for hexanucleotides with five guanines at their 5'-terminus (G₅A and G₅T, lanes #5–8) in the presence of 10 mM K⁺ (complex A). G₃TG₂ serves as a negative control for five consecutive Gs (lanes #9 and 10). **(B)** 500 MHz ¹H NMR spectra of G₅T in the absence and presence of 50 mM K⁺ (top and middle, respectively) and, for comparison, a monomeric ssDNA hexanucleotide (CCGGAG, bottom). The resonances between 10 and 11 ppm (middle spectrum) are indicative of imino groups protected by hydrogen bonding. Spectra were recorded at 30°C, pH 7.5.

between G2-H8 and H1' of G1 and G1* G1 and G1* in conjunction with an exceedingly weak cross peak between G1 and G1* resonances indicates slow exchange between two quadruplex forms containing either G1 or G1*.

Starting from the H8 resonances assignment of imino peaks of the two quadruplex forms was straightforward (Figure 6B). The most prominent diagonal imino peak at 10.78 ppm (asterisks in Figure 6B) contained the G5-H1 and G4-H1 diagonal peaks as well as the G1*-H1 diagonal peak. The remaining guanine H1 peaks were sufficiently resolved to observe cross peaks between G1-H1 and G2-H1 as well as between G2-H1 and G3-H1. The additional cross peak between G3-H1 and the G5, G4 and G1* overlapping H1 resonances could arise from the sequential connectivity between G3-H1 and G4-H1. No cross peak between the T6-H6 and any imino resonance was observed, suggesting that the thymine 3'-end residue is not involved in the hydrogen bond network.

The assigned imino–imino cross peaks are all consistent with a parallel stranded quadruplex structure in which only sequential H1–H1 cross peaks would be expected. In this interpretation, the cross peak between G2-H1 and the G5, G4 and G1* overlapping H1 resonances would be a G2–G1* imino–imino cross peak. However, it is also possible that this peak is in fact a G2–G4 imino–imino cross peak whose existence would point to an antiparallel quadruplex structure in which G2 of one strand would be hydrogen bonded to G4 of another strand. In such a scenario, G1–G5 imino–imino cross peaks would also be expected. While there is clearly no such cross peak between G5 and G1, nothing can be said about G5 and G1* because a cross peak between these imino resonances would be too close to the diagonal to be resolved. Because no cross peaks between G4 and G2 or G5 and G1 are observed in the aromatic region (Figure 6A), we consider an antiparallel strand orientation

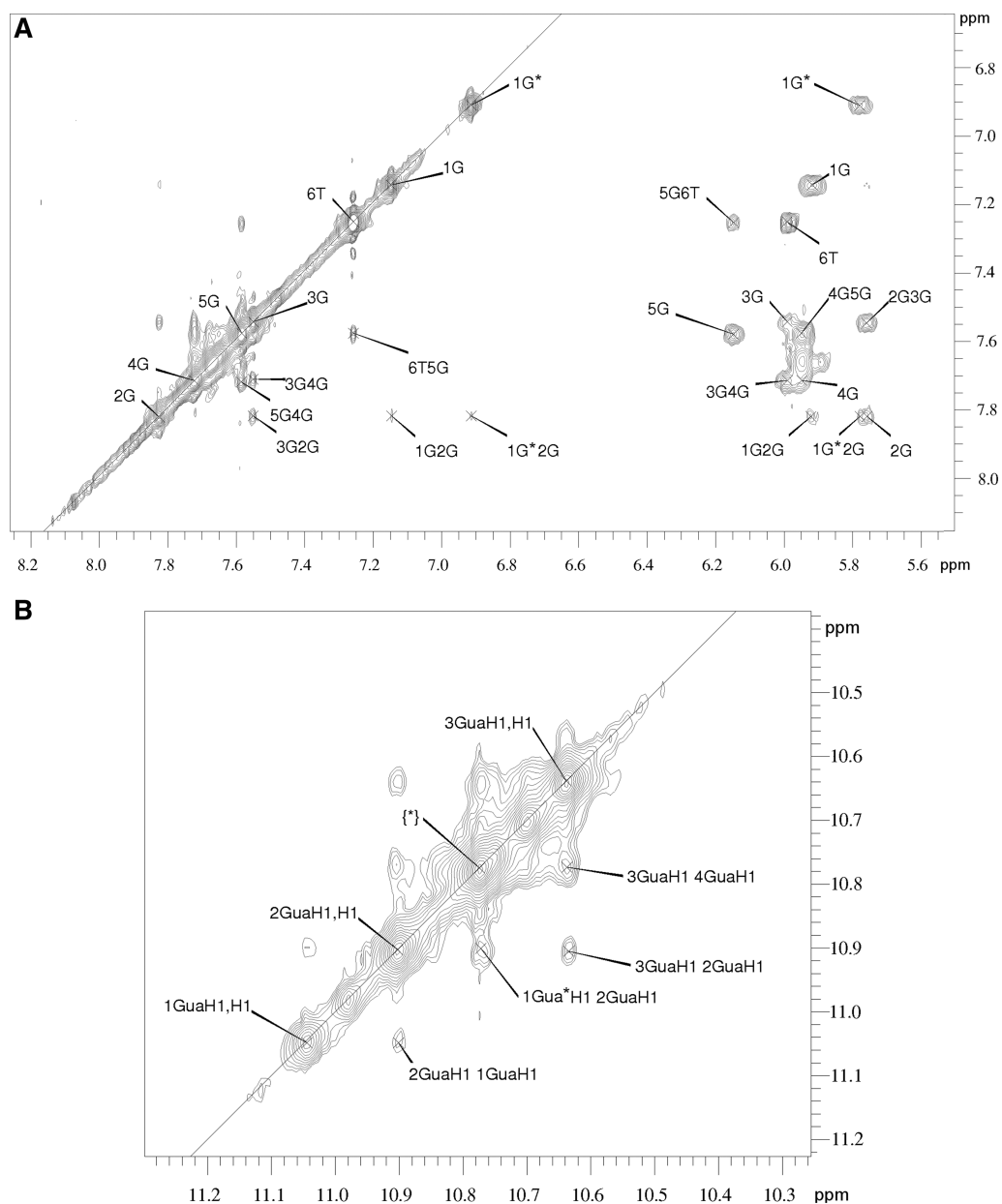


Figure 6. (A) Assignment of G₅T (aromatic and anomeric regions). G₅T ¹H, ¹H-NOESY spectrum (30°C, 250 ms mixing time, 100 mM K⁺, 10% D₂O). Left: Aromatic region. For clarity, the assignments are abbreviated (3G stands for G3-H8 diagonal peak, 3G4G for cross peak between G3-H8 and G4-H8). Cross peaks between aromatic protons and both their 5'- and 3'-neighboring aromatic protons are observed. G2-H8 exhibits weak cross peaks with both G1 and G1* (see text). Right: Anomeric region showing cross peaks with G5-H8 and T6-H6. For every guanine anomeric proton, cross peaks with the aromatic proton belonging to the same nucleotide and with the 3'-neighboring aromatic proton are observed. Assignments are again abbreviated (5G stands for the G5-H1'/G5-H8 cross peak, 5G6T for the cross peak between G5-H1' and T6-H6). Unassigned cross peaks belong to a second, unidentified conformation. (B) Assignment of G₅T (imino region). G₅T ¹H, ¹H-NOESY spectrum (30°C, 250 ms mixing time, 100 mM K⁺, 10% D₂O), imino region. The peak arising from an overlap of the G5, G4 and G1* diagonal peaks is denoted as asterisks. The two unassigned diagonal peaks belong to the unidentified conformation. While the unassigned resonance at 10.96 ppm has no further cross peaks in the imino region, the unassigned resonance at 10.70 ppm may form cross peaks with G3-H8 and the G5, G4 and G1* overlapping peak, pointing to a possible second conformation of one of these residues.

in G₅T unlikely. However, given the existence of unidentified peaks and the low spectral dispersion in the imino diagonal it appears that more sophisticated NMR experiments, including those targeted at ¹⁵N and ³¹P nuclei, would be needed to unambiguously resolve the multiple G₅T conformations. The spectral dispersion in G₅T for both the aromatic and imino peaks is poor, which is in

accordance with the previous observation that the absence of thymine loops in quadruplex structures leads to decreased spectral resolution of guanine protons due to the absence of thymine ring currents (31).

At a 10-fold excess of G₅T over 3C^{Pro}, there is no indication that the higher order structure is disrupted and monomers are formed (Supplementary Figure S2), but

since under these conditions the ^1H NMR signals reflect mainly the free form of G_5T we cannot completely rule out that the overall structure of G_5T is affected upon binding to 3C^{pro} . Slight signal broadening and small chemical shift changes were observed for the imino resonances upon binding to the protein. In agreement with previous studies, which identified monovalent cations at the center of guanine-quartets (32,33), addition of 10 mM MgCl_2 did not produce detectable changes in the G_5T 1D spectrum, nor did it alter the ^1H , ^{15}N -HSQC spectrum of the protein-DNA complex, suggesting that, unlike in other nucleic acid-protein complexes, no divalent cations are necessary to mediate the HAV 3C^{pro} - G_5T interaction.

G_5T forms quadruplex dimers in solution

Addition of G_5T to 3C^{pro} in low salt buffer had a dramatic effect on the protein's spectral line width, leading to severe signal loss in the ^1H , ^{15}N -HSQC spectrum (Supplementary Figure S3). Signals that were observable at a protein: G_5T (monomeric concentration) ratio of 1:4 in the presence of 20 mM NaCl belonged predominantly to the flexible side chains of asparagine and glutamine residues and the protein's flexible termini. Comparable spectra are commonly observed for proteins of high molecular weight in the absence of deuteration and pulse programs optimized for large molecular weights. To shed light on the unexpected increase in molecular weight upon 3C^{pro} - G_5T complex formation, DOSY measurements were recorded and translational diffusion coefficients extracted for free G_5T and the 3C^{pro} - G_5T complex (Table 1). Surprisingly, the translational diffusion coefficient D of free G_5T at 30°C was found to be $12.4 \pm 0.2 \times 10^{-7} \text{ cm}^2/\text{s}$ for a 1.5 mM sample containing 200 mM KCl, slightly smaller than the free protein's diffusion coefficient which was determined to $12.8 \pm 0.2 \times 10^{-7} \text{ cm}^2/\text{s}$. At a 1:4 protein:DNA ratio (concentration relating to the monomeric form of G_5T) and a salt concentration of 20 mM NaCl, the apparent D for the complex (analyzed using the protein methyl groups) decreased to $11.4 \pm 0.4 \times 10^{-7} \text{ cm}^2/\text{s}$. The apparent D for the complex in a sample with the same molar ratios of 3C^{pro} and G_5T but 150 mM NaCl was $12.1 \pm 0.3 \times 10^{-7} \text{ cm}^2/\text{s}$, and the respective ^1H , ^{15}N -HSQC spectrum showed line widths that were comparable to those observed for the free protein. Continuous chemical shift changes for 3C^{pro} amide cross peaks were observed with the high-salt sample upon titration with G_5T as well as a decrease in

the apparent D ($11.6 \pm 0.4 \times 10^{-7} \text{ cm}^2/\text{s}$ for a 1:20 ratio of protein:monomeric G_5T).

These results suggest that the 3C^{pro} - G_5T complex is in fast exchange in the presence of 150 mM NaCl and that the complex is weakened by the addition of salt. Therefore, ionic interactions may be the main driving force for the interaction. This observation is in agreement with the salt dependency observed in the *in vitro* inhibition of the protease activity. Furthermore, and on the basis of previously reported diffusion coefficients for G-quadruplexes (31,33), we concluded that G_5T may in fact form a quadruplex dimer in solution (see 'Discussion' section). Presumably, the association of two G_5T quadruplexes is promoted by the fact that the 5'-end of G_5T is part of a G-tetrad that is not obstructed by preceding 'unstructured' nucleotides. It is thus feasible that two parallel G_5T -quadruplexes align 'head to head' with their respective 5'-end orientated toward one another, giving rise to extensive aromatic stacking. In such a scenario, the additional resonances observed in the ^1H , ^1H -NOESY spectra could originate from NOEs between the two G_5T quadruplexes within such a dimer.

NMR assignment of 3C^{pro} at physiological pH and ^1H , ^{15}N -HSQC-monitored binding of G_5T

NMR spectra of HAV 3C^{pro} C24S displayed signs of a well-folded protein with good dispersion in the ^1H and ^{15}N dimensions. DOSY measurements yielded a translational diffusion coefficient D of $12.8 \pm 0.2 \times 10^{-7} \text{ cm}^2/\text{s}$ for a 110 μM sample at 30°C , suggesting that the protein is predominantly monomeric in the concentration range chosen.

NMR is well-suited for the analysis of weak ligand binding to small and intermediate size proteins via chemical shift changes and differential line broadening caused by direct binding events and allosteric effects (34). We conducted a ^1H , ^{15}N -HSQC-monitored titration of ^{15}N - 3C^{pro} with G_5T in the presence of 150 mM NaCl. 1D ^1H and 2D ^1H , ^{15}N -HSQC spectra were recorded with 0, 4, 20 and 50 equivalents of G_5T (monomeric concentrations). Both chemical shift changes and selective broadening of backbone NH cross peaks were observed during this titration experiment (Figure 7A). An assignment of 3C^{pro} backbone resonances has been deposited in the BMRB databank prior to this work (25). However, the experimental conditions used for this assignment were very different from the conditions used in the present study. In particular, the original assignment was performed at pH 5.4 where 3C^{pro} exhibits <30% proteolytic activity. The maximum activity was observed between pH 8 and 9, and at pH 7.5 around 80% activity was measured (data not shown). Because of the difficulties in transferring the original assignment to pH 7.5, a reassignment of 3C^{pro} was attempted, using ^2H , ^{13}C and ^{15}N -labeled protein and TROSY variants of two pairs of backbone 3D spectra. A sequential assignment strategy was employed as far as possible, yielding assignment of 70% of the backbone NH resonances (cf. Supplementary Figure S4 for a representative walk through the protein backbone based on HNCACB and HN(CO)CACB spectra).

Table 1. Translational DOSY diffusion coefficients of 3C^{pro} , G_5T and 3C^{pro} - G_5T complexes at 30°C

Sample	$D/10^{-7} \text{ cm}^2/\text{s}$
3C^{pro}	12.8 ± 0.2
G_5T	12.4 ± 0.2
3C^{pro} : G_5T 1:4, 20 mM NaCl	11.4 ± 0.4
3C^{pro} : G_5T 1:4, 150 mM NaCl	12.1 ± 0.3
3C^{pro} : G_5T 1:20, 150 mM NaCl	11.6 ± 0.4

Relative G_5T amounts refer to monomeric concentrations.

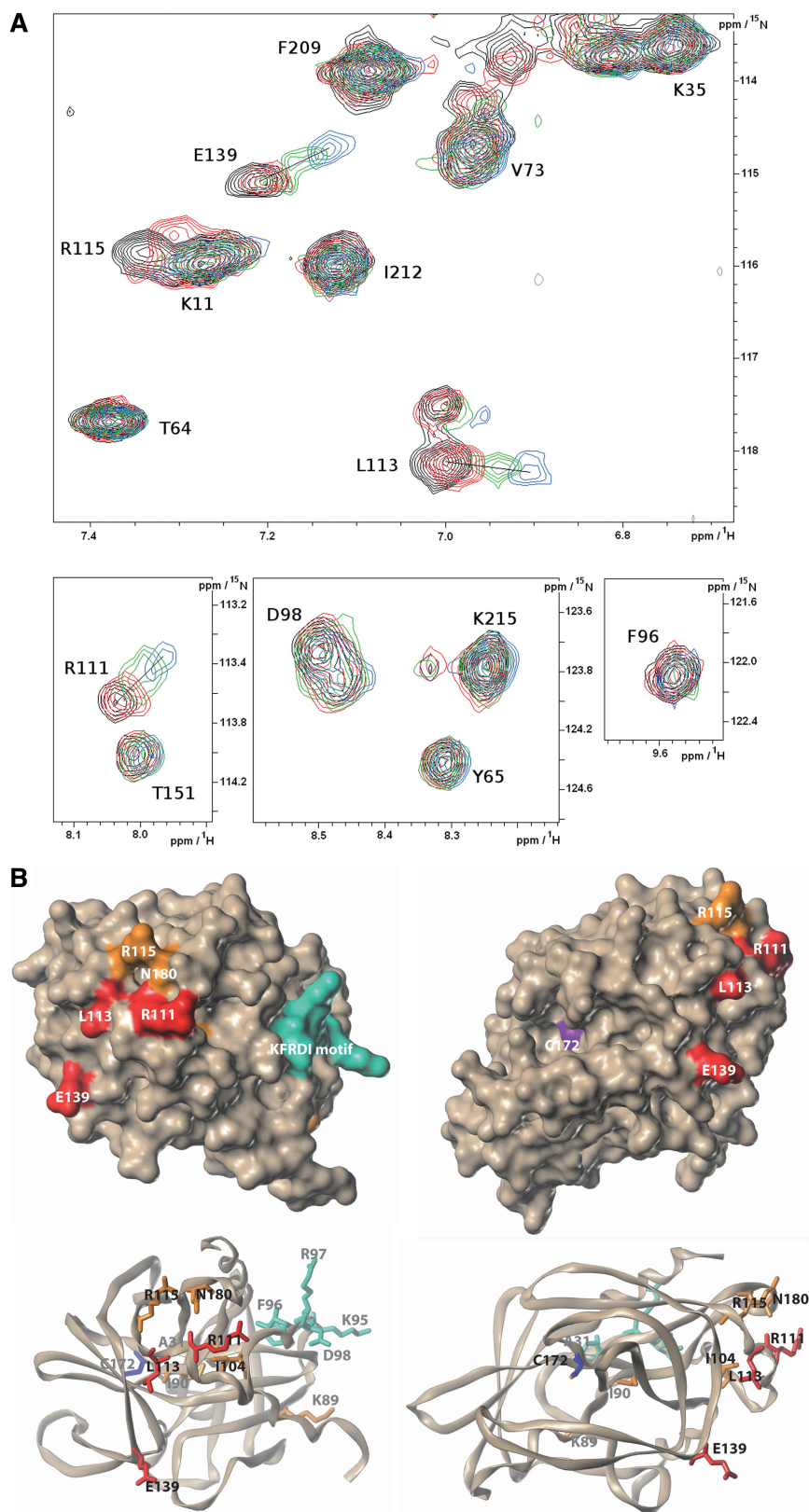


Figure 7. (A) G_5T binding site on $3C^{Pro}$ mapped by NMR. Details of the 1H , ^{15}N -HSQC spectra of $3C^{Pro}$ with different concentrations of G_5T . Ratios were 1:0 (black), 1:4 (red), 1:20 (green) and 1:40 (blue) equivalents of G_5T (monomeric G_5T concentrations). Examples of residues which undergo chemical shift changes are shown (R111, L113 and R115) as well as unaffected residues from the KFRDI motif (F96 and D98). Resonances experiencing large chemical shifts are connected by black lines. (B) Chemical shift changes mapped onto $3C^{Pro}$: Mapping of chemical shift changes on $3C^{Pro}$ induced by 10 equivalents G_5T on the HAV $3C^{Pro}$ crystal structure (PDB entry 1QA7). Red: large shift changes, orange: medium shift changes, turquoise: KFRDI motif, purple: catalytic C172. The changes in the 1H , ^{15}N -HSQC spectrum are mostly confined to the C-terminal β -barrel and preceding helix. Figure rendered using Sybyl Software (Tripos Assoc.).

This strategy was supplemented by the use of a MTSL spin label and measurements of paramagnetic relaxation enhancements in conjunction with the HAV 3C^{PRO} crystal structures deposited in the pdb, which allowed another 5% of residues to be unambiguously assigned. The C^α, C^β, C', N and H backbone resonances were referenced using CHECKshift (35,36) and deposited in the BMRB database under the accession code 16837.

Chemical shift changes and/or selective peak broadening were observed for a set of 3C^{PRO} amide NH resonances upon addition of G₅T, indicating their direct or indirect involvement in complex formation. Most cross peaks that experienced significant chemical shift changes were also broadened. The residues with the largest chemical shift changes, where an assignment at pH 7.5 was available, were R111, L113, R115 and E139 (Figure 7A and Supplementary Figure S5). Only for the first two titration steps, cross peaks were observed for R115, which broadened beyond detection during subsequent titration steps. Smaller but significant shift changes were also observed for A31, K89, I104, N180 and I190. Mapping of these residues onto the crystal structure of 3C^{PRO} (15) revealed that they are almost all part of the C-terminal β-barrel or the preceding α-helix (Figure 7B). Only A31 and I190 are in proximity of the catalytic residue C172, which belongs to the C-terminal β-barrel that is affected by binding of G₅T. Residues of the conserved KFRDI motif, on the other hand, did not experience significant chemical shift changes or line broadening during titration with G₅T, as can be seen clearly for F96 and D98 (Figure 7A).

DISCUSSION

In this study, we identified a small number of hexanucleotides with measurable binding affinity to HAV 3C^{PRO} using a hexanucleotide array. Hexanucleotide binding does not compete with binding of viral RNA to HAV 3C^{PRO}. For one of the hexanucleotides identified, G₅T, a detailed analysis of its interaction with HAV 3C^{PRO} was performed.

From gel mobility assays and ¹H NMR spectroscopy, it became evident that this hexanucleotide forms a higher order structure with features indicative of G-quadruplex formation. Therefore, we studied this possibility using DOSY NMR before entering NMR binding studies. DOSY experiments delivered translational diffusion coefficient of $12.4 \pm 0.2 \times 10^{-7}$ cm²/s (at 30°C) for free G₅T. This value is surprisingly small when compared to 3C^{PRO} ($12.8 \pm 0.2 \times 10^{-7}$ cm²/s), given that the molecular weight of 3C^{PRO} (~24 kDa) is roughly four times the weight of the G₅T quadruplex. However, it has to be taken into consideration that in dilute solutions translational diffusion reflects on hydrodynamic radii rather than molecular weights. A comparison of the crystal structures of HAV 3C^{PRO} [pdb entries for example 2A4O, 1HAV, 2CXV (15,37,38)] and [TG₄T]₄ [pdb entry 244D, (32)] demonstrates that the size of a parallel stranded hexanucleotide G-quadruplex is larger than its molecular weight suggests, reaching to roughly half the size of 3C^{PRO}. This observation, together with literature values for

similar quadruplexes, leads us to conclude that G₅T might in fact form a four-stranded quadruplex-dimer in solution. The diffusion coefficient for the 32-nt G-quadruplex [T₂G₄T₂]₄ was reported to be 13.4×10^{-7} cm²/s at 20°C (33). In another report (31), a diffusion coefficient of 14.2×10^{-7} cm²/s was found for the 24-nt G-quadruplex [G₄T₄G₄]₂ at 25°C. G₅T, also forming a 24-nt G-quadruplex ([G₅T]₄), would thus diffuse slower at 30°C than [G₄T₄G₄]₂ at 25°C although an increase in temperature generally promotes diffusion. Similarly, the higher diffusion coefficient for the slightly larger G-quadruplex [T₂G₄T₂]₄ at 20°C is not in accordance with the assumption that G₅T is a quadruplex [G₅T]₄. Rather, these comparisons as well as the similarity in the diffusion times of G₅T and 3C^{PRO} suggest that two G₅T-quadruplexes [G₅T]₄ associate in solution to form a dimer of quadruplexes, [[G₅T]₄]₂. The apparent molecular weights of G₅T and 3C^{PRO} were found to be 33 and 22.8 kDa, respectively, in a gel filtration experiment (data not shown). While the latter value is in good agreement with the calculated molecular weight for 3C^{PRO} (23.9 kDa), G₅T does again appear much larger than this would be expected for an individual quadruplex [G₅T]₄. We speculate that this unusual behavior of G₅T is a consequence of the uncapped 5'-end guanosine. This residue is likely part of an easily accessible G-tetrad at the 5'-end of a parallel quadruplex, allowing a second quadruplex to undergo favorable stacking through its own 5'-end. It is noteworthy that such a stacking is observed in the crystal structure of TG₄T despite the presence of 'loose' thymine residues at either end of the quadruplex (32).

NMR binding studies were based on ¹H,¹⁵N-HSQC spectra of 3C^{PRO} in the presence of G₅T. This allowed identification of amino acids that are directly or indirectly affected by binding of the hexanucleotide (Figure 7B). Not all of the amino acids that display chemical shift changes are necessarily in direct contact with G₅T. Some of the effects may be due to conformational changes of 3C^{PRO} upon binding to G₅T. For example, the changes in the N-terminal β-barrel could be due to disturbances of a hydrogen bonding network connecting both barrels. Such conformational changes could also be the reason for a decreased proteolytic activity in the presence of G₅T. According to the ¹H,¹⁵N-HSQC spectra, the proteolytic cleft itself is not affected by binding of G₅T. This suggests that the mechanism of inhibition is non-competitive. More recently, a competitive allosteric mechanism has been described (39) that would also be in accordance with G₅T binding to a site distant from the proteolytic cleft. Our attempts to discriminate between these two alternatives by kinetic analysis (cf. Supplementary Figure S1) were not successful. Also, it has to be taken into consideration that chemical shift changes could be followed only for a subset of amide cross peaks, leaving the possibility that some shift changes closer to the proteolytic cleft were overlooked. Crystallization attempts are underway to further elucidate the mechanism of inhibition.

It is instructive to compare our results with two recent studies which have addressed the binding of viral RNA to 3C^{PRO} of poliovirus and rhinovirus (21,22).

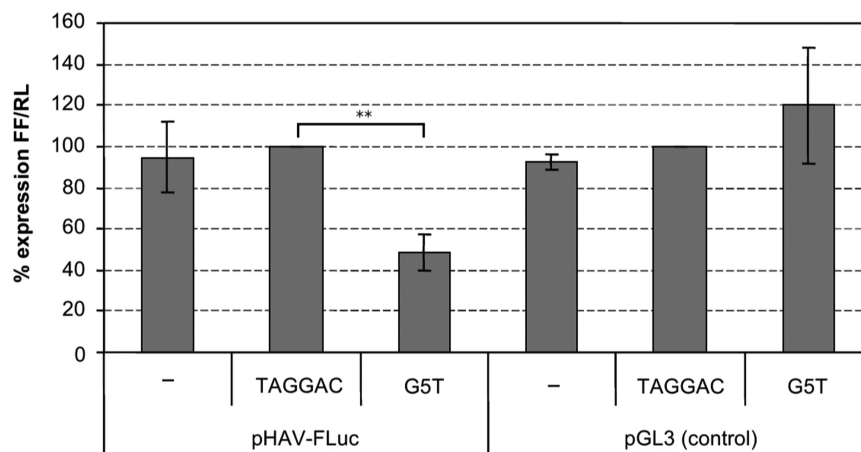


Figure 8. Inhibition of HAV replication by G₅T. A luciferase-based monitor system for replication of HAV (pHAV-Luc) shows 60% inhibition of the luciferase expression by G₅T (bar #3) in comparison to a control hexanucleotide (bar #2) and control plasmid (bars #4–6). Data were compared using paired *t* test and the program Prism 4 (GraphPad Software) (*n* = 3). Indicated are mean values ± SEM; ***P* < 0.01.

In poliovirus 3C^{pro}, the conserved KFRDI motif along with the N-terminal α -helix, exhibited chemical shift changes when titrated with viral RNA. Similarly, this motif was also found to be affected in an NMR titration of rhinovirus 3C^{pro} with the untranslated 5'-end of the viral RNA. These studies provide further evidence that the conserved KFRDI motif, which has previously been shown to be involved in the binding of the viral RNA, is indeed of central importance for the role of HAV 3C^{pro} in genome replication.

In HAV 3C^{pro}, residues 95–99 contain the KFRDI motif, located (as in poliovirus 3C) in the interdomain connection loop between the two β -barrel domains. In this study, we found this stretch in HAV 3C^{pro} not affected by the addition of G₅T in ¹H, ¹⁵N-HSQC spectra. This observation was further substantiated by gel shift assays in which addition of G₅T to the viral RNA/3C^{pro} complex caused a super-shift, indicating that G₅T does not compete with viral RNA for binding to 3C^{pro}.

We found that interaction of 3C^{pro} and G₅T strongly depends on ionic strength, suggesting that most molecular contacts are mediated by ionic side chains of the protein and the anionic backbone of the G-quadruplex. To render the complex less sensitive to physiological salt concentrations, it should be possible to 'supplement' the complex with additional hydrophobic interactions, for instance, by the addition of thymine loops to the oligonucleotide. Because structural prediction of G-quadruplex conformations is not reliable to date, a systematic array-based screening with marginally longer G-rich sequences and higher thymine content could be a step toward the design of a tighter binding ligand starting from G₅T. It is noteworthy that, despite some structural ambiguity, the complex between the thrombin binding aptamer (TBA) and thrombin can also be characterized as mediated by a number of ionic interactions with the phosphate backbone, supplemented by aliphatic interactions with single thymine and guanine residues that reach out of the TBA helical core (40).

Regarding the biology and relevance of the interaction between G₅T and HAV 3C^{pro}, it is noteworthy that among the 3C^{pro}-binding hexanucleotides identified in this study two sequences are present within HAV 5'-UTR, 5'-CCGGAG-3' and 5'-AGGCTA-3' (Figure 1). In the context of the viral 5'-UTR RNA, these sequence segments are both thought to be involved in intramolecular duplex formation (41), and therefore would not be available for interactions with 3C^{pro} in the same way as they would do as single-stranded oligonucleotides. This is consistent with the observed independent binding of *in vitro* transcribed HAV 5'-UTR RNA and G₅T; viral RNA and G₅T recognize different sites of 3C^{pro}. Both the primary sequence of G₅T and its propensity to form G-quartets are important for site-specific binding and interference with the enzymatic activity of 3C^{pro}. While other hexanucleotides were also able to bind HAV 3C^{pro} (Figure 1), these interactions were not classified as functional with respect to inhibition of the proteolytic activity. The species CCGGAG (group I, pos. 25–30 of 5'-UTR RNA; Figure 1) and AGGCTA (group III, pos. 91–96 of 5'-UTR RNA) have been included in gel shift assays (Figure 2) where they did not provide any evidence for the formation of higher order complexes such as G-quartets. ¹H NMR spectra of these compounds also showed all characteristics of monomeric single-stranded oligonucleotides (shown for CCGGAG in Figure 5B).

To test whether G₅T could suppress viral replication, we studied the influence of G₅T on viral functions in cell culture-based assays using a luciferase-based monitor system (42). An HAV control region-driven and 3C^{pro}-dependent luciferase expression plasmid (pHAV-Fluc) was transfected in permissive Huh-T7 cells in the presence of G₅T or other hexanucleotides. For control purposes, we also replaced pHAV-Fluc by the HAV-lacking standard plasmid for luciferase pGL3. A significant suppression of HAV-controlled gene expression was observed by G₅T but not when other hexanucleotides were used instead (Figure 8). This is in line with preliminary experiments indicating that in the presence of G₅T

but not in the presence of other hexamers the formation of infectious recombinant HAV particles in HAV-infected Huh-T7 cells was significantly decreased (data not shown). These experiments strongly indicate that G₅T acts as an inhibitor of HAV-specific gene expression, presumably via its inhibitory effects on 3C^{pro} as suggested by the *in vitro* studies described above. Based on these findings, G₅T might be considered as a starting point for the development of pharmaceutically relevant anti-HAV compounds.

Two recent studies reported the use of small oligonucleotides as inhibitors of picornaviral and hepatitis C virus (HCV) internal ribosomal entry sites (IRES)-mediated translation (43,44). In one of these studies (43), poly(rC) binding proteins 1 and 2 (PCBP1 and 2) were found to interact with a short C-rich oligonucleotide (CCCCCTT), thereby inhibiting picornaviral IRES-dependent translation. Two yet uncharacterized cellular proteins which contribute to IRES-mediated translation in HCV were quarried in the second study (44). Both studies demonstrate how short DNA oligonucleotides can be employed to identify non-canonical cellular factors which are part of the viral translation machinery. While we cannot prove that such cellular factors do not exist for HAV 3C^{pro}, on the basis of the extensive characterization of this interaction *in vitro* we think that it is likely that the inhibition observed in this study is due to a direct interaction between G₅T and 3C^{pro}.

CONCLUSION

We employed a range of biochemical and biophysical techniques to thoroughly characterize the interaction between HAV 3C^{pro} and a small synthetic oligonucleotide. Our results, together with previous work in this direction (24) support the hypothesis that small nucleotide fragments such as the DNA hexamers studied here have a potential as specific ligands that interfere with protein function. We suggest that further modification of such ligands may lead to novel inhibitors of virus replication *in vivo*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–5.

FUNDING

German Research Council (DFG Pe494/1-1 and DFG Pe494/1-2 to V.G.M. and T.P., DFG ME1830 and DFG Pe494/8-1 to T.P.); European Social Funds and the state of Schleswig-Holstein within a competence center [ASH2000/32/19535 (KDDTM) to G.S. and T.P.]. We also acknowledge support by the EU within the SILVER program (WP3). Funding for open access charge: Institute of Molecular Medicine and Deutsche Forschungsgemeinschaft grant DFG Pe494/8-1.

Conflict of interest statement. None declared.

REFERENCES

- Barton, J.L., Bunka, D.H., Knowling, S.E., Lefevre, P., Warren, A.J., Bonifer, C. and Stockley, P.G. (2009) Characterization of RNA aptamers that disrupt the RUNX1-CBFBeta/DNA complex. *Nucleic Acids Res.*, **37**, 6818–6830.
- Hall, B., Micheletti, J.M., Satya, P., Ogle, K., Pollard, J. and Ellington, A.D. (2009) Design, synthesis, and amplification of DNA pools for *in vitro* selection. *Curr. Protoc. Mol. Biol.*, **Chapter 24**, Unit 24 22; October 2009 (doi:10.1002/0471142727.mb2402588; epub ahead of print).
- Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
- Jeon, S.H., Kayhan, B., Ben-Yedidia, T. and Arnon, R. (2004) A DNA aptamer prevents influenza infection by blocking the receptor binding region of the viral hemagglutinin. *J. Biol. Chem.*, **279**, 48410–48419.
- Joyce, G.F. (1994) *In vitro* evolution of nucleic acids. *Curr. Opin. Struct. Biol.*, **4**, 331–336.
- Schneider, D., Tuerk, C. and Gold, L. (1992) Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. *J. Mol. Biol.*, **228**, 862–869.
- Brodin, P., Pinskaya, M., Buckle, M., Parsch, U., Romanova, E., Engels, J., Gottikh, M. and Mouscadet, J.F. (2002) Disruption of HIV-1 integrase-DNA complexes by short 6-oxocytosine-containing oligonucleotides. *Biochemistry*, **41**, 1529–1538.
- Buchko, G.W., Tung, C.S., McAteer, K., Isern, N.G., Spicer, L.D. and Kennedy, M.A. (2001) DNA-XPA interactions: a (31)P NMR and molecular modeling study of dCCAATAACC association with the minimal DNA-binding domain (M98-F219) of the nucleotide excision repair protein XPA. *Nucleic Acids Res.*, **29**, 2635–2643.
- Mescalchin, A., Wunsche, W., Laufer, S.D., Grohmann, D., Restle, T. and Szakiel, G. (2006) Specific binding of a hexanucleotide to HIV-1 reverse transcriptase: a novel class of bioactive molecules. *Nucleic Acids Res.*, **34**, 5631–5637.
- Wyatt, J.R., Vickers, T.A., Roberson, J.L., Buckheit, R.W. Jr, Klimkait, T., DeBaets, E., Davis, P.W., Rayner, B., Imbach, J.L. and Ecker, D.J. (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proc. Natl Acad. Sci. USA*, **91**, 1356–1360.
- Andino, R., Rieckhof, G.E., Achacoso, P.L. and Baltimore, D. (1993) Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *EMBO J.*, **12**, 3587–3598.
- Birtley, J.R., Knox, S.R., Jalent, A.M., Brick, P., Leatherbarrow, R.J. and Curry, S. (2005) Crystal structure of foot-and-mouth disease virus 3C protease. New insights into catalytic mechanism and cleavage specificity. *J. Biol. Chem.*, **280**, 11520–11527.
- Matthews, D.A., Smith, W.W., Ferre, R.A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J.E., Janson, C.A., McElroy, H.E., Gribskov, C.L. *et al.* (1994) Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*, **77**, 761–771.
- Mosimann, S.C., Cherney, M.M., Sia, S., Plotch, S. and James, M.N. (1997) Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.*, **273**, 1032–1047.
- Bergmann, E.M., Cherney, M.M., McKendrick, J., Frommann, S., Luo, C., Malcolm, B.A., Vederas, J.C. and James, M.N. (1999) Crystal structure of an inhibitor complex of the 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein processing in HAV. *Virology*, **265**, 153–163.
- Bjorn Dahl, T.C., Andrew, L.C., Semenchenko, V. and Wishart, D.S. (2007) NMR solution structures of the apo and peptide-inhibited human rhinovirus 3C protease (Serotype 14): structural and dynamic comparison. *Biochemistry*, **46**, 12945–12958.
- Matthews, D.A., Dragovich, P.S., Webber, S.E., Fuhrman, S.A., Patick, A.K., Zalman, L.S., Hendrickson, T.F., Love, R.A., Prins, T.J., Marakovits, J.T. *et al.* (1999) Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple

- rhinovirus serotypes. *Proc. Natl Acad. Sci. USA*, **96**, 11000–11007.
18. Zunszain, P.A., Knox, S.R., Sweeney, T.R., Yang, J., Roque-Rosell, N., Belsham, G.J., Leatherbarrow, R.J. and Curry, S. (2010) Insights into cleavage specificity from the crystal structure of foot-and-mouth disease virus 3C protease complexed with a peptide substrate. *J. Mol. Biol.*, **395**, 375–389.
 19. Shih, S.R., Chiang, C., Chen, T.C., Wu, C.N., Hsu, J.T., Lee, J.C., Hwang, M.J., Li, M.L., Chen, G.W. and Ho, M.S. (2004) Mutations at KFRDI and VGK domains of enterovirus 71 3C protease affect its RNA binding and proteolytic activities. *J. Biomed. Sci.*, **11**, 239–248.
 20. Walker, P.A., Leong, L.E. and Porter, A.G. (1995) Sequence and structural determinants of the interaction between the 5'-noncoding region of picornavirus RNA and rhinovirus protease 3C. *J. Biol. Chem.*, **270**, 14510–14516.
 21. Amero, C.D., Arnold, J.J., Moustafa, I.M., Cameron, C.E. and Foster, M.P. (2008) Identification of the oril-binding site of poliovirus 3C protein by nuclear magnetic resonance spectroscopy. *J. Virol.*, **82**, 4363–4370.
 22. Claridge, J.K., Headey, S.J., Chow, J.Y., Schwalbe, M., Edwards, P.J., Jeffries, C.M., Venugopal, H., Trehwella, J. and Pascal, S.M. (2009) A picornaviral loop-to-loop replication complex. *J. Struct. Biol.*, **166**, 251–262.
 23. Ohlenschlager, O., Wohner, J., Bucci, E., Seitz, S., Hafner, S., Ramachandran, R., Zell, R. and Gorlach, M. (2004) The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure*, **12**, 237–248.
 24. Mescalchin, A., Wunsche, W. and Sczakiel, G. (2011) Specific recognition of proteins by array-bound hexanucleotides. *Angew. Chem. Int. Ed. Engl.*, **50**, 1052–1054.
 25. Bjorn Dahl, T.C., Watson, M.S., Slupsky, C.M., Spyropoulos, L., Sykes, B.D. and Wishart, D.S. (2001) Complete 1H, 13C and 15N backbone assignments for the hepatitis A virus 3C protease. *J. Biomol. NMR*, **19**, 187–188.
 26. Chang, K.H., Brown, E.A. and Lemon, S.M. (1993) Cell type-specific proteins which interact with the 5' nontranslated region of hepatitis A virus RNA. *J. Virol.*, **67**, 6716–6725.
 27. Peters, H., Kusov, Y.Y., Meyer, S., Benie, A.J., Bauml, E., Wolff, M., Rademacher, C., Peters, T. and Gauss-Muller, V. (2005) Hepatitis A virus proteinase 3C binding to viral RNA: correlation with substrate binding and enzyme dimerization. *Biochem. J.*, **385**, 363–370.
 28. Jing, N., Xiong, W., Guan, Y., Pallansch, L. and Wang, S. (2002) Potassium-dependent folding: a key to intracellular delivery of G-quartet oligonucleotides as HIV inhibitors. *Biochemistry*, **41**, 5397–5403.
 29. Aboul-ela, F., Murchie, A.I. and Lilley, D.M. (1992) NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG4T). *Nature*, **360**, 280–282.
 30. Lim, K.W., Alberti, P., Guedin, A., Lacroix, L., Riou, J.F., Royle, N.J., Mergny, J.L. and Phan, A.T. (2009) Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res.*, **37**, 6239–6248.
 31. Cevc, M. and Plavec, J. (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry*, **44**, 15238–15246.
 32. Laughlan, G., Murchie, A.I., Norman, D.G., Moore, M.H., Moody, P.C., Lilley, D.M. and Luisi, B. (1994) The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, **265**, 520–524.
 33. Bolten, M., Niermann, M. and Eimer, W. (1999) Structural analysis of G-DNA in solution: a combination of polarized and depolarized dynamic light scattering with hydrodynamic model calculations. *Biochemistry*, **38**, 12416–12423.
 34. Takeuchi, K. and Wagner, G. (2006) NMR studies of protein interactions. *Curr. Opin. Struct. Biol.*, **16**, 109–117.
 35. Ginzinger, S.W., Gerick, F., Coles, M. and Heun, V. (2007) CheckShift: automatic correction of inconsistent chemical shift referencing. *J. Biomol. NMR*, **39**, 223–227.
 36. Ginzinger, S.W., Skocibusic, M. and Heun, V. (2009) CheckShift improved: fast chemical shift reference correction with high accuracy. *J. Biomol. NMR*, **44**, 207–211.
 37. Bergmann, E.M., Mosimann, S.C., Chernaia, M.M., Malcolm, B.A. and James, M.N. (1997) The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J. Virol.*, **71**, 2436–2448.
 38. Yin, J., Bergmann, E.M., Cherney, M.M., Lall, M.S., Jain, R.P., Vederas, J.C. and James, M.N. (2005) Dual modes of modification of hepatitis A virus 3C protease by a serine-derived beta-lactone: selective crystallization and formation of a functional catalytic triad in the active site. *J. Mol. Biol.*, **354**, 854–871.
 39. Hu, D.D., White, C.A., Panzer-Knodle, S., Page, J.D., Nicholson, N. and Smith, J.W. (1999) A new model of dual interacting ligand binding sites on integrin alphaIIb beta3. *J. Biol. Chem.*, **274**, 4633–4639.
 40. Padmanabhan, K. and Tulinsky, A. (1996) An ambiguous structure of a DNA 15-mer thrombin complex. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 272–282.
 41. Brown, E.A., Day, S.P., Jansen, R.W. and Lemon, S.M. (1991) The 5' nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation in vitro. *J. Virol.*, **65**, 5828–5838.
 42. Gauss-Muller, V. and Kusov, Y.Y. (2002) Replication of a hepatitis A virus replicon detected by genetic recombination in vivo. *J. Gen. Virol.*, **83**, 2183–2192.
 43. Choi, K., Kim, J.H., Li, X., Paek, K.Y., Ha, S.H., Ryu, S.H., Wimmer, E. and Jang, S.K. (2004) Identification of cellular proteins enhancing activities of internal ribosomal entry sites by competition with oligodeoxynucleotides. *Nucleic Acids Res.*, **32**, 1308–1317.
 44. Li, X., Mueller, S. and Wimmer, E. (2009) Inhibition of hepatitis C virus IRES-mediated translation by oligonucleotides. *Virus Res.*, **146**, 29–35.
 45. Kusov, Y.Y. and Gauss-Muller, V. (1997) In vitro RNA binding of the hepatitis A virus proteinase 3C (HAV 3Cpro) to secondary structure elements within the 5' terminus of the HAV genome. *RNA*, **3**, 291–302.
 46. Battiste, J.L. and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry*, **39**, 5355–5365.
 47. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
 48. Allison, S.A. and Tran, V.T. (1995) Modeling the electrophoresis of rigid polyions: application to lysozyme. *Biophys. J.*, **68**, 2261–2270.