



**HAL**  
open science

## Structure and target interaction of a G-quadruplex RNA-aptamer

Kristina Szameit, Katharina Berg, Sven Kruspe, Erica Valentini, Eileen Magbanua, Marcel Kwiatkowski, Isaure Chauvot de Beauchêne, Dmitri Svergun, Hartmut Schlüter, Martin Zacharias, et al.

► **To cite this version:**

Kristina Szameit, Katharina Berg, Sven Kruspe, Erica Valentini, Eileen Magbanua, et al.. Structure and target interaction of a G-quadruplex RNA-aptamer. *RNA Biology*, 2016, 13 (10), pp.973 - 987. 10.1080/15476286.2016.1212151 . hal-01505860

**HAL Id: hal-01505860**

**<https://hal.science/hal-01505860>**

Submitted on 1 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Structure and target interaction of a G-quadruplex RNA-aptamer

Kristina Szameit,<sup>1</sup> Katharina Berg,<sup>1</sup> Sven Kruspe,<sup>2</sup> Erica Valentini,<sup>3</sup> Eileen Magbanua,<sup>1</sup> Marcel Kwiatkowski,<sup>4</sup> Isaure Chauvot de Beauchêne,<sup>5</sup> Dmitri I Svergun,<sup>3</sup> Hartmut Schlüter,<sup>4</sup> Martin Zacharias,<sup>5</sup> Ulrich Hahn<sup>1,\*</sup>

<sup>1</sup>Institute for Biochemistry and Molecular Biology; Department of Chemistry; University of Hamburg; Hamburg, Germany

<sup>2</sup>present address: Carver College for Biomedical Research; Department of Internal Medicine; University of Iowa; Iowa City, IA USA

<sup>3</sup>European Molecular Biology Laboratory; Hamburg Unit; Hamburg, Germany

<sup>4</sup>University Medical Center Hamburg-Eppendorf; Department of Clinical Chemistry; Hamburg, Germany

<sup>5</sup>Physics Department; Technical University Munich; Garching, Germany

\* Correspondence to: Ulrich Hahn; Email: [uli.hahn@uni-hamburg.de](mailto:uli.hahn@uni-hamburg.de)

### Keywords

Aptamers, G-quadruplexes, Interleukin-6 receptor, molecular modelling, protein-RNA interaction, SAXS

### Abbreviations

CD, circular dichroism; DMEM, Dulbecco's modified Eagle's medium; GQ, Guanine quadruplex; hIL-6R, human interleukin-6 receptor; ILP, In-line probing; IL-6, interleukin-6; sIL-6R, soluble interleukin-6 receptor; FBS, fetal bovine serum; gp130, glycoprotein 130; 2'-F-Py, 2'-deoxy-2'-fluoro pyrimidines; FRA, filter retention assay; MST, Microscale Thermophoresis; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline; SELEX, Systematic Evolution of Ligands by Exponential enrichment

## Abstract

G-quadruplexes have recently moved into focus of research in nucleic acids, thereby evolving in scientific significance from exceptional secondary structure motifs to complex modulators of gene regulation. Aptamers (nucleic acid based ligands with recognition properties for a specific target) that form G-quadruplexes may have particular potential for therapeutic applications as they combine the characteristics of specific targeting and G-quadruplex mediated stability and regulation. We have investigated the structure and target interaction properties of one such aptamer: AIR-3 and its truncated form AIR-3A. These RNA aptamers are specific for human interleukin-6 receptor (hIL-6R), a key player in inflammatory diseases and cancer, and have recently been exploited for *in vitro* drug delivery studies. With the aim to resolve the RNA structure, global shape, RNA:protein interaction site and binding stoichiometry, we now investigated AIR-3 and AIR-3A by different methods including RNA structure probing, Small Angle X-ray scattering and microscale thermophoresis. Our findings suggest a broader spectrum of folding species than assumed so far and remarkable tolerance towards different modifications. Mass spectrometry based binding site analysis, supported by molecular modeling and docking studies propose a general G-quadruplex affinity for the target molecule hIL-6R.

## Introduction

Guanine quadruplexes (GQs) represent three-dimensional structure motifs of guanine-rich regions in nucleic acids built from stacked G-tetrads. Their importance has been reported for a variety of biological processes where GQs can function through different mechanisms. Such examples comprise telomere maintenance by genomic DNA-GQs<sup>1,2</sup> or telomeric regulation by the counterpart transcript or telomeric RNA-GQs (TERRA).<sup>3-5</sup> Furthermore GQs hold regulative potential for gene expression either on the transcriptional level by being part of or close to promoters in genomic DNA (e.g. *c-MYC* or *bcl2* promoter)<sup>6,7</sup> or as regulators of splicing<sup>8</sup> and translation<sup>9</sup> (inhibitory or enhancing) when present in mRNAs and finally as regulative elements by binding of certain effector proteins.<sup>3,10</sup> It has been shown that due to this fact GQs are also crucial elements in the development or progression of malignancy such as neurodegenerative disorders<sup>11</sup> and cancer.<sup>5</sup> One property that is important in this regard is the increased stability of GQs compared to differently structured DNA or RNA. This stability relies however on the presence of cations, complexed by the G-tetrads and is therefore transient. Due to their intrinsic coherence, GQs can form in nucleic acids regardless of their origin (naturally occurring or emerging from chemical synthesis) as long as the abundance and distribution of guanines allows for quadruplex formation. As such, also several aptamers were found to display GQ characteristics. Aptamers are short monomolecular oligonucleotides that display specific affinity towards a certain target molecule for which they were selected. The selection is an iterative process

termed Systematic Evolution of Ligands by Exponential enrichment (SELEX).<sup>12</sup> Since first mentioned in 1990,<sup>12, 13</sup> aptamers have gained more and more importance as versatile tools applied in molecular biology,<sup>14</sup> nanotechnology,<sup>15</sup> and medicinal sciences.<sup>16</sup> It can be assumed that aptamer selections for targets that display general affinity to G-quadruplexes have high probability for GQ forming aptamers to emerge. The idea of combining the benefits of specific targeting by aptamers and intrinsic regulatory potential of GQs is striking. However, the number of identified GQ-binding proteins constantly increases while common features of their binding property often remain unclear and categorization is rather conducted according to the type of GQ bound than on the level of protein characteristics.<sup>17</sup> One such characteristic connected with GQ affinity in some proteins is the presence of the stabilizing Arg-Gly-Gly (RGG) repeats as found for example in fragile X mental retardation protein,<sup>18</sup> nucleolin (a multifunctional heterogeneous nuclear ribonucleoprotein (hnRNP) belonging to the family of RNA binding proteins)<sup>19</sup> or engineered proteins.<sup>20</sup>

GQ forming aptamers include for example nucleolin-specific DNA aptamer AS1411,<sup>21</sup> thrombin binding aptamer,<sup>22</sup> HIV-1 aptamer T30177,<sup>23</sup> aptamers for prion proteins<sup>24-26</sup> or fluorophore binding aptamer spinach.<sup>27</sup> Another example is given by aptamer AIR-3 which has specificity for human interleukin-6 receptor (hIL-6R). This 106 nt RNA aptamer could be truncated to its GQ forming binding motif AIR-3A (19 nt) and has recently been shown effective in targeted drug delivery.<sup>28, 29</sup>

While characterization of AIR-3(A) binding, cell internalization and therapeutic potential was successfully conducted, the knowledge on its structure is limited to the fact that it forms a parallel GQ<sup>30</sup> and that almost all nucleotides in the binding motif are essential to maintain affinity.<sup>31</sup> Here we report on the structural analysis performed on AIR-3 and AIR-3A. Characterization of the comprised GQ was realized by circular dichroism (CD)-detected titration experiments and fragment analyses of in-line probing (ILP) and RNase protection assays. As amendment to former studies<sup>31</sup> we performed SAXS measurements and *ab initio* modeling, yielding a model for templated dimerization of AIR-3A. The stoichiometry of RNA and receptor protein was further investigated by Microscale Thermophoresis (MST). Also, the tolerance of AIR-3A integrity to nucleotide modifications was examined probing nucleoside analog gemcitabine as a possible candidate for future drug delivery approaches. Finally the binding site of AIR-3A on its target protein, hIL-6R, was revealed as being part of the immune globulin like (Ig-like) domain of the receptor molecule. These findings give rise to further investigations on the GQ- affinity of structurally related proteins such as cytoskeleton proteins.

## Results

To determine the structure of aptamer AIR-3A and the basis of its specificity to the protein hIL-6R (and its soluble form sIL-6R), we undertook several approaches. First, we investigated GQ formation by CD-detected titrations of potassium- ( $K^+$ ), lithium- ( $Li^+$ ) and barium-cations ( $Ba^{2+}$ ) to assess affinity and minimum concentrations required for folding.

### GQ-Formation requires potassium

We examined the influence of cation concentration and identity on GQ formation in AIR-3A. As shown by CD, the RNA aptamer adopts an all parallel GQ folding with a positive peak at  $\sim 260$  nm and a negative peak at  $\sim 240$  nm. We performed CD-detected titration of  $K^+$ , plotting the changes in intensity at 262 nm as a function of  $K^+$ -concentration (100 nM to 1 mM) and fitting to a two-state model. The values in a  $K^+_{1/2}$  of  $159 \pm 12 \mu M$  with a Hill coefficient of  $n = 1.5 \pm 0.1$  (Figure 1A + B). While curve progression of the plot indicates one transition, which means that AIR-3A folds into a two-tier G-quadruplex with one cation coordinated, the Hill-coefficient of 1.5 might also hint at the formation of AIR-3A dimers where each quadruplex complexes one  $K^+$  and one additional cation is placed between the two molecules (in total 3  $K^+$  per 2 AIR-3A).

To test the possibility of monovalent cations to be complexed instead of  $K^+$ , titrations were performed with  $Li^+$  and  $Ba^{2+}$ , whereas  $Li^+$  is known to interfere with GQ folding. For AIR-3A in presence of  $Li^+$  concentrations ranging from 1 to 100 mM, no difference to the background CD intensities was observed (Figure 1).

In contrast, with  $Ba^{2+}$  present in solution, a concentration dependent increase was detected (Figure 1C). The resulting  $Ba^{2+}_{1/2}$  value of  $3.35 \pm 0.35 \mu M$  (Figure 1D) as well as the Hill-coefficient are only hypothetical and could not accurately be determined due to the fact that not more than one data point close to the point of inflection was available. Yet, we observed that saturation with  $Ba^{2+}$  was reached at lower concentrations compared to  $K^+$ . In conclusion,  $Ba^{2+}$  exchanges with  $K^+$  and is coordinated with higher affinity, probably due to the twofold positive charge which mediates stronger interaction. In addition comparative CD analysis in presence of  $Na^+$  was conducted to assess GQ formation caused by remnant  $Na^+$  cations (Figure S5). While background GQ signals were obtained, no concentration dependence was observable.

### AIR-3A forms dimers in solution and binds to hIL-6R with a molar ratio of 2:1

Next, we examined stoichiometry of AIR-3A and hIL-6R. Several hints for oligomerization of the aptamer were noticed, such as bands migrating slower in native PAGE analyses appearing in presence of  $K^+$  compared to  $Li^+$  (Figure 2A) and formation of aggregates observed in dynamic light scattering (see 1). For clarification, MST was performed on a dilution series of hIL-6R at constant AIR-3A

concentrations (200 nM). According to the data, the displayed curve reached saturation at about 100 nM hIL-6R (1B). This indicates a stoichiometry of 2:1. These data hint at AIR-3A binding to hIL-6R as a dimer whereas it is also possible that it targets two different sites. We further analyzed, whether molecularity changes occurred at different aptamer concentrations. By UV-detected melting analyses, GQ specific  $T_m$  values obtained for AIR-3A (2.5 – 10  $\mu$ M) were constant indicating equal molecularity over this concentration range.

### **Structure probing reveals polymorphic GQ folding and hints at induced fit mechanism**

Next, we examined the influence of these different cations on the single nucleotide positions in the GQ of AIR-3 and AIR-3A by probing their flexibility and protection from cleavage by ILP and RNase T1 protection assays respectively. Cleavage with RNase T1 targets only Gs and is a powerful tool to elucidate G specific protection.<sup>11, 32</sup> For AIR-3A, overall protection of Gs as well as an increasing protection of the whole oligonucleotide with increasing  $K^+$  concentrations was revealed (Figure 1Figure 3). Total cleavage in dependence of other effectors decreased in the order hIL-6R >  $K^+$  >  $Li^+$  >  $Ba^{2+}$  >  $NH_4^+$  while cleavage ratios for all positions were ranging in similar quantities except for positions G1 and G2 which seem to be favorable for cleavage in general. Position G19, representing the full length and 5'-end of AIR-3A, could not be probed in this setup due to the lack of a cleavage site for RNase T1. Thus, no information on specific G-positions involved in tetrad formation could be obtained.

To investigate the flexibility of all nt positions, we used ILP on AIR-3A. This probing technique monitors the susceptibility of nucleotides in a structured RNA for hydrolysis (so called in-line cleavage) due to long term incubation under alkaline conditions (36 h at pH 8.3).<sup>33</sup> It is applied on GQs to identify increased cleavage in loop regions in presence of  $K^+$  compared to  $Li^+$  and represents a valuable complement to RNase protection. ILP was conducted also on AIR-3, to include positions G1 and G19 and probe the GQ in its original flanking sequences. In AIR-3, the 19 nt GQ element is found within nt positions 32 to 50. ILP was conducted in presence of  $Li^+$ ,  $K^+$  and target protein hIL-6R. Data obtained from ILP experiments indicate the presence of a loop region between the second and third G-tract with strong cleavage of position U40, exceeding the threshold of 2.0, an indication for significant increase of in-line cleavage accessibility (Figure 1Figure 4B). This is in accordance with the hypothesized folding, predicting a loop region from position C39 to U42 (Figure 4Figure 1C). ILP reactivities for the other positions are >0.6 except for A36 showing medium reactivity (Figure 4Figure 1B+C). In presence of hIL-6R however, lower accessibility of positions for A36, G37, U42, G43, U45, G46 and A47 is represented by decreased reactivity (Figure 4Figure 1B+C). Therefore, it can be assumed that these nucleotides mediate interaction between RNA and protein.

Also, in the regions preceding and following the GQ, a  $K^+$  dependent protection is observed for positions C26 to U28, A52, A54 and U55. In presence of target protein, this is shifted to protection of U27 to U30 and increased cleavage of the region following the GQ (Figure 4 Figure 1C). This may point at a conformational change of AIR-3 in presence of hIL-6R.

### **Affinity of AIR-3 variants for hIL-6R**

Recently, successful AIR-3 mediated delivery of cytostatic nucleoside analog 5-FdU was presented.<sup>29</sup> Next to former studies dealing with stabilization by 2'-F-pyrimidines this presented another possibility to post selectively modify hIL-6R specific aptamers with retained affinity.<sup>31, 34</sup> In order to further investigate the tolerance to nucleotide modifications, we identified different aptamer variants containing nucleoside analogs gemcitabine (2',2'-Difluorodeoxycytidine) and 5-FdU which maintained binding capacity to hIL-6R and investigated their stability.

In Figure 1 Figure 5 results from binding studies are compared. AIR-3 variants incorporating gemcitabine AIR-3-Gem ( $K_d = 103 \pm 10$  nM), AIR-3-Gem-2'FU ( $K_d = 73 \pm 5$  nM), AIR-3-Gem-5FdU retained affinity for hIL-6R ( $K_d = 167 \pm 52$  nM) although decreased compared to unmodified AIR-3A ( $K_d = 55 \pm 5$  nM) as confirmed by filter retention assay (Figure 5 Figure 1A). Affinity for hIL-6R presented by target cells (BaF3\_hIL-6R) was also assayed by monitoring the ability of our variants for displacement of fluorescently labeled AIR-3A. We observed the same tendencies (increase in  $K_d$  compared to unmodified aptamer) with a generally higher affinity to the native receptor than to recombinant protein (Figure 5 Figure 1B). The stability of modified AIR-3 derivatives was increased from half-lives of several seconds (unmodified AIR-3A and AIR-3-Gem) to 51 min for AIR-3-Gem-5FdU and >200 min for AIR-3-Gem-2'FU (Figure 5 Figure 1C). Taken together, these data represent another example for AIR-3 tolerance to modifications with therapeutic potential. For additionally tested AIR-3 and AIR-3A variants containing deoxy-pyrimidines dC and dU, as well as for the all DNA analog of AIR-3, a total loss of affinity was observed (data not shown).

### **Small Angle X-ray scattering (SAXS)**

To analyze the overall structure of AIR-3A in solution, SAXS measurements were conducted and the experimental scattering data, resulting molecular models and their fits are displayed in Figure 6. The geometrical parameters of the particle determined from the SAXS data are radius of gyration  $R_g = 1.9 \pm 0.1$  nm, maximum size  $D_{max} = 6.5 \pm 0.5$  nm and the excluded volume  $V_p = 13 \pm 3$  nm<sup>3</sup>. The Guinier plot of the SAXS data (Figure 6C, insert) shows a clear linear dependence strongly suggesting monodispersity of the solution, i.e. presence of a single major conformation of the solute. Given that for nucleic acids  $V_p$  in nm<sup>3</sup> is equal to the molecular mass (MM) in kDa and taking into account that the MM of monomeric AIR-3A is 6.4 kDa, these results indicate that AIR-3A forms dimers in solution.

The distance distribution function  $p(r)$  in Figure 6B has a skewed profile indicating that the particle has an elongated shape. This is corroborated by the *ab initio* shape reconstruction of AIR-3A (Figure 6A) depicting an elongated particle. To get further insights into the structural organization of AIR-3A, molecular modelling was conducted using the predicted model of AIR-3A generated by molecular dynamics simulations. The rigid body refinement using two copies of predicted AIR-3A and assuming a P2 symmetry provided a hybrid quasi-atomic model with the dimerization occurring through the stacking of the tetrads of the two AIR-3A molecules. The hybrid model overlaps well with the *ab initio* shape (Figure 6A) and both models yield good fits to the experimental data (Figure 6C, D). The good agreement between the two independently constructed models lends credit to the structural modelling further corroborating the finding that the solutions studied contain a single folded form of the AIR-3.

### **Binding site analysis**

In order to investigate the binding site of AIR-3A on hIL-6R we performed an approach of Mass Spectrometry analysis of tryptic peptides from hIL-6R. Prior to protease digestion hIL-6R was covalently cross linked to AIR-3A by means of UV-irradiation.<sup>35-37</sup> Purified samples of this covalent adduct, termed herein as hIL-6R:AIR-3A, were subjected to a standard trypsin digestion protocol and the final peptides were analyzed by LC-MS. Samples of hIL-6R UV-irradiated in the absence of AIR-3A served as a reference. From two independent cross linking experiments at least five single analytical samples yielded the data sets for comparison of the identified peptides. Peptides with cross linked aptamer RNA are expected to be of lower abundance in the analysis or missing at all. Therefore, the presence of all peptides from hIL-6R, in particular of domain 1 and domain 2 (D1 and D2), were investigated from these data sets.

Table 1 shows the assigned peptides belonging to D1 and D2 of hIL-6R that were identified from hIL-6R:AIR-3A and the reference. The whole sequence area of D1 and D2 could be covered by the analysis for the reference. In contrast, hIL-6R:AIR-3A yielded coverage of this area lacking two peptides. These sequence sections are in close proximity to each other in D1 of hIL-6R and are suspected to be the contact area of AIR-3A on the hIL-6R surface (Figure 1Figure 7).

To examine if this area is suitable for GQ-binding we performed a molecular docking simulation. Since an exact structure for AIR-3A was not available we made use of the fact that another GQ, namely DNA-aptamer AID-1<sup>41</sup>, binds the same epitope of hIL-6R. It is therefore reasonable to assume that both GQ share similar structure and binding site on hIL-6R. Thus we performed the docking simulation with the structure of AID-1's homodimer.



The DNA-protein interface shows a dozen of hydrogen-bonds, mainly implying the DNA backbone (invariant between DNA and RNA), and either the protein backbone (residues 116-118) or arginine side-chains (R54, R76, R86 and R114). The tyrosine Y78 establishes a hydrogen bond with a DNA phosphate and a stacking interaction with a flipped-out thymine base (dT8, see 1). The protein binding site, defined as the residues with at least one atom at 4 Å from any DNA atom, comprises residues 12-15, 29, 42, 44-50, 52, 54, 62, 74, 76, 78, 82-86, 90 and 123-124. The large overlap between the interface in our docking model and the interacting protein residues revealed by mass spectrometry analysis reinforces our confidence in our model.

## Discussion

In the present study we gave insight into the structural properties and target specificity of hIL-6R specific RNA aptamer AIR-3A. Although this GQ forming aptamer has been dealt with in earlier publications and was subjected to intensive characterization, we now focused on clarification of GQ folding and target binding interaction. It turned out that several features had been undiscovered so far. In this regard, we found that no specific structuring can be determined and that AIR-3A is likely to oligomerize in solution as well as in presence of target molecule hIL-6R.

By CD-detected titration experiments, we observed that the quadruplex formed by AIR-3A has affinity for different cations. As such, high affinity for Ba<sup>2+</sup> was shown reaching saturation at concentrations of 320 μM. This finding is promising with regard to future applications in which barium could function as an intrinsic scattering agent to investigate AIR-3A structure (X-ray analysis) or cellular trafficking by in-liquid TEM<sup>38</sup> if retained target affinity can be confirmed. Further, for potassium the K<sup>+</sup><sub>1/2</sub> value for AIR-3A could be determined to 159 ± 12 μM with the suggestion of one transition. The number of K<sup>+</sup> binding sites, given by the Hill-coefficient was n = 1.5, which gives rise to two assumptions. (1) Although the curve progression indicates a two state transition and hints at the formation of a two tier GQ, also concomitant binding of two cations could take place resulting in formation of a three-tier GQ with equal K<sup>+</sup><sub>1/2</sub> for both sites. In general, this might be possible as after energy minimization in molecular modeling structures resulted that allowed for formation of a third tier in the GQ (formed by G1, G10, G15 and G19). (2) However, it may also be that two AIR-3A molecules dimerize, coordinating each one K<sup>+</sup> sandwiched between their two tetrads and one additional cation between the two aptamers. Finally, both assumptions could also be combined. If the number of complexed cations was known, the number of tetrads could be deduced. This could be determined using for example native mass spectrometry.

Indications for dimerization were also gained from other experiments as for example by MST. Here, stoichiometric analyses pointed at a RNA:protein-ratio of 2:1. Also electrophoretic mobility shift

assays (EMSA) in presence of  $K^+$  led to formation of additional bands with higher migration behavior. With  $Li^+$  present this could not be observed which indicates G-quadruplex association.<sup>39</sup> Finally, DLS measurements in preparation of the SAXS experiment proved severe aggregation of AIR-3A in solution under standard buffer conditions. Even after buffer optimization yielding homogeneous particle size, SAXS data suggested a global shape of AIR-3A that fitted with the molecular model of a dimeric two-tier GQ.

Taking into account the stoichiometric ratio and that in solution oligomerization of AIR-3A was observed by PAGE, DLS and SAXS it can be assumed that the binding species is oligomeric and possibly dimeric. As for hIL-6R specific RNA aptamer RAID3, Small Angle X-ray scattering (SAXS) provided hints for dimerization, dimeric AIR-3A is likely to form as well. SAXS would therefore also be an option for AIR-3A analysis to prove this assumption. In Figure 6, the calculated molecular model for a possible dimer of AIR-3A is displayed in the superimposition.

Remarkably, in our structure probing analyses we observed a general protection of Gs in the GQ region and similar ILP reactivities. This gives rise to the assumptions that either a) all Gs take part in GQ formation, forming a three-tier bulged GQ which allows for the respective protection, or b) different GQ foldings co-exist at once of which protection/reactivity patterns overlay in the probing, resulting in similar signals for all Gs.

In conclusion, we suppose that the hypothesized GQ structure (Figure 1Figure 4C) is only one of several possible foldings. These could be analyzed and differentiated by further structural analyses of aptamer variants with different G exchanges that still bind to hIL-6R.<sup>31</sup>

Despite the difficulties to elucidate the detailed involvement of guanines in tetrad formation, our data however confirmed the presence of the hypothesized loop region formed by AIR-3 nucleotides C39 to U42 (C8-U11 in AIR-3A). Moreover, the reactivity profile indicated predicted loop positions A36, U42, U45, G46 and A47 as well as tetrad associated G37 and G43 to take part in target interaction. Another new observation was the shift in reactivity for GQ flanking sites C26 to U30 and A52 to U55 in presence of hIL-6R. A possible explanation would be an induced fit mechanism, causing the G-quadruplex in AIR-3 to adopt an energetically favorable conformation when interacting with the target protein. For AIR-3A, dimer formation might be favorable, as it lacks these flanking sites.

Our studies on AIR-3 derivatization with modified nucleotides confirmed former findings on the remarkable tolerance of this aptamer toward pyrimidine modification.<sup>12, 29</sup> Taken together, all pyrimidines could be exchanged for 2'-F modified building blocks or nucleoside analogs gemcitabine and 5-FdU with retained target affinity and increased stability toward serum nucleases. The derivatives presented herein could next be tested for therapeutic effectivity in targeted drug

delivery. As derivatization with deoxy pyrimidines resulted in loss of affinity, overall structural investigations of AIR-3 derivatives might reveal features important for hIL-6R interaction that the deoxy variants lack.

Concerning interaction with target protein hIL-6R our binding site analysis indicated domain D1 in the extracellular region of hIL-6R to be involved in AIR-3A binding. As depicted in Figure 1Figure 7, peptides that were identified as mediators of aptamer recognition by LC-MS analysis (yellow) are located in domain D1. Considering the electrostatic surface potential of the receptor, this site has high probability for nucleic acid binding as it is congruent with a high density of positively charged residues (blue). This accounts for G-quadruplexes in particular as they create a polyanionic shape with high negative charge density.<sup>40</sup>

Against this background, one considerable aspect for future analyses is the investigation of hIL-6R affinity towards other G-quadruplexes as, next to AIR-3(A), two other aptamers (AID-1 and RAID3) independently selected for hIL-6R, were shown to form GQs.<sup>34, 41</sup> Additionally, AID-1 also showed affinity for gp120 as well as HIV-1 integrase which in turn is known to bind GQs.<sup>42</sup>

The number of identified GQ binding proteins constantly increases,<sup>17</sup> especially as during the past years the regulatory potential of GQs in biological processes moved into focus of research.<sup>43</sup> There are numerous questions that remain to be solved concerning the common properties of GQ binding proteins as well as the endogenous functions of the GQs targeting them and their interplay.

It may therefore be that hIL-6R is a GQ binding protein as well, which might in conclusion hint at a GQ mediated regulation of hIL-6R associated processes.

With the aim of investigating binding site characteristics, our docking studies of DNA GQ AID-1 on hIL-6R D1 confirmed the general suitability of this proposed region for GQ interaction. Next to hydrogen bonding that is mediated by the phosphate backbone and several amino acid residues in proximity, also stacking interactions were shown to be possible as for tyrosine Y78 and thymine base dT8 which in the model both stick out (1Figure S2). However, docking models can never be fully reliable, and this model should be a starting point for further experimental validations, for instance by directed mutagenesis on residues establishing key interactions in our model (e.g. the four reported arginines).

The impact of GQs interacting with hIL-6R remains to be specified. However, a natural predisposition for GQ affinity on the protein side was shown. With respect to AIR-3A, a predicted mRNA for RASSF1, a gene encoding Ras association domain-containing protein 1, might link to future experiments on the topic. This predicted mRNA contains a variant of AIR-3A (AIR-3A\_C8A\_U14A) which may

represent a natural equivalent of the aptamer. If verifiable, it could be investigated with regard to potential functions as a control element for example for Ras protein signal transduction.

## Materials and Methods

### Chemicals and buffers

Chemicals were purchased from Sigma-Aldrich if not stated otherwise. All buffers were prepared using 2 x deionized water (ddH<sub>2</sub>O) from a Millipore system. For binding studies, 1 x PBS (137 mM NaCl, 2.7 mM KCl, 6.5 mM NaHPO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.5) was used and supplemented as needed with Tween (0.1% (SAXS), 0.05% (MST)), BSA (1 mg/mL (MST)), MgCl<sub>2</sub> (3 mM (FRA)). For native gel EMSA 1 x TBE (100 mM Tris, 80 mM Boric acid, 1 mM EDTA, pH 8.4) was used, whereas for sequencing gels we applied 1 x TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.4). Assays monitoring monovalent cation dependence were conducted using Tris-HCl (10 mM, pH 7.5).

### Proteins

IL-6 for cell culture was purchased from PeproTech sIL-6R (here referred to as hIL-6R) was kindly provided from Conaris Research Institute AG. RNase A and RNase T1 was purchased from Thermo Scientific.

### Nucleotides

Nucleoside analogs gemcitabine and 5-FdU were purchased as triphosphates from Jena Bioscience. Radio labeled [ $\gamma$ -<sup>32</sup>P]-ATP was purchased from Hartmann Analytic. DNA oligonucleotides (Table 1) were purchased from Life Technologies.

AIR-3A (Table 1) was customarily synthesized with optional modification (5'-AlexaFluor<sup>®</sup> 647) by IBA or Metabion.

### *In vitro* transcription

AIR-3 and derivatives (Table 1) containing nt modifications were synthesized according to Kruspe *et al.*<sup>29</sup> as follows: template DNA from unpurified PCR product (approx. 0.5  $\mu$ M) was used in transcription reaction containing 40 mM Tris-HCl (pH 8.1), 20 mM MgCl<sub>2</sub>, 2 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 1.5% PEG 6,000, pyrophosphatase (1 U/mL, Thermo Scientific, EF0221) NTPs as needed (GTP, ATP CTP, UTP 2 mM each (Carl Roth, K049.1)) or partially exchanged for gemcitabine-triphosphate (1 mM, Jena Bioscience, NU-1607S), 5-FdUTP (1 mM, Jena Bioscience, NU-154S) or 2'-F-UTP (1.25 mM, TriLink BioTechnologies, N-1010). T7 RNAP Y639F (100 U, in house prepared) was added followed by incubation for 4 to 8 hours at 37 °C. RNA was purified by 10% denaturing PAGE, excision and diffusion elution (0.3 M NaOAc, pH 5.2) for two hours at 50 °C and

subsequent ethanol precipitation. Precipitates were air dried, dissolved in nuclease free water and stored at -20 °C.

### **Cell culture**

BaF3/gp130/IL6R/TNF (referred to as BaF3\_hIL-6R) cell line was cultured in Dulbecco's modified Eagle medium (PAA, E15-843), penicillin (60 mg/L, PAA, P11-010) and streptomycin (100 mg/L, PAA, P11-010) at 37 °C and 5% CO<sub>2</sub>. Culture medium was further supplemented with human IL-6 (10 ng/mL, PeproTech, 200-06). BaF3/gp130/IL6R/TNF cell line was kindly provided from the Rose-John lab (University of Kiel, Germany).

### **Molecular modeling of the AIR-3A RNA quadruplex structure**

A modified version of the JUMNA program<sup>44, 45</sup> was used in combination with the Amber 4.1 force field<sup>46</sup> for all calculations. The JUMNA approach utilizes a combination of helicoidal and internal coordinates to describe nucleotide placement and flexibility of a DNA or RNA molecule. The helicoidal description allows easy nucleotide addition and substitution in a given structural motif. Stepwise modeling was performed on the sequence 5'-GGGGAGGCUGUGGUGAGGG-3' following a previously designed protocol.<sup>47</sup> The placement of guanine residues in experimentally known quadruplex structures (PDB: 1MY9 and PDB: 3IBK) served as template for generated a quadruplex core for AIR-3A consisting of the eight underlined guanine nucleotides. Additional nucleotides and the short predicted stem were initially placed close to neighboring nucleotides in the structure. The structure was energy minimized by keeping only the added nucleotides and neighboring nucleotides mobile and locking helical coordinates of the central core formed by guanine nucleotides. Finally, the entire model including a potassium ion placed in the central pore of the quadruplex was energy minimized resulting in a structure with stable quadruplex structure. Interestingly, the energy minimization resulted in structures that allowed also formation of a third layer of guanines (formed by G1, G10, G15 and G19) as a quadruplex.

### **Small Angle X-ray scattering**

Proper folding was ensured by heating and successive cooling in 5 - 10 °C steps from 95 to 20 °C (40 µM in 1 x PBS with 0.1% Tween). Absence of aggregated aptamer was verified by dynamic light scattering.

Synchrotron SAXS measurements were performed at the European Molecular Biology Laboratory (EMBL) on the storage ring PETRA III (DESY, Hamburg) on the EMBL P12 beamline<sup>48</sup> equipped with a robotic sample changer and a 2D photon counting pixel X-ray detector Pilatus 2M (DECTRIS, Switzerland). The scattering intensity,  $I(s)$ , was recorded in the range of the momentum transfer  $0.05 < s < 4.5 \text{ nm}^{-1}$ . The measurements were carried out in water at 10° C using continuous flow operation

over a total exposure time of 1 s divided into 20 x 50 ms individual frames to monitor for potential radiation damage (no radiation effects were detected). One solute concentration of 0.10 mg/ml was measured. The data were normalized to the intensity of the transmitted beam, radially averaged, corrected for the solvent scattering and processed using standard procedures.<sup>49</sup>

The particle radius of gyration,  $R_g$ , was calculated from the experimental SAXS pattern using the Guinier approximation<sup>50</sup> and the maximum particle dimension,  $D_{max}$  was assessed through the distance distribution function  $p(r)$  computed by the program GNOM.<sup>51</sup> The excluded volume was calculated from the Porod invariant.<sup>52</sup> The low-resolution shape of AIR-3A was reconstructed *ab initio* using the program DAMMIN<sup>53</sup> that employs simulated annealing to generate a compact dummy atom (bead) model fitting the experimental data. Hybrid rigid body modelling was conducted using the program SASREF<sup>54</sup> to fit the scattering data by a dimeric model consisting of two AIR-3A molecules.

The experimental SAXS data and models were deposited in the SASBDB database ([www.sasbdb.org](http://www.sasbdb.org)), entry SASDAV7.

### **Circular dichroism (CD) spectroscopy and data fitting**

RNAs were desalted from remnant  $\text{Na}^+$  or  $\text{K}^+$  by denaturation in presence of LiOAc and subsequent ethanol precipitation, resuspended in ddH<sub>2</sub>O and prepared at 5  $\mu\text{M}$  in 10 mM Tris-HCl (pH 7.5) with or without salts added at different concentrations ranging from 100 nM to 100 mM (KCl, LiCl, BaCl<sub>2</sub>, NH<sub>4</sub>OAc or NaCl), heated to 80 °C for 3 min and allowed to cool on ice for renaturation until measured. CD spectra were recorded using a Jasco J-815 CD spectrometer. Spectra were acquired at 25 °C every 1 nm from 220 to 320 nm. Each spectrum is an average of ten scans. Data were buffer subtracted and normalized to provide molar ellipticity values.

Titration with cations were conducted in a similar approach as described by Mullen *et al.*<sup>55</sup> To assess the  $\text{K}^+$  concentration at which GQ formation can take place, ellipticity data at wavelength of maximum signal change were plotted as a function of  $\text{K}^+$  and fit with OriginPro 8.5G (OriginLab Corporation software) according to the two state Hill equation:

$$\epsilon = \epsilon_F + (\epsilon_U - \epsilon_F) / ([\text{K}^+] / [\text{K}^+_{1/2}])$$

with  $\epsilon_U$  being the normalized ellipticity corresponding to unfolded and  $\epsilon_F$  to fully folded GQ;  $[\text{K}^+]$  is potassium concentration and  $[\text{K}^+_{1/2}]$  potassium concentration needed to obtain half the RNA in a folded state.

### **UV spectroscopy**

For UV-melting experiments, AIR-3A RNA was dissolved in 10 mM TRIS-HCl (pH 7.5) containing 5 mM KCl. RNA concentrations ranged from 2.5  $\mu$ M to 10  $\mu$ M. UV-melting studies were conducted on a Varian Cary Bio 300 UV-Visible Spectrophotometer with a temperature controller. Each sample (1,000  $\mu$ L) was filled into a quartz cuvette (1 cm path length), transferred to the spectrophotometer, heated to 80 °C and cooled down to 20 °C for four times with a heating or cooling rate of 0.5 °C min<sup>-1</sup>. Absorbance was recorded at 295 nm every 30 seconds. Each melting curve was analyzed using the method of van't Hoff to determine the  $T_m$  value.<sup>56, 57</sup>

### **Electrophoretic mobility shift assay (EMSA)**

For investigation of potassium dependent oligomerization, native gel electrophoresis was conducted.<sup>39</sup> 15% polyacrylamide gels were prepared with 1 x TBE (100 mM Tris, 80 mM Boric acid, 1 mM EDTA, pH 8.4) with 10 mM KCl added to gels and running buffer. 5'-end labeled RNA (1 pmol) was added to 4  $\mu$ M unlabeled RNA in a final volume of 25  $\mu$ L. Samples were heated to 95 °C for 3 min, chilled and allowed to cool to room temperature. Incubation at room temperature for 20 h in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 100 mM of either KCl or LiCl. For PAGE analysis, one volume of loading buffer (Tris-EDTA buffer, 30% glycerol) was added to each sample. After electrophoresis for 6 h at 4 °C and 6.5 V/cm, gels were frozen and exposed to a phosphor imaging screen (BioRad, 170-7841) until analyzed using Molecular Imager FX (BioRad) with Quantity One 4.6.6 and Image Lab 5.1 software (BioRad).

### **Dynamic light scattering (DLS)**

The size and size distribution of particles in aqueous solution were measured using a SpectroSize 300 instrument (XtalConcepts) at a wavelength of 660 nm and a scattering angle of 90°, a measurement temperature of 20 °C and a medium viscosity of 1.006 mPa·s. Before DLS measurement, proper folding was ensured by heating and successive cooling in 5 - 10 °C steps from 95 °C to 20 °C (40  $\mu$ M AIR-3A in 1 x PBS containing no, 0.5% or 0.1% TWEEN). Samples were loaded in a quartz cuvette and eight measurements were performed, for which the mean results were recorded.

### **ILP and RNase T1 protection assays**

For in-line probing<sup>33</sup> (ILP) 5'-end labeled RNA (< 1 nM) in 5  $\mu$ L of nuclease free water was heated to 95 °C for 1 min and allowed to refold by cooling to room temperature for 5 min. Then, one volume of 2 x ILP reaction buffer (50 mM Tris-HCl, pH 8.3, 10 mM MgCl<sub>2</sub> and either 200 mM LiCl, 200 mM KCl or 200 mM KCl and additional 200 nM hIL-6R) was added. Incubation was carried out at room temperature for 36 h.

For RNase T1 protection assays,<sup>55</sup> samples were prepared with 5'-end labeled RNA (1 pmol) in 10 mM Tris-HCl (pH 7.5) and 0.5 mM EDTA, heated to 95 °C for 1 min and cooled to room temperature for 10 min. To a final volume of 10  $\mu$ L, KCl, LiCl, BaCl<sub>2</sub>, NH<sub>4</sub>Cl or hIL-6R was added at desired concentrations or 10 mM 18-crown-6 (to complex possibly contaminating potassium) as a no K<sup>+</sup> control. Cleavage under native folding conditions was initiated by addition of RNase T1 (Thermo Scientific, AM2280) to a final concentration of 0.005 U/ $\mu$ L and further incubation for 1 h.

To assign occurring bands to the nt position, several ladders were generated from ~1 pmol radio labeled RNA each. An alkaline hydrolysis ladder allowing nt resolution was obtained by incubating labeled RNA in 5  $\mu$ L hydrolysis buffer (100 mM sodium carbonate, 2 mM EDTA, pH 9.2) at 95 °C for 2 min and quenching by addition of 5  $\mu$ L Tris buffer (100 mM, pH 7.5). For a denaturing RNase T1 ladder identifying all G positions, RNA was incubated in denaturing reaction buffer (7 M Urea, 20 mM sodium citrate, pH 5) at 60 °C for 2 min. 0.5 U RNase T1/ $\mu$ L (dilution freshly prepared) was added and the mixture was incubated at 50 °C for 3 min. An RNase A ladder (marking C and U positions) was produced from RNA in 10  $\mu$ L Tris buffer (100 mM, pH 7.5) by addition of 10 pg RNase A/ $\mu$ L (dilution freshly prepared, Thermo Scientific, AM2274). After 2 min incubation at room temperature, reaction was stopped by addition of 2 x loading buffer.

Following the respective treatment, samples were mixed with an equal volume of 2 x loading buffer (95% formamide, 0.025% SDS, 0.5 mM EDTA and bromophenol blue and xylene cyanol dyes for tracking) and applied on a 12.5 or 15% denaturing sequencing gel (8 M urea). Ladders were heated to 95 °C for 2 min prior to loading. Gels were run at 50 W for 3 to 6 h. Frozen and exposed to a phosphor imaging screen and scanned as described above.

For analysis, bands were quantified with rectangular boxes around and signals were normalized to total counts per lane. For ILP data, signal intensities for each nt position in the K<sup>+</sup> or hIL-6R lane were divided by the intensity of the respective position in the Li<sup>+</sup> lane and plotted against nt position considering the twofold threshold as significant ILP reactivity. RNase T1 protection was evaluated by relating G position intensities to full length intensity (G19 = 1).

### **MST measurements**

MST experiments for stoichiometric investigation of AIR-3A and hIL-6R were performed on a Monolith NT.115 system (NanoTemper Technologies) using 10% LED- and 40% IR-laser power. Laser on and off times were set at 30 s and 5 s, respectively. A twofold dilution series for unlabeled recombinant target protein hIL-6R (Conaris Research Institute AG) was prepared and equal volumes of AlexaFluor® 647 labeled AIR-3A (400 nM) were added resulting in protein concentrations ranging from 10 to 250 nM with a constant AIR-3A concentration of 200 nM in 1 x PBS supplemented with



0.05% Tween and 1 mg BSA/mL. Samples were incubated at room temperature for 15 min followed by centrifugation for 15 min to exclude aggregates. Samples were filled in standard treated capillaries (NanoTemper Technologies) for measurement. Saturation concentration was indicated by the kink around 100 nM. Experiments were performed in triplicate.

### Binding assays

Filter retention assay (FRA) was used to investigate binding of RNAs to recombinant protein. RNAs obtained from *in vitro* T7 transcription were dephosphorylated with FastAP (Thermo Scientific, EF0654) following the standard protocol. After inactivation of FastAP, radioactive 5'-end labeling was performed by phosphorylation using T4 polynucleotide kinase (Thermo Scientific, EK0031) and [ $\gamma$ -<sup>32</sup>P]-ATP (3,000 Ci/mmol, Hartmann Analytic, SCP 301). After PAGE purification, constant amounts of labeled RNA (< 1 nM) were incubated with dilution series (0 – 500 nM or up to 2  $\mu$ M) of protein in SELEX buffer (1 x PBS, 3 mM MgCl<sub>2</sub>). Complex formation was allowed to take place by incubation for 30 min at room temperature. Subsequently, samples were filtrated through a pre-equilibrated (for RNA: 20% Methanol, 40 mM 6-aminohexanoic acid, for DNA: KOH) nitrocellulose membrane (0.45  $\mu$ M, Carl Roth) in a vacuum manifold (Minifold® I Dot-Blot-System, Schleicher & Schuell) and washed twice with SELEX buffer. The membrane was dried and exposed to a phosphor imaging screen (BioRad, 170-7841), read out with Molecular Imager FX and quantified using Image Lab 5.1 software (BioRad). From the amount of labeled RNA on the filter related to total amount applied, dissociation constant  $K_d$  and maximum binding ( $B_{max}$ ) values were calculated using OriginPro 8.5G software (OriginLab Corporation) according to the equation:  $RNA_{bound} = (B_{max} \times C_{protein}) / (K_d + C_{protein})$ .

Affinity to cellular presented hIL-6R was determined by flow cytometry using a FACSCalibur flow cytometer with CellQuestPro software (BD Biosciences). Briefly, cells were harvested, washed and incubated (10<sup>5</sup> cells per sample) with different concentrations of AlexaFluor® 647 labeled AIR-3A (ranging from 0 to 62.5 nM) on ice for 15 minutes. Then cells were washed twice with 1 x PBS and used for analysis. Per sample 10,000 cells were counted and viable populations were gated. Mean fluorescence intensities were compared and normalized for maximum mean fluorescence intensity.

To assess cellular binding of AIR-3 derivatives to BaF3/gp130/IL6R/TNF target cells, competition with fluorescently labeled AIR-3A was determined as described before.<sup>29</sup> Cells were incubated with serial concentrations of AIR-3 derivative (Table 1) or AIR-3A as a control for 15 min on ice. Then cells were washed with 1 x PBS and incubated for 15 min on ice with 50 nM fluorescently labeled AIR-3A. Finally, cells were washed again, suspended in PBS and analyzed by flow cytometry.  $K_d$  values and maximum mean fluorescence intensity ( $B_{max}$ ) were determined from fitting mean fluorescence intensity ( $F$ ) to corresponding concentration of RNA as described by:

$$F = (B_{\max} \times C_{\text{RNA}}) / (K_d + C_{\text{RNA}}).$$

### Mass spectrometry analysis of tryptic peptides from cross linked hIL-6R:AIR-3A

In order to link AIR-3A covalently to hIL-6R a small fraction of the RNA was used with 5'-end radiolabel. 3 pmol RNA were incubated with 15 pmol hIL-6R in 15  $\mu$ L PBS for 20 min at room temperature to assure complex formation. Subsequently this mixture was pipetted on a coverslip placed on ice and irradiated for 12 min by a UV-lamp (254 nm, 4 W) placed above the coverslip at 3 cm distance. The sample was resolved on a 8% non-reducing SDS-PAGE to separate formed AIR-3:hIL-6R from free RNA and free protein. The band of the covalent adduct was detected by autoradiography, cut and subjected to in gel digestion by Trypsin following a standard protocol. As a control a sample lacking the RNA was treated in the same way. The yielded peptide samples were analyzed by MALDI-TOF with a hydro benzoic acid matrix. The peptide identification was carried out using Mascot software and SwissProt databank.

### Molecular docking of DNA quadruplex AID-1 to hIL-6R

A dimer structure of G-quadruplex (PDB: 2LE6<sup>58</sup>) was docked on the structure of the hIL-6R protein (PDB: 1N26<sup>59</sup>) with the ATTRACT engine.<sup>60</sup> Both partners, protein and quadruplex, were first considered as rigid and converted into coarse-grain representation.<sup>61</sup> The quadruplex was placed at 10,000 random starting positions and orientations toward the protein, then the intermolecular energy was minimized in ATTRACT knowledge-based force-field.<sup>61</sup> The experimentally assessed proximity of the quadruplex and the protein residue Y78 in the complex was used as a high-ambiguity maximal distance restraint during docking. The 1,000 best-scored poses were then clustered at 3 Å RMSD, and the best-scored structure in each of the the 300 most populated clusters was kept for further refinement. This refinement was performed as a series of minimizations and molecular dynamics simulations with the AMBER simulation package in the AMBER-ff10 force-field. The 300 docking models of the complex were converted back into all-atoms representation, and hydrogens were added with the leap module of AMBER. A 500-step minimization *in vacuo* was followed by a 1,000 steps minimization in implicit water (GB model<sup>62</sup>). Then two short molecular dynamics simulations were performed, first at high temperature (400 K, 20 ps) then at mid temperature (350 K, 10 ps), with weak positional restraints on the C-alpha and phosphate atoms. The structures were finally minimized in 5,000 steps and ranked by the DNA-protein binding energy assessed by computing the free energy of the complex and of each partner and subtracting the second ones from the first one. The best-scored model was visually inspected and retained as the most probable structure of the complex.

## Acknowledgments

K.S. gratefully acknowledges support by the International Max Planck Research School for Ultrafast Imaging & Structural Dynamics (IMPRS-UFAST).

## References

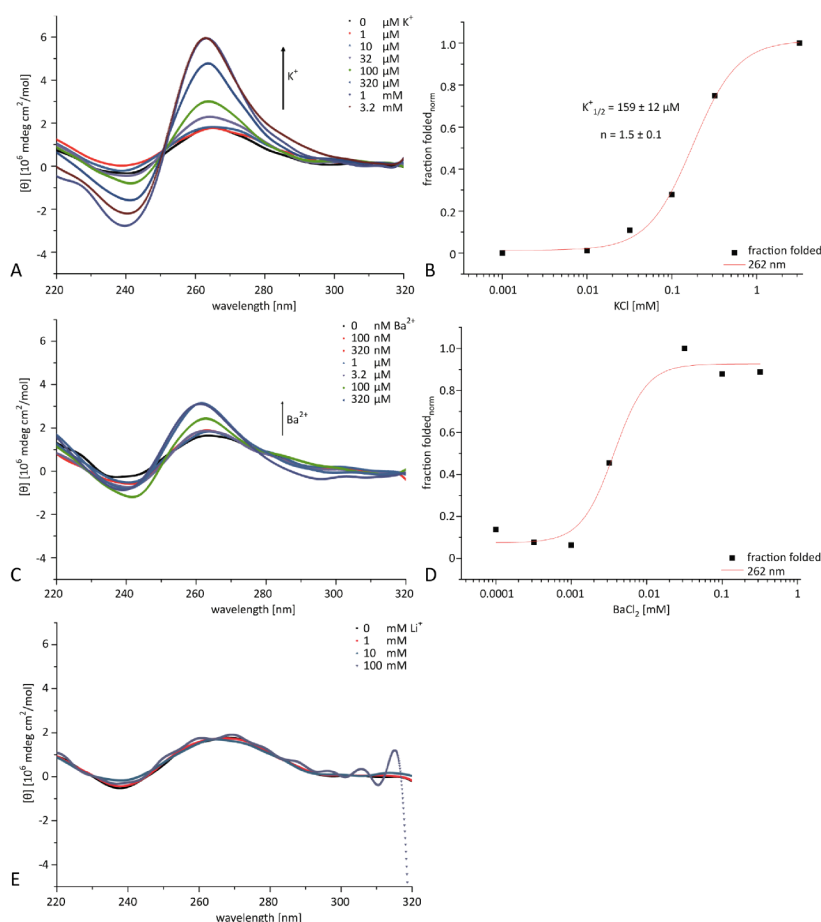
1. Makarov VL, Hirose Y, Langmore JP. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* 1997; 88:657-66.
2. Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, et al. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nature structural & molecular biology* 2011; 18:478-85.
3. Biffi G, Tannahill D, Balasubramanian S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *Journal of the American Chemical Society* 2012; 134:11974-6.
4. Redon S, Reichenbach P, Lingner J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic acids research* 2010; 38:5797-806.
5. Hirashima K, Seimiya H. Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic acids research* 2015; 43:2022-32.
6. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America* 2002; 99:11593-8.
7. Onyshchenko MI, Gaynutdinov TI, Englund EA, Appella DH, Neumann RD, Panyutin IG. Stabilization of G-quadruplex in the BCL2 promoter region in double-stranded DNA by invading short PNAs. *Nucleic acids research* 2009; 37:7570-80.
8. Fiset JF, Montagna DR, Mihailescu MR, Wolfe MS. A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *Journal of neurochemistry* 2012; 121:763-73.
9. Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic acids research* 2012; 40:4727-41.
10. von Hacht A, Seifert O, Menger M, Schutze T, Arora A, Konthur Z, et al. Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic acids research* 2014; 42:6630-44.
11. Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 2014; 507:195-200.
12. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990; 249:505-10.
13. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature* 1990; 346:818-22.
14. Cho EJ, Lee JW, Ellington AD. Applications of aptamers as sensors. *Annu Rev Anal Chem (Palo Alto Calif)* 2009; 2:241-64.
15. Levy-Nissenbaum E, Radovic-Moreno AF, Wang AZ, Langer R, Farokhzad OC. Nanotechnology and aptamers: applications in drug delivery. *Trends in biotechnology* 2008; 26:442-9.
16. Keefe AD, Pai S, Ellington A. Aptamers as therapeutics. *Nature Reviews Drug Discovery* 2010; 9:537-50.
17. Brazda V, Haronikova L, Liao JC, Fojta M. DNA and RNA quadruplex-binding proteins. *Int J Mol Sci* 2014; 15:17493-517.

18. Bole M, Menon L, Mihailescu MR. Fragile X mental retardation protein recognition of G quadruplex structure per se is sufficient for high affinity binding to RNA. *Molecular bioSystems* 2008; 4:1212-9.
19. Hanakahi LA, Sun H, Maizels N. High affinity interactions of nucleolin with G-G-paired rDNA. *The Journal of biological chemistry* 1999; 274:15908-12.
20. Takahama K, Miyawaki A, Shitara T, Mitsuya K, Morikawa M, Hagihara M, et al. G-Quadruplex DNA- and RNA-Specific-Binding Proteins Engineered from the RGG Domain of TLS/FUS. *ACS chemical biology* 2015; 10:2564-9.
21. Bates PJ, Kahlon JB, Thomas SD, Trent JO, Miller DM. Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. *The Journal of biological chemistry* 1999; 274:26369-77.
22. Macaya RF, Schultze P, Smith FW, Roe JA, Feigon J. Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proceedings of the National Academy of Sciences of the United States of America* 1993; 90:3745-9.
23. Mukundan VT, Do NQ, Phan AT. HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges. *Nucleic acids research* 2011; 39:8984-91.
24. Olsthoorn RC. G-quadruplexes within prion mRNA: the missing link in prion disease? *Nucleic acids research* 2014; 42:9327-33.
25. Proske D, Gilch S, Wopfner F, Schatzl HM, Winnacker EL, Famulok M. Prion-protein-specific aptamer reduces PrPSc formation. *Chembiochem : a European journal of chemical biology* 2002; 3:717-25.
26. Mashima T, Nishikawa F, Kamatari YO, Fujiwara H, Saimura M, Nagata T, et al. Anti-prion activity of an RNA aptamer and its structural basis. *Nucleic acids research* 2013; 41:1355-62.
27. Huang H, Suslov NB, Li NS, Shelke SA, Evans ME, Koldobskaya Y, et al. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature chemical biology* 2014; 10:686-91.
28. Kruspe S, Meyer C, Hahn U. Chlorin e6 Conjugated Interleukin-6 Receptor Aptamers Selectively Kill Target Cells Upon Irradiation. *Molecular therapy Nucleic acids* 2014; 3:e143.
29. Kruspe S, Hahn U. An Aptamer Intrinsically Comprising 5-Fluoro-2'-deoxyuridine for Targeted Chemotherapy. *Angew Chem Int Ed Engl* 2014; 53:10541-4.
30. Meyer C, Eydelor K, Magbanua E, Zivkovic T, Piganeau N, Lorenzen I, et al. Interleukin-6 receptor specific RNA aptamers for cargo delivery into target cells. *RNA biology* 2012; 9:67-80.
31. Meyer C, Berg K, Eydelor-Haeder K, Lorenzen I, Grotzinger J, Rose-John S, et al. Stabilized Interleukin-6 receptor binding RNA aptamers. *RNA biology* 2014; 11:57-65.
32. Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. Probing the structure of RNAs in solution. *Nucleic acids research* 1987; 15:9109-28.
33. Kwok CK, Ding Y, Shahid S, Assmann SM, Bevilacqua PC. A stable RNA G-quadruplex within the 5'-UTR of Arabidopsis thaliana ATR mRNA inhibits translation. *The Biochemical journal* 2015; 467:91-102.
34. Mittelberger F, Meyer C, Waetzig GH, Zacharias M, Valentini E, Svergun DI, et al. RAID3 - An interleukin-6 receptor-binding aptamer with post-selective modification-resistant affinity. *RNA biology* 2015; 12:1043-53.
35. Bley CJ, Qi X, Rand DP, Borges CR, Nelson RW, Chen JJ. RNA-protein binding interface in the telomerase ribonucleoprotein. *Proceedings of the National Academy of Sciences of the United States of America* 2011; 108:20333-8.
36. Christian H, Hofele RV, Urlaub H, Ficner R. Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. *Nucleic acids research* 2014; 42:1162-79.
37. Meisenheimer KM, Koch TH. Photocross-linking of nucleic acids to associated proteins. *Crit Rev Biochem Mol* 1997; 32:101-40.

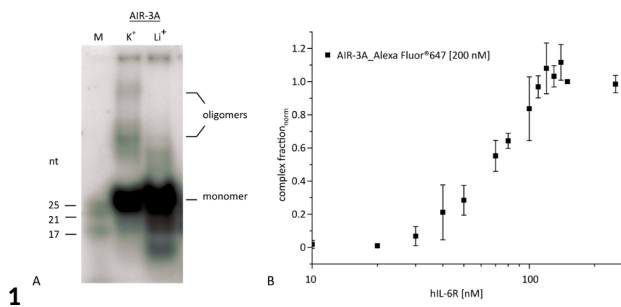
38. Keskin S, Besztejan S, Kassier G, Manz S, Bucker R, Riekeberg S, et al. Visualization of Multimerization and Self-Assembly of DNA-Functionalized Gold Nanoparticles Using In-Liquid Transmission Electron Microscopy. *The journal of physical chemistry letters* 2015; 6:4487-92.
39. Piekna-Przybylska D, Sullivan MA, Sharma G, Bambara RA. U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence. *Biochemistry* 2014; 53:2581-93.
40. Simonsson T. G-quadruplex DNA structures--variations on a theme. *Biol Chem* 2001; 382:621-8.
41. Magbanua E, Zivkovic T, Hansen B, Beschorner N, Meyer C, Lorenzen I, et al. d(GGGT)(4) and r(GGGU)(4) are both HIV-1 inhibitors and interleukin-6 receptor aptamers. *RNA biology* 2013; 10:216-27.
42. Li T, Shi L, Wang E, Dong S. Multifunctional G-quadruplex aptamers and their application to protein detection. *Chemistry* 2009; 15:1036-42.
43. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic acids research* 2015; 43:8627-37.
44. Lavery R, Zakrzewska K, Sklenar H. Jumna (Junction Minimization of Nucleic-Acids). *Comput Phys Commun* 1995; 91:135-58.
45. Zacharias M. Conformational analysis of DNA-trinucleotide-hairpin-loop structures using a continuum solvent model. *Biophys J* 2001; 80:2350-63.
46. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *Journal of the American Chemical Society* 1996; 118:2309-.
47. Homann M, Lorger M, Engstler M, Zacharias M, Goring HU. Serum-stable RNA aptamers to an invariant surface domain of live African trypanosomes. *Combinatorial Chemistry & High Throughput Screening* 2006; 9:491-9.
48. Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM, et al. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *Journal of applied crystallography* 2015; 48:431-43.
49. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of applied crystallography* 2003; 36:1277-82.
50. Guinier A. *La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques*. Paris.; Univ. de Paris., 1939:3 p.l., 80, 2 p., 1 l.
51. Svergun DI. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of applied crystallography* 1992; 25:495-503.
52. Porod G. General theory. In: Glatter O, Kratky O, eds. *Small-Angle X-Ray Scattering*. New York: Academic Press, 1982:17-51.
53. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 1999; 76:2879-86.
54. Petoukhov MV, Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 2005; 89:1237-50.
55. Mullen MA, Assmann SM, Bevilacqua PC. Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *Journal of the American Chemical Society* 2012; 134:812-5.
56. Bugaut A, Balasubramanian S. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* 2008; 47:689-97.
57. Marky LA, Breslauer KJ. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 1987; 26:1601-20.

58. Do NQ, Lim KW, Teo MH, Heddi B, Phan AT. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic acids research* 2011; 39:9448-57.
59. Varghese JN, Moritz RL, Lou MZ, Van Donkelaar A, Ji H, Ivancic N, et al. Structure of the extracellular domains of the human interleukin-6 receptor alpha -chain. *Proceedings of the National Academy of Sciences of the United States of America* 2002; 99:15959-64.
60. de Vries SJ, Schindler CE, Chauvot de Beauchene I, Zacharias M. A web interface for easy flexible protein-protein docking with ATTRACT. *Biophys J* 2015; 108:462-5.
61. Setny P, Zacharias M. A coarse-grained force field for Protein-RNA docking. *Nucleic acids research* 2011; 39:9118-29.
62. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins-Structure Function and Bioinformatics* 2004; 55:383-94.
63. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI. ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *Journal of applied crystallography* 2007; 40:S223-S8.

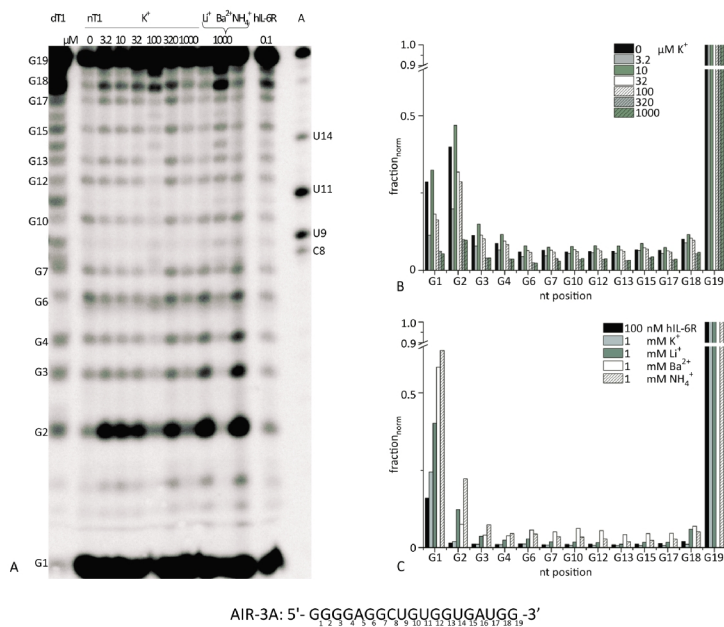
## Figure legends



**Figure 1: CD spectroscopic analysis of AIR-3A GQ formation in dependence of different cations.** CD spectra (A, C, E) and fraction folded plot (B, D) of 5  $\mu\text{M}$  AIR-3A in a background of 10 mM Tris-HCl (pH 7.5) in the range of 220 to 320 nm. A)  $\text{K}^+$  induced GQ formation on AIR-3A indicates parallel topology by positive peak at  $\sim 260 \text{ nm}$  and negative peak at  $\sim 240 \text{ nm}$ . B) Molar ellipticity at  $\lambda_{\text{max}} = 262 \text{ nm}$  from (A) as a function of  $\text{K}^+$  concentration fit to a two-state Hill equation yielded  $K_{1/2}$  value of  $159 \mu\text{M}$  at  $n = 1.5$ . C)  $\text{Ba}^{2+}$  concentration dependent change of amplitude at  $\sim 260 \text{ nm}$  represents D) Molar ellipticity at  $\lambda_{\text{max}} = 263 \text{ nm}$  from (C) as a function of  $\text{Ba}^{2+}$  concentration fit to a two-state Hill equation yielded  $K_{1/2}$  value of  $3.35 \pm 0.35 \mu\text{M}$  at  $n = 2.4 \pm 0.48$ , which represent hypothetic values as only one data point close to the inflection point was determinable. (E)  $\text{Li}^+$  inhibits GQ formation of AIR-3A as indicated by preservation of background signals independent from  $\text{Li}^+$  concentration.

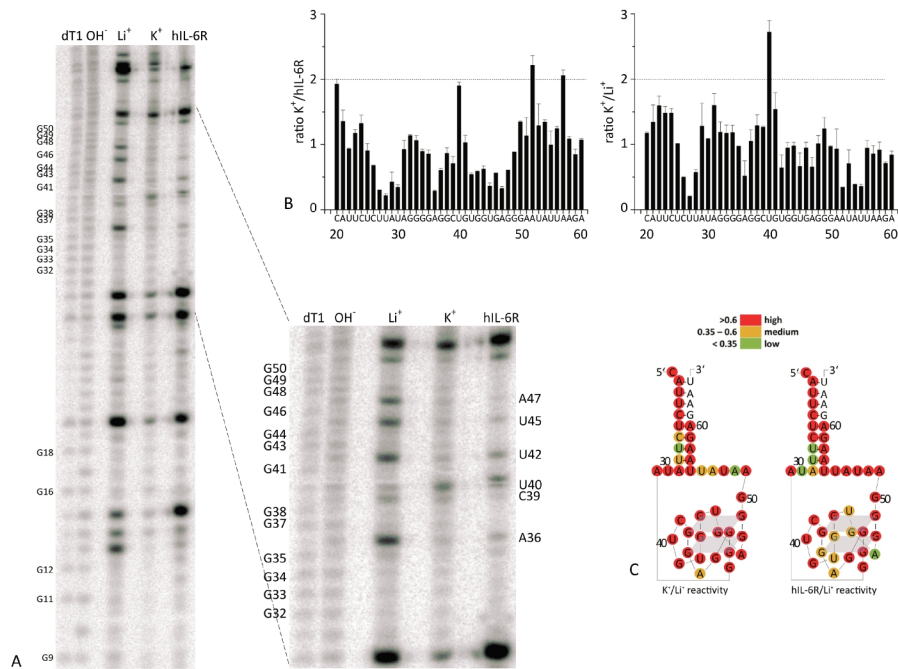


**11Figure 2: Aggregation and binding stoichiometry of AIR-3A.** A) Native gel mobility analysis (15% polyacrylamide, 1 x TBE, 10 mM  $K^+$ ) of AIR-3A (4  $\mu$ M, 5'-end labeled with  $^{32}$ P) supports formation of GQ oligomers in presence of  $K^+$  compared to  $Li^+$  (100 mM). B) Stoichiometric investigation of fluorescently labeled AIR-3A by MST indicating a ratio of 2:1 of RNA:protein. Plotted data represent mean values from three experiments.

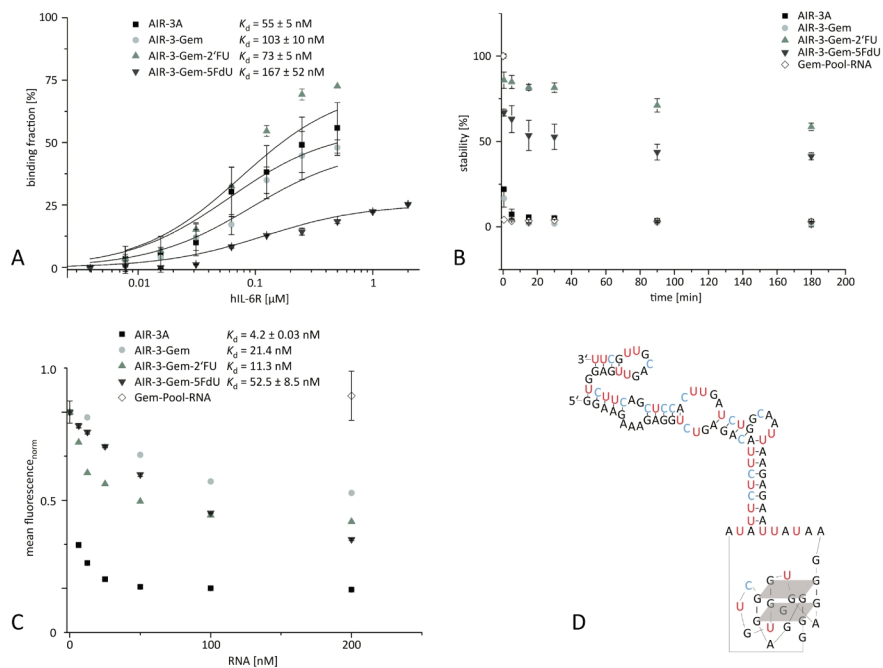


**Figure 13: RNase T1 protection assay reveals strong protection of all Gs in AIR-3A.** A) G-specific T1 cleavage of AIR-3A (5'-end labeled with  $^{32}$ P) under native folding conditions (nT1; 10 mM Tris-HCl, pH 7.5, 0.5 mM EDTA) in presence of different  $K^+$  concentrations, cations and target protein hIL-6R (concentrations as indicated). A denaturing G-specific RNase T1 ladder (dT1) and C- and U-specific RNase A ladder (A) were applied for band assignment. B) T1-Protection of different G positions visualized as fractions cleaved in dependence of  $K^+$  revealed overall protection in presence of  $K^+$  but no significant differences among positions G3 to G18. C) T1-protection of G positions in presence of different effectors. Protection increased in the order  $NH_4^+ < Ba^{2+} < Li^+ < K^+ < hIL-6R$  whereas no difference among the single positions was observed. For G1 and G2 higher cleavage rates were observed under all conditions. Data were normalized for total counts per lane and in relation to full length present (G19 = 1).



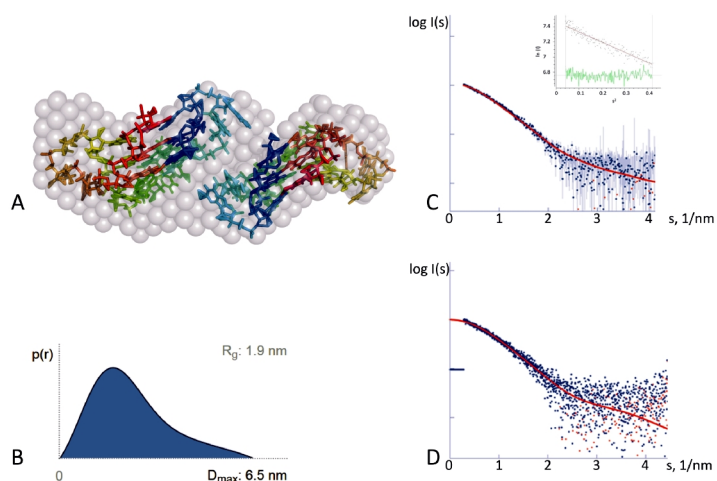


**Figure 14: In-line probing (ILP) supports loop formation in GQ region of AIR-3 and conformational change in presence of target hIL-6R.** **A)** ILP on relevant regions of AIR-3 (5'-end labeled with  $^{32}\text{P}$ , GQ motif formed by nt 32 to 50). Reference ladders dT1 (G-specific) and  $\text{OH}^-$  (alkaline hydrolysis, all positions) were used for band assignment. GQ identity is given by different cleavage patterns in presence of  $\text{Li}^+$  and  $\text{K}^+$  respectively, with further differences by addition of hIL-6R. ILP conditions were 25 mM Tris-HCl, pH 8.3, 5 mM  $\text{MgCl}_2$  at room temperature for 36 h with 100 mM  $\text{Li}^+$  or  $\text{K}^+$  (without or with addition of 200 nM hIL-6R) present. The GQ region is highlighted as expanded region with Gs assigned on the left and nucleotides in between on the right. **B)** Analysis of ILP results representing ratios for each nt position in lane hIL-6R or  $\text{K}^+$  to lane  $\text{Li}^+$ . The typical twofold threshold indicating significant ILP reactivity is shown as dotted line. **C)** Representation of ILP ratios from **(B)** color coded for  $\text{K}^+/\text{Li}^+$  (left) and hIL-6R/ $\text{Li}^+$  (right) for GQ region flanking areas. Illustration shows hypothesized GQ within the secondary structure of AIR-3 according to Mfold prediction, forcing nt 32 - 50 to be single-stranded. Data were normalized for total counts per lane.

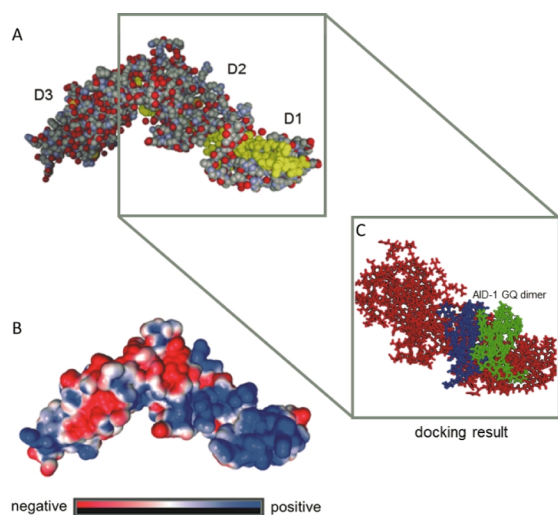


**Figure 15: AIR-3 variants containing nucleoside analogs exhibit target affinity and increased serum stability.**

**A)** Binding was assayed by filter retention of RNA bound hIL-6R. Constant amounts of  $^{32}\text{P}$  5'-end labeled RNA (< 1 nM) were incubated with increasing amounts of protein. Data points represent mean values of bound RNA,  $K_d$  values are given in the figure (experiments performed in duplicates and triplicates). **B)** Stability of  $^{32}\text{P}$  5'-end labeled AIR-3 variants in DMEM supplemented with 10% FBS at 37 °C. At indicated time points aliquots were taken and immediately frozen in liquid  $\text{N}_2$ . PAGE separation and analysis by autoradiography revealed remnant amounts of full length RNA. Intensities were normalized to total counts per lane. **C)** Cellular binding was analyzed by flow cytometry on BaF3\_hIL-6R target cells. Cells were incubated with AIR-3 variants on ice for binding, washed and subsequently stained with Alexa Fluor® 647 labeled AIR-3A. Concentration dependent decrease of fluorescence was a measure for binding capacity. **D)** Schematic of AIR-3 secondary structure as predicted by Mfold with hypothesized GQ depicting sites of exchange with gemcitabine (blue C) and 5-FdU or 2'-FU (red U).



**Figure 6: SAXS data for AIR-3A suggest dimer formation in solution.** **A)** Superposition of the *ab initio* model (gray semi-transparent spheres) and the hybrid model containing a dimer of predicted molecules of AIR-3A (rainbow-colored sticks). **B)** Distance distribution computed from the experimental data. **C)** and **D)** Fits of the experimental data (dots) by the scattering computed from the *ab initio* and rigid body models, respectively (solid lines). Insert in Figure 8C depicts the Guinier plot automatically computed by the program AutoRG<sup>63</sup> (black dots, experimental data, red line, linear fit, green line, residuals). Both models agree well with the experimental data yielding the goodness-of-fit of  $\chi^2 = 1.12$  for the *ab initio* model and  $\chi^2 = 1.04$  for the rigid body model. The intensities  $I(s)$  are displayed as a function of momentum transfer  $s = 4\pi \sin(\theta)/\lambda$ , where  $\lambda = 0.124$  nm is the X-ray wavelength and  $2\theta$  is the scattering angle).



**Figure 17: Proposed binding site of AIR-3A on hIL-6R (PDB: 1N26).** **A)** Sequential part derived from peptide MS analysis (yellow) that is involved in contact to AIR-3A. **B)** Electrostatic surface potential of hIL-6R. **C)** Binding site model for DNA-GQ AID-1 according to the performed docking studies.

## 11Tables

**Table 1: Tryptic peptides found for hIL-6R and hIL-6R:AIR-3A (listed are only peptides of D1 and D2 of hIL-6R)**

Position (start-end)	Mass (expected)*	Peptide Sequence	reference hIL-6R	hIL-6R:AIR-3A
SP** + 1-4	2208.186	MLAVGCALLAALLAAPGAALAPR.R + Oxidat. (M)	2208.096- 2208.194	2208.142- 2208.190
5-13	1086.102	R.RCPAQEVAR.G	1068.075- 1086.109	1086.189- 1086.320
6-13	930.059	R.CPAQEVAR.G	930.067-930.112	930.054- 930.098
14-44	3318.656	R.GVLTSLPGDSVTLTCPGVEPEDNATV HWVLR.K	not found***	not found***
45-54	1007.538	R.KPAAGSHPSR.W	1007.052- 1007.171	1007.045- 1007.170
55-60	677.054	R.WAGMGR.R	677.061-677.105	677.122- 677.186
55-60	693.047	R.WAGMGR.R + Oxidat. (M)	693.054-693.085	693.088- 693.159
55-61	833.160	R.WAGMGRR.L	833.146-833.168	833.208- 833.302
62-65	514.187	R.LLLR.S	514.144-514.194	514.242- 514.445
66-79	1628,712	R.SVQLHDSGNYSYR.A	1628.218 1628.258	- <b>not found</b>
83-104	2461.331	R.PAGTVHLLVDVPPEEPQLSCFR.K	2461.042 2461.190	- <b>not found</b>
105-118	1628.207	R.KSPLSNVCEWGPR.S	1628.214- 1628.436	1628.288- 1628.492
106-118	1500,161	K.SPLSNVCEWGPR.S	1500.023- 1500.169	1500.152- 1500.399
119-126	834.439	R.STPSLTK.A	834.419-834.434	834.027- 834.400
127-132	670.215	K.AVLLVR.K	670,223-670.307	670.243- 670.306
133-154	2674.098	R.KFQNSPAEDFQEPQYSQESQK.F	2674.079- 2674.105	2674.109- 2674.534
134-154	2546.181	K.FQNSPAEDFQEPQYSQESQK.F	2546.122- 2546.189	2546.175- 2546.314
155-182	3012.686	K.FSCQLAVPEGDSSFIVSMCVASSVGS K.F	3012.472- 3012,693	3012,235- 3012.347
183-210	3017.545	K.FSKTQTFQGCILQDPPANITVAVA R.N	3017.583- 3018.107	3017,943- 3018.042

\* Cysteins alkylated (carboxyamidomethyl) + 57.033 g/mol, presented is the peptide's monoisotopic major product

\*\* SP = signal peptide, 19AA N-terminal

\*\*\* underrepresented due to peptide size and composition

**Table 1: AIR-3A**

Oligonucleotide	Sequence (5' - 3' direction)
AIR-3A	GGGGAGGCUGUGGUGAGGG

**Table 1: AIR-3 sequence and derivatives synthesized by T7 transcription**

RNA	Sequence (5' - 3' direction)
AIR-3	GGAAGAAAGAGGUCUGAGACAUUCUCUUAUAGGGGAGGCUGUGGUGAGGGAAU AUUAAGAGAAUUAACGGUCUAGUUCACCU CGACUUCUGGAGUUGACGUUGCUU
AIR-3-Gem	AIR-3 derivative containing gemcitabine instead of cytidine nucleotides
AIR-3-Gem-2'-FU	AIR-3 derivative containing gemcitabine for cytidine nucleotides and 2'-FU for uridine nucleotides
AIR-3-Gem-5FdU	AIR-3 derivative containing gemcitabine for cytidine nucleotides and 5-FdU for uridine nucleotides
Gem-Pool-RNA	GGAAGAAAGAGGUCUGAGACAUUCU-N60-CUUCUGGAGUUGACGUUGCUU Unselected library (N60 = randomized region) containing gemcitabine for cytidine nucleotides