Hexanucleotide Arrays

DOI: 10.1002/anie.201003462

Specific Recognition of Proteins by Array-Bound Hexanucleotides**

*Alessandra Mescalchin, Winfried Wünsche, and Georg Sczakiel**

Hexanucleotides may function as aptamer-like very short oligonucleotides. Their specific binding to proteins suggests exploiting them as biological tools and potential pharmaceutical lead compounds. Systematic screening of hexanucleotide–protein recognition at high-throughput is possible by a novel array-based platform. This technology considers the complete hexanucleotide sequence space and reveals high target specificity indicated by species-selectively recognizing the highly conserved polymerases of HIV-1 and HIV-2.

Classically, specific high-affinity binding of folded oligomeric single-stranded nucleic acids to target molecules can be studied by SELEX-like protocols and results in the selection of aptamers.[1,2] Surprisingly, very short oligonucleotides have also been shown to bind tightly and specifically to target proteins, thereby affecting their biological activity.[3–5] In functional terms this includes hexanucleotides for which almost aptamer-like characteristics have been indicated recently.[3] In terms of pharmaceutical drug development one might consider hexanucleotides as novel nucleic acid based lead compounds. This is supported by the identification of species that have been shown to suppress viral functions including HIV-1[3] and HAV.[10] Hexanucleotides are attractive because their chemical synthesis and modification are far advanced and the modeling of short nucleic acids is established.

Once hexanucleotide-based lead compounds have been identified, subsequent drug development for the increase of affinity and efficacy can conceivably be based on rational approaches such as docking studies[6] combined with site-specific synthetic oligonucleotide chemistry, or for example, saturation transfer difference nuclear magnetic resonance (STD NMR) spectroscopy.[7]

Hexanucleotides do not adopt stable intramolecular secondary structures and, in the case of the target HIV-1 reverse transcriptase (HIV-1 RT),[3] they apparently enter the process of target recognition as a quasi-one-dimensional chain which, subsequent to binding, adopts a structure that sterically and chemically fits with the local target structure. Alternatively, higher-order structures formed by hexanucleotides may be biologically active complexes such as the G-quadruplex-forming G_5T in the case of the 3C proteinase of HAV.[10] The complete relevant sequence space to be considered for hexanucleotides comprises 4096 ($=4^6$) linear species. Selecting target-binding species from this library in solution in a one-round approach is conceivable and should be much less labor-intensive than amplification-based methodologies. However, combinatorial approaches in solution are hampered by fundamental disadvantages that cannot be excluded or even reasonably controlled, including the formation of higher binding-competent complexes among individual species, interference of hexanucleotides in the formation of specific hexamer–target complexes, unequal representation of the sequence space, and technical hurdles including cloning steps, which are elaborate for short oligonucleotides.

Since hexanucleotides are sufficient for specific binding to proteins at K_d values in solution in the micromolar range,[3] it is attractive to consider a combinatorial setup in which the complete hexanucleotide sequence space is immobilized on a solid support in an arraylike fashion. A hexanucleotide-based array would conceivably bypass disadvantages related to combinatorial approaches in solution. Further, an array-based approach would reduce the time, cost, and computational and quantitative assessment of the experimental results.

Here we set out to design and evaluate a hexanucleotide-based array to study oligonucleotide–protein interactions. Firstly, a control array was produced onto which a set of known oligonucleotides was printed; these had been investigated previously in the use of HIV-1 RT as a target protein in solution[3] (see the Supporting Information, Table S1). This set includes RT-binding and nonbinding hexanucleotides, longer oligonucleotides mimicking a polymerase template strand, and an RT-directed DNA aptamer.[8] The oligonucleotides were attached to the solid support by means of non-nucleotidic linkers of varying length (see the Supporting Information, Figure S1a) in order to avoid additional unspecific interactions with the target protein of interest. Initially, deoxyribonucleotides were chosen which showed patterns comparable with ribonucleotides. Only minor differences have been observed between the two types of strands, including chemically modified derivatives thereof. One exception is phosphorothioate derivatives which led to higher affinity but loss of specificity.[3] It should also be considered that downstream optimization steps of hexanucleotide-based leads will include a substantial variation of the nucleic acid unit including the sugar moiety.

[*] Dr. A. Mescalchin, W. Wünsche, Prof. Dr. G. Sczakiel
Institut für Molekulare Medizin
Zentrum für medizinische Struktur- und Zellbiologie (ZMSZ)
Universität zu Lübeck und Universitätsklinikum Schleswig-Holstein
Ratzeburger Allee 160, 23538 Lübeck (Germany)
Fax: (+49) 451-500-2729
E-mail: sczakiel@imm.uni-luebeck.de
Homepage: http://www.molmed.uni-luebeck.de
Dr. A. Mescalchin, Prof. Dr. G. Sczakiel
Kompetenzzentrum für Drug Design/Target Monitoring (KDDTM)
Maria-Goeppert-Strasse 1, 23562 Lübeck (Germany)

[**] We thank T. Restle for kindly providing purified polymerases of HIV-1 and HIV-2, and P. Zeltz (BadenBioTec GmbH) for helpful advice on the use of arrays. This work was supported by the Medical Faculty of the University of Lübeck (grant to A.M.; FKZ: A26-2007), the Possehl-Stiftung (Lübeck), and EC (grant LSHG-CT-2003-503480).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201003462.

In a first round of studies, control arrays were exposed to HIV-1 RT at concentrations between 5 nM and 5 μM, and the best signal strength and signal/noise ratio were observed at protein concentrations higher than 5 nM (see the Supporting Information, Figure S1b). Those concentrations are surprisingly low; this represents a substantial technical advantage when one considers, for example, the apparent K_d of 5.7 μM for Hex-S3 for binding with HIV-1 RT in solution.[3] Most importantly, the binding pattern of the hexanucleotides tested reflects their binding behavior with RT in solution in a qualitative manner.[3] On the technical level it is noteworthy that the best results were obtained with a linker length of 21 and 38 atoms, whereas longer linkers were related to loss of signal (Figure S1c). It is important to note that a similar dependence on linker length was also observed when bound proteins were detected by antibodies after binding to array-bound hexanucleotides. The two independent sets of consistent signal data indicate that processes such as quenching do not have a major influence on signal strength. In sum, these control array-based experiments provide strong evidence for the feasibility of exploiting the complete array-bound hexanucleotide sequence space to systematically screen for species that bind sequence-specifically to HIV-1 RT.

Next, an array consisting of 4096 hexanucleotides was processed as depicted in Figure 1a. Prior to protein binding studies the array and the number of species per printed spot was monitored by labeling their 3′-end with terminal deoxynucleotidyl transferase or by staining with SYBR-Gold. Well-characterized hexanucleotides that bind HIV-1 RT, and controls, such as longer oligonucleotides and non-binders,[3] were included (see the Supporting Information, Table S1). A typical binding experiment with fluorescently labeled HIV-1 RT (Figure 1b) shows a large dynamic range of signals indicating clear discrimination of RT for the printed nucleotide sequences. This pattern is compatible with the pattern generated by controls used for validation on the control array and in solution binding studies. It should also be noted that in some cases we never identified a hexanucleotide binding to a protein target. For example, no binding was observed for array-attached hexanucleotides to firefly luciferase. This is consistent with a SELEX experiment in solution that did not reveal a luciferase-specific aptamer.

To study even more rigorously the specificity of the binding of HIV-1 RT to the hexanucleotide we investigated the signal pattern of the RT of HIV-2. The polymerases of both viruses are highly homologous in primary and tertiary structure;[9] thus, HIV-2 RT may serve as one of the most stringent controls for the specificity of hexamer/HIV-1 RT interactions. The signal pattern produced for the two related target proteins, however, is clearly different (Fig-

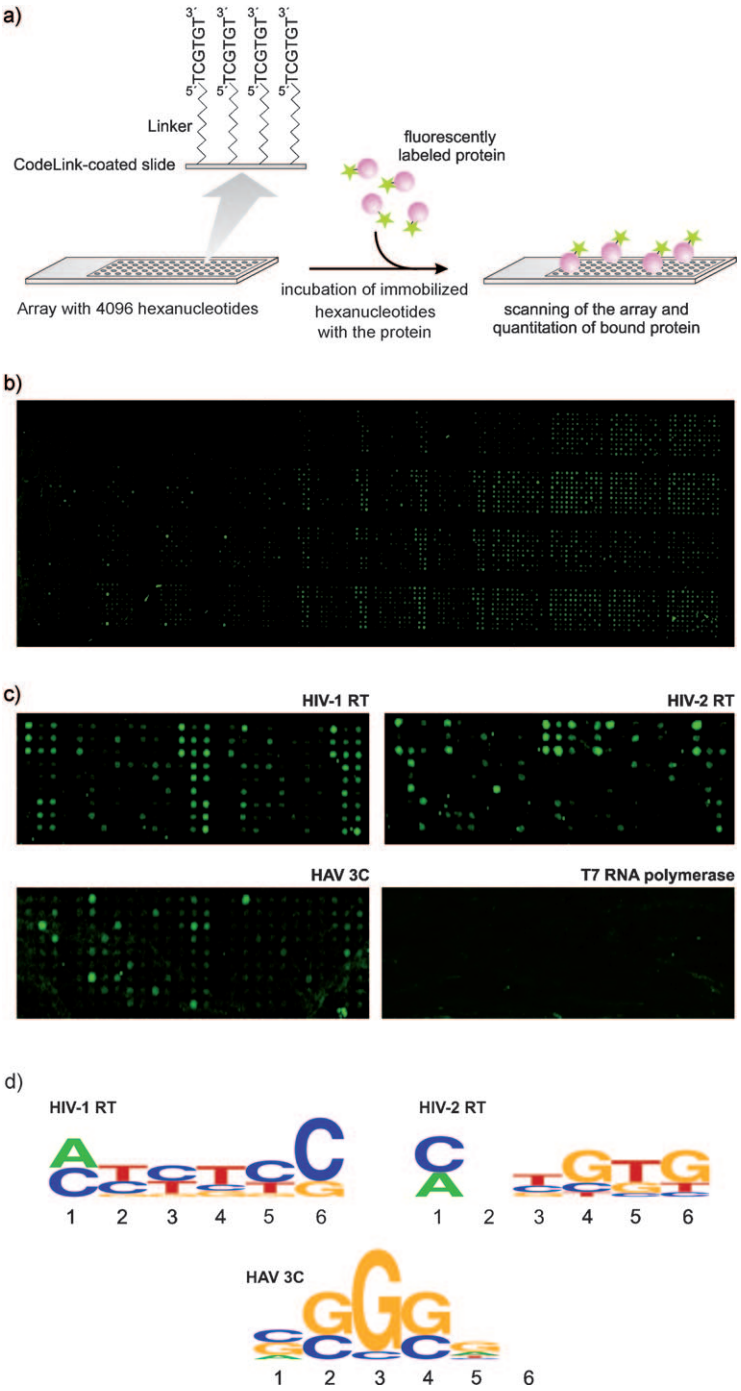

Figure 1. a) Array-based screen for protein–hexanucleotide interactions. Hexanucleotides are covalently attached to the surface through a nonnucleic acid linker (see the Supporting Information, Figure S1a) thereby avoiding non-sequence-specific interactions. In order to visualize hexanucleotide–target interactions, proteins may be labeled fluorescently, radioactively, by biotin, or an antigenic domain for subsequent immunodetection. Here, hexanucleotide binding was monitored by using fluorescently labeled protein. After a blocking and binding step, the arrays were washed and dried. Signals were detected by a Typhoon 8600 scanner. b) Signal pattern obtained with an array treated with 100 nM p66/p51^{K281C}Alexa^{488} (upper panel). c) Examples of target-specific signatures of signals in an enlargement of a specific subarea. The targets HIV-1 RT, HIV-2 RT, HAV 3C, and T7 RNA-polymerase are indicated on the upper right of each field. d) Consensus binding sequences (5′ to 3′) were derived from the analysis of the complete signal pattern of the 4096 array for HIV-1 RT, HIV-2 RT, and HAV 3C. A consensus sequence could not be obtained for T7 RNA polymerase.

ure 1b). The analysis of the tightest binders (see the Supporting Information, Table S2) reveals consensus sequences that are not identical for the two targets. This difference is most obvious for positions 4 and 6 of the hexanucleotides (Figure 1b).

Additional target proteins and their binding hexanucleotides are listed in Table 1. One should note the small number of positive signals for protein targets recognized by the arrays consisting of 4096 hexanucleotides, as well as the complete lack of signals in some instances including for T7 RNA polymerase. These observations indicate a high degree of specificity for the hexanucleotide–protein interactions.

Table 1: Protein targets studied for interactions with the hexanucleotide arrays.

Protein[a]	Array[b] (4096)	Array[b] (control)	Number of binders	Consensus sequence[c]
HIV-1 RT	tested	tested	>15	MYYYYC
HIV-2 RT	tested	tested	>15	MNYSKK
HAV 3C	tested	n.t.	14	SGGGNN
T7 RNA polymerase	tested	tested	0	–
$3CL^{pro}$	tested	n.t.	4	CTCTYN
ICAM	tested	n.t.	0	–
firefly luciferase	tested	tested	0	–
BSA	n.t.	tested	0	–

[a] HAV 3C, hepatitis A virus protease 3C; 3CLpro, coronavirus main proteinase. [b] n.t., not tested. [c] Nucleotide codes: M (A or C), Y (C or T), N (any nucleotide), S (G or C), K (G or T).

This array-based approach allows an extremely fast and reliable identification of specific hexanucleotide–target protein interactions in a single experimental round. This technology can be adapted to other short-chain oligonucleotides and targets. Moreover, the immobilization of hexanucleotides to the array circumvents the possible interference of different species in target binding. This approach proved to be sensitive (see the Supporting Information, Figure S1b), and obviates steps that might put bias into results, such as for example cloning or amplification steps.

Considering the natural predisposition of RT to interact with nucleic acids, one would expect a major number of sequences to be highlighted. On the contrary, different sequences displayed different specificities towards the target protein, strongly indicating that the interactions identified here are mostly dependent on the pattern of functional groups of nucleobases. This finding suggests that this kind of array-based technology can serve as a powerful tool for the systematic search for hexanucleotide binders which, subsequently, can be optimized by adjusting the chemical modifications on the sugar moiety, the nucleobases, and the internucleotide phosphates or additional substituents. In this way the limited complexity of oligonucleotides in the first array-based screen, that is, 4096 species, can be substantially extended by additional chemical modifications and might approach the complexity of systematic combinatorial methods that start with more species in the starting pools. Short nucleic acids slightly different in length, including pentamers or heptamers might also be used; these groups would be easier and less expensive to synthesize than long-chain oligonucleotides including classical aptamers.

The technology described here could support clinical drug discovery at the stage of lead identification, and it could conceivably be exploited for diagnostics and laboratory purposes where hexanucleotide binders can be envisaged as aptamer-like tools. Further, this approach could also be exploited for new applications, such as the identification of regulatory sequences recognized by specific proteins, the analysis of changes in the binding ability of proteins under various environmental conditions, and for studying interactions between hexanucleotides and other possibly more complex and structured nucleic acids.

Received: June 7, 2010
Revised: August 4, 2010
Published online: December 22, 2010

Keywords: HIV-1 · microarrays · oligonucleotides · proteins · reverse transcriptase

[1] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.
[2] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505–510.
[3] A. Mescalchin, W. Wünsche, S. D. Laufer, D. Grohmann, T. Restle, G. Sczakiel, *Nucleic Acids Res.* **2006**, *34*, 5631–5637.
[4] M. Pinskaya, E. Romanova, E. Volkov, E. Deprez, H. Leh, J. C. Brochon, J. F. Mouscadet, M. Gottikh, *Biochemistry* **2004**, *43*, 8735–8743.
[5] J. R. Wyatt, T. A. Vickers, J. L. Roberson, R. W. Buckheit, T. Klimkait, E. DeBaets, P. W. Davis, B. Rayner, J. L. Imbach, D. J. Ecker, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1356–1360.
[6] A. D. van Dijk, R. Boelens, A. M. Bonvin, *FEBS J.* **2005**, *272*, 293–312.
[7] B. Meyer, T. Peters, *Angew. Chem.* **2003**, *115*, 890–918; *Angew. Chem. Int. Ed.* **2003**, *42*, 864–890.
[8] D. J. Schneider, J. Feigon, Z. Hostomsky, L. Gold, *Biochemistry* **1995**, *34*, 9599–9610.
[9] J. Ren, L. E. Bird, P. P. Chamberlain, G. B. Stewart-Jones, D. I. Stuart, D. K. Stammers, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14410–14415.
[10] Blaum et al., unpublished results.

