Bioorganic & Medicinal Chemistry Letters 20 (2010) 3793–3796

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Evolution of specific RNA motifs derived from pan-protein interacting precursors

Marie-Sophie L. Ahmed, Günter Mayer *

Strathclyde Institute for Pharmacy and Biological Sciences, University of Strathclyde, 27 Taylor Street, Glasgow G4 0NR, Scotland, UK
Life and Medical Sciences, University of Bonn, Gerhard-Domagk-Str. 1, 53123 Bonn, Germany

ARTICLE INFO

Article history:
Received 2 March 2010
Revised 9 April 2010
Accepted 13 April 2010
Available online 18 April 2010

ABSTRACT

In vitro evolution of nucleic acid aptamers is a powerful tool to investigate the structure–function relationship of natural occurring RNA–protein interaction motifs. Otherwise, it also allows the identification of novel RNA-based ligands that can be used to investigate a target's function in its native environment. However, artifacts have been described during in vitro selection procedures hampering the successful enrichment of aptamers. Here we describe a novel observation, namely the enrichment of pan-protein binding RNA sequences. We demonstrate that evolution of specific target binding sequences originating from a pan-protein binding RNA precursor is possible in general. Our data demonstrate that the mutual co-variation of an ancestor molecule can be applied for the evolution of specific target binding RNA sequences. These results might have implications in the context of the RNA world theory, exemplifying a possible evolutionary route towards protein-specific RNA molecules from a common ancestor.

In vitro selection is a powerful procedure for developing nucleic acid-based ligands, so-called aptamers that interact with various molecules, ranging from small molecules, peptides, proteins to living cells.[1–3] Aptamers have been employed for various applications being either of diagnostic, functional or therapeutic relevance.[4–6] The in vitro selection process comprises distinct steps, namely (i) the incubation of a diverse nucleic acid library with the target molecule, (ii) separation of bound from non-bound sequences, (iii) elution of bound molecules and the (iv) replication and amplification of the eluted species.[7,8] Inherent properties of the employed nucleic acid library, such as the length of the random region, the primary sequence of the constant regions and its suitability for the in vitro transcription/replication process can have a strong influence on the outcome of the selection experiment. However, artifacts have been described that are mainly associated with the replication step of the entire process.[9] The appearance of so-called molecular parasites, nucleic acids that frequently are either shorter or longer than the molecules of the nucleic acid library, has been reported previously.[10] It is anticipated that these parasitic nucleic acids evolve due to their sophisticated replication properties. Their detailed origin and mechanism of replication though remains elusive. Here, we report on a different incident observed during in vitro selection procedures. We describe the identification of RNA species that interact with various proteins in a non-specific manner, although these proteins are unrelated according to their primary sequence and structure. These types of RNA molecules share an important poly-C-tract motif. Furthermore, we demonstrate that such RNA species can be employed as ancestors for the evolution of a protein-specific RNA molecule.

Initially we performed an in vitro selection experiment targeting the Sec7 domain of cytohesin-1, a cytoplasmic regulatory molecule.[11,12] Therefore, we used a RNA library comprising 40 random nucleotides, which has been used successfully for previous selection experiments.[13] The N-terminally His-tagged Sec7 domain was expressed in *Escherichia coli* and purified via Ni^{2+}-ion chelating affinity chromatography to almost homogeneity. The purity of the protein was determined to be greater than 95% by SDS–PAGE analysis (data not shown). Aliquots of the purified protein were stored at −80 °C. Prior to in vitro selection the target was biotinylated as described previously using a fivefold molar excess of *sulfo*-LC-NHS-biotin.[14–16] The biotinylated protein was coupled to streptavidin-coated magnetic particles, and the derivatized beads were used in the selection process. After six selection cycles the enriched RNA library was analyzed for Sec7 binding and an increased affinity of the enriched RNA library compared to the starting library was observed (Fig. 1B). The RNA library was cloned and the sequences of nine monoclonal RNA sequences were determined (Fig. 1A).

Among them two sequences, #5.5 and #5.6 were found twice whereas all other sequences were unique. Noteworthy, all sequences share poly-C-tract motifs (Fig. 1A, gray boxes) albeit only the sequences #5.2, #5.3, and #5.5 revealed target binding (Fig. 1B). Filter retention analysis revealed that these sequences display a remarkable unspecific interaction behavior and besides binding to the target used for the selection process, interaction

* Corresponding author.
E-mail addresses: gmayer@uni-bonn.de, gunter.mayer@strath.ac.uk (G. Mayer).

0960-894X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2010.04.053

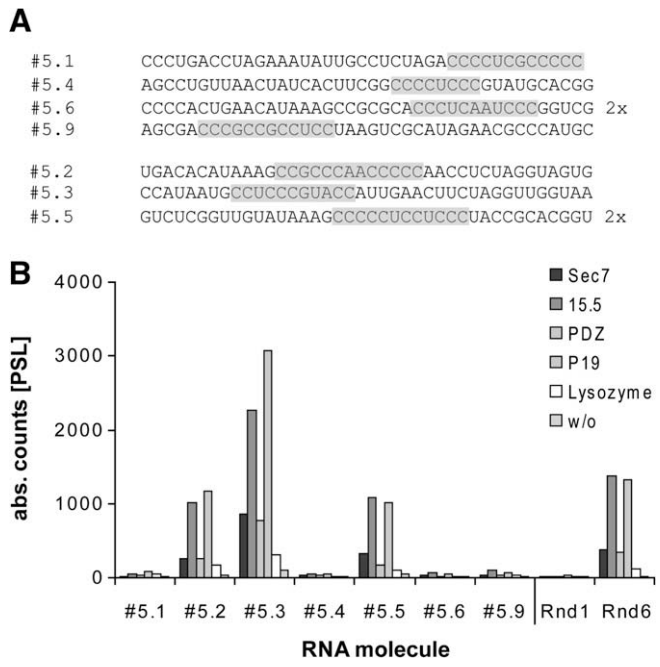

Figure 1. (A) Sequences obtained from the selection targeting the Sec7 domain of cytohesin-1. Only the sequence that corresponds to the initial random region is shown. (B) Filter retention analysis of the monoclonal aptamers shown in (A) and the RNA library from selection cycles 1 (Rnd 1) and 6 (Rnd 6) against the indicated proteins [500 nM].

with the splicosomal protein 15.5, the PDZ domain of cytip, P19 and even, albeit less pronounced, with lysozyme was observed (Fig. 1B and Supplementary Table 1). This binding performance is even more astonishing since the proteins do not share any common motifs neither at the primary nor at the 3D-structure level (Supplementary information). Since 15.5 does not have one, interaction of the RNA molecules with the his-tag of the recombinant proteins can be excluded. Thus, the interaction must be ascertained by a different mode of action. One reasonable explanation might be given through an anticipated interaction of the RNA molecule with the peptide-bond backbone of the proteins. However, this remains elusive yet and needs to be clarified in an ongoing study.

We next determined a minimal binding motif. Therefore, we chose the sequence #5.5 as one representative of the pan-protein binding RNAs and synthesized the shortened variants #5.5.1 to #5.5.4 (Fig. 2A). These variants were defined according to exclude the primer binding sites and on a rational basis using the mfold RNA secondary prediction program (Fig. 2D).[17] Filter retention analysis revealed that the RNA molecule #5.5.1 still interacts with the target proteins (Fig. 2B and C), whereas further truncation (#5.5.2 to #5.5.4) resulted in a loss of binding (data not shown). Remarkably, the minimal variant #5.5.1 had a similar unspecific interaction pattern as revealed by its parental full-length RNA molecule #5.5 (Fig. 2B). A K_D-value of 415 nM of the RNA molecule #5.5.1 regarding its interaction with 15.5 was observed (Fig. 2C and Table 1). All other proteins that interact with #5.5.1 revealed similar affinities within the sub-micromolar range (Supplementary Table 1).

Having the RNA molecule with its remarkably general protein binding pattern in hand, we wondered whether it can serve as an origin for the evolution of specific protein-binding RNA molecules. If successful, this might have implications for the RNA world theory regarding the role of poly-C-tract bearing RNA sequences as molecular RNA fossils that guided the evolution of specific protein interacting motifs.[18–20] To test this hypothesis we synthesized a biased nucleic acid library based on the #5.5.1 RNA molecule, but embedded between the original 5′-primer and a novel 3′-primer binding site. This design was chosen to reasonably prevent the re-evolution of previously identified RNA molecules. The motif #5.5.1 was degenerated to contain an 85% nucleotide identity at each position whereas each of the other three nucleotides was incorporated with 5 % likelihood. This library (S85) did not interact with any of the investigated proteins and was subjected to various selection schemes. First, we started to reselect RNA molecules targeting 15.5. We also included the MAP kinase p38α as target, since it has been shown that it binds to #5.5 only with a very low activity (Supplementary Table 1). Within each selection cycle a pre-selection step against either the matrix (streptavidin-coated beads) or another protein that is bound by the general protein-binding motif, for example as the SOS domain that has also been shown to interact with #5.5 (Supplementary Table 1), was included. This set-up was chosen to reduce the likelihood of recurrence of aforementioned pan-protein binding RNA molecules, such as #5.5. After 3–8 selection cycles we analyzed the enriched libraries by filter retention analysis for protein binding. Interestingly and independently of the pre-selection strategy (matrix or SOS), the p38α targeting selection solely yielded RNA molecules with general protein binding characteristics, whereas no enrichment of specific p38α

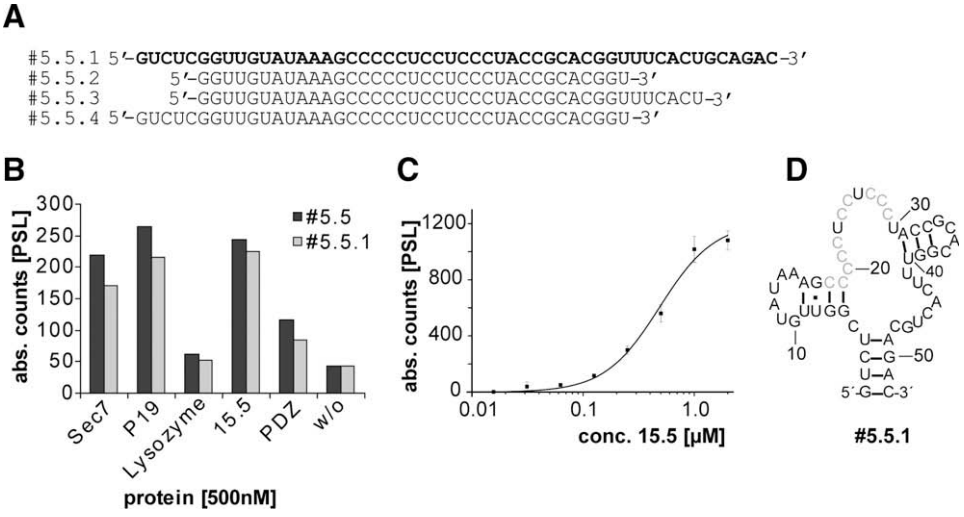

Figure 2. (A) Sequences of the truncated variants of the RNA molecule #5.5. #5.5.1 (bold) represents the minimal binding motif. B) Filter retention analysis of the parent aptamer #5.5 and its truncated minimal motif #5.5.1. (C) Determination of the dissociation constant of #5.5.1 regarding its interaction with the protein 15.5. (D) Putative secondary structure of the RNA sequence #5.5.1. Poly-C-tract cytidines are shown in orange.[17]

Table 1
Dissociation constants of the RNA motif and its truncated and mutated variants towards the protein 15.5

Aptamer	K_D[a]
#5.5	468 (±59)
#5.5.1	415 (±25)
17	578 (±59)
23	308 (±35)
29	257 (±17)
29.1	348 (±40)
29.2	n.d.

[a] Values given in nM and are means of two experiments, mean ± half of range is given in parentheses (n.d., not detectable).

binding sequences was detectable. Cloning and sequencing of the cycle 3 RNA derived from the p38α selection revealed that the enriched library mainly consisted of #5.5 related sequences (Fig. 3A).

The sequences revealed either none (11), 3 (10), or 4 (4) point mutations compared to #5.5 and sequence 4 additionally bears a deletion of five nucleotides in the 3′-part of the original #5.5.1 motif (Fig. 3A and C). Filter retention analysis revealed that these RNA molecules display a binding behavior that is very similar to the one observed by #5.5 (Fig. 3B). In contrast, the selection targeting the spliceosomal protein 15.5 yielded a population of RNA species that specifically interacted with 15.5. Cloning and sequencing revealed three RNA molecules that interact with 15.5 with high affinity and specificity (Fig. 4A and B). Remarkably, two sequences, 23 and 29,

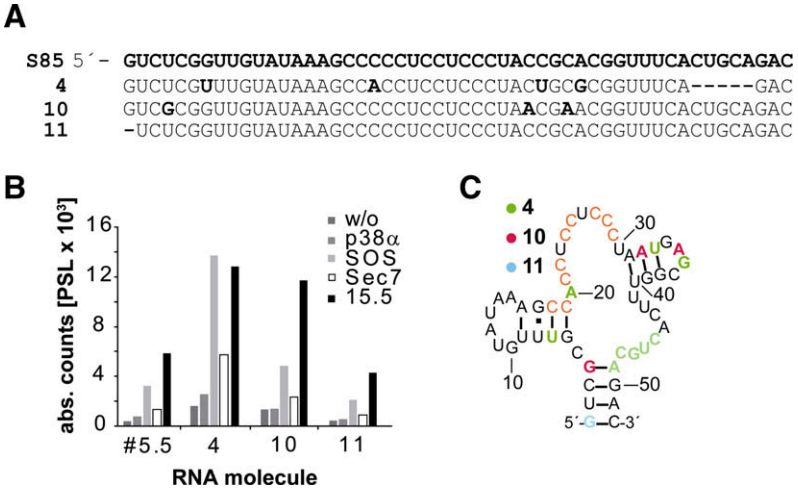

Figure 3. (A) Sequences obtained from the selection targeting p38α with the degenerated RNA library S85 (top, bold). Only the initial degenerated sequence region is given. (B) Filter retention analysis of the monoclonal RNA sequences 4, 10, and 11 (A) and the general protein-binding RNA molecule #5.5 as reference against the indicated proteins [500 nM]. (C) Secondary structure of the sequences 4, 10, and 11 and the position of the mutated nucleotides compared to the sequence #5.5.1. Color code: mutants found in four are shown in green (transparent nucleotides 45–49 in green correspond to deletions), those in 10 shown in magenta and the deletion of 11 is shown in cyan.

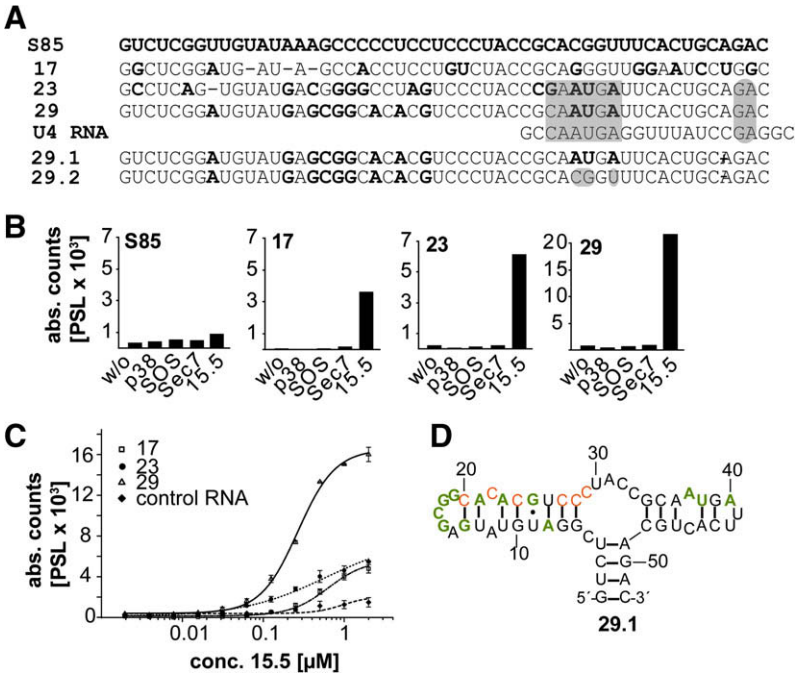

Figure 4. (A) Sequences obtained from the selection targeting 15.5 with the degenerated RNA library S85. Only the initial degenerated sequence region is given. 29.1 and its mutated variant 29.2 represent minimal motifs derived from 29. + indicates 15.5 binding whereas – indicates non-target binding. (B) Filter retention analysis of the monoclonal RNA sequences 17, 23, and 29 (A) and the degenerated RNA library S85 as reference against the indicated proteins [500 nM]. (C) Determination of the dissociation constant of 17, 23, 29 and the control RNA (S85) regarding their interaction with the splicosomal protein 15.5. (D) Putative secondary structure of 29.1. Mutated nucleotides compared to #5.5.1 (Figure 2D) are shown in green.[17] Remaining residues belonging to the poly-C-tract from #5.5.1 are in orange.

bear three mutations that form the conserved consensus motif of the wild-type U4 RNA, which naturally interacts with 15.5 (Fig. 4A, gray boxes).[21,22] The minimal variant of the RNA molecule 29, namely 29.1, also revealed specific binding to 15.5 (Fig. 4A). Interestingly, mutation of the U4 RNA-like consensus motif resulted in a loss of binding, indicating the importance of this motif for 15.5 interaction (Fig. 4A and B). The mutations of 17, 23 and 29 were all found to disrupt the poly-C-tract motif (Fig. 4A and D) a further indication that the poly-C-tract is required for a rather general protein binding behavior.

While nucleic acid-based ligands have a variety of biomedical applications, in vitro selection procedures also allow questions regarding evolutionary chemistry and biology models to be addressed. For example the evolution of ribozymes that catalyze the formation of chemical bonds and the synthesis of metabolite precursors is a major issue.[23,24] Here, we add a new piece to the evolution of specific RNA–protein recognition motifs. We describe a poly-C-tract containing RNA molecule, able to interact with various unrelated proteins (Supplementary Fig. 1). Similar polypyrimidine tracts have been found in viral RNA and were shown to be necessary to control translation in eukaryotes,[25–28] indicating that such motifs play important roles in biology. We employed the poly-C-tract RNA as a molecular precursor for the evolution of specific protein-binding RNA motifs. Most remarkably, we isolated a RNA sequence that specifically interacts with the spliceosomal protein 15.5 and resembles an important conserved core motif homologous with the natural occurring U4 RNA sequence. All other mutations were found to interfere with poly-C-tract formation. This strongly indicates that the original poly-C-tract motif is a pre-requisite albeit not sole indicator for general protein binding properties. In turn a similar evolution experiment with the degenerated S85 library targeting the MAP kinase p38α resulted in the re-evolution of the general protein-binding RNA motif #5.5.1. This finding is surprising since no co-enrichment of specific p38α binders was observed and the binding activity of the general protein-binding motif towards p38α has been shown to be rather weak.

In conclusion we demonstrated that pan-protein binding RNA species can be further evolved by in vitro selection experiments yielding specific protein-binding RNA molecules. These findings may have implications for the RNA world theory, indicating that very first RNA molecules with an unspecific protein interaction pattern might represent common ancestors from which specific protein recognition motifs were achieved through co-evolutionary processes. Furthermore, the results of this study suggest conducting experiments to address the detailed molecular mechanism by which the pan-protein interaction motif facilitates protein recognition.

Acknowledgments

The authors thank Prof. Famulok, Limes Institute University of Bonn, for kind provision of several proteins. The plasmid for expression and purification of p38α was kindly provided by Prof. Han, The Scripps Research Institute. This work was made possible by grants of the Fonds der Chemischen Industrie to G.M. G.M. acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) and donations by the Scottish Universities Life Science Association (SULSA).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2010.04.053.

References and notes

1. Mayer, G. *Angew. Chem. Int. Ed.* **2009**, *48*, 2672.
2. Mayer, G.; Jenne, A. *BioDrugs* **2004**, *18*, 351.
3. Schurer, H.; Stembera, K.; Knoll, D.; Mayer, G.; Blind, M.; Forster, H. H.; Famulok, M.; Welzel, P.; Hahn, U. *Bioorg. Med. Chem.* **2001**, *9*, 2557.
4. Famulok, M.; Mayer, G. *Chembiochem* **2005**, *6*, 19.
5. Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611.
6. Berezovski, M. V.; Lechmann, M.; Musheev, M. U.; Mak, T. W.; Krylov, S. N. *J. Am. Chem. Soc.* **2008**, *130*, 9137.
7. Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
8. Tuerk, C.; Gold, L. *Science (New York, NY)* **1990**, *249*, 505.
9. Musheev, M. U.; Krylov, S. N. *Anal. Chim. Acta* **2006**, *564*, 91.
10. Marshall, K. A.; Ellington, A. D. *J. Mol. Evol.* **1999**, *49*, 656.
11. Mayer, G.; Lohberger, A.; Butzen, S.; Pofahl, M.; Blind, M.; Heckel, A. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6561.
12. Mayer, G.; Blind, M.; Nagel, W.; Bohm, T.; Knorr, T.; Jackson, C. L.; Kolanus, W.; Famulok, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4961.
13. Cox, J. C.; Ellington, A. D. *Bioorg. Med. Chem.* **2001**, *9*, 2525.
14. Mayer, G.; Wulffen, B.; Huber, C.; Brockmann, J.; Flicke, B.; Neumann, L.; Hafenbradl, D.; Klebl, B. M.; Lohse, M. J.; Krasel, C.; Blind, M. *RNA* **2008**, *14*, 524.
15. Mayer, G.; Hover, T. *Methods Mol. Biol.* **2009**, *535*, 19.
16. Muller, J.; El-Maarri, O.; Oldenburg, J.; Potzsch, B.; Mayer, G. *Anal. Bioanal. Chem.* **2008**, *390*, 1033.
17. Zuker, M.; Stiegler, P. *Nucleic Acids Res.* **1981**, *9*, 133.
18. Joyce, G. F. *Cold Spring Harb. Symp. Quant. Biol.* **2009**.
19. Cech, T. R. *Cell* **2009**, *136*, 599.
20. Bartel, D. P.; Unrau, P. J. *Trends Cell Biol.* **1999**, *9*, M9.
21. Schultz, A.; Nottrott, S.; Watkins, N. J.; Luhrmann, R. *Mol. Cell. Biol.* **2006**, *26*, 5146.
22. Nottrott, S.; Hartmuth, K.; Fabrizio, P.; Urlaub, H.; Vidovic, I.; Ficner, R.; Luhrmann, R. *EMBO J.* **1999**, *18*, 6119.
23. Fusz, S.; Eisenfuhr, A.; Srivatsan, S. G.; Heckel, A.; Famulok, M. *Chem. Biol.* **2005**, *12*, 941.
24. Lincoln, T. A.; Joyce, G. F. *Science (New York, NY)* **2009**, *323*, 1229.
25. Zhang, B.; Seitz, S.; Kusov, Y.; Zell, R.; Gauss-Muller, V. *Biochem. Biophys. Res. Commun.* **2007**, *364*, 725.
26. Kuo, J. S.; Veale, R.; Maxwell, B.; Sive, H. *Gene* **1996**, *176*, 17.
27. Duke, G. M.; Osorio, J. E.; Palmenberg, A. C. *Nature* **1990**, *343*, 474.
28. Duke, G. M.; Palmenberg, A. C. *J. Virol.* **1989**, *63*, 1822.