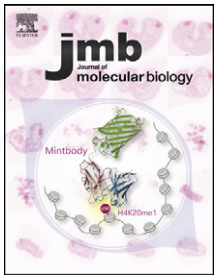

Specific Recognition of a Single-Stranded RNA Sequence by a Synthetic Antibody Fragment

Yaming Shao[1], Hao Huang[2], Daoming Qin[3], Nan-Sheng Li[1], Akiko Koide[1], Jonathan P. Staley[3], Shohei Koide[1], Anthony A. Kossiakoff[1] and Joseph A. Piccirilli[1,2]

*1 - **Department of Biochemistry and Molecular Biology,** The University of Chicago, Chicago, IL 60637, USA*
*2 - **Department of Chemistry,** The University of Chicago, Chicago, IL 60637, USA*
*3 - **Department of Molecular Genetics and Cell Biology,** Molecular Genetics and Cell Biology, The University of Chicago, Chicago, IL 60637, USA*

***Correspondence to Joseph A. Piccirilli:** Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA. jpicciri@uchicago.edu*
http://dx.doi.org/10.1016/j.jmb.2016.08.029
Edited by Philip C Bevilacqua

Abstract

Antibodies that bind RNA represent an unrealized source of reagents for synthetic biology and for characterizing cellular transcriptomes. However, facile access to RNA-binding antibodies requires the engineering of effective Fab libraries guided by the knowledge of the principles that govern RNA recognition. Here, we describe a Fab identified from a minimalist synthetic library during phage display against a branched RNA target. The Fab (BRG) binds with 20 nM dissociation constant to a single-stranded RNA (ssRNA) sequence adjacent to the branch site and can block the action of debranchase enzyme. We report the crystal structure in complex with RNA target at 2.38 Å. The Fab traps the RNA in a hairpin conformation that contains a 2-bp duplex capped by a tetraloop. The paratope surface consists of residues located in four complementarity-determining regions including a major contribution from H3, which adopts a helical structure that projects into a deep, wide groove formed by the RNA. The amino acid composition of the paratope reflects the library diversity, consisting mostly of tyrosine and serine residues and a small but significant contribution from a single arginine residue. This structure, involving the recognition of ssRNA via a stem–loop conformation, together with our two previous structures involving the recognition of an RNA hairpin loop and an RNA tertiary structure, reveals the capacity of minimalist libraries biased with tyrosine, serine, glycine, and arginine to form binding surfaces for specific RNA conformations and distinct levels of RNA structural hierarchy.

Introduction

As key protein components of the mammalian immune systems, antibodies serve critical roles in identifying, neutralizing, and eliminating foreign substances. Within the molecular context of the immunoglobulin architecture, the variation of the complementarity-determining region (CDR) amino acid composition allows the immune system to generate antibodies that bind to virtually any target with high affinity and exquisite specificity. This vast, functional capacity has led to a rapidly growing number of antibodies for medical applications as diagnostic tools and therapeutic agents and has empowered a variety of methodologies for research applications ranging from imaging techniques to chaperone-assisted crystallization [1–3]. Their versatility and widespread utility has fueled the development of methods to obtain antibodies, including hybridoma approaches and display technologies using natural repertoires derived from immune systems or man-made repertoires derived from synthetic oligonucleotides of designed diversity [4,5].

Most antibodies used for medical or research applications target proteins of the terrestrial proteome and associated post-translational modifications [6,7]. Nevertheless, biological systems also possess a highly complex and dynamic cellular RNA population,

collectively known as the transcriptome, and antibodies that target RNA could provide equally valuable reagents for medicine and research. Transcriptomes play significant roles in gene regulation, cell growth, differentiation, and disease [8–11]. A small fraction of these RNAs serve as mRNAs for protein synthesis, but the biological role for the majority of transcripts is still unclear [12]. Despite their biological and medical importance, our understanding of transcriptomes and associated post-transcriptional modifications remains in its infancy, due in part to limited methods for elucidating the transcriptome's structure and function. Currently, large-scale identification and analysis of these non-coding RNAs (ncRNAs) rely predominantly on deep-sequencing data [13], sometimes following immunoprecipitation of specific proteins [14,15]. Following ncRNA identification, other conventional techniques such as hybridization and antisense can be deployed to locate, image, or isolate specific RNAs for further research [16,17]. Nevertheless, many ncRNAs mediate their functions through diverse secondary and tertiary structures, which cannot be characterized readily using available techniques [10]. Conformation-specific RNA-binding antibodies, as highly engineerable proteins with diversity in their CDRs, can potentially serve this purpose well.

The blood serum from human patients or animal models with autoimmune disease may contain antibodies that bind to nucleic acids [18–24]. However, healthy immune systems deploy components of the innate immune system, such as Toll-like receptors and protein kinases, in response to the presence of non-self RNA [25]. Thus, injection of structured RNA into the blood of a host animal would likely not elicit the production of antibodies that target the injected RNA. Moreover, the presence of nucleases in blood serum would likely limit the lifetime of the injected RNA, precluding the use of immunization approaches to obtain antibodies.

In the recent years, display technologies have enabled access to recombinant antibodies without the need for host immunization, providing an opportunity to circumvent exposure to nucleases and thereby offering a platform to engineer and isolate antibodies that bind to specific RNAs [6]. These display technologies have used natural immune repertoires [26], but minimalist repertoires built from designed diversity have also proven effective against protein antigens [1,27]. These synthetic libraries have undergone successive rounds of optimization based on information from antibody: protein interactions derived from analysis of functional antibody sequences and structures [28,29]. For RNA targets, our recent work has demonstrated that minimalist libraries within the Fab 4D5 framework can yield Fabs that bind RNAs with high affinity and specificity [30]. We have also established proof of principle that these Fabs can serve as chaperones for RNA crystallization [30–32]. Nevertheless, compared with selections against protein antigens using optimized libraries, selections against RNA targets have yielded relatively few high-affinity Fabs, reflecting the limited knowledge of library design principles with respect to CDR sequence diversity, length, and structure.

The limitations of currently available libraries underscore the need to investigate structural and energetic principles of Fab:RNA complexes for a range of RNA types in order to elucidate the principles and minimum requirements underlying protein–RNA recognition and to inform future library design. Previously, we described the structures of two Fab–RNA complexes, one revealing the recognition of an RNA tertiary structure and the other revealing the recognition of an RNA hairpin loop [30,32]. In this study, we describe a Fab capable of recognizing a stem–loop RNA conformation of a single-stranded RNA (ssRNA). Using a minimalist library, we obtained a Fab that binds to a 10-nt ssRNA sequence with high affinity and specificity, and we solved the crystal structure of the Fab:RNA complex at 2.38-Å resolution. The new structure together with the two previous structures reveals the principles of RNA recognition that govern binding surfaces generated from minimalist libraries enriched in tyrosine (Y), serine (S), glycine (G), and arginine (R) and that suggest strategies to improve the design of future libraries.

Results

Selection and characterization of Fabs directed to ssRNA

We initially set out to isolate Fabs that bind to branched RNAs, which arise from the chemistry of nuclear pre-mRNA splicing catalyzed by the spliceosome. We used a deoxyribozyme to construct a branched RNA corresponding in sequence to the branch region of the pre-mRNA intron derived from the *YBL059W* gene from *Saccharomyces cerevisiae* [33,34]. For target immobilization, strand L of the branched RNA was biotinylated on the 3′-end during solid-phase oligonucleotide synthesis (Fig. 1a). For *in vitro* selection using phage display, we deployed two synthetic antibody libraries that carry a “reduced genetic code” [29,32]. One library, termed YSGR, contains Y, S, G, and R and encodes equal proportions of Y and S at variable positions in CDR-L3, CDR-H1, and CDR-H2. CDR-H3 encodes 38% Y, 25% S, 25% G, and 12% R [28,30]. The second library, termed YSGRKX, encodes diversity in all six CDRs. CDR-L1 and L2 contain equal proportions of Y and S at variable positions; CDR-H1 and-H2 contain equal proportions of Y, F, and S; and CDR-L3 and H3 encode 25% Y, 15% S, 10% G, 12.5% R, 7.5% K and 30% X, where X represents all other amino acids except C, I, and M. In addition, four of the CDRs