Nucleic Acids Research, 2025, **53**, gkaf072
https://doi.org/10.1093/nar/gkaf072
Advance access publication date: 14 February 2025
Chemical Biology and Nucleic Acid Chemistry

Cancer cell target discovery: comparing laboratory evolution of expanded DNA six-nucleotide alphabets with standard four-nucleotide alphabets

Sharpkate Shaker[1,†], Jun Li[1,†], Shuo Wan[2,*,†], Hong Xuan[1], Jinchen Long[1], Haiyan Cao[1], Tongxuan Wei[1], Qinguo Liu[1], Da Xu[3,*], Steven A. Benner[2,*], Liqin Zhang[1,*]

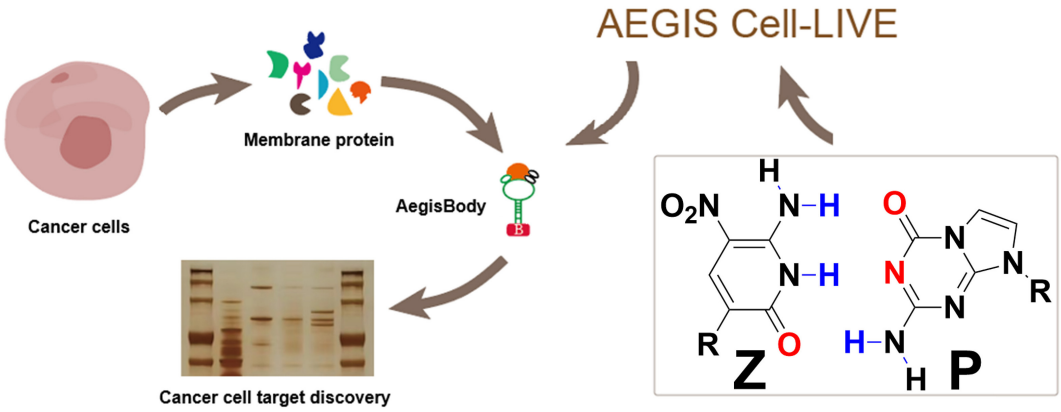
[1]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China
[2]Foundation for Applied Molecular Evolution, Alachua, FL 32615, United States
[3]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Hepatopancreatobiliary Surgery Department I, Peking University Cancer Hospital & Institute, Beijing 100142, China

[*]To whom correspondence should be addressed. Email: sbenner@ffame.org
Correspondence may also be addressed to Da Xu. Email: daxu@bjmu.edu.cn
Correspondence may also be addressed to Shuo Wan. Email: shuowan2chem@gmail.com
Correspondence may also be addressed to Liqin Zhang. Email: lqzhang@hsc.pku.edu.cn
[†]These authors are Joint First Authors.

Abstract

Anthropogenic evolvable genetic information systems (AEGIS) are DNA-like molecules that can be copied, support laboratory *in vitro* evolution (LIVE), and evolve to give AegisBodies, analogs of antibodies. However, unlike DNA aptamers built from four different nucleotides, AegisBodies are currently built from six different nucleotides. Thus, six-letter AEGIS–LIVE delivers AegisBodies with greater stability in biological mixtures, more folds, and enhanced binding and catalytic power. Unlike DNA however, AEGIS has not benefited from 4 billion years of biological evolution to create AEGIS-specialized enzymes, but only a decade or so of human design. To learn whether AEGIS can nevertheless perform as well as natural DNA as a platform to create functional molecules, we compared two six-letter AegisBodies (LZH5b and LZH8) with a single standard four-letter aptamer, both evolved to bind specific cancer cells with ∼10 cycles of LIVE. Both evolved ∼50 nM affinities. Both discovered proteins on their cancer cell surfaces thought to function only inside of cells. Both can be internalized. Internalizing of LZH5b attached to an AEGIS nanotrain brings attached drugs into the cell. These data show that AEGIS–LIVE can do what four-letter LIVE can do at its limits of performance after 4 billion years of evolution of DNA-specialized enzymes, and better by a few metrics. As synthetic biologists continue to improve enzymology and analytical chemistry to support AEGIS-LIVE, this technology shoud prove increasingly useful as a tool, especially in cancer research.

Graphical abstract



Introduction

It has been over 30 years since Gold [1, 2], Szostak [3], Joyce [4], and others proposed using DNA and RNA (nucleic acids, NA) as scaffolds in "laboratory *in vitro* evolution" (LIVE) experiments to evolve binders and catalysts from NA libraries. Their approach was inspired by the discovery that

Received: August 14, 2024. **Revised:** January 7, 2025. **Editorial Decision:** January 20, 2025. **Accepted:** January 29, 2025

RNA could catalyze chemical reactions [5–8]. This revived the concept of an "RNA replicase"—an RNA molecule capable of polymerizing other RNA molecules [9–11]. This concept, in turn, added support to the "RNA World" hypothesis [12, 13], which holds that early life on Earth might have used RNA as its only encoded component of biological catalysis, with similar systems possibly existing on other planets [14].

The logic behind LIVE was that nucleic acids might directly evolve, under the guidance of a scientist, to give functional molecules, by random variation and natural selection applied to a library of NA sequences, just as they might have evolved in an "RNA World" [10, 15]. In this process, a library of NA molecules, typically with $\sim 10^{14}$ variants with a randomized sequence of 25–40 nucleotides, is synthesized. These libraries are then exposed to a target molecule or cell, with nonbinding molecules washed away. The surviving nucleic acids are amplified and selected again, using a counter-target to refine binding specificity through multiple rounds of selection. After enough rounds, the most selective binders are sequenced and analyzed as potential "aptamers".

Aptamers were proposed to offer several advantages over antibodies and, more generally, binding molecules [16, 17] based on any protein scaffold. First, they were seen as cheaper to develop, easier to manufacture, and more reproducible [18]. Further, they have low immunogenicity [19], and can be readily modified to carry drugs, reporters, or other aptamers in multifunctional constructs [20]. Their nearly universal ability to dissolve in water, a contrast with the intrinsic propensity of proteins to precipitate, was another advantage [21].

Essentially, all of these advantages are realizable. Nevertheless, not withstanding pegaptanib sodium and avacincaptad pegol as approved drugs [22, 23], aptamers have not yet displaced antibodies in most of their application areas, neither in research, diagnostics, nor the pharmacopeia.

The reasons for this lie in the limitations of natural NAs as scaffolds for evolving functional molecules. As a first limitation, they have low information density. With only four nucleotides, NAs have poor control over folding, a limit to their intrinsic binding and catalytic potentials [24]. Additionally, their repeating backbone charges, while critical for solubility [25], restrict their ability to form compact tertiary structures. Only one motif, the G-quadruplex, can reliably form such structures [26]. Moreover, adding functional groups to enhance binding and catalysis [27] leads to "over-functionalization," where many added functional groups reduce solubility, create nonspecific binding, and otherwise cause the heavily modified NAs to no longer behave "like nucleic acids" [28]. This is especially true when those added groups are hydrophobic [29, 30].

Efforts to overcome these challenges have focused on improving selection workflows to identify the most useful sequences faster [31, 32]. Representative innovations include single-molecule and dual-tag selection strategies, a variety of tools based on modifications [33–37].

Anticipating these problems in 1987, the Benner group suggested the creation of a new biopolymer with higher information density and an ability to incorporate protein-like functional groups without over-functionalization. This led to the development of "Anthropogenic Evolvable Genetic Information Systems" (AEGIS, Fig. 1), biopolymers with up to 12 replicable nucleotides and the capacity for more diverse base pairing than natural nucleic acids [38].

The AEGIS concept is simple yet required extensive efforts across multiple fields to rebuild DNA from the ground up [39–46]. These included developing chemistry to make AEGIS nucleosides and oligonucleotides, engineering polymerases to replicate AEGIS DNA, methods to sequence AEGIS DNA [47–49], performing thermodynamic studies to define the biophysics of AEGIS DNA, [50–54], and solving the structures of AEGIS DNA and its complexes [55]. Ongoing work seeks to democratize AEGIS-LIVE to let other laboratories use it easily [56].

With these, AEGIS has proven effective in overcoming many limitations of standard NAs. Its higher information density has supported diagnostic products with over $1 billion in lifetime sales [57–60]. AEGIS supports new folding motifs beyond the G-quadruplex, such as pentaplexes, fat and skinny pairs, and the fZ motif, the last exploiting pairing between Z and deprotonated Z [61-63].

Exploiting these features, AEGIS–LIVE has generated functional AegisBodies, such as those that bind anthrax toxin [64], with fewer selection cycles due to the added diversity of functional groups [56]. The higher information density allows for the sparse addition of functional groups, avoiding over-functionalization, and improving catalytic power. For example, an AegisZyme catalyzes ribonuclease-type reactions using the general acid–general base properties of one of its added building blocks [65].

Nevertheless, AEGIS–LIVE faces resistance from research, diagnostics, and therapeutics communities who either use standard 4-letter LIVE, or who know the limitations of standard LIVE and therefore do not consider LIVE useful to support their work. These individuals, especially in cancer-related communities, must be persuaded that AEGIS DNA offers advantages over standard DNA as an evolving platform.

To address these, we report here a comparison between AEGIS–LIVE and standard LIVE using a well-known product of the second, the C10 aptamer from Raddatz *et al.* [66]. The Raddatz C10 aptamer was selected in a 4-letter DNA LIVE experiment to bind Ramos lymphoma B cells. The C10 aptamer showed "state-of-the-art" performance for 4-letter aptamers in general, with the benefits and limitations expected from several decades of experience with these.

Here, we show that AEGIS–LIVE can achieve binding AegisBodies that perform as well (or better, by some metrics) than C10, even though the molecular biology and analytical chemistry supporting AEGIS-LIVE are far less advanced than the biology and chemistry supporting standard LIVE. This suggests that as its supporting molecular biology and chemistries are improved, AEGIS–LIVE will surpass the limitations of natural DNA in evolving high-affinity, specific binders and catalysts.

Materials and methods

Cell lines, cell cultures, and reagents

Human hepatocellular carcinoma HepG2 cells were obtained from the American Type Culture Collection (ATCC). They were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum with 1% penicillin–streptomycin. HepG2 cells were cultured at 37°C in a humid atmosphere with 5% CO_2.

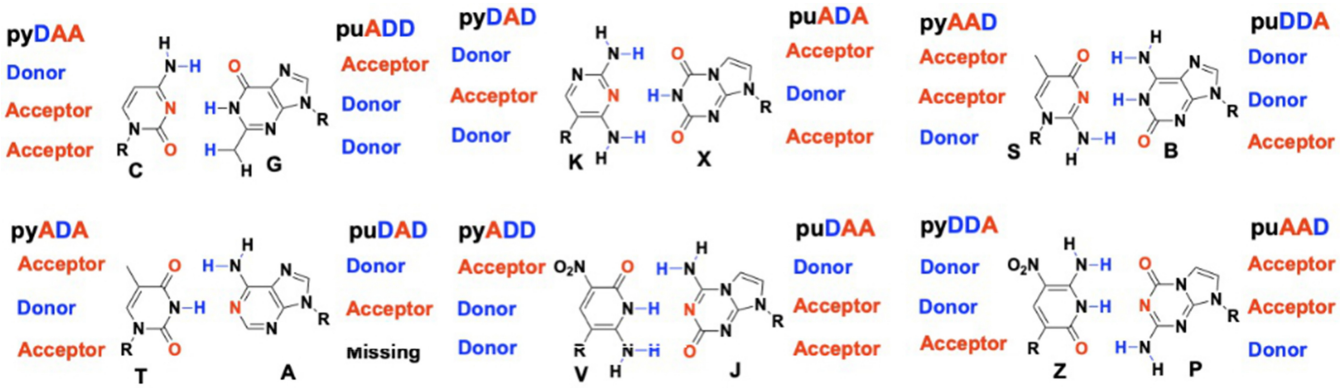


Figure 1. Twelve independently replicable nucleotides form six pairs within a Watson–Crick geometry as an AEGIS. The GACTZP six-letter DNA system studied here incorporates both standard and anthrropogenic nucleotides, adhering to two key complementarity rules: (**A**) size complementarity, where small pyrimidine analogs pair with large purine analogs, and (**B**) hydrogen bonding complementarity with hydrogen bond donors (H) pairing with acceptors (O or N). The sequences of the two AegisBodies examined in this work, with the anthropogenic AEGIS P nucleotide highlighted in bold, are: LZH5b: GTGACGCAGCAGCTAC**P**TGGGCCCTGGT**P**TCTGTGCTGGACAC LZH8: ATCCAGAGTGACGCAGCATATTAGTACGGCTTAACCC**P**CATGGTGGACACGGTGGCTTAGT. These AegisBodies were evolved through AEGIS-LIVE to selectively bind HepG2 liver cancer cells, while avoiding binding to untransformed liver cells [56].

DNA sequences and buffers

Random sequence: GTGACGCAGCAGCTGCCTGTACATGGGCTATCTGGCTGGACAC;

LZH5b: GTGACGCAGCAGCTAC**P** TGGGCCCTGGT**P** TCTGTGCTGGACAC;

LZH8: ATCCAGAGTGACGCAGCATATTAGTACGGC TTAACCC**P** CATGGTGGACACGGTGGCTTAGT;

Scrambled LZH5b: GGGCGCAGCTCTACPTGGACTA CCGGTGGAGT**P** TCTCGCCTGA;

LZH8 P-G: ATCCAGAGTGACGCAGCATATTAGTACG GCTTAACCCGCATGGTGGACACGGTGGCTTAGT;

LZH8 P-A: ATCCAGAGTGACGCAGCATATTAGTACG GCTTAACCCACATGGTGGACACGGTGGCTTAGT;

LZH5b P-G: GTGACGCAGCAGCTACGTGGGCCCTG GTGTCTGTGCTGGACAC;

LZH5b P-A: GTGACGCAGCAGCTACATGGGCCCTG GTATCTGTGCTGGACAC

Washing buffer: Glucose (4.5 g) and 5 ml of 1 M $MgCl_2$ were added to Dulbecco's Phosphate-Buffered Saline (DPBS) [11] and stored at 4°C. Binding buffer was prepared by adding to 1 l of DPBS, 4.5 g of glucose, 100 mg of transfer RNA (tRNA), 1 g of bovine serum albumin (BSA), and 5 ml of 1 M $MgCl_2$. This was also stored at 4°C. Hypotonic buffer was prepared by adding to washing buffer, 10 mM phenylmethylsulfonyl fluoride (PMSF), 50 mM Tris–HCl, and 1× of a protease inhibitor cocktail.

Flow cytometric analysis

To compare the binding abilities of AegisBodies with those of anti-XRCC5 and anti-GRP78 antibodies that bind targets discovered by LZH5b and LZH8 (see below), 200 nM of FITC-LZH5b/LZH8 and the antibodies were prepared in triplicate and incubated with the target cells in 200 µl of binding buffer at 4°C for 1 h. After incubation, non-bound AegisBodies or antibodies were washed out from cells, and the cells were then resuspended in 500 µl of binding buffer. Flow cytometry (Beckman Cytoflex) was used to measure the intensity of fluorescent labeling of cells within each sample.

Competition experiment

The anti-GRP78 BiP antibody (ab21685) (rabbit polyclonal to GRP78 BiP; Reactivity: mouse, rat, human, and Chinese hamster; Isotype: IgG) and the recombinant anti-Ku80 antibody (ab80592) [rabbit monoclonal (EPR3468) to Ku80) were used as primary antibodies. The secondary antibody was goat anti-rabbit IgG H&L (ab150080) (fluorophore: Alexa Fluor® 594, excitation: 590 nm, emission: 617 nm). The cells were collected, washed, and resuspended in cold phosphate buffered saline (PBS) at a concentration of ~1–5 × 10^6 cells/ml. Primary antibodies were added to a final concentration of 2 µg/ml, and the mixture was incubated at 4°C for 30 min in the dark. The cells were then washed three times by centrifugation at 400 × *g* for 5 min and resuspended in cold PBS. Fluorescein (FAM) labeled aptamers were subsequently added to achieve a final concentration of 1 or 5 µM. After incubating at 4°C for 30 min in the dark, the cells were washed and resuspended. Fluorescently labeled secondary antibody was then added, and the mixture was incubated again at 4°C for 30 min in the dark. Following incubation, the cells were washed three times and resuspended in cold PBS. Finally, the cells were analyzed using the flow cytometry to measure the signal intensities of the FAM-labeled aptamer and the fluorescent secondary antibody.

Determining the target of the evolved AegisBodies

To determine the target for the LZH8 AegisBody, flow cytometry was used. First, we detached HepG2 cells with 0.1 mg/ml Proteinase K or 0.25% trypsin-EDTA, respectively. After being washed three times with DPBS, 3 × 10^5 cells were incubated with 200 nM of FITC-labeled random sequences (as a control) or FITC-labeled LZH5b or LZH8 AegisBodies in 200 µl of binding buffer at 4°C for 1 h. After resuspending, cells were washed in 500 µl of binding buffer, and the fluorescence was analyzed by a flow cytometry (Beckman Cytoflex) with FlowJo (v10.8) software.

Western blotting

The cultured cells were rinsed with PBS and lysed in Thermo Scientific's radioimmunoprecipitation assay (RIPA) buffer

with protease inhibitor cocktail. Protein concentrations were determined using a Bio-Rad protein assay. Protein samples (30 μg/well) were used for electrophoresis in sodium dodecylsulfate (SDS)/10% polyacrylamide gels. Proteins then were electrophoretically transferred to 0.22-μm nitrocellulose (NC) membranes using a Bio-Rad Western blot transfer system for 15 min at 25 V. The NC membranes were washed with Tris-buffered saline (TBST) containing 0.5% Tween 20 and blocked with 5% BSA by incubating for 1 h at room temperature. Thereafter, the membranes were washed three times with TBST and incubated with rabbit polyclonal primary antibody specific to XRCC5 or/and GRP78 diluted 1:1000 overnight at 4°C. After being washed, membranes were incubated with goat anti-rabbit HRP-conjugated secondary antibody (Life Technologies) and diluted 1:1000 for 1 h at room temperature, followed by detection using ChemiDoc XRS+ (Bio-Rad).

Pull-down assays

HepG2 cells (3×10^5 cells) were detached using enzyme-free cell dissociation buffer and mixed with 6 ml of precooling hypotonic buffer on ice. Cytosolic and nuclear proteins were removed by the centrifugation. Cell debris containing the membrane protein fractions was further treated with 6 ml of lysis buffer. Finally, crude membrane fractions were centrifuged at $4000 \times g$ to remove insoluble impurities. The resulting supernatant was blocked with BSA and tRNA. Then, the membrane protein solution was incubated with 600 pmol of biotin-conjugated AegisBodies and Random sequences, respectively. A volume of 100 μl of streptavidin-coated magnetic beads was added to bind to the DNA–protein complexes for incubation. After washing, the magnetic beads were mixed with 0.1% SDS and then denatured at 95°C. The proteins were separated using a 10% SDS/PAGE gel. Subsequently, the gel was visualized using silver staining.

RNA interference experiments

The lipofectamine 8000 transfection reagent was from Beyotime. The sequences of the small interfering RNA (siRNA) molecules were, for XRCC5, siRNA1 sense were ACUUGCUGUAGAUGAAGAA; siRNA2 sense, CUUCCAGACUAACAACAGA; and a nonspecific siRNA sense, CACGCUCGGUCAAAAGGUUUU, and for XRCC6, siRNA3 sense AGGAAACAGAAGAGCUAAATT; siRNA4 sense, CUUCCAGACUAACAACAGA; and a nonspecific siRNA sense, GGAAAGUUACCAAGAGAAATT. They were obtained from General BIOL. HepG2 cells were transfected with siRNA by Lipofectamine RNAiMAX transfection reagent for 24 h. The cells then were harvested and analyzed by the flow cytometry, and protein extracts were detected by western blot.

Overexpression experiments

Lipofectamine 8000 transfection reagent and the gene sequence of GRP78 (HSPA5) were from Beyotime. HEK293T cells were transfected with transfection reagent for 72 h. The cells then were harvested and analyzed by the flow cytometry.

Study on cell uptake mechanism of LZH5b mediated by XRCC5

The effect of endocytosis inhibitors on HepG2 cell viability

HepG2 cells were seeded in 48-well plates and incubated for 24 h. Different concentrations of endocytosis inhibitors were then added to the cells. The inhibitors used in the experiment included amiloride (250–1000 μM), chlorpromazine (2.5–10 μg/ml), and genistein (20–80 μg/ml). After 6 h of incubation, Cell counting kit-8 (CCK-8) solution, mixed with fluorescence active medium in a 1:10 ratio, was added to the cells and incubated for 1 h. The absorbance at 450 nm was measured using a microplate reader. The formula for calculating relative cell viability was: $[(A_s - A_b)/(A_c - A_b)] \times 100\%$, where A_s, A_c, and A_b represent the absorbance of the sample with inhibitor, the sample without inhibitor, and the blank control, respectively. Each concentration was tested in triplicate.

Effect of endocytosis inhibitors on LZH5b cellular uptake

HepG2 cells were seeded in 48-well plates and allowed to grow to 80%–90% confluency. Different inhibitors were then preincubated with the cells at 37°C for 30 min. Cy5-labeled LZH5b aptamer was added to the medium to a final concentration of 5 μM. After 6 h of incubation, the cells were washed with PBS, trypsinized for 5 min to remove surface-bound aptamers, and centrifuged at 1000 rpm for 3 min. The cells were then resuspended in prechilled DMEM, and the uptake of LZH5b aptamer was detected by fluorescence activated cell sorting (FACS) using a CytoFLEX flow cytometer (Beckman Coulter).

Confocal miscroscopy to determine the time-resolved subcellular localization of LZH5b in cells

Confocal microscopy experiments

HepG2 cells were seeded in 20-mm confocal culture dishes. After 24 h, Cy5-labeled LZH5b was added to the cell culture medium to a final concentration of 5 μM. The cells were incubated at 37°C for 2, 4, 6, and 8 h. Subsequently, the cells were washed three times with PBS. Hoechst 33342 was used for nuclear staining by incubating the cells at 37°C for 20 min, followed by three washes with PBS. LysoTracker Green was used for lysosome staining by incubating the cells at 37°C for 30 min. After removing the supernatant, DMEM medium was added, and the cellular distribution and lysosomal colocalization of Cy5-labeled LZH5b were observed using a Nikon AXR high-resolution laser confocal system.

Results

Development of the experiment

We began with a set of 15 AegisBodies that bound unknown targets on the surface of HepG2 liver cancer cells. These were obtained from 11 rounds of AEGIS–LIVE with a six-letter AEGIS DNA GACT**ZP** (Fig. 1) library containing ~10^{14} components. Selection for HepG2 binders was followed by counterselection against untransformed liver cells [56]. In comparison, standard four-letter LIVE (without FACS, an innovative workflow to get C10) required 19 rounds to create a set of aptamers with affinities ranging from 64 to 349 nM [69]. However, the number of cycles used to generate our AegisBodies was comparable to the number of FACS rounds needed to get aptamer C10. Sequencing was performed by an early version of transliteration .

Two of these AegisBodies, LZH5b and LZH8, had properties similar to those displayed by the standard the C10.36. First, like C10.36, neither had known surface targets. Nevertheless, data showed that LZH5b was internalized like C10.36. LZH8 was visualized as binding to exosomes.

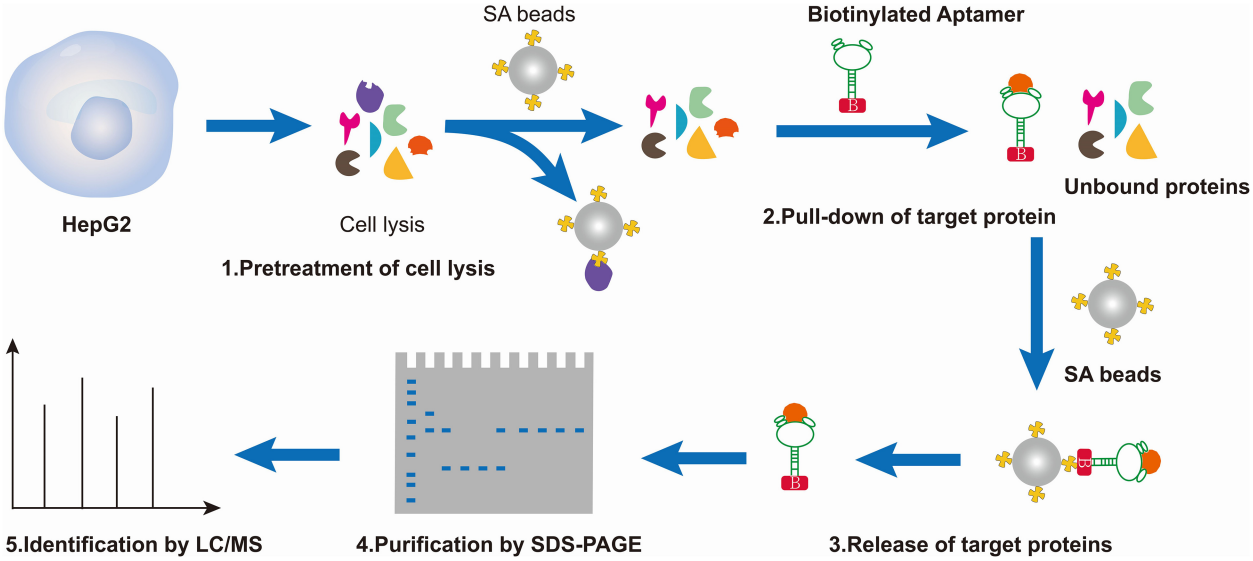


Figure 2. Schematic of the procedure used in the "pull-down" experiments to identify the surface proteins that the LZH5b and LZH8 AegisBodies bound.

Further, while other aptamers in the LZH set included AEGIS **Z**, LZH5b and LZH8 contained only AEGIS **P**, together with the standard four nucleotides. Experiments showed that when the **P** AEGIS base was replaced with A or G, the binding abilities of both AegisBodies were compromised (Supplementary Fig. S1). Thus, AEGIS can contribute to the structure and function of LZH5b and LZH8 only by adding information density, similar to results seen in the AegisBody that binds anthrax protective antigen [64]. They could not gain from added "universal binding" functionality, such as a nitro group or an aliphatic hydrophobic group.

Thus, it appeared appropriate to compare AegisBodies LZH5b and LZH8 with aptamer C10.36 as biotechnological tools. Here, we asked whether LZH5b and LZH8 could be used to discover their protein binding targets on liver cancer cells, just like C10.36 was used to discover its protein binding target on Ramos lymphoma B cells.

LZH5b and LZH8 were used to pull down their cell surface bound targets

To this end, LZH5b and LZH8 were used in separate pull-down experiments by the workflow graphically illustrated in Fig. 2. HepG2 cells were disrupted, and the materials treated with biotinylated AegisBodies. The AegisBody–biotin conjugates were then recovered on streptavidin beads. The material was released, resolved by SDS/polyacrylamide gel electrophoresis, and submitted to proteomic analysis using liquid chromatography and mass spectrometry (LC-MS). Sequences were scored against a sequence database.

Two distinct bands were pulled down from the experiments with AegisBody LZH5b. Their sequence proteomic analysis (Supplementary Table S1) matched the X-ray repair cross complementing proteins 5 and 6 (XRCC5 and XRCC6, also known as Ku80 and Ku70, Fig. 3A). This generated the hypothesis that XRCC5 and/or XRCC6 are/is expressed on the surface of liver cancer cells but not the normal cells used in the counterselection. The XRCC5/Ku80 protein forms part of a Ku heterodimer, which binds to DNA double-strand break ends. It best defined role is thus in the nucleus of the cell, not on its surface. As with the spliceosome target discovered by standard aptamer C10.36, one might (again perhaps naively) expect it to be found in the nucleus, but not on a cell surface, since it helps manage double stranded breaks in nuclear DNA. This surprise is analyzed below.

Only one distinct band was pulled down with LZH8 aptamers. Proteomic analysis identified it as glucose-regulated protein 78 (GRP78, also known as BiP, Fig. 3B). GRP78 is well known in the lumen of the endoplasmic reticulum (ER), where it binds newly synthesized proteins as they are translocated into the ER. It maintains them in a dissolved state competent for subsequent folding and oligomerization. This discovery drove the hypothesis that GRP78 is a second protein that is expressed on liver cancer cells but not on the normal cells used in the counter-selection. Here, finding GRP78 on the surface was perhaps less surprising, since the ER can have direct contact to the cell surface.

Analysis of the discovery of XRCC5 as an AegisBody target

We can compare the performance, as part of a discovery workflow, of the standard four-letter LIVE that generated a standard aptamer C10.36, to that of the AEGIS–LIVE that generated AegisBody LZH5b. In both cases, the target discovered was surprising. Just as one might not *a priori* expect the Ku80 protein (the target for AegisBody LZH5a) to be found on a cell surface, so might one not *a priori* expect a spliceosome complex (the target for aptamer C.10.36) to be on a cell surface.

Thus, a series of experiments was done to ensure that our observation of XRCC5 on the surface of cancer cells was not an artifact. First, HepG2 cells were treated with trypsin for 10 min. Fluorescence flow cytometry was then used with fluorescein-tagged LZH5b AegisBody. Consistent with its binding to protein targets on the surface, LZH5b lost its ability to bind HepG2 cells upon trypsinization (Supplementary Fig. S2).

Separately, commercial anti-XRCC5 antibodies were used to confirm that XRCC5 was indeed on the surface of intact cells. These antibodies fluorescently labeled the HepG2 cells

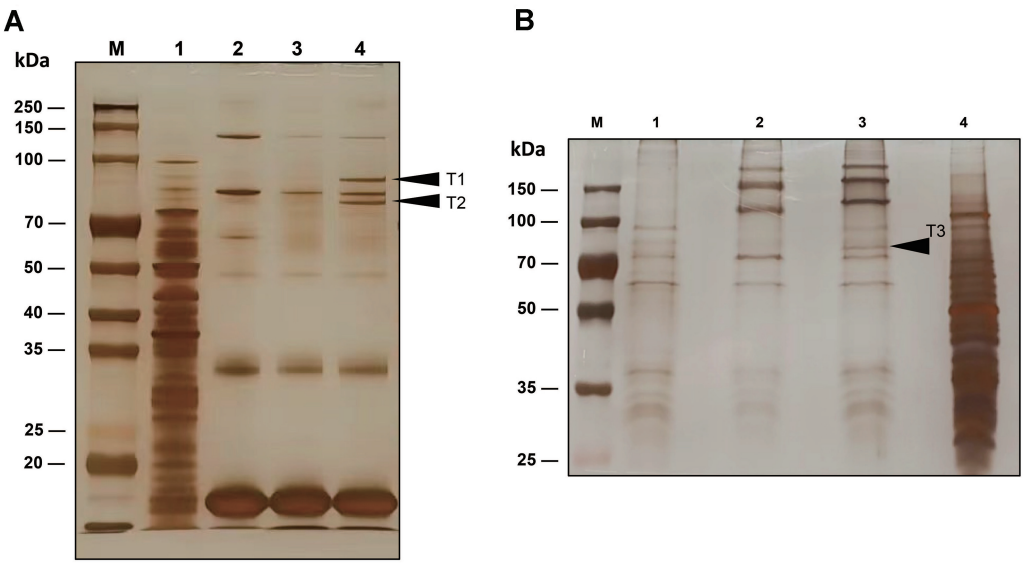


Figure 3. Silver stained SDS/PAGE (10%) was used to identify proteins pulled down by the LZH5b and LZH8 AegisBodies. The black triangles point to the AegisBody bound protein bands. (**A**) Pull-down experiments using AegisBody LZH5b discovered two proteins (T1 and T2) expressed on liver cancer cells but evidently not on untransformed liver cells. Proteomics identified these as X-ray repair cross complementing proteins 5 and 6 (XRCC5 and XRCC6) (Supplementary Table S1). This was confirmed by various follow-on experiments described in the text. Lane M, molecular marker; Lane 1, total protein of HepG2 cell lysis; lane 2, proteins captured by scrambled LZH5b from HepG2 cell lysis; lane 3, beads only; lane 4, proteins captured by AegisBody LZH5b from HepG2 cell lysis. (**B**) Pull-down experiments using AegisBody LZH8 discovered a protein (T3) that was expressed on liver cancer cells but again not on untransformed liver cells. Proteomics identified this band as GRP78 (Supplementary Table S2). Lane M, molecular marker; Lane 1, beads only; lane 2, proteins captured by scrambled LZH8 from HepG2 cell lysis; lane 3, proteins captured by AegisBody LZH8 from HepG2 cell lysis; lane 4, total protein of HepG2 cell lysis.

(Supplementary Fig. S3), a standard proof of the presence of an antibody target on the cell surface. Further, the XRCC5 binding target that was extracted by LZH5b was confirmed by western blotting, using anti-XRCC5 and anti-XRCC6 antibodies, respectively (Supplementary Fig. S4). This competition experiment also offered preliminary evidence that the LZH5b aptamer competes with the XRCC5 antibody on the cell surface (Supplementary Fig. S5), suggesting overlapping binding sites.

Molecular biology approaches were then used to further validate the discovered target. Here, HepG2 cells were transfected with target siRNA molecules to suppress the expression of the putative XRCC5 target. The resulting cell extracts were analyzed by SDS/PAGE using western blotting. These showed that the expression of XRCC5 was downregulated by the siRNA molecule (Fig. 4A). Lowering the level of expression of XRCC5 also diminished AegisBody binding to the cells, as judged by a fluorescence flow cytometry. Lowering of the level of expression of XRCC6 did not (Fig. 4B). This suggests that AegisBody LZH5b interacts with XRCC5 as its primary target, with XRCC6 being pulled down because of the well-known interactions between XRCC5 and XRCC6 [70]. As an additional discovery statement, it appears as if the binding of AegisBody LZH5b to XRCC5 is distant from the epitopes that allow XRCC5 to interact with XRCC6.

The target validation was then completed by measuring the direct interactions between the LZH5b AegisBody and recombinant XRCC5 protein (Fig. 4C) using biolayer interferometry. These experiments gave a dissociation constant of 59.5 ± 10 nM for the complex between AegisZyme LZH5b and isolated XRCC5, while a control sequence showed no binding signal (Supplementary Fig. S6). This was comparable to the affinity of the standard aptamer C10.36 for its target.

This work confirms that the target of AegisBody LZH5b is XRCC5 on the surface of the cells. The associated protein is elevated in HepG2 cells. Interestingly, literature corroborates this. For example, Liu *et al.* reported that high expression of XRCC5 is associated with metastasis through the Wnt signaling pathway and predicts poor prognosis in patients with hepatocellular carcinoma having high levels of this protein [71]. Similar suggestions were made by two online databases: TIMER2.0 (http://timer.comp-genomics.org/timer/) and GEPIA2.0 (http://gepia.cancer-pku.cn/). Here, the expression level of XRCC5 is significantly higher in hepatocellular carcinoma tissues than in adjacent tissues in a single biopsy sample (Supplementary Fig. S7). High expression of XRCC5 gene is negatively correlated with the survival of liver cancer patients (Supplementary Fig. S8). This suggests that AegisBody LZH5b might serve as a biomarker for poor prognosis in this class of liver cancer patients.

Analysis of the discovery of GRP78 as an AegisBody target

The AEGIS–LIVE workflow generates sets of AegisBodies, where each set binds a different protein on a target cell, again avoiding proteins present on the counter-selected cell. Illustrating this for HepG2 cells, the target for AegisBody LZH8, generated in the same cell-AEGIS–LIVE experiment as LZH5b, was used in pull-down experiments (Fig. 2). LZH8 was found to pull down a different protein from HepG2 cells. Proteomics assigned this second target as GRP78 (also known as BiP) (Supplementary Table S2). That this might appear on the surface of a cell is less surprising, as it is a chaperone protein active in the ER.

Both biophysics and cell biology experiments confirmed this assignment. Flow cytometry analysis with a commercial

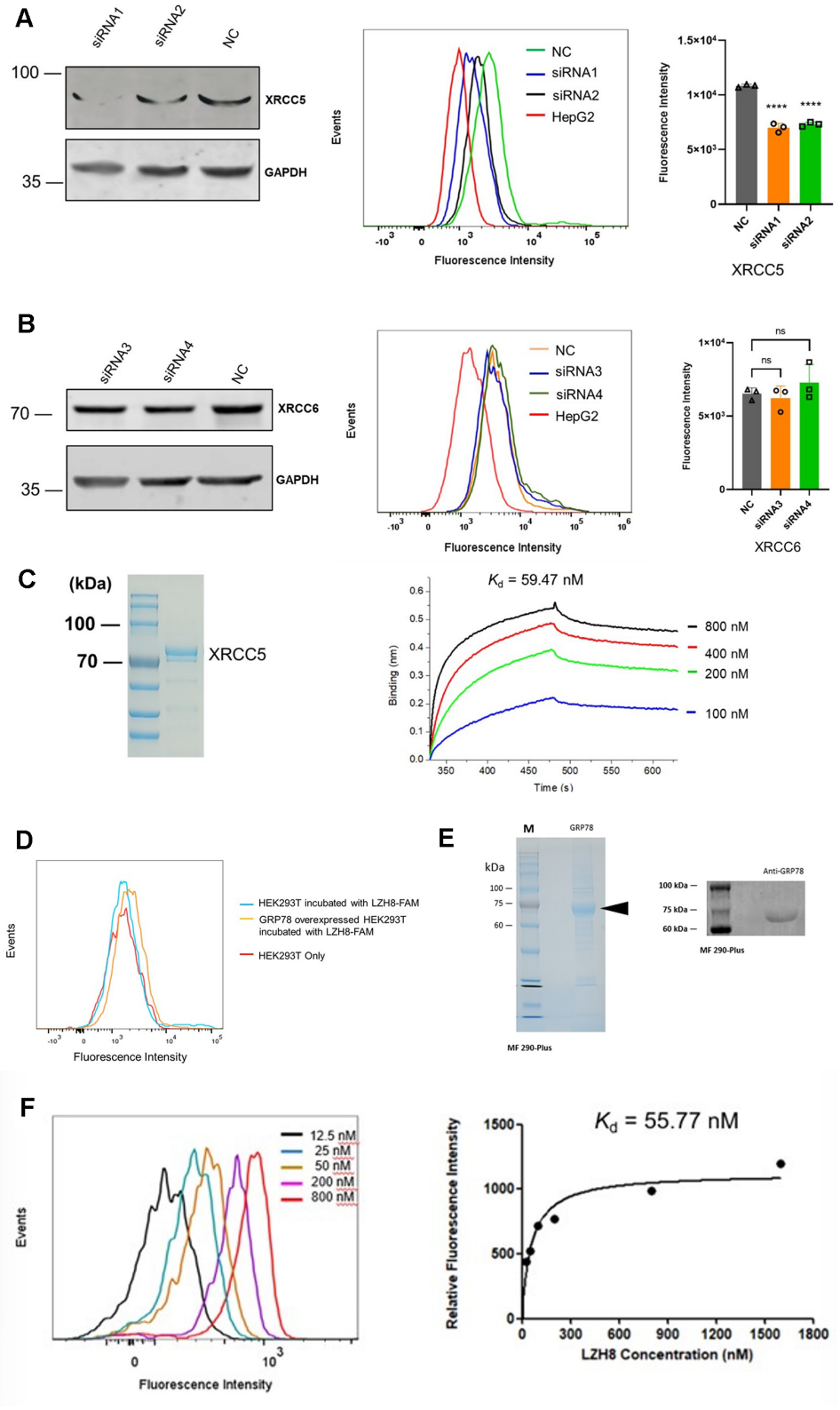


Figure 4. Molecular biology experiments to validate the discovered protein targets of LZH5b and LZH8 AegisBodies. (**A**) HepG2 cells were transfected with the indicated siRNAs. The resulting cell extracts were analyzed using western blot with anti-XRCC5 antibody (left). Flow cytometry results showed that cells down-regulating XRCC5 will results in less internal binding to HepG2 (middle and right). (**B**) HepG2 cells were transfected with the indicated siRNAs. The resulting cell extracts were analyzed using western blotting with anti-XRCC6 antibody (left). Flow cytometry results showed the downregulating XRCC6 did not affect the binding of AegisBody LZH5b to HepG2 cells (middle and right). (**C**) *Escherichia coli* recombinant XRCC5 protein. Coomassie brilliant blue stained SDS/PAGE (10%) was used to analyze the recombinant XRCC5 (left). LZH5b interacted with recombinant XRCC5 measured by biolayer interferometry (right). (**D**) Flow cytometry analysis showing that overexpressing GRP78 results in increased binding of LZH8. (**E**) Eukaryotic expressed GRP78 was analyzed by 10% SDS/PAGE (left). Eukaryotic expressed GRP78 was analyzed by western blotting with anti-GRP78 antibody (right). (**F**) Eukaryotic expressed GRP78 immobilized on beads surface and stained by different concentrations of fluorescence-labeled LZH8, measured by the flow cytometry. The dissociation constant was estimated to be 55.8 nM.

anti-GRP78 antibody showed that GRP78 is expressed on the surface of HepG2 cells (Supplementary Fig. S9). The competitive experiment preliminarily confirmed that the LZH5b aptamer competes with the GRP78 antibody on the cell surface (Supplementary Fig. S10). Overexpressing GRP78 in the HepG2 cells increased the extent to which fluorescently tagged LZH8 AegisBody illuminated the cells (Fig. 4D). As a final target validation, biolayer interferometry measured a dissociation constant of 55 ± 10 nM for GRP78 bound to LZH8 (Fig. 4E and F). This was comparable to the affinity of the standard aptamer C10.36 for its target.

This work confirmed that the target of AegisBody LZH8 is GRP78, a protein that plays roles in maintaining protein stability, regulates protein folding, and induces apoptosis autophagy. Again, GRP78/BiP is connected to liver cancer in other literature [68]. For example, GRP78 has been a target for cancer therapy [72]. Likewise, the databases cited above show that the expression of GRP78 is significantly higher in hepatocellular carcinoma tissues than in adjacent tissues in single biopsies (Supplementary Fig. S11). Expression of the GRP78 gene is inversely correlated with the survival of liver cancer patients (Supplementary Fig. S12). These analyses suggest that GRP78 may also serve as a biomarker for liver cancer.

LZH8 was observed to bind to exosomes. In the ExoCarta exosomes database (exocarta.org/index.html), GRP78 is listed in the exosomes of colorectal cancer cells, breast cancer cells, prostate cancer cells, and medulloblastoma cells. As shown in Supplementary Table S3, GRP78 was mentioned as an exosome protein in 43 papers (2004 to 2013). The vesiclepedia database (http://microvesicles.org/extracellular_vesicle_markers) ranked GRP78 (HSPA5) as one of the top 100 EV proteins, indicating that the target protein of unnatural base containing aptamer LZH8 is overexpressed in the exosomes of cancer cells. Thus, LZH8 may find use as a tool for detect cryptic cancer.

Exploiting the discovery of XRCC5 as a target for LZH5b to deliver drugs to cancer cells

Preliminary studies [20] showed that LZH5b was taken up into the HepG2 cells. Here again, it resembled C10.36, with similar affinity. Thus, the LZH5b AegisBody from AEGIS–LIVE was compared to the C10.36 aptamer from standard LIVE as tools for selective uptake into their respective target cancer cells. To explore entry paths, the effects of several inhibitors known to impact specific endocytosis mechanisms were measured. Thus, the diuretic amiloride inhibits Na+/H + exchange, reduces sub-membrane pH, disrupts Rac1 and Cdc42 signaling, and is known to affect macropinocytosis. In contrast, the antipsychotic chlorpromazine causes the assembly of clathrin lattices on endosomal membranes, preventing the appearance of coated pits on the cell surface, and therefore impacts clathrin-mediated endocytosis. The isoflaone genistein inhibits tyrosine kinase and is an inhibitor of caveolae-mediated endocytosis.

Recognizing that some of these compounds might damage the viability of the target liver cancer cells, we first evaluated the cytotoxicity of these different endocytosis inhibitors using a CCK-8 assay. The results (Supplementary Fig. S13) show that the three concentrations of these compounds used in the uptake experiments did not significantly affect the viability of HepG2 cells.

We then measured the cellular uptake of LZH5b after treatment with these three inhibitors (Supplementary Fig. S14). Both amiloride and genistein modestly inhibited LZH5b uptake, suggesting that LZH5b uptake may rely on macropinocytosis and/or caveolae-mediated endocytosis. Chlorpromazine did not inhibit LZH5b uptake. This is different from the four-letter C10.36 aptamer, which was reported to be taken up by clathrin-mediated endocytosis.

This rationalized our observations that the LZH5b AegisBody can be part of an “AegisBody Conjugated Drug” (ABCD) delivery system to deliver molecules to the inside of cells, one application of the C10.36 aptamer suggested by Mayer *et al*. in their work. In our reported work, LZH5b was coupled to a trigger sequence that initiated the assembly of an AEGIS-containing nanoassembly (Supplementary Fig. S15) [20]. Termed a "nanotrain", this delivery system carried >50 molecules of the anticancer drug doxorubicin into HepG2 cells, selectively killing them. This is proposed to occur by co-internalization of the XRCC5 protein together with the bound ABCD construct covalently attached to a nanotrain that was loaded with doxorubicin. Presumably, entry exploits macropinocytosis and/or caveolae-mediated endocytosis, different from C10.36.

Discussion

An inter-generational effort is required before we can replace, in biotechnology workflows, natural molecular biology by anthropogenic molecular biology. Natural molecular biology emerged historically from a limited set of prebiotic chemical reactions [11, 15], followed by optimization over four billion years within the constraints of Darwinian evolution. Primary among those constraints is the exclusion from Darwinian evolution of "prospective mutations", those that anticipate future fitness outcomes. Darwinian evolution also excludes design, of course.

Prebiotic chemistry was sharply constrained by available geology and energy sources. Thus, there is no reason to expect that the molecular features of modern Terran molecular biology represent optimal solutions to problems posed by life, including the solutions that the Terran biosphere uses to meet the informational needs of Darwinian evolution.

Unlike natural Darwinian evolution, anthropogenic biologists have the benefit of chemical theory. Thus, as we design our own molecular biology, we need not be constrained by the limits of prebiotic chemistry nor non-prospective Darwinian evolution. We can exploit design. We can also explore possibilities more widely in structure “space” than the natural Terran biosphere had the opportunity to explore. Thus, our solutions have the *potential* to be better.

Nevertheless, at first, anthropogenic molecular biology will not perform as well as natural molecular biology. First, our design theory is incomplete. Further, anthropogenic molecular biology must recruit, at least at the outset, enzymes and other tools from natural molecular biology. These enzymes and tools are adapted to work with natural systems, not anthropogenic systems. Thus, even if the anthropogenic systems, in the long run, will prove to be more powerful scaffolds for molecular evolution, campaigns of research in synthetic chemistry, analytical chemistry, and biotechnology are needed to get them to the point where they are selected by biotechnologists over natural systems.

This creates a "resource problem." Biotechnologists expert in natural molecular biology will (and do) object to anthropogenic molecular biology, even when the anthropogenic replacement might be better theoretically. They can argue that because those improvements have *notyetbeen* realized because standard molecular biology tools are not optimal to manage anthropogenic molecules, they *mightneverbe realized*.

This establishes a vicious circle. Resources are needed to improve anthropogenic molecular biology to the point where it can be proven to exploit its intrinsic (but theoretical) benefits over standard molecular biology. But until the anthropogenic molecular biology is proven to be better, resources will not be available to develop the anthropogenic molecular biology.

This resource obstacle must be managed by a research program that has two steps. Step (i) shows that the anthropogenic molecular biology can perform as well as, or somewhat better than, standard molecular biology. Step (ii) uses this showing to recruit resources to anthropogenic molecular biology that can perform better than standard molecular biology.

This paper is presented as part of Step (i). As noted in the introduction, these steps are already somewhat out of order. Anthropogenic molecular biology has already been shown to be better than standard molecular biology in many areas, including human diagnostics [57–60], as a way of creating ultra-clean polymerase chain reactions (PCR) [47], especially in multiplexed form, and catalysts improved by a factor of $\sim 10^5$ [48].

Here, we take step (i) as it relates to a specific class of medical applications, cancer target discovery. We show that AEGIS–LIVE does a bit better than standard LIVE, using an elegant example from Mayer *et al.* [66–68] as a representative of what excellent natural molecular biology applied to four-letter DNA can deliver by way of aptamers.

In this reference example, natural molecular biology delivered, after 10 cycles of FACS–LIVE, a 50 nM actionable aptamer. This was used to identify a molecular target unexpectedly present on the surface of a cancer cell line. It was internalized by recruitment of the clathrin system. That internalization perhaps led to necrotic cell death, perhaps by causing global changes in splicing patterns.

In our case, anthropogenic molecular biology delivered, after 11 cycles, two actionable 50 nM AegisBodies, LZH5b and LZH8. These were used to identify not one, but *two* molecular targets unexpectedly presented on the surface of a cancer cell line. One was internalized by macropinocytosis and/or caveolae-mediated endocytosis to bring a toxic molecule into a cell. Here, the toxin was doxorubicin, and the delivery system was itself an AEGIS construct, an AEGIS nanotrain. This is, of course, a general toxicity mechanism, representing already a benefit of the AEGIS system. Although the cells and the proteins mediating internalization differ, this comparison demonstrates that AegisBodies can achieve the similar goal of intracellular delivery.

So why should we develop AEGIS–LIVE if it performed only modestly better than standard LIVE,? The answer depends on one's vision.

The standard aptamer C10.36 may represent a limit, of a sort, of what can be done with standard four-letter LIVE. It is the product of decades of improvement of the workflow in standard four-letter LIVE.

However, the LZH5b and LZH8 AegisBodies are the products of nascent LIVE technology. They did not exploit the elegant and innovative FACS workflow that discovered C10.36. Nor do they represent the limits of AEGIS as a chemical system. The LZH5b and LZH8 AegisBodies used only a part of AEGIS, just five nucleotides (GACTP). AEGIS **P**, shown to be essential for function by mutation experiments, likely performing by controlling folding, as with the AegisBody raised against anthrax toxin [64]. Neither LZH5b nor LZH8 exploited the ability of AEGIS to deliver single functional groups sparingly, without over-functionalization, neither a nitro group nor a special hydrophobic moiety, which can deliver picomolar affinity.

For that matter, the LZH5b and LZH8 AegisBodies were not obtained by exploiting the generations of tools generated by campaigns in synthesis, enzyme engineering, and analytical chemistry since 2017. These campaigns improved sequencing [73], created polymerases with higher PCR amplification fidelity [45, 74], discovered new AEGIS folds [63], delivered new structural biology [75], improved tautomeric ratios in some in AEGIS components [53], added better predictive thermodynamic parameters [51], and gave new droplet isolation systems for directed evolution [76]. In contrast, the LZH5b and LZH8 AegisBodies were invented with pre-COVID tools. The DNA polymerases then available slowly lost **P**:**Z** pairs. Sequencing of GACT**ZP** DNA products was laborious.

Investment in anthropogenic technologies is justified in two ways: the optimization process in Earth's 4-billion-year natural history, governed by Darwinian evolution, relies on random variations that lack foresight, resulting in only local optimization. Natural DNA remains constrained by its origins in prebiotic chemistry. In contrast, human engineering enables intentional, forward-looking variations. Through synthesis, synthetic biologists can transcend the limitations of prebiotic chemistry and access a new "universe" of evolvable systems.

The choice thus depends on one's assessment of the relative tempo of two processes to create new organic matter with new and desired properties: Darwinian evolution based on systems derived from Darwinian evolution or Darwinian evolution based on systems derived from chemists' design. The data reported here make the case for the second.

Acknowledgements

The authors extend our thanks to the staff at the University of Florida, the Foundation for Applied Molecular Evolution (Alachua FL), the Peking University Medical and Health Analysis Center, and the State Key Laboratory of Natural and Biomimetic Drugs for their invaluable assistance with instrumental analysis.

Author contributions: Conceptualization: S.W., S.A.B., and L.Z. Methodology: S.S., J. Li., S.W., S.A.B., and L.Z. Formal analysis: S.S., J. Li., S.W., and H.X. Investigation: S.S., J. Li., S.W., H.X., J. Long., H.C., T.W., and Q.L. Funding acquisition: D.X., S.A.B., and L.Z. Supervision: S.W., D.X., S.A.B., and L.Z. Validation: S.S., J. Li., and S.W. Visualization: S.S., J.Li., and S.W. Writing—original draft: S.A.B. and L.Z. Writing—review & editing: S.A.B. and L.Z.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

One of the authors (S.A.B.) owns patents covering AEGIS and AEGIS–LIVE.

Funding

This work was supported by the National Key Research and Development Program of China [2023YFC3405100 to D.X. and 2022YFA1304501 to L.Z.]; National Natural Science Foundation of China [22227805, 22374004 to L.Z. and 82303727 to D.X.]; Clinical Medicine Plus X–Young Scholars Project of Peking University, the Fundamental Research Funds for the Central Universities [PKU2024LCXQ026 to D.X. and L.Z.]; Science Foundation of Peking University Cancer Hospital [JC202404 to D.X. and L.Z.]; the United States National Science Foundation (NSF MCB-2123995 to S.A.B.); and the National Institutes of Health [R01GM141391 to S.A.B.]. Funding to pay the Open Access publication charges for this article was provided by the National Institutes of Health.

Data availability

Mass spectrometry proteomics data have been deposited in PRIDE (https://www.ebi.ac.uk/pride/) under accession numbers: PXD055107 (LZH5b-T1), PXD055115 (LZH5b-T2), and PXD055130 (LZH8). Other data will be shared on reasonable request to the corresponding author.

References

1. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment—RNA ligands to bacteriophage-T4 DNA-polymerase. *Science* 1990;**249**:505–10. https://doi.org/10.1126/science.2200121
2. Brody EN, Gold L. Aptamers as therapeutic and diagnostic agents. *Rev Mol Biotechnol* 2000;**74**:5–13. https://doi.org/10.1016/S1389-0352(99)00004-5
3. Lorsch JR, Szostak JW. Chance and necessity in the selection of nucleic acid catalysts. *Acc Chem Res* 1996;**29**:103–10. https://doi.org/10.1021/ar9501378
4. Joyce GF. *Invitro* evolution of nucleic-acids. *Curr Opin Struct Biol* 1994;**4**:331–36. https://doi.org/10.1016/S0959-440X(94)90100-7
5. Cech TR. Self-splicing of group I introns. *Annu Rev Biochem* 1990;**59**:543–68. https://doi.org/10.1146/annurev.bi.59.070190.002551
6. Altman S. A view of RNase P. *Mol Biosyst* 2007;**3**:604–7. https://doi.org/10.1039/b707850c
7. Noller HF, Hoffarth V, Zimniak L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 1992;**256**:1416–19. https://doi.org/10.1126/science.1604315
8. Ban N, Nissen P, Hansen J *et al.* The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 2000;**289**:905–20. https://doi.org/10.1126/science.289.5481.905
9. Rich A. On the problems of evolution and biochemical information transfer. In: *Horizons in Biochemistry*. 1962, 103–26.
10. Benner SA, Kim HJ, Biondi E. Prebiotic chemistry that could not not have happened. *Life* 2019;**9**:84. https://doi.org/10.3390/life9040084
11. Benner SA, Biondi E, Bell EA *et al.* When did life likely emerge on Earth in an RNA-first process? *ChemSystemsChem* 2020;**2**:e1900035. https://doi.org/10.1002/syst.201900035
12. Gilbert W. Origin of life—the RNA world. *Nature* 1986;**319**:618. https://doi.org/10.1038/319618a0
13. Sankaran N. The RNA world at thirty: a look back with its author. *J Mol Evol* 2016;**83**:169–75. https://doi.org/10.1007/s00239-016-9767-3
14. Carrier B, Beaty D, Meyer M *et al.* Mars extant life. What's next? Conference report. *Astrobiology* 2020;**20**:785–814. https://doi.org/10.1089/ast.2020.2237
15. Jerome CA, Kim HJ, Mojzsis SJ *et al.* . Catalytic synthesis of polyribonucleic acid on prebiotic rock glasses. *Astrobiology* 2022;**22**:629–36. https://doi.org/10.1089/ast.2022.00
16. Stumpp MT, Dawson KM, Binz HK. Beyond antibodies: the DARPin® drug platform. *BioDrugs* 2020;**34**:423–33. https://doi.org/10.1007/s40259-020-00429-8
17. Chen HF, Terrett JA. Transient receptor potential ankyrin 1 (TRPA1) antagonists: a patent review (2015–2019). *Expert Opin Ther Pat* 2020;**30**:643–57. https://doi.org/10.1080/13543776.2020.1797679
18. Baker M. Blame it on the antibodies. *Nature* 2015;**521**:274–6. https://doi.org/10.1038/521274a
19. Wang T, Chen CY, Larcher LM *et al.* Three decades of nucleic acid aptamer technologies: lessons learned, progress and opportunities on aptamer development. *Biotechnol Adv* 2019;**37**:28–50. https://doi.org/10.1016/j.biotechadv.2018.11.001
20. Zhang L, Wang S, Yang Z *et al.* An aptamer-nanotrain assembled from six-letter DNA delivers doxorubicin selectively to liver cancer cells. *Angew Chem Int Ed* 2020;**59**:663–68. https://doi.org/10.1002/anie.201909691
21. Acharya VV, Chaudhuri P. Modalities of protein denaturation and nature of denaturants. *IJPSRR* 2021;**69**:19–24. https://doi.org/10.47583/ijpsrr.2021.v69i02.002
22. Katz B, Goldbaum M. Macugen (pegaptanib sodium), a novel ocular therapeutic that targets vascular endothelial growth factor (VEGF). *Int Ophthalmol Clin* 2006;**46**:141–54. https://doi.org/10.1097/01.iio.0000212130.91136.31
23. Patel SS, Lally DR, Hsu JS *et al.* Avacincaptad pegol for geographic atrophy secondary to age-related macular degeneration: 18-month findings from the GATHER1 trial. *Eye* 2023;**37**:3551–7. https://doi.org/10.1038/s41433-023-02548-2
24. Carrigan MA, Ricardo A, Ang DN *et al.* Quantitative analysis of a RNA-cleaving DNA catalyst obtained via *in vitro* selection. *Biochemistry* 2004;**43**:11446–59. https://doi.org/10.1021/bi049898l
25. Benner SA. Detecting darwinism from molecules in the Enceladus plumes, Jupiter's moons, and other planetary water lagoons. *Astrobiology* 2017;**17**:840–51. https://doi.org/10.1089/ast.2016.1611
26. Riccardi C, Napolitano E, Platella C *et al.* G-quadruplex-based aptamers targeting human thrombin: discovery, chemical modifications and antithrombotic effects. *Pharmacol Therapeut* 2021;**217**:107649. https://doi.org.10.1016/j.pharmthera.2020.107649
27. Wang YJ, Ng N, Liu EK *et al.* Systematic study of constraints imposed by modified nucleoside triphosphates with protein-like side chains for use in selection. *Org Biomol Chem* 2017;**15**:610–8. https://doi.org/10.1039/C6OB02335E
28. Wolk SK, Mayfield WS, Gelinas AD *et al.* Modified nucleotidesmay have enhanced early RNA catalysis. *Proc Natl Acad Sci USA* 2020;**117**:8236–42. https://doi.org/10.1073/pnas.1809041117
29. Kimoto M, Nakamura M, Hirao I. Post-ExSELEX stabilization of an unnatural-base DNA aptamer targeting VEGF toward pharmaceutical applications. *Nucleic Acids Res* 2016;**44**:7487–94. https://doi.org/10.1093/nar/gkw619
30. Gold L, Ayers D, Bertino J *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLOS ONE* 2010;**5**:e15004. https://doi.org/10.1371/journal.pone.0015004
31. Cox JC, Hayhurst A, Hesselberth J *et al.* Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. *Nucleic Acids Res* 2002;**30**:e108. https://doi.org/10.1093/nar/gnf107

32. Wu D, Gordon CKL, Shin JH *et al.* Directed evolution of aptamer discovery technologies. *Acc Chem Res* 2022;**55**:685–95. https://doi.org/10.1021/acs.accounts.1c00724
33. Battersby TR, Ang DN, Burgstaller P *et al.* Quantitative analysis of receptors for adenosine nucleotides obtained via *in vitro* selection from a library incorporating a cationic nucleotide analog. *J Am Chem Soc* 1999;**121**:9781–9. https://doi.org/10.1021/ja9816436
34. Vaught JD, Bock C, Carter J *et al.* Expanding the chemistry of DNA for *in vitro* selection. *J Am Chem Soc* 2010;**132**:4141–51. https://doi.org/10.1021/ja908035g
35. Zhou C, Avins JL, Klauser PC *et al.* DNA-catalyzed amide hydrolysis. *J Am Chem Soc* 2016;**138**:2106–9. https://doi.org/10.1021/jacs.5b12647
36. Kraemer S, Vaught JD, Bock C *et al.* From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PLoS One* 2011;**6**:e26332. https://doi.org/10.1371/journal.pone.0026332
37. Tolle FB, Brändle GM, Matzner D *et al.* A versatile approach towards nucleobase-modified aptamers. *Angew Chem Int Ed* 2015;**54**:10971–74. https://doi.org/10.1002/anie.201503652
38. Benner SA, Allemann RK, Ellington AD *et al.* Natural-selection, protein engineering, and the last riboorganism—rational model-building in biochemistry. *Cold Spring Harbor Symp Quant Biol* 1987;**52**:53–63. https://doi.org/10.1101/SQB.1987.052.01.009
39. von Krosigk U, Benner SA. Expanding the genetic alphabet: pyrazine nucleosides that support a donor-donor-acceptor hydrogen-bonding pattern. *Helv Chim Acta* 2004;**87**:1299–324. https://doi.org/10.1002/hlca.200490120
40. Wellington KW, Benner SA. Synthesis of aryl C-glycosides via the Heck coupling reaction. *Nucleosides Nucleotides Nucleic Acids* 2006;**25**:1309–33. https://doi.org/10.1080/15257770600917013
41. Jurczyk SC, Kodra JT, Park JH *et al.* Synthesis of 2′-Deoxyisoguanosine 5′-Triphosphate and 2′-Deoxy-5-methylisocytidine 5′-Triphosphate. *Helv Chim Acta* 1999;**82**:1005–15. https://doi.org/10.1002/(SICI)1522-2675(19990707)82:7⟨1005::AID-HLCA1005⟩3.0.CO;2-5
42. Jurczyk SC, Horlacher J, Devine KG *et al.* Synthesis and characterization of oligonucleotides containing 2′-deoxyxanthosine using phosphoramidite chemistry. *Helv. Chimica Acta* 2000;**83**:1517–24. https://doi.org/10.1002/1522-2675(20000705)83:7⟨1517::AID-HLCA1517⟩3.0.CO;2-S
43. Jurczyk SC, Kodra JT, Rozzell JD *et al.* Synthesis of oligonucleotides containing 2′-deoxyisoguanosine and 2′-deoxy-5-methyliso-cytidine using phosphoramidite chemistry. *Helv Chim Acta* 1998;**81**:793–811. https://doi.org/10.1002/hlca.19980810502
44. Singh I, Laos R, Hoshika S *et al.* Snapshots of an evolved DNA polymerase pre- and post-incorporation of an unnatural nucleotide. *Nucleic Acids Res* 2018;**46**:7977–88. https://doi.org/10.1093/nar/gky552
45. Ouaray Z, Benner SA, Georgiadis MM *et al.* Building better polymerases: engineering the replication of expanded genetic alphabets. *J Biol Chem* 2020;**295**:17046–59. https://doi.org/10.1074/jbc.REV120.013745
46. Oh J, Shan Z, Hoshika S *et al.* A unified Watson–Crick geometry drives transcription of six-letter expanded DNA alphabets by RNA polymerase. *Nat Commun* 2023;**14**:8219. https://doi.org/10.1038/s41467-023-43735-9
47. Yang ZY, Chen F, Alvarado JB *et al.* Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J Am Chem Soc* 2011;**133**:15105–12. https://doi.org/10.1021/ja204910n
48. Wang B, Bradley KM, Kim MJ *et al.* Enzyme-assisted high throughput sequencing of an expanded genetic alphabet at single base resolution. *Nat Commun* 2024;**15**:4057. https://doi.org/10.1038/s41467-024-48408-9
49. Thomas CA, Craig JM, Hoshika S *et al.* Assessing readability of an 8-letter expanded deoxyribonucleic acid alphabet with nanopores. *J Am Chem Soc* 2023;**145**:8560–68. https://doi.org/10.1021/jacs.3c00829
50. Hoshika S, Leal NA, Kim MJ *et al.* Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science* 2019;**363**:884–7 https://doi.org/10.1126/science.aat0971
51. Pham TM, Miffin T, Sun HY *et al.* DNA structure design is improved using an artificially expanded alphabet of base pairs including loop and mismatch thermodynamic parameters. *ACS Synth Biol* 2023;**12**:2750–63. https://doi.org/10.1021/acssynbio.3c00358
52. Karalkar NB, Khare K, Molt R *et al.* Tautomeric equilibria of isoguanine and related purine analogs. *Nucleosides Nucleotides Nucleic Acids* 2017;**36**:256–74. https://doi.org/10.1080/15257770.2016.1268694
53. Eberlein L, Beierlein FR, Hommes NJRV *et al.* Tautomeric equilibria of nucleobases in the Hachimoji expanded genetic alphabet. *J Chem Theory Comput* 2020;**16**:2766–77. https://doi.org/10.1021/acs.jctc.9b01079
54. Singh I, Kim MJ, Molt RW *et al.* Structure and biophysics for a six letter DNA alphabet that includes imidazo[1,2-a]-1,3,5-triazine-2(8H)-4(3H)-dione (X) and 2,4-diaminopyrimidine (K). *ACS Synth Biol* 2017;**6**:2118–29. https://doi.org/10.1021/acssynbio.7b00150
55. Shukla MS, Hoshika S, Benner SA *et al.* Crystal structures of 'ALternative isoinformational ENgineered' DNA in B-form. *Phil Trans R Soc B* 2023;**378**:20220028 https://doi.org/10.1098/rstb.2022.0028
56. Zhang L, Yang Z, Sefah K *et al.* Evolution of functional six-nucleotide DNA. *J Am Chem Soc* 2015;**137**:6734–37. https://doi.org/10.1021/jacs.5b02251
57. Elbeik T, Markowitz N, Nassos P *et al.* Simultaneous runs of the Bayer VERSANT HIV-1 version 3.0 and HCV bDNA version 3.0 quantitative assays on the system 340 platform provide reliable quantitation and improved work flow. *J Clin Microbiol* 2004;**42**:3120–27. https://doi.org/10.1128/JCM.42.7.3120-3127.2004
58. Elbeik T, Surtihadi J, Destree M *et al.* Multicenter evaluation of the performance characteristics of the Bayer VERSANT HCV RNA 3.0 assay (bDNA). *J Clin Microbiol* 2004;**42**:563–69. https://doi.org/10.1128/JCM.42.2.563-569.2004
59. Nolte FS, Marshall DJ, Rasberry C *et al.* MultiCode-PLx system for multiplexed detection of seventeen respiratory viruses. *J Clin Microbiol* 2007;**45**:2779–86. https://doi.org/10.1128/JCM.00669-07
60. Lee WM, Grindle K, Pappas T *et al.* High-throughput, sensitive, and accurate multiplex PCR-microsphere flow cytometry system for large-scale comprehensive detection of respiratory viruses. *J Clin Microbiol* 2007;**45**:2626–34. https://doi.org/10.1128/JCM.02501-06
61. Chaput JC, Switzer C. A DNA pentaplex incorporating nucleobase quintets. *Proc Natl Acad Sci USA* 1999;**96**:10614–19. https://doi.org/10.1073/pnas.96.19.10614
62. Hoshika S, Singh I, Switzer C *et al.* . "Skinny" and "fat" DNA: two new double helices. *J Am Chem Soc* 2018;**140**:11655–60. https://doi.org/10.1021/jacs.8b05042
63. Wang B, Rocca JR, Hoshika S *et al.* A folding motif formed with an expanded genetic alphabet. *Nat Chem* 2024; **16**:1715–22. https://doi.org/10.1038/s41557-024-01552-7
64. Biondi E, Lane JD, Das D *et al.* Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res* 2016;**44**:9565–77. https://doi.org/10.1093/nar/gkw890
65. Jerome CA, Hoshika S, Bradley KM *et al. In vitro* evolution of ribonucleases from expanded genetic alphabets. *Proc Natl Acad Sci USA* 2022;**119**:e2208261119. https://doi.org/10.1073/pnas.2208261119

66. Raddatz MSL, Dolf A, Endl E *et al.* Enrichment of cell-targeting and population-specific aptamers by fluorescence-activated cell sorting. *Angew Chem Int Ed* 2008;**47**:5190–93. https://doi.org/10.1002/anie.200800216
67. Opazo F, Eiden L, Hansen L *et al.* Modular assembly of cell-targeting devices based on an uncommon G-quadruplex aptamer. *Mol Ther Nucleic Acids* 2015;**4**:e251. https://doi.org/10.1038/mtna.2015.25
68. Tonapi SS, Pannu V, Duncan JE *et al.* Translocation of a cell surface spliceosomal complex induces alternative splicing events and lymphoma cell necrosis. *Cell Chem Biol* 2019;**26**:764–56. https://doi.org/10.1016/j.chembiol.2019.02.016
69. Xu JH, Teng IT, Zhang LQ *et al.* Molecular recognition of human liver cancer cells using DNA aptamers generated via cell-SELEX. *PLoS One* 2015;**10**:e0125863. https://doi.org/10.1371/journal.pone.0125863
70. Walker JR, Corpina RA, Goldberg J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 2001;**412**:607–14. https://doi.org/10.1038/35088000
71. Liu ZH, Wang N, Wang FQ *et al.* High expression of XRCC5 is associated with metastasis through Wnt signaling pathway and predicts poor prognosis in patients with hepatocellular carcinoma. *Eur Rev Med Pharmaco* 2019;**23**:7835–47. https://doi.org/10.26355/eurrev_201909_18993
72. Lu G, Luo H, Zhu X. Targeting the GRP78 pathway for cancer therapy. *Front Med* 2020;**7**: 351. https://doi.org/10.3389/fmed.2020.00351
73. Kawabe H, Thomas CA, Hoshika S *et al.* Enzymatic synthesis and nanopore sequencing of 12-letter supernumerary DNA. *Nat Commun* 2023;**14**:6820. https://doi.org/10.1038/s41467-023-42406-z
74. Laos R, Shaw R, Leal NA *et al.* Directed Evolution of Polymerases To Accept Nucleotides with Nonstandard Hydrogen Bond Patterns. *Biochemistry* 2013;**52**:5288–94. https://doi.org/10.1021/bi400558c
75. Hoshika S, Shukla MS, Benner SA *et al.* Visualizing "alternative isoinformational engineered" DNA in A- and B-forms at high resolution. *J Am Chem Soc* 2022;**144**:15603–11. https://doi.org/10.1021/jacs.2c05255
76. Laos R, Benner S. Fluorinated oil-surfactant mixtures with the density of water: artificial cells for synthetic biology. *PLoS One* 2022;**17**:e0252361. https://doi.org/10.1371/journal.pone.0252361

Received: August 14, 2024. **Revised:** January 7, 2025. **Editorial Decision:** January 20, 2025. **Accepted:** January 29, 2025