doi:10.2533/chimia.2026.108 Chimia 80 (2026) 108–119 © M. Rabar, A. Zenz, S. Kath-Schorr*

Adding to the Genetic Script: Extra Letters for New Functions

Matea Rabar, Andre Zenz, and Stephanie Kath-Schorr*

Abstract: Oligonucleotides, both RNA and DNA, are fundamental to life despite being composed of a limited set of simple molecular building blocks. Chemists have long strived to add additional components, especially orthogonal, unnatural base pairs (UBPs). These increase the informational content of nucleic acids and provide site-specific anchors for labelling, enabling applications in aptamer enhancement, RNA structure elucidation, pathway tracing, sequencing, and the construction of semi-synthetic organisms. For this, suitable enzymes and techniques are required to incorporate and later analyse expanded alphabet genetic material. In this review we aim to outline some challenges, achievements, and possibilities that this field encompasses.

Keywords: Expanded alphabet · Nucleic acid labelling · Oligonucleotides · Sequencing - SELEX · Unnatural base pairs

Matea Rabar obtained her Bachelor's degree in chemistry at the University of Zagreb, Croatia, and then transferred to the University of Cologne, Germany, for her Master's studies. She joined Prof. Stephanie Kath-Schorr's research group for her Master's thesis and obtained her MSc degree in 2023. Currently she is a doctoral candidate in the same group, working on synthesis of modified canonical and unnatural nucleotides for postsynthetic labelling of nucleic acids *via* click chemistry.

Andre Zenz started studying chemistry at the University of Cologne, pursuing his BSc with Prof. M. Breugst and afterwards joined the group of Prof. Kath-Schorr for his Master's studies and since 2024 for his doctoral work. He is focused on the modification of oligonucleotides with unnatural building blocks, primarily with novel hydrophobic base pairs and with threose sugars in the backbone. For this he synthesizes both nucleoside triphosphates for enzymatic incorporation and phosphoramidites for solid phase oligonucleotide synthesis.

Stephanie Kath-Schorr received her PhD in 2010 in the laboratory of Thomas Carell. Following postdoctoral research with David Lilley in Dundee UK, she returned to Germany to establish an independent program as a Liebig Fellow of the Fonds der Chemischen Industrie at the LIMES Institute, University of Bonn in 2013. She received the Plus 3 Perspectives Programme award from the Boehringer Ingelheim Foundation in 2018. In 2020 she became Professor of Organic Chemistry at the University of Cologne and since 2025 she has served as Full Professor. Her research focuses on the synthesis and functional analysis of chemically modified ribonucleic acids, including artificial systems with sensing and catalytic functions.

*Correspondence: Prof. Dr. S. Kath-Schorr, E-mail: skathsch@uni-koeln.de
Department of Chemistry and Biochemistry, University of Cologne, Germany

1. Introduction

Nucleic acids, namely DNA and RNA, are fundamental to life and have long been a central focus of modern biotechnology, from the elucidation of the DNA double helix,[1] to sequencing the human genome,[2] and routine genome editing *via* CRISPR (clustered regularly interspaced short palindromic repeats).[3] Despite these transformative advances, fundamental questions persist, particularly regarding the function and regulatory pathways of many RNAs,[4,5] and the molecular ancestry in the context of the origin of life.[6]

The pursuit of answers to these fundamental questions is closely linked to translational applications, including the treatment of cancers,[7] and hereditary diseases,[8] the improvement or fortification of crops,[9] and the development of vaccines against viral infections.[10]

In nature, oligonucleotides are composed of only four canonical bases arranged in two complementary pairs, **C-G** and **T-A** in DNA or **U-A** in RNA. This representation is, however, an oversimplification, as numerous naturally occurring base modifications are known to modulate nucleic acid processing and function as epigenetic markers.[11] Motivated by this inherent chemical plasticity, scientists have long hypothesized that the deliberate introduction of unnatural bases and base pairs into nucleic acids could expand the genetic alphabet, a goal that has been successfully realized in multiple systems over the past decades. Through this alphabet expansion, unnatural nucleotides are placed site-specifically within nucleic acids, that can be designed to convey any kind of desired function that natural nucleotides might lack. Thereby detection of nucleic acids can be simplified or interactions with other molecules tuned for diagnostic and therapeutic purposes. In this review, we highlight key contributions to this field and discuss emerging applications that build upon expanded genetic alphabets.

Several research programs have had a particularly strong impact on the development of UBPs. Steven Benner and his team pioneered base analogues that pair through alternative arrangements of hydrogen bond donors and acceptors, thereby extending the canonical rules of base pairing. In contrast, Ichiro Hirao demonstrated that selective pairing can also be achieved using predomi-

nantly hydrophobic base analogues with complementary shapes. This concept was further advanced by Floyd E. Romesberg who developed hydrophobic UBPs through iterative screening of large libraries and rational optimisation guided by structure-activity-relationships. Beyond these three seminal efforts, important contributions were also made by Eric T. Kool, who established non-hydrogen-bonding base analogues and clarified the role of shape and stacking in replication fidelity.[12,13]

The successful implementation of a new unnatural base pair is a multifaceted endeavour that requires coordinated optimisation across chemical design, enzymatic compatibility, and biological function, key aspects of which are outlined below.

2. Enzymatic Recognition and Processing of UBPs

The development of novel UBPs must be guided by their compatibility with enzymatic processing, most notably by polymerases that are required to recognise and incorporate the corresponding nucleoside triphosphates. Polymerases have evolved to be highly precise in processing natural nucleotides, which innately limits their ability to accommodate unnatural nucleobase analogues. To successfully incorporate a pairing base triphosphate, the resulting base pair geometry has to fit into the polymerase active site. In **A-T** and **C-G**, coplanar geometry is conveyed *via* hydrogen bonding, but overall size and sugar-to-sugar interstrand angles are very similar. Unnatural base pairs that mimic this behaviour by their molecular shape and interaction forces can be processed analogously.[14]

Polymerase performance is not solely determined by base structure but can be strongly influenced by reaction conditions, including buffer composition and the identity and concentration of divalent metal ions such as magnesium, manganese, or calcium.[15,16]

When incorporation is achievable, the primary parameter requiring optimisation is fidelity, defined by the frequency of misincorporation events, while catalytic efficiency and extension speed represent secondary but important considerations. A recurring challenge is the trade-off between fidelity and substrate promiscuity, such that a polymerase optimised for efficient incorporation of one unnatural base pair may be poorly suited for others. Conversely, the identification or engineering of alternative polymerases may in some cases convert a previously low performing unnatural base pair into an enzymatically compatible system.

Unnatural base pairs are often developed *via* iterative improvements starting from a hit in screening combinations of nucleoside candidates / polymerases. Therefore, the three most established efficient UBPs are commonly only used with a small set of matching enzymes. Namely d**Ds**-d**Px** is well incorporated by DeepVent DNA polymerase,[17] or AccuPrime Pfx,[18] d**Z**-d**P** by Taq and Phusion polymerase,[19] and the series of d**TPT3**, d**5SICS**, and d**MMO2**, all pairing to d**NaM** efficiently by DeepVent,[20] OneTaq[21] and T7 RNA polymerase.[22]

As additional examples, T7 RNAP can also process d**s**-d**y**,[23] HIV-1 reverse transcriptase variants have been tested for early variants of Benner's bases[24] and the Kath-Schorr group has investigated reverse transcription of r**TPT3**-r**NaM** using a selection of transcriptases.[25]

It is challenging to give a full overview, since many possible combinations of UBPs and polymerases were either not tested or negative results remain unpublished. Furthermore, the observed fidelity of a UBP can be dependent on the specific experimental modalities and definitive conclusions between different studies are therefore hard to draw.

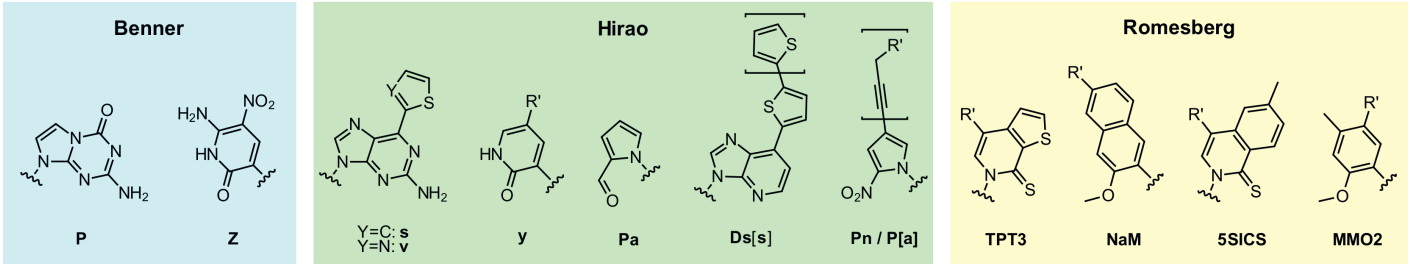
Departing from natural polymerase versions, several polymerases have been specifically engineered to achieve high fidelity with unnatural nucleotides. A widely used strategy is compartmentalised self-replication, in which libraries encoding polymerase variants are expressed and isolated in emulsion droplets together with primers for self-amplification, thereby enriching functional variants over successive selection cycles while poorly performing enzymes are eliminated.[26] The Benner group used this approach for the d**P**-d**Z** pair, which is a UBP structurally similar to **G**-**C** with an altered hydrogen bonding pattern (Scheme 1), and applied primers containing d**P** such that only polymerase variants capable of incorporating d**Z** opposite d**P** could amplify their own genes. This selection strategy led to the identification of Taq DNA polymerase mutants that exhibited reduced pausing when challenged with d**Z** containing templates.[27]

Beyond canonical templated reactions, unnatural base pairs can also be introduced into DNA by non-templated enzymatic processes, thereby avoiding the need to synthesise modified templates by solid-phase synthesis. The Chen group showed that terminal deoxynucleotidyl transferase can extend single stranded DNA at the 3′ end with d**NaM** or d**TPT3**, including derivatives bearing reactive handles, although control over the number of incorporated unnatural nucleotides remained limited.[28] In subsequent work, the authors demonstrated the generation of DNA containing a single unnatural base pair by exploiting polymerase promiscuity through the formation of unnatural base mispairs. This was accomplished either by a gap filling strategy using bacteriophage T4 DNA polymerase (D219A mutant) and T7 ligase,[29] or by a one pot approach based on controlled pausing and restarting of primer extension with the Klenow fragment (exo-) DNA polymerase.[30] These strategies enabled the preparation of UBP containing DNA oligonucleotides without the need for solid-phase synthesis, albeit with the limitation that only a single UBP could be introduced at the position immediately downstream of the primer 3′ end.

Access to a broader range of enzymes capable of processing UBP containing nucleic acids would enable more efficient and versatile manipulation strategies. Accordingly, expansion of the enzymatic toolbox represents a key prerequisite for the widespread application of unnatural base pairs.

3. Emerging Applications of Expanded Genetic Alphabets

Expansion of the genetic alphabet inherently broadens the scope of nucleic-acid-based applications. The most prominent examples include site-specific labelling of nucleic acids, particu-



Scheme 1. Collection of nucleobase analogues. R′ denotes positions that can accommodate additional modifications.

larly RNA, the development of aptamers, and the construction of semi-synthetic organisms, which constitute the primary focus of this review. Additional applications span DNA encryption,[31] biosensing,[32] enhancing DNA nanostructures,[33] detection of epigenetic modifications,[34] identification of DNA lesions,[35] and in diagnostics for detection of viral RNA.[36–38]

3.1 *Site-specific Labelling of Nucleic Acids*

Labelling of nucleic acids (NA) is widely employed in basic research to probe structure and function through structural mapping,[39] analysis of binding interactions with proteins or other NA,[40,41] cellular localisation and quantification,[42] as well as in development of new technologies, for example in next generation sequencing[43] and in real-time polymerase chain reaction (real-time PCR)[44] such as TaqMan and Scorpion probes or molecular beacons. Enzymatic synthesis of modified nucleic acids relying exclusively on canonical nucleobases is generally restricted to random label incorporation or modification at multiple sites within a sequence. In contrast, site-specific labelling can be achieved by solid-phase oligonucleotide synthesis, which is largely limited to short oligonucleotides,[45,46] or by post-synthetic modification using methyltransferases, a strategy with inherently narrow scope.[47] Expansion of the genetic alphabet enables the enzymatic incorporation of UBP analogues at predefined positions within NAs, thereby allowing precise label placement and facilitating studies of a broader range of substrates, including functional RNAs (Fig.1).

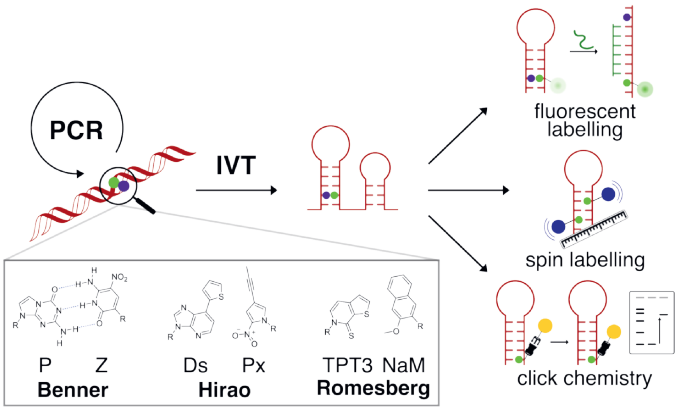


Fig. 1. Site-specific labelling of RNA can be achieved through modification of unnatural base pairs. Key applications include fluorescent and spin labelling for functional and structural studies. Click chemistry provides a versatile strategy for the introduction of larger labels.

The Hirao group demonstrated the utility of their unnatural base pairs for site-specific labelling of both DNA and RNA. In this system, the unnatural base **y** was functionalised with various fluorophores, enabling site-specific RNA labelling by *in vitro* transcription (IVT) from DNA templates containing **s** or **v** (Scheme 1).[48] Using this approach, the anti-(Raf-1) RNA aptamer was fluorescently labelled and its dissociation constant for binding to the Ras binding domain of Raf 1 was determined, illustrating the value of precise fluorescent labelling for detecting aptamer protein interactions. In the same study, site-specific labelling of a theophylline-binding RNA aptamer allowed correlation of fluorescence intensity changes with ligand concentration, reflecting conformational rearrangements upon theophylline binding.

Building on the principle of shape complementarity and following expansion of their UBP system, the Hirao group achieved incorporation of the fluorescent base **s** into RNA by *in vitro* transcription using DNA templates containing **Pa** (Scheme 1).[49] Its fluorescence is sensitive to stacking interactions with neighbouring bases, providing an additional means to monitor structural changes in RNA. A base exhibiting even stronger fluorescence, **Dss** is incorporated opposite **Pa** during replication and transcription with high specificity, yet can pair with any natural base within a duplex, a property that in principle allows fluorescent labelling of duplex nucleic acids of arbitrary sequence by replacement of a natural base with **Dss**.[50] The fluorescence of **Dss** is efficiently quenched upon pairing with the 2-nitropyrrole analogues **Pn** or **Px** (Scheme 1) in a DNA duplex. This property was exploited in molecular beacon designs by placing the **Dss-Pn** pair within the stem region and was further applied in real-time quantitative PCR using the **Dss-Px** pairing system.[51] **Px** also partially quenches the fluorescence of fluorophores such as Cy3 and Cy5 when conjugated *via* its alkynyl side chain. Upon pairing of the modified **Px** opposite **Ds** (Scheme 1) in a DNA duplex during replication, fluorescence increases as a result of diminished quenching, enabling this base pair system to be applied in real-time quantitative PCR as well.[52] While the **Ds-Px** pair offers efficient replication during PCR in DNA, the **Ds-Pa** pair performs better in IVT and was therefore used to introduce a biotin-**Pa** conjugate to a 260 nucleotide (nt) long RNA with selectivity ranging from 72 to 92% depending on the sequence context.[53] The incorporation efficiency of modified **Pa** in IVT was diminished for larger groups such as TAMRA, FAM and digoxigenin.[54]

As an alternative strategy, post transcriptional labelling of alkynyl modified **Pa** bearing either an ethynyl group or a C4 linker was achieved *via* copper-catalysed azide-alkyne cycloaddition (CuAAC) chemistry. Using this approach, a 75 nt tRNA was site-specifically labelled with AF488-, AF594-, biotin- and 3-(7-hydroxycoumarinyl)- azides.[55] Furthermore, an azide-modified **Pa** was incorporated into 17-, 76- and 260-mer RNA by IVT and copper-free post-transcriptional labelling achieved with dibenzocyclooctyne (DIBO) conjugates of AF488, AF594 and biotin.[56] The efficiency of incorporation of more bulky labels (TAMRA, Cy3) during IVT was increased by using the T7 mutant VRS-M5 which also enabled simultaneous incorporation of 2′-fluoro pyrimidine nucleotides.[57]

The Romesberg group achieved site-specific modification of DNA and RNA using the **5SICS-MMO2** UBP system (Scheme 1) by incorporating functionalised **5SICS** and **MMO2** (d)NTPs during PCR and IVT.[20] Functionalisation relied on a propargylamine linker, in which the amine was protected as a dichloroacetate, conjugated to biotin through a polyethylene glycol (PEG) linker or left unprotected. PCR proceeded more efficiently with the protected amine, and subsequent deprotection enabled post synthetic labelling using biotin *N*-hydroxysuccinimidyl (NHS) ester conjugates. It was further observed that the **MMO2** scaffold tolerates chemical modification more readily. *In vitro* transcription with T7 RNA polymerase using modified unnatural ribonucleotides bearing either protected or unprotected propargylamine linkers was also successful, although a modest reduction in transcription fidelity for pairing of d**NaM** with **5SICS** (Scheme 1) was observed when the free amine linker was used.[20] Further studies by the Romesberg group showed that alkyne-modified **5SICS** was efficiently incorporated into RNA opposite **dNaM** during IVT and could be fluorescently labelled post-synthetically *via* copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC).[58] On the other hand, neither propargylamine- or alkyne-modified **NaM** analogues were efficiently transcribed. By combining amine modified **MMO2** with alkyne modified **5SICS**, doubly site-specifically modified RNA was generated enzymatically for the first time. Cy3 and Cy5 fluorophores were conjugated to the resulting RNA *via* NHS amide coupling and CuAAC, respectively, and the labelled constructs were applied in fluorescence resonance energy transfer (FRET) studies of a 243 nt fragment of the central domain of *Thermus thermophilus* 16S rRNA. These experiments revealed conformational changes induced by binding of multiple riboso-

mal proteins, underscoring the power of UBP-enabled site-specific labelling for probing RNA structure and dynamics.[58]

Although CuAAC is widely used for post synthetic labelling of nucleic acids due to its robustness and functional group tolerance,[59] its applicability *in vivo* is limited by the cytotoxicity associated with Cu(I).[60] Consequently, the development of truly biorthogonal labelling strategies is essential for broader biological use of UBP containing constructs. In this context we established a site-specific RNA labelling strategy by employing a cyclopropene or norbonene modified **TPT3** (Scheme 1) and catalyst-free inverse electron demand Diels Alder (IEDDA) ligation to a tetrazine derivative bearing a fluorophore.[61,62] Several RNAs of varying length, notably a 77 nt tRNA, the 89 nt CPEB3 ribozyme, the 185 nt glmS ribozyme, the 401 nt A-region of the long-noncoding RNA Xist, and an mRNA encoding the mCherry protein, were modified with a cyclopropene reactive handle without detectable loss of RNA activity and were subsequently fluorescently labelled *via* IEDDA chemistry.[63,64] For example, the CPEB3 ribozyme retained its self-cleavage activity, and mRNA labelled in the 3′-UTR was translated into the mCherry protein in cells. In this case, cyclopropene modified mCherry mRNA was produced by *in vitro* transcription, transfected into HeLa cells, and labelled intracellularly in live cells by addition of AF488 tetrazine to the culture medium, enabling confocal microscopy visualization of both the mRNA and the expressed protein.[64] In addition to fluorophores, nitroxide spin labels were introduced into RNA, allowing distance measurements between modification sites by electron paramagnetic resonance spectroscopy (EPR).[65,66] Transcription using an expanded genetic alphabet enabled site-specific spin labelling of RNA through incorporation of cyclopropene modified nucleotides followed by tetrazine spin labelling, as well as direct spin labelling using a nitroxide modified **TPT3** analogue, **TPT3^{NO}**. Owing to its increased structural rigidity, **TPT3^{NO}** yielded sharper distance distributions in pulsed EPR experiments compared to the posttranscriptional labelling approach. Using **TPT3^{NO}**, complexly folded RNAs including the 185 nt glmS ribozyme and the 377 nt A-region of the Xist lnc RNA were modified at predetermined positions, and the experimentally determined distances by pulsed EPR were consistent with molecular dynamics simulations.

Fang and coworkers later used a propargylamine-modified **TPT3** and NHS chemistry to label 97 nt 3′SL RNA and a 719 nt mini genomic RNA (DENV-mini) with gold nanoparticles.[67] Interparticle distances measured by X-ray scattering interferometry provided experimental evidence for RNA conformational rearrangements in the 3′ SL element associated with genome circularization mediated by long range 5′-3′ RNA interactions. The authors also employed an acetylene-modified **TPT3** (**TPT3^{CO}**) for spin labelling of the 419 nt ribonuclease P *via* CuAAC with an azide-modified nitroxide spin-label and measured pulsed EPR to obtain distance distributions which were consistent with the crystal structure[68] and enabled long-range measurements up to approximately 14 nm using perdeuterated RNA.[69] The same nucleotide was used for site-specific fluorescent labelling of several exonuclease-resistant RNAs (xrRNAs) to study their folding dynamics by single-molecule FRET. This was achieved by incorporating the **TPT3^{CO}** at the desired position by IVT, introducing the Cy5 dye by CuAAC and annealing the extended 3′-end of the RNA to a DNA oligonucleotide with a 3′-end Cy3 modification and 5′-end biotin.[70] These studies revealed that resistance to exonuclease Xrn1 depends on RNA conformational dynamics modulated by magnesium concentration and the length of pseudoknot 2. In addition, the single-molecule FRET strategy was improved with dual orthogonal labelling by combining propargylamine modified **TPT3** with ethynyl modified **NaM**, which were incorporated into the T99 aptamer domain of a translational T box riboswitch, enabling detailed analysis of its folding energy landscape.[71]

Beyond nucleobase modification, UBPs also enable site-specific labelling at the oligonucleotide backbone through incorporation of thiophosphate nucleotides. Using this strategy, the Romesberg group attached a biotin tag to DNA by using an iodoacetyl-biotin label,[20] while thiol-reactive maleimide probes were used to conjugate gold nanoparticles, spin-labels, and fluorophores for applications including X-ray scattering interferometry, EPR, and single-molecule FRET.[72]

Though labelling of nucleic acids has been developed using a variety of modification strategies, the site-specific labelling was explored only to a limited extent due to the restricted polymerase compatibility and increased synthetic effort to access the modified monomers. Nevertheless, introducing further click reactions or widening the range of modified UBPs may be used to extend the scope of site-specific labelling strategies by genetic alphabet expansion, which provides a unique opportunity for structural studies of NAs of any length, particularly of long RNAs with important cellular functions.

3.2 *In vitro Selection Using Modified Nucleic Acids*

The term aptamer was introduced by Andrew Ellington and Jack Szostak, who demonstrated that specific RNA binders could be isolated from a highly diverse random library of approximately 10^{10} sequences using an iterative *in vitro* selection process.[73]

This methodology is now commonly referred to as systematic evolution of ligands by exponential enrichment (SELEX), a concept independently described in the same year by Tuerk and Gold.[74] In a typical SELEX experiment, a randomised oligonucleotide library is incubated with an immobilised target such as a protein or small molecule. Weakly binding sequences are removed during washing steps, whereas high affinity binders are retained, recovered, and amplified by PCR to regenerate an enriched pool for subsequent rounds of selection (Fig. 2).

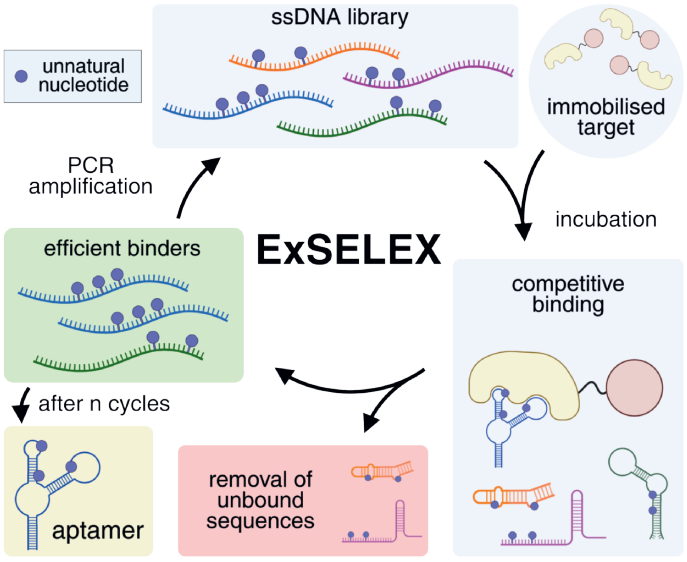


Fig. 2. Typical steps in aptamer selection *via* systematic evolution of ligands by exponential enrichment including an expanded genetic alphabet (ExSELEX).

In its simplest form, DNA libraries can be used for aptamer selection, but RNA and chemically modified nucleic acids (XNA) are also accessible substrates. Selection with RNA or modified nucleic acids requires intermediate transcription and reverse transcription steps between amplification and binding cycles. RNA is often preferred because its greater conformational diversity supports more intricate three-dimensional folds, which can enhance

binding affinity and adaptability in cellular contexts. This reflects a broader consensus that structural plasticity of RNA contributes to the identification of high affinity aptamers in SELEX experiments.

At a fundamental level, target binding arises from optimal balances of electrostatic, steric, and other noncovalent interactions between the folded aptamer and its target.[75]

This restricted chemical space limits the range of functional groups available for target recognition, reducing the likelihood of isolating high affinity binders for structurally complex targets. Incorporation of chemical modifications expands the available interaction chemistries and has been shown to yield aptamers with enhanced binding affinities, increased nuclease resistance, and broader functional capabilities.[76]

One effective strategy to expand aptamer functionality is the chemical modification of individual nucleobases. For example, in a large collaborative effort, aptamers were generated against the majority of human blood protein targets, including many for which selections using natural nucleobases had previously failed, by employing hydrophobically modified uridines to produce slow off rate modified aptamers, known as SOMAmers.[77]

In vitro selection can likewise be directed toward catalytic function, enabling the isolation of DNAzymes and RNAzymes. To further enhance the chemical diversity of DNA libraries for such applications, Perrin and coworkers synthesized four nucleoside analogues that pair analogously to canonical **C-G** and **T-A** base pairs but carry protein-like side chains. These modifications were designed to reduce the dependence of RNA cleaving DNAzymes on divalent metal ions, thereby broadening their catalytic potential.[78]

For *in vitro* selection to be effective, chemical modifications must be faithfully retained during enzymatic amplification and, where required, transcription. At the same time, the limited informational density of nucleic acids, arising from only four canonical nucleobases arranged in two base pairs, results in relatively small energetic differences between alternative folded states. As a consequence, nucleic acid libraries often populate ensembles of rapidly interconverting conformations, of which only a minor fraction adopts the structure required for the desired binding or catalytic function. Expanding the chemical diversity through stable modifications can increase both the structural and energetic differentiation of functional conformers, thereby improving selection efficiency.[79]

The theoretical diversity of a nucleic acid library is given by b^n, where b denotes the number of distinct nucleotides and n is the sequence length, for example $4^{20}=10^{12}$ for a library composed of 20 nt long nucleic acids of only the canonical bases. Expanding the genetic alphabet from two to three base pairs increases b from four to six, resulting in a dramatic rise in sequence space ($6^{20}=10^{15}$). This substantial increase in sequence diversity, combined with the added chemical functionality and site-selectivity provided by unnatural base pairs, renders UBPs particularly attractive for the generation of high affinity and functionally enhanced aptamers.

One notable body of work was reported by Hirao's group, who developed the so called XenoAptamers incorporating the hydrophobic base **Ds**, originally introduced as a component of the d**Ds**-d**Px** UBP system.[18,80] Using this approach, aptamers were selected against the vascular endothelial cell growth factor 165 and interferon-γ with affinities of 0.65 pM and 0.038 nM. A key challenge in this work was that many UBPs are incompatible with standard sequencing methodologies, complicating identification of selected nucleic acid sequences after *in vitro* selection. This limitation was initially addressed by restricting incorporation of modified nucleosides to predefined positions that could be decoded using sequence encoded tags. Remarkably, even aptamers containing no more than three **Ds** residues exhibited affinity enhancements exceeding two orders of magnitude. Subsequent advances in sequencing technologies enabled the use of fully randomized libraries containing UBPs, which allowed further improvements in aptamer affinity.[81] More recently, Hirao and coworkers selected additional XenoAptamers against the A1-domain of the von Willebrand factor as well as against the NS1 proteins of four dengue virus serotypes, obtaining dissociation constants in the picomolar range. In parallel, they analysed the theoretical success probabilities for identifying high affinity binders from these informationally expanded sequence pools, providing quantitative insight into the advantages conferred by genetic alphabet expansion.[82] By further derivatising the **Ds** base within the selected sequences, the binding affinities of the resulting aptamers could be tuned and optimised.[83] Their latest report discusses the effects of the unnatural bases on DNA folding and interaction with the target *via* obtained cryo-electron microscopy (cryo-EM) structures.[84] Therein the authors could pinpoint a site in which the hydrophobic **Px/Pa**' residues directly interact with a hydrophobic pocket in the target. Additionally, the **Ds** bases help to stabilize the structural core of the aptamer to interact efficiently with the target's surface.[84] This is a specific example on how the addition of foreign moieties increases the binding strength and specificity of UBP-containing aptamers in contrast to natural nucleotide limited ones.

In a biotechnological context, aptamers are often regarded as alternatives to antibodies, since they can fold into defined three-dimensional structures that recognize target molecules with affinities and specificities comparable to protein antibodies. Oligonucleotide aptamers can be readily synthesised and chemically modified, and their binding domains can be conjugated to other functional modalities, enabling applications in targeted delivery, sensing, and other engineered functions.

Another major contribution to this area comes from the Benner group, which introduced the d**Z**-d**P** UBP (Scheme 1) as part of a systematic effort to explore alternative hydrogen bonding patterns beyond the canonical **C-G / T-A**.[85] They demonstrated that their six letter artificially expanded genetic information system (AEGIS) can be efficiently amplified by PCR and applied it to the selection of an aptamer binding to a breast cancer cell line with an affinity of approximately 30 nM.[19] A key advance in this work was the development of polymerases capable of faithfully amplifying **GACTZP** libraries, combined with conversion PCR and deep sequencing for sequence identification.[86] Using this platform, the group subsequently selected aptamers targeting an anthrax antigen,[87] and the cancer associated surface protein Glypican 3, which is upregulated in several liver cancers.[88] These aptamers were further assembled into so called nanotrains that can be loaded with the cytotoxic drug doxorubicin and deliver it selectively to cancer cells. In this context, expansion of the genetic alphabet confers additional conformational rigidity, as all-natural nucleic acids cannot hybridise to the expanded sequences, by which unwanted strand invasion and premature drug release is mitigated.[89]

A complementary strategy to increase chemical diversity in aptamer libraries involves the incorporation of clickable handles into nucleic acids, followed by modular attachment of diverse functional groups by click-chemistry during selection (click-SELEX). A key advantage of this approach is the convenient diversification of chemical functionality by varying the click-partners. Mayer *et al*. showed this technique by replacing thymines with ethynyl-uracil (Ed**U**) in the evolution of a DNA aptamer targeting a mutant of green fluorescent protein (Cycle3 GFP) and systematically comparing binding affinities after attachment of six different side chains *via* click-functionalisation.[90] Using a split-click-combine strategy, multiple functional groups could be evaluated in parallel,[91] although all click-sites within a given strand carried the same modification.

It is conceivable to combine click-based diversification with an expanded genetic alphabet by employing click-handle modified,

or directly functionalised, unnatural base pairs such as **TPT3**.[92] This would simultaneously increase informational sequence diversity and chemical diversity within aptamer libraries.

Expanded genetic alphabets can likewise be applied to the evolution of RNA aptamers, which requires a reverse-transcription step prior to cDNA amplification. To address this requirement, our group developed a series of quantitative assays using the **TPT3-NaM** unnatural base pair as a model system, enabling systematic evaluation of reverse-transcription efficiency and sequence fidelity across a range of different reverse-transcriptases.[25]

3.3 *Sequencing of Expanded Genetic Alphabets*

Since its introduction, Sanger sequencing[93] has served as the benchmark for accurate DNA sequencing. The method relies on primer extension in the presence of a low fraction of dye-labelled dideoxynucleoside analogues of the canonical bases, which lack a 3′-hydroxyl group and therefore terminate chain elongation. Separation of the resulting fragments by size by automated capillary electrophoresis with optical detection allows base calling by colour at single nucleotide resolution and yields the full sequence.[94] Over the past decades more advanced sequencing techniques like Illumina-[95], ion-torrent-[96], and nanopore sequencing[97,98] gained widespread adoption owing to reduced cost, high throughput, and seamless integration into automated platforms.[99] These approaches enable deep-sequencing, in which large ensembles of oligonucleotides are read in parallel, providing multiple reads per sequence variant and allowing detection of rare mutations within complex populations. However, the foreign chemical nature of unnatural nucleobases often precludes their direct compatibility with these established sequencing technologies, posing a major challenge for sequence determination of oligonucleotides containing expanded genetic alphabets.

One indirect strategy for sequencing nucleic acids containing UBPs is replacement or conversion PCR, which has been employed by Hirao for their d**Ds**-d**Px**[100] and by Benner for their d**Z**-d**P** pair.[19] In this approach, UBP-containing templates are amplified by PCR in the absence of the corresponding unnatural triphosphates, resulting in characteristic misincorporation of natural nucleotides at positions originally occupied by the UBP. Such misincorporation occurs with reproducible base-specific biases, for example replacement of d**Ds** predominantly by **A**, occasionally by **T**, and rarely by **G**. Upon deep sequencing of the amplified material, positions that originally contained an unnatural base can be identified by these diagnostic substitution patterns, whereas positions containing only natural bases retain high fidelity replication.

Replacement of the unnatural nucleobases depends on the local sequence context, specifically the identities of the two preceding and two succeeding nucleotides. In sequence contexts containing more than two d**Ds**-d**Px** pairs within a six-nucleotide window, PCR amplification was found to be inefficient, highlighting intrinsic context dependent limitations of this conversion-based sequencing approach.[100]

3.3.1 *Nanopore Sequencing of UBPs*

One of the most promising approaches for future sequencing of xenonucleic acids is nanopore sequencing.[101] In this technique, two electrolyte filled chambers are separated by a lipid membrane containing a single transmembrane pore formed by a transporter protein whose diameter accommodates only single stranded nucleic acids. Upon application of an electric potential, buffer ions and nucleic acid strands are driven through the pore, resulting in characteristic modulations of the ionic current (Fig. 3).[102] These current changes depend on the geometry and electronic properties of the nucleobases occupying the pore and are therefore sensitive to sequence composition. Because the measured signal reflects not an individual nucleotide but a local sequence context, base identification relies on computational models trained on reference sequences. This intrinsic sensitivity to chemical and structural features makes nanopore sequencing particularly attractive for reading expanded genetic alphabets. For accurate signal acquisition, it is advantageous to reduce the translocation speed of nucleic acids through the pore, which can be achieved by coupling auxiliary enzymes such as helicases or polymerases to the nanopore to control strand movement.

Nanopore sequencing remains a rapidly evolving field with several conceptual variants under active development. One related approach employs a nuclease positioned upstream of the pore to sequentially cleave individual nucleotides, whose passage through the pore generates characteristic current signals, enabling base identification on a single nucleotide level.[103]

The promise of these approaches is twofold. First, because they do not inherently rely on polymerase mediated processing, unnatural nucleotides can in principle be directly detected (provided appropriate calibration, signal interpretation, and modelling of the recorded data are available). Second, nanopore based methods are intrinsically deep sequencing techniques, enabling quantitative assessment of UBP fidelity and detailed analysis of replacement and misincorporation behaviour by natural nucleotides.

In 2020, Romesberg demonstrated nanopore sequencing of the d**TPT3**-d**NaM** pair using a MspA protein nanopore and Hel308 helicase. In this system, Hel308 drives single stranded DNA through the pore in discrete, stepwise motions, enabling high resolution current measurements.

The authors analysed 65-mer DNA sequences containing a centrally positioned unnatural base pair in two distinct sequence contexts and assessed replication fidelity by PCR.[104] A specifically designed double stranded DNA construct with an overhang region was used to recruit and orient the helicase, such that the

Table 1. Comparison of conversion PCR and nanopore technique to sequence UBPs in DNA.

	Conversion PCR	Nanopore sequencing
Sample preparation	PCR with dye-labelled triphosphates	No additional preparation step
Read length	Typically 100–500 bp	Common long-reads of 10–30 kbp
Detection of UBPs	Indirect *via* probabilistic replacement PCR, requires additional deep-sequencing	Direct recording and recognition via training-data
Hardware requirements	PCR combined with deep-sequencing	Nanopore sequencer
Required knowledge	Context-specific UBP replacement probabilities	Context-specific variation of ion-current by UBP through nanopore
UBP density	Low (due to deconvolution limitations)	Unlimited (provided appropriate training library exists)

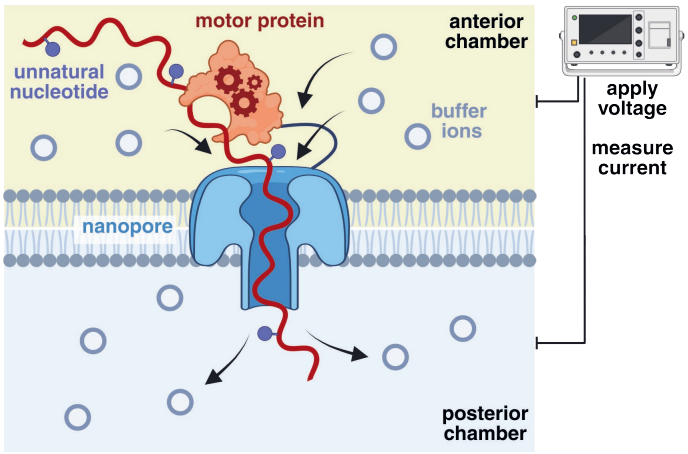


Fig. 3. Overview of a typical nanopore sequencing setup. Nucleic acids are driven through a protein nanopore by an applied voltage across a lipid membrane separating two chambers. The resulting ionic current carried by negatively charged buffer ions is measured. As the nucleic acid translocates, nucleotides of different size and structure partially block the pore in a characteristic manner, generating a sequencing signal. An adenosine triphosphate (ATP) dependent motor protein can control the direction and rate of translocation, slowing movement to enable higher sensitivity and resolution.

motor enzyme pulled the single stranded DNA against the direction of ionic current through the nanopore. Strand translocation was powered by ATP hydrolysis, allowing controlled and reproducible sequencing of the UBP containing region. Very recently Benner reported an application of the same nanopore sequencing architecture to the ALternative Isoinformational ENgineered (ALIEN) DNA system, which comprises the orthogonal d**Z**-d**P** and d**S**-d**B** pairs.[105] Sequencing reliability was systematically evaluated by accounting for the fact that multiple nucleotides simultaneously occupy the sensing region of the pore and collectively modulate the ionic current. Consequently, reference signals had to be acquired for all relevant local sequence contexts spanning four to six nucleotides. Using a training set comprising 256 distinct four nucleotide contexts, correct base calling rates of 63% (**B**), 87% (**P**), 78% (**S**) and 89% (**Z**) were achieved, demonstrating the feasibility and promise of nanopore sequencing for expanded genetic alphabets despite the limited size of the training data.[106]

Another recent study by the Hirao lab employed the commercially available Oxford Nanopore Technologies sequencing platform.[101] A library of 1024 oligonucleotides containing all possible 6-mer sequences with a single UB and a read-splicing based data augmentation technique were used to train the basecaller algorithm, which was able to identify UBs with high accuracy (>80%) and specificity (99%).

A major challenge of the nanopore technology is sequencing of oligonucleotides with multiple UBs in close proximity since the signal data depends on sequence context, therefore an oligonucleotide library with diverse and complex sequences in different contexts is required for basecalling training. While insertion of one UB could be performed enzymatically, sequences with multiple UBs within a short section need to be synthesised chemically due to lower enzyme processivity, which at a higher level of complexity becomes prohibitively expensive. Hirao's work shows that data augmentation achieved by splicing authentic and artificial read data effectively trains basecalling algorithms to recognize expanded genetic alphabets. Continued refinement of these deep learning architectures could facilitate high-accuracy sequencing even when constrained by limited empirical training datasets.

Another challenge arises from the interaction between motor enzyme and UB-containing DNA. For example, premature nanopore or helicase dissociation in regions with high UB density led to many incomplete reads,[107] which in the future may be addressed by enzyme engineering and nanopore design.

Since sequencing *via* replacement PCR always involves a secondary deep-sequencing step, nanopore technology is already a cost-competitive technology today. The biggest hurdle to overcome is the chemical synthesis of UBP-modified nucleic acid libraries and the training of models on these, although software for this has been made publicly available.[101] Once trained, scalability and practicality are the same as for natural oligonucleotides and it seems that nanopore sequencing will only get more affordable and prevalent as time moves on.

3.4 *Development of a Semi-synthetic Organism*

Soon after introducing their respective unnatural base pair systems in the 1990s and early 2000s, Benner and Hirao independently demonstrated that information encoded in an expanded genetic alphabet can be retrieved and translated *in vitro*. In these systems, ribosomal synthesis of proteins containing non-canonical amino acids was achieved with mRNA coding the unnatural information together with chemically synthesized tRNAs bearing UBP-containing anticodons and charged with the desired amino acids.[108,109] However, the construction of a semi-synthetic organism capable of stably maintaining, replicating, and expressing an expanded genetic alphabet across generations required overcoming additional challenges related to nucleotide uptake, replication fidelity, transcription, translation, and long term genetic stability.

First, the information encoded by unnatural base pairs must be replicated by endogenous polymerases with sufficient fidelity, and UBP containing DNA must evade removal by the host DNA repair machinery to ensure stable inheritance of the expanded genetic information. A central challenge in this context is the intracellular availability of the corresponding unnatural nucleoside triphosphates. One strategy to address this limitation is supplementation of the growth medium with the unnatural nucleosides and reliance on the cellular nucleotide salvage pathways to phosphorylate them to the active triphosphate forms. Although the deoxynucleoside kinases are known to have high specificity,[110] Benner[111] and Romesberg[112] showed that unnatural nucleosides could be monophosphorylated by *Drosophila melanogaster* deoxyribonucleoside kinase or engineered variants thereof. However, subsequent phosphorylation steps are inefficient for most unnatural substrates, and efforts to compensate by overexpression of endogenous nucleoside diphosphate kinase in *E. coli* were found to impair cellular growth.[113] Consequently, direct delivery of the corresponding unnatural nucleoside triphosphates to the cell becomes necessary. Because these triphosphates are not membrane permeable, they must be introduced either in a masked form or *via* a dedicated membrane transporter. In addition, the unnatural nucleotides must be well tolerated by the host organism and must not compromise cellular viability or proliferation.

In 2014, the Romesberg group reported the first semi-synthetic organism capable of propagating expanded genetic information encoded in a plasmid.[113] This was achieved in *E. coli* heterologous expression of the algal nucleotide triphosphate transporter *PtNTT2*, which enabled cellular import of the d**5SICS** and d**NaM** triphosphates. Although high retention of the unnatural base pair was observed during both exponential and stationary growth phases, overall cell growth was impaired and prolonged culturing led to loss of the unnatural information due to mispairing events during replication.

Subsequent improvements addressed these limitations by replacing the original pair with the *in vitro* optimised d**TPT3**-d**NaM** system, truncating the *PtNTT2* transporter at the *N*-terminus to

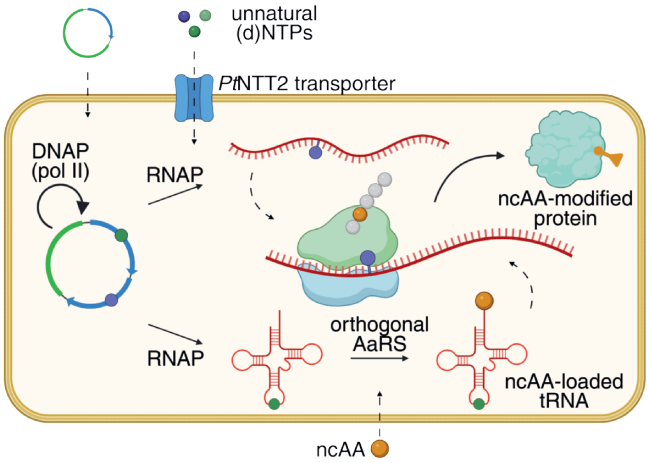


Fig. 4. Components of a semi-synthetic organism. Unnatural base (deoxy)nucleotide triphosphates ((d)NTPs) are introduced into the *E. coli* cell through a transporter and take part in plasmid amplification by endogenous DNA polymerases (DNAP). The unnatural information is transcribed into mRNA and tRNA bearing the UBP-containing codon and anticodon, respectively, by heterologous T7 RNA polymerase (RNAP). The orthogonal aminoacyl-tRNA synthetase (AaRS) charges the tRNA with the non-canonical amino acid (ncAA), which is added onto the growing polypeptide chain to produce the specifically modified protein.

improve performance, and introducing a CRISPR Cas9 based selection mechanism. In this strategy, Cas9 guided by an appropriate single guide RNA selectively eliminated plasmids carrying the most common substitution or deletion mutations at UBP sites, thereby conferring effective immunity against loss of the expanded genetic alphabet.[114] *In vivo* transcription and translation were successfully demonstrated in *E. coli* cells expressing T7 RNA polymerase that carried a plasmid encoding a superfolder GFP gene in which one natural codon was replaced by an unnatural codon, together with a tRNA gene bearing the corresponding anticodon (Fig. 4).

Using established orthogonal aminoacyl-tRNA synthetase systems to charge the modified tRNA, either serine or the non-canonical amino acid *N*-ε-propargyl-L-lysine was incorporated at the predetermined position with high fidelity, and decoding of unnatural codons was shown to be more efficient than conventional amber suppression.[115] This work underscores that expanded genetic alphabets can be translated *in vivo* to produce proteins with new building blocks, completing the information retrieval cycle in a semi-synthetic organism.

Subsequent optimisation of the semi-synthetic organism (SSO) revealed that recombinational repair of stalled replication forks, mediated by RecA, constitutes the primary pathway leading to loss of unnatural base pairs during DNA replication. In contrast, DNA polymerase II was identified as the major contributor to efficient replication of UBPs. Accordingly, deletion of RecA combined with overexpression of polymerase II significantly reduced UBP loss, enabling high level retention of the expanded genetic information not only on plasmids but also within chromosomal DNA across diverse sequence contexts.[116] *In vivo* structure-activity relationship studies brought forward several nucleobases which are more efficiently retained than **NaM**: the monocylic **MMO2**, **ClMO**, **CNMO** and **5FM**, and the benzothiophene derivatives **MTMO** and **PTMO**.[117,118]

These improvements are attributed to subtle differences in hydrophobic surface and packing interactions leading to a reduced tendency towards cross-strand intercalation and enhanced polymerase recognition under cellular replication conditions, even though the pairs still rely on hydrophobic interactions rather than Watson–Crick complementarity.

Further optimisation revealed that the d**CNMO**–d**TPT3** pair exhibits especially high retention *in vivo* and overall performance for storage and retrieval of expanded genetic information, making it one of the best performing pairs identified to date.[119] In parallel, systematic evaluation of transcription identified **NaM** paired with the **TAT1** variant, a **TPT3** analogue with reduced toxicity, as particularly effective for incorporation into RNA, reflecting sequence-dependent biases in polymerase processing of UBPs.[119] These findings underscore that variations in nucleobase substituents can modulate hydrophobic interactions, steric complementarity, and local packing within the developing duplex, thereby influencing both replication fidelity and retention of expanded alphabets *in vivo*. Moreover, the unnatural nucleotides can be used as substrates by endogenous *E. coli* RNA polymerase, which could provide better control of RNA production.[120] Progress toward an eukaryotic semi-synthetic organism was also reported by demonstrating protein expression in cells transfected with UBP-containing mRNA and tRNA together with a plasmid encoding an orthogonal aminoacyl tRNA synthetase.[121] Finally, systematic interrogation of codons containing **NaM** paired with **TPT3** anticodons identified nine codon anticodon combinations that support efficient protein synthesis, three of which functioned orthogonally. This expansion effectively increased the number of usable codons for protein synthesis from 64 to 67.[122]

Production of site-specifically modified proteins in semi-synthetic organisms enables precise interrogation of biological function and has attracted significant interest in drug discovery, particularly in immunology and oncology. A promising candidate is THOR-707,[123] a recombinant human cytokine interleukin (rhIL-2) in which Pro65 is replaced by *N*-ε-azido-L-lysine and site-specifically pegylated using biorthogonal strain promoted azide-alkyne cycloaddition (SPAAC) reaction. In murine models, this modification resulted in an extended circulating half-life relative to unmodified rhIL 2 and a dose dependent reduction in tumour growth. However, phase II clinical trials evaluating THOR-707 in combination with other anticancer agents were discontinued. Following the acquisition of Synthorx by Sanofi, the current clinical pipeline derived from this platform comprises a single Synthorin™ therapeutic, SAR444336, an autoimmune drug candidate presently in phase I clinical trials.[124,125]

Table 2. Challenges and the required advancements to overcome them in the development of the semi-synthetic organism.

challenges	advancements
UBP import	transporter engineering
UBP retention	RecA deletion and pol II overexpression
toxicity of UBPs	improved UBP structure design
retrieval of unnatural information	orthogonal aminoacyl tRNA synthetases

Expansion of the genetic code offers precise control over the modification of ribosomally synthesised polypeptides with far reaching impact for production of therapeutics such as antibody-drug conjugates, antimicrobial peptides and covalent warheads, enabling optimisation of efficacy, pharmacokinetics and safety.[126,127] It could also lead to development of enzymology and biocatalysis through the targeted probing and engineering of enzymatic activity, leading to higher activity, selectivity, and stability.[128] Using the expanded genetic alphabet to expand the genetic code would not only increase the number of available co-

dons to introduce non-canonical amino acids, but also address the issue of lower efficiency of protein production in conventional genetic code expansion strategies due to the competition of engineered components and host machinery. On the other hand, there are several other challenges for scalable semi-synthetic organism production. Industrial applications of the semi-synthetic organism could suffer from higher operational costs due to more complex cell growing conditions, and the observed toxicity of dNTPs which may hinder optimal cell growth and protein production yields. Further development of the semi-synthetic organism could entail engineering endogenous biocatalytic machinery which could synthesise unnatural dNTPs and the non-canonical amino acids from simple starting materials. This would significantly reduce the cost but does surely raise ethical questions as the semi-synthetic organism becomes self-sufficient.

4. Future Directions for UBP Applications

As outlined above, by using semi-synthetic organisms the expansion of chemical space in nucleic acids can be translated into an expanded chemical space in protein in a predictable programmable way. Most commercial pharmaceuticals target protein[129] and a gradually increasing number of them are proteins themselves.[130] Driven by recent advances in delivery mechanisms and the exquisite specificity of protein-target interactions, one can readily envision the immense potential for drug discovery through a pipeline that surveys the vast chemical space accessible to proteins incorporating unnatural amino acids synthesized *via* SSOs. Bacterial synthesis of drugs is commonplace and can be cost-effective,[131] but to include SSOs in such processes, their robustness must certainly be improved upon.

Another application in which UBPs may emerge as an important factor is gene editing with CRISPR/Cas systems. Since CRISPR/Cas cuts dsDNA programmed by guide RNA, target-guide combinations containing UBPs might be able to markedly decrease off-target cutting. Alternatively, the technology might be used to introduce UBP-modified genes. These could then, for example, be efficiently targeted by modified siRNA. Paradoxically, the comparatively limited fidelity of current UBPs might be interesting for gene-therapy, in which the permanent manipulation of a patient's genome is cause for safety concerns. If an introduced UBP can stop genetic variants from expressing their pathologic phenotype, but are removed by natural repair mechanisms over time, this could result in a 'softer' version of gene therapy.

Although guide RNAs are often modified on the backbone, primarily to increase stability, introducing unnatural bases may also be advantageous. As aptamer design has already shown, UBs enhance the binding of oligonucleotides to proteins, and therefore inclusion of UBs in the guide RNA could prove to be beneficial for its binding to the Cas protein. Moreover, guide RNAs can include aptamer insertions which can bind other proteins or fluorophores (*e*.g. Broccoli aptamer),[132,133] which could be expanded by use of UBs. Similarly, the scope of fluorescent labelling of guide RNA,[134] which is used for mechanistic studies of chromosomal dynamics and genomic mapping, and of optical or chemical control of gene editing,[135–138] which facilitates dynamic regulation of gene expression, could be expanded by use of unnatural base pairs.

5. Conclusions

Over the past three decades, the development of unnatural base pairs has transformed nucleic acids from information carriers with fixed chemical composition into programmable polymers with tuneable information and chemical diversity. What began as a conceptual exploration of alternative base pairing principles has matured into a robust toolkit that supports enzymatic replication, transcription, translation, sequencing, and functional deployment in increasingly complex biological contexts.

Expanded genetic alphabets have enabled precise site-specific labelling of DNA and RNA, overcoming long standing limitations of random modification and solid phase synthesis, and have opened new avenues for structural biology, single molecule spectroscopy, and in cell imaging. In parallel, their integration into *in vitro* selection strategies has yielded aptamers with markedly enhanced affinities and functional performance, illustrating how increased sequence space and chemical diversity directly translate into improved molecular recognition. Advances in sequencing, particularly nanopore-based approaches, are beginning to close the remaining technological gap by enabling direct readout of expanded alphabets, thereby completing the experimental cycle from design and selection to analysis.

Most notably, the construction of semi-synthetic organisms capable of maintaining, expressing, and translating expanded genetic information demonstrates the feasibility of operating expanded alphabets in living cells, leading to establishment of frameworks with substantial commercialisation potential, and considerable space for further development.

Taken together, these advances establish unnatural base pairs as a mature and versatile platform for chemical biology and synthetic biology. Continued progress in enzyme engineering, cellular compatibility, and sequencing accuracy is expected to further extend the molecular language of life beyond its natural limits and enable increasingly sophisticated control over biological information and function.

Key challenges remain: At the molecular level, polymerases and other enzymes must be further optimised to support long-term, high-fidelity processing of UBPs across diverse sequence contexts. In parallel, sequencing technologies need to mature toward routine, direct, and high-accuracy readout of expanded alphabets. At the cellular level, improved nucleotide uptake, metabolism, and broader host compatibility will be essential, particularly for extending semi-synthetic organisms beyond bacterial systems.

Looking forward, the major opportunity lies in moving beyond incremental extension toward purposeful design of genetic systems with new functions, including applications that extend *e.g.* into regenerative medicine. Expanded genetic systems may enable regenerative strategies that surpass current biomaterials and gene therapy approaches by supporting programmable control of cell fate, dynamic tissue patterning, and orthogonal signalling networks insulated from endogenous regulation. More broadly, expanded alphabets offer the potential to encode new chemical functionality, orthogonal regulatory layers, and programmable behaviours inaccessible to natural nucleic acids. As these capabilities mature, unnatural base pairs may redefine how biological information is stored, processed, and exploited, paving the way for molecular systems and applications beyond those permitted by the natural genetic code.

Acknowledgements

M. R. thanks the Fonds of the Chemical Industry for a Kekule PhD fellowship.

Author Contributions

M. R. and A. Z. contributed equally. M. R. and A. Z. prepared figures, M. R., A. Z. and S. K.-S. wrote the manuscript.

Received: January 15, 2026

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737, https://doi.org/10.1038/171737a0.
[2] International Human Genome Sequencing Consortium, *Nature* **2001**, *409*, 860, https://doi.org/10.1038/35057062.
[3] M. Jinek, K. Chylinski, I Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, *Science* **2012**, *337*, 816, https://doi.org/10.1126/science.1225829.

[4] L.-L. Chen, V. N. Kim, *Cell* **2024**, *187*, 6451, https://doi.org/10.1016/j.cell.2024.10.024.
[5] K. Nemeth, R. Bayraktar, M. Ferracin, G. A. Calin, *Nat. Rev. Genet*. **2024**, *25*, 211, https://doi.org/10.1038/s41576-023-00662-1.
[6] N. Kitadai, S. Maruyama, *Geosci. Front*. **2018**, *9*, 1117, https://doi.org/10.1016/j.gsf.2017.07.007.
[7] D. Bartolucci, A. Pession, P. Hrelia, R. Tonelli, *Pharmaceutics* **2022**, *14*, 1453, https://doi.org/10.3390/pharmaceutics14071453.
[8] M. Kaltak, P. de Bruijin, D. Piccolo, S.-E. Lee, K. Dulla, T. Hoogenboezem, W. Beumer, A. R. Webster, R. W. J. Collin, M. E. Cheetham, G. Platenburg, J. Swildens, *Mol. Ther.-Nucleic Acids* **2022**, *31*, 674, https://doi.org/10.1016/j.omtn.2023.02.020.
[9] C. Gao, *Cell* **2021**, *184*, 1621, https://doi.org/10.1016/j.cell.2021.01.005.
[10] E. Fang, X. Liu, M. Li, Z. Zhang, L. Song, B. Zhu, X. Wu, J. Liu, D. Zhao, Y. Li, *Signal Transduction Targeted Ther*. **2022**, *7*, 94, https://doi.org/10.1038/s41392-022-00950-y.
[11] W.-Y. Meng, Z.-X. Wang, Y. Zhang, Y. Hou, J.-H. Xue, *J. Biol. Chem*. **2024**, *300*, 106791, https://doi.org/10.1016/j.jbc.2024.106791.
[12] J. C. Morales, E. T. Kool, *Nat. Struct. Mol. Biol*. **1998**, *5*, 950, https://doi.org/10.1038/2925.
[13] T. W. Kim, E. T. Kool, *J. Org. Chem*. **2005**, *70*, 2048, https://doi.org/10.1021/jo048061t.
[14] A. Marx, K. Betz, *Chem. Eur. J*. **2020**, *26*, 3446, https://doi.org/10.1002/chem.201903525.
[15] A. Vaisman, H. Ling, R. Woodgate, W. Yang, *EMBO J*, **2005**, *24*, 2957, https://doi.org/10.1038/sj.emboj.7600786.
[16] W.-J. Wu, W. Yang, M.-D. Tsai, *Nat. Rev. Chem*. **2017**, *1*, 0068, https://doi.org/10.1038/s41570-017-0068.
[17] I. Okamoto, Y. Miyatake, M. Kimoto, I. Hirao, *ACS Synth. Biol*. **2016**, *5*, 1220, http://doi.org/10.1021/acssynbio.5b00253.
[18] R. Yamashige, M. Kimoto, Y. Takezawa, A. Sato, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res*. **2012**, *40*, 2793, https://doi.org/10.1093/nar/gkr1068.
[19] Z. Yang, F. Chen, J. B. Alvarado, S. A. Benner, *J. Am. Chem. Soc*. **2011**, *133*, 15105, https://doi.org/10.1021/ja204910n.
[20] Y. J. Seo, D. A. Malyshev, T. Lavergne, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc*. **2011**, *133*, 19878, https://doi.org/10.1021/ja207907d.
[21] K. Dhami, D. A. Malyshev, P. Ordoukhanian, T. Kubelka, M. Hocek, F. E. Romesberg, *Nucleic Acids Res*. **2014**, *42*, 10235, https://doi.org/10.1093/nar/gku715.
[22] Y. J. Seo, S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc*. **2009**, *131*, 5046, https://doi.org/10.1021/ja9006996.
[23] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa, S. Yokoyama, *Nat. Biotechnol.* **2002**, *20*, 177, https://doi.org/10.1038/nbt0202-177.
[24] A. M. Sismour, S. Lutz, J.-H. Park, M. J. Lutz, P. L. Boyer, S. H. Hughes, S. A. Benner, *Nucleic Acids. Res*. **2004**, *32*, 2, https://doi.org/10.1093/nar/gkh241.
[25] E. S. Hoffmann, M. C. De Pascali, L. Neu, C. Domnick, A. Solda, S. Kath-Schorr, *RSC Chem. Biol*. **2024**, *5*, 556, https://doi.org/10.1039/d4cb00084f.
[26] F. J. Ghadessi, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. U. S. A*. **2001**, *98*, 4552, https://doi.org/10.1073/pnas.071052198.
[27] R. Laos, R. Shaw, N. A. Leal, E. Gaucher, S. Benner, *Biochemistry* **2013**, *52*, 5288, https://doi.org/10.1021/bi400558c.
[28] G. Wang, C. He, J. Zou, J. Liu, Y. Du, T. Chen, *ACS Synth. Biol*. **2022**, *11*, 4142, https://doi.org/10.1021/acssynbio.2c00456.
[29] J. Bai, J. Zou, Y. Cao, Y. Du, T. Chen, *ACS Synth. Biol*. **2023**, *12*, 2676, https://doi.org/10.1021/acssynbio.3c00250.
[30] Y. Cao, J. Bai, J. Zou, Y. Du, T. Chen, *ACS Synth. Biol*. **2023**, *12*, 2691, https://doi.org/10.1021/acssynbio.3c00258.
[31] X. Huang, Z. Hou, W. Qiang, H. Wang, X. Wang, X. Chen, X. Hu, J. Dai, L. Li, G. Zhao, *Natl. Sci. Rev*. **2025**, *12*, nwae469, https://doi.org/10.1093/nsr/nwae469.
[32] S. Manna, M. Kimoto, J. Truong, P. Bommisetti, A. Peitz, I. Hirao, M. C. Hammond, *ACS Sens*. **2023**, *8*, 4468, https://doi.org/10.1021/acssensors.3c01266.
[33] Q. Liu, G. Liu, T. Wang, J. Fu, R. Li, L. Song, Z.-G. Wang, B. Ding, F. Chen, *ChemPhysChem* **2017**, *18*, 2977, https://doi.org/10.1002/cphc.201700809.
[34] D. Schmidl, S. M. Becker, J. M. Edgerton, S. Balasubramanian, *Nat. Chem*. **2025**, 1, https://doi.org/10.1038/s41557-025-01925-6.
[35] J. Riedl, Y. Ding, A. M. Fleming, C. J. Burrows, *Nat. Commun*. **2015**, *6*, 8807, https://doi.org/10.1038/ncomms9807.
[36] L. G. Glushakova, B. W. Alto, M.-S. Kim, D. Hutter, A. Bradley, K. M. Bradley, N. D. Burkett-Cadena, S. A. Benner, *BMC Infect. Dis.* **2019**, *19*, 418, https://doi.org/10.1186/s12879-019-3998-z.
[37] L. G. Glushakova, N. Sharma, S. Hoshika, A. C. Bradley, K. M. Bradley, Z. Yang, S. A. Benner, *Anal. Biochem*. **2015**, *489*, 62, https://doi.org/10.1016/j.ab.2015.08.015.
[38] L. G. Glushakova, A. Bradley, K. M. Bradley, B. W. Alto, S. Hoshika, D. Hutter, N. Sharma, Z. Yang, M.-J. Kim, S. A. Benner, *J. Virol. Methods* **2015**, *214*, 60, https://doi.org/10.1016/j.jviromet.2015.01.003.
[39] B. Strauss, A. Nierth, M. Singer, A. Jäschke, *Nucleic Acids Res*. **2012**, *40*, 861, https://doi.org/10.1093/nar/gkr733.
[40] J. M. Pagano, C. C. Clingman, S. P. Ryder, *RNA* **2011**, *17*, 14, https://doi.org/10.1261/rna.2428111.
[41] F. D. Steffen, M. Khier, D. Kowerko, R. A. Cunha, R. Börner, R. K. O. Sigel, *Nat. Commun*. **2020**, *11*, 104, https://doi.org/10.1038/s41467-019-13683-4.
[42] X. Long, J. Colonell, A. M. Wong, R. H. Singer, T. Lionnet, *Nat. Methods* **2017**, *14*, 703, https://doi.org/10.1038/nmeth.4309.
[43] J. Shendure, H. Ji, *Nat. Biotechnol*. **2008**, *26*, 1135, https://doi.org/10.1038/nbt1486.
[44] I. M. Artika, Y. P. Dewi, I. M. Nainggolan, J. E. Siregar, U. Antonjaya, *Genes* **2022**, *13*, 2387, https://doi.org/10.3390/genes13122387.
[45] L. Flemmich, R. Bereiter, R. Micura, *Angew. Chem. Int. Ed*. **2024**, *63*, e202403063, https://doi.org/10.1002/anie.202403063.
[46] H. Depmeier, E. Hoffmann, L. Bornewasser, S. Kath-Schorr, *ChemBioChem* **2021**, *22*, 2826, https://doi.org/10.1002/cbic.202100161.
[47] T. R. Fischer, L. Meidner, M. Schwickert, M. Weber, R. A. Zimmermann, C. Kersten, T. Schirmeister, M. Helm, *Nucleic Acids Res.* **2022**, *50*, 4216, https://doi.org/10.1093/nar/gkac224.
[48] R. Kawai, M. Kimoto, S. Ikeda, T. Mitsui, M. Endo, S. Yokoyama, I. Hirao, *J. Am. Chem. Soc*. **2005**, *127*, 17286, https://doi.org/10.1021/ja0542946.
[49] M. Kimoto, T. Mitsui, Y. Harada, A. Sato, S. Yokoyama, I. Hirao, *Nucleic Acids Res*. **2007**, *35*, 5360, https://doi.org/10.1093/nar/gkm508.
[50] M. Kimoto, T. Mitsui, S. Yokoyama, I. Hirao, *J. Am. Chem. Soc*. **2010**, *132*, 4988, https://doi.org/10.1021/ja100806c.
[51] M. Kimoto, T. Mitsui, R. Yamashige, A. Sato, S. Yokoyama, I. Hirao, *J. Am. Chem. Soc*. **2010**, *132*, 15418, https://doi.org/10.1021/ja1072383.
[52] R. Yamashige, M. Kimoto, T. Mitsui, S. Yokoyama, I. Hirao, *Org. Biomol. Chem*. **2011**, *9*, 7504, https://doi.org/10.1039/C1OB06118F.
[53] M. Kimoto, R. Yamashige, S. Yokoyama, I. Hirao, *J. Nucleic Acids* **2012**, *2012*, 230943, https://doi.org/10.1155/2012/230943.
[54] N. Morohashi, M. Kimoto, A. Sato, R. Kawai, I. Hirao, *Molecules* **2012**, *17*, 2855, https://doi.org/10.3390/molecules17032855.
[55] T. Ishizuka, M. Kimoto, A. Sato, I. Hirao, *Chem. Commun*. **2012**, *48*, 10835, https://doi.org/10.1039/C2CC36293G.
[56] T. Someya, A. Ando, M. Kimoto, I. Hirao, *Nucleic Acids Res*. **2015**, *43*, 6665, https://doi.org/10.1093/nar/gkv638.
[57] M. Kimoto, A. J. Meyer, I. Hirao, A. D. Ellington, *Chem. Commun*. **2017**, *53*, 12309, https://doi.org/10.1039/C7CC06661A.
[58] T. Lavergne, R. Lamichhane, D. A. Malyshev, Z. Li, L. Li, E. Sperling, J. R. Williamson, D. P. Millar, F. E. Romesberg, *ACS Chem. Biol*. **2016**, *11*, 1347, https://doi.org/10.1021/acschembio.5b00952.
[59] N. Z. Fantoni, A. H. El-Sagheer, T. Brown, *Chem. Rev*. **2021**, *121*, 7122, https://doi.org/10.1021/acs.chemrev.0c00928.
[60] J. Han, *Dalton Trans*. **2023**, *52*, 15277, https://doi.org/10.1039/D3DT02166A.
[61] F. Eggert, S. Kath-Schorr, *Chem. Commun*. **2016**, *52*, 7284, https://doi.org/10.1039/C6CC02321E.
[62] C. Domnick, F. Eggert, S. Kath-Schorr, *Chem. Commun*. **2015**, *51*, 8253, https://doi.org/10.1039/C5CC01765C.
[63] F. Eggert, K. Kulikov, C. Domnick, P. Leifels, S. Kath-Schorr, *Methods* **2017**, *120*, 17, https://doi.org/10.1016/j.ymeth.2017.04.021.
[64] L. Bornewasser, C. Domnick, S. Kath-Schorr, *Chem. Sci*. **2022**, *13*, 4753, https://doi.org/10.1039/D2SC00670G.
[65] C. Domnick, G. Hagelueken, F. Eggert, O. Schiemann, S. Kath-Schorr, *Org. Biomol. Chem*. **2019**, *17*, 1805, https://doi.org/10.1039/C8OB02597E.
[66] Domnick, F. Eggert, C. Wuebben, L. Bornewasser, G. Hagelueken, O. Schiemann, S. Kath-Schorr, *Angew. Chem. Int. Ed*. **2020**, *132*, 7965, https://doi.org/10.1002/ange.201916447.
[67] Y. Wang, Y. Chen, Y. Hu, X. Fang, *Proc. Natl. Acad. Sci. U. S. A*. **2020**, *117*, 22823, https://doi.org/10.1073/pnas.2005217117.
[68] Y. Wang, V. Kathiresan, Y. Chen, Y. Hu, W. Jiang, G. Bai, G. Liu, P. Z. Qin, X. Fang, *Chem. Sci*. **2020**, *11*, 9655, https://doi.org/10.1039/D0SC01717E.
[69] B. Endeward, Y. Hu, G. Bai, G. Liu, T. F. Prisner, X. Fang, *Biophys. J.* **2022**, *121*, 37, https://doi.org/10.1016/j.bpj.2021.12.007.
[70] X. Niu, R. Sun, Z. Chen, Y. Yao, X. Zuo, C. Chen, X. Fang, *Nat. Commun*. **2021**, *12*, 6417, https://doi.org/10.1038/s41467-021-26616-x.
[71] X. Niu, Z. Xu, Y. Zhang, X. Zuo, C. Chen, X. Fang, *Nat. Commun*. **2023**, *14*, 7394, https://doi.org/10.1038/s41467-023-43232-z.
[72] Y. Hu, Y. Wang, J. Singh, R. Sun, L. Xu, X. Niu, K. Huang, G. Bai, G. Liu, X. Zuo, C. Chen, P. Z. Qin, X. Fang, *ACS Chem. Biol.* **2022**, *17*, 2448, https://doi.org/10.1021/acschembio.2c00199.
[73] A. D. Ellington, J. W. Szostak, *Nature* **1990**, 346, https://doi.org/10.1038/346818a0.
[74] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505, https://doi.org/10.1126/science.2200121.
[75] T. Herman, D. J. Patel, *Science* **2000**, *287*, 820, https://doi.org/10.1126/science.287.5454.820.
[76] U. Akhtar, W. He, C. Gasse, J. A. Tanner, M. Hollenstein, *Trends Chem*. **2025** (preprint), https://doi.org/10.1016/j.trechm.2025.11.006.

[77] L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T. Fitzwater, D. Flather, A. Forbes, T. Foreman, C. Fowler, B. Gawande, M. Goss, M. Gunn, S. Gupta, D. Halladay, J. Heil, J. Heilig, B. Hicke, G. Husar, N. Janjic, T. Jarvis, S. Jennings, E. Katilius, T. R. Keeney, N. Kim, T. H. Koch, S. Kraemer, L. Kroiss, N. Le, D. Levine, W. Lindsey, B. Lollo, W. Mayfield, M. Mehan, R. Mehler, S. K. Nelson, M. Nelson, D. Nieuwlandt, M. Nikrad, U. Ochsner, R. M. Ostroff, M. Otis, T. Parker, S. Pietrasiewicz, D. I. Resnicow, J. Rohloff, G. Sanders, S. Sattin, D. Schneider, B. Singer, M. Stanton, A. Sterkel, A. Stewart, S. Stratford, J. D. Vaught, M. Vrkljan, J. J. Walker, M. Watrobka, S. Waugh, A. Weiss, S. K. Wilcox, A. Wolfson, S. K. Wolk, C. Zhang, D. Zichi, *PLoS One* **2010**, *5*, 12, https://doi.org/10.1371/journal.pone.0015004.

[78] M. Hollenstein, C. J. Hipolito, C. H. Lam, D. M. Perrin, *ChemBioChem* **2009**, *10*, 1988, https://doi.org/10.1002/cbic.200900314.

[79] O. C. Uhlenbeck, *RNA* **1995**, *1*, 4.

[80] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*, 729, https://doi.org/10.1038/nmeth915.

[81] K.-i. Matsunaga, M. Kimoto, I. Hirao, *J. Am. Chem. Soc*. **2017**, *139*, 324, https://doi.org/10.1021/jacs.6b10767.

[82] M. Kimoto, H. P. Tan, Y. S. Tan, N. A. B. M. Mislan, I. Hirao, *Philos. Trans. R. Soc., B* **2023**, *378*, 1871, https://doi.org/10.1098/rstb.2022.0031.

[83] M. Kimoto, H. P. Tan, K.-i. Matsunaga, N. A. B. M. Mislan, G. Kawai, I. Hirao, *J. Am. Chem. Soc*. **2023**, *145*, 20432, https://doi.org/10.1021/jacs.3c06122.

[84] K. Sawada, M. Kimoto, K.-i. Matsunaga, H. P. Tan, R. Takayanagi, F. K. Sano, Y. Shuto, H. Hirano, T. Kusakizako, Y. Itoh, Y. Kise, I. Hirao, O. Nureki, *Nat. Commun*. **2026**, *17*, https://doi.org/10.1038/s41467-025-67486-x.

[85] Z. Yang, D. Hutter, P. Sheng, A. M. Sismour, S. A. Benner, *Nucleic Acids Res*. **2006**, *34*, 6095, https://doi.org/10.1093/nar/gkl63.3

[86] K. Sefah, Z. Yang, K. M. Bradley, S. Hoshika, E. Jimenez, L. Zhang, G. Zhu, S. Shanker, F. Yu, D. Turex, W. Tan, S. A. Benner, *Proc. Natl. Acc. Sci. U. S. A*. **2014**, *111*, 1449, https://doi.org/10.1073/pnas.1311778111.

[87] E. Biondi, J. D. Lane, D. Das, S. Dasgupta, J. A. Piccirilli, S. Hoshika, K. M. Bradley, B. A. Krantz, S. A. Benner, *Nucleic Acids Res.* **2016**, *44*, 9565, https://doi.org/10.1093/nar/gkw890

[88] L. Zhang, Z. Yang, T. Le Trinh, I.-T. Teng, S. Wang, K. M. Bradley, S. Hoshika, Q. Wu, S. Cansiz, D. J. Rowold, C. McLendon, M.-S. Kim, Y. Wu, C. Cui, Y. Liu, W. Hou, K. Stewart, S. Wan, C. Liu, S. A. Brenner, W. Tan, *Angew. Chem. Int. Ed.* **2016**, *55*, 12372, https://doi.org/10.1002/anie.201605058.

[89] L. Zhang, S. Wang, Z. Yang. S. Hoshika, S. Xie, J. Li, X. Chen, S. Wan, L. Li, S. A. Benner, W. Tan, *Angew. Chem. Int. Ed*. **2020**, *59*, 663, https://doi.org/10.1002/anie.201909691.

[90] F. Tolle, G. M. Brändle, D. Matzner, G. Mayer, *Angew. Chem. Int. Ed*. **2015**, *54*, 10971, https://doi.org/10.1002/anie.201503652.

[91] O. Plückthun, J. Siegl, L. L. Bryant, G. Mayer, *Chem. Sci*. **2020**, *11*, 9577, https://doi.org/10.1039/D0SC01952F

[92] F. Eggert, S. Kath-Schorr, *Chem. Commun*. **2016**, *52*, 7284, https://doi.org/10.1039/c6cc02321e

[93] F. Sanger, S. Nicklein, A. R. Coulson, *Proc. Natl. Acad. Sci. U. S. A*. **1977**, *74*, 5463, https://10.1073/pnas.74.12.5463

[94] B. L. Karger, A. Guttmann, *Electrophoresis*, **2009**, *30*, 196, https://doi.org/10.1002/elps.200900218

[95] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. K. Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. J. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. D. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. C. E. Catenazzi, S. Chang, R. N. Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. S. Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. H. Jones, G.-D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. L. Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. C. Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. C. Rodriguez, P. M. Roe, J. Rogers, M. C. R. Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. E. S. Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. vandeVondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. W. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, A. J. Smith, *Nature*, **2008**, *456*, 53, https://doi.org/10.1038/nature07517

[96] J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, J. Bustillo, *Nature* **2011**, *475*, 348, https://doi.org/10.1038/nature10242

[97] J. J. Kasianowicz, E. Brandin, D. Branton, D. W. Dreamer, *Proc. Natl. Acad. Sci. U. S. A*. **1996**, *93*, 13770, https://doi.org/10.1073/pnas.93.24.13770

[98] J. Clarke, H.-C. Wu, L. Jayasinge, A. Patel, S. Reid, H. Bayley, *Nat. Nanotechnol*. **2009**, *4*, 265, https://doi.org/10.1038/nnano.2009.12.

[99] S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat. Rev. Genet*. **2016**, *17*, 333, https://doi.org/10.1038/nrg.2016.49.

[100] K. Hamashima, Y. T. Soong, K.-i. Matsunaga, M. Kimoto, I. Hirao, *ACS Synth. Biol*. **2019**, *8*, 1401, https://doi.org/10.1021/acssynbio.9b00087.

[101] M. Perez, M. Kimoto, P. Rajakumar, C. Suphavilai, R. Peres da Silva, H. P. Tan, N. T. X. Ong, H. Nicholas, I. Hirao, W. L. Chew, N. Nagarajan, *Nat. Commun*. **2025**, *16*, 6980, https://doi.org/10.1038/s41467-025-62347-z.

[102] D. Deamer, M. Akeson, D. Branton, *Nat. Biotechnol*. **2016**, *34*, 518, https://doi.org/10.1038/nbt.3423.

[103] J. Clarke, H.-J. Wu, L. Jayasinghe, A. Patel, S. Reid, H. Bayley, *Nat. Nanotechnol.* **2009**, *4*, 265, https://doi.org/10.1038/nnano.2009.12.

[104] M. P. Ledbetter, J. M. Craig, R. J. Karadeema, M. T. Noakes, H. C. Kim, S. J. Abell, J. R. Huang, B. A. Anderson, R. Krishnamurthy, J. H. Gundlach, F. E. Romesberg, *J. Am. Chem. Soc*. **2020**, *142*, 2110, https://doi.org/10.1021/jacs.9b09808

[105] S. Hoshika, M. S. Shukla, S. A. Benner, M. M. Georgiadis, *J. Am. Chem. Soc*. **2022**, *144*, 15603, https://doi.org/10.1021/jacs.2c05255

[106] C. A. Thomas, H. Brinkerhoff, J. M. Craig, S. Hoshika, D. Mihaylova, A. M. Pfeffer, M. C. Franzi, S. J. Abell, J. D. Carrasco, J. H. Gundlach, S. A. Benner, A. H. Laszlo, *Nat. Commun*. **2025**, *16*, 7240, https://doi.org/10.1038/s41467-025-61991-9

[107] C. A. Thomas, J. M. Craig, S. Hoshika, H. Brinkerhoff, J. R. Huang, S. J. Abell, H. C. Kim, M. C. Franzi, J. D. Carrasco, H.-J. Kim, D. C. Smith, J. H. Gundlach, S. A. Benner, A. H. Laszlo, *J. Am. Chem. Soc*. **2023**, *145*, 8560, https://doi.org/10.1021/jacs.3c00829.

[108] J. D. Bain, C. Switzer, R. Chamberlin, S. A. Benner, *Nature* **1992**, *356*, 537, https://doi.org/10.1038/356537a0.

[109] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa, S. Yokoyama, *Nat. Biotechnol*. **2002**, *20*, 177, https://doi.org/10.1038/nbt0202-177.

[110] S. Eriksson, B. Munch-Petersen, K. Johansson, H. Ecklund, *Cell. Mol. Life Sci.* **2002**, *59*, 1327, https://doi.org/10.1007/s00018-002-8511-x.

[111] F. Chen, Y. Zhang, A. B. Daugherty, Z. Yang, R. Shaw, M. Dong, S. Lutz, S. A. Benner, *PLoS ONE* **2017**, *12*, e0174163, https://doi.org/10.1371/journal.pone.0174163.

[112] Y. Wu, M. Fa, E. L. Tae, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc*. **2002**, *124*, 14626, https://doi.org/10.1021/ja028050m.

[113] D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Corrêa, F. E. Romesberg, *Nature* **2014**, *509*, 385, https://doi.org/10.1038/nature13314.

[114] Y. Zhang, B. M. Lamb, A. W. Feldman, A. X. Zhou, T. Lavergne, L. Li, F. E. Romesberg, *Proc. Natl. Acad. Sci. U. S. A*. **2017**, *114*, 1317, https://doi.org/10.1073/pnas.1616443114.

[115] Y. Zhang, J. L. Ptacin, E. C. Fischer, H. R. Aerni, C. E. Caffaro, K. San Jose, A. W. Feldman, C. R. Turner, F. E. Romesberg, *Nature* **2017**, *551*, 644, https://doi.org/10.1038/nature24659.

[116] M. P. Ledbetter, R. J. Karadeema, F. E. Romesberg, *J. Am. Chem. Soc*. **2018**, *140*, 758, https://doi.org/10.1021/jacs.7b11488.

[117] A. W. Feldman, F. E. Romesberg, *J. Am. Chem. Soc*. **2017**, *139*, 11427, https://doi.org/10.1021/jacs.7b03540.

[118] V. T. Dien, M. Holcomb, A. W. Feldman, E. C. Fischer, T. J. Dwyer, F. E. Romesberg, *J. Am. Chem. Soc*. **2018**, *140*, 16115, https://doi.org/10.1021/jacs.8b08416.

[119] A. W. Feldman, V. T. Dien, R. J. Karadeema, E. C. Fischer, Y. You, B. A. Anderson, R. Krishnamurthy, J. S. Chen, L. Li, F. E. Romesberg, *J. Am. Chem. Soc*. **2019**, *141*, 10644, https://doi.org/10.1021/jacs.9b02075.

[120] K. Hashimoto, E. C. Fischer, F. E. Romesberg, *J. Am. Chem. Soc*. **2021**, *143*, 8603, https://doi.org/10.1021/jacs.1c03860.
[121] A. X.-Z. Zhou, K. Sheng, A. W. Feldman, F. E. Romesberg, *J. Am. Chem. Soc*. **2019**, *141*, 20166, https://doi.org/10.1021/jacs.9b09080.
[122] E. C. Fischer, K. Hashimoto, Y. Zhang, A. W. Feldman, V. T. Dien, R. J. Karadeema, R. Adhikary, M. P. Ledbetter, R. Krishnamurthy, F. E. Romesberg, *Nat. Chem. Biol*. **2020**, *16*, 570, https://doi.org/10.1038/s41589-020-0507-z.
[123] J. L. Ptacin, C. E. Caffaro, L. Ma, K. M. San Jose Gall, H. R. Aerni, N. V. Acuff, R. W. Herman, Y. Pavlova, M. J. Pena, D. B. Chen, L. K. Koriazova, L. K. Shawver, I. B. Joseph, M. E. Milla, *Nat. Commun*. **2021**, *12*, 4785, https://doi.org/10.1038/s41467-021-24987-9.
[124] 'Our Product Pipeline | Sanofi', can be found under https://www.sanofi.com/en/our-science/our-pipeline, accessed December 5, 2025.
[125] J. L. Ptacin, L. Ma, C. E. Caffaro, N. V. Acuff, K. Germar, P. Severy, Y. Qu, J.-L. Vela, X. Cai, K. M. San Jose, H. R. Aerni, D. B. Chen, E. Esche, T. K. Ismaili, R. Herman, Y. Pavlova, M. J. Pena, J. Nguyen, L. K. Koriazova, L. K. Shawver, I. B. Joseph, J. Mooney, M. Peakman, M. E. Milla, *Commun. Med*. **2024**, *4*, 58, https://doi.org/10.1038/s43856-024-00485-z.
[126] Y. Huang, P. Zhang, H. Wang, Y. Chen, T. Liu, X. Luo, *Chem. Rev*. **2025**, *125*, 523, https://doi.org/10.1021/acs.chemrev.4c00216.
[127] M. L. Pigula, P. G. Schultz, *Curr. Opin. Chem. Biol*. **2024**, *83*, 102537, https://doi.org/10.1016/j.cbpa.2024.102537.
[128] Z. Birch-Price, F. J. Hardy, T. M. Lister, A. R. Kohn, A. P. Green, *Chem. Rev*. **2024**, *124*, 8740, https://doi.org/10.1021/acs.chemrev.4c00120.
[129] R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Gersey, T. I. Oprea, J. P. Overington, *Nat. Rev. Drug Discovery* **2017**, *16*, 19, https://doi.org/10.1038/nrd.2016.230.
[130] V. Kumar, A. Barwal, N. Sharma, D. S. Mir, P. Kumar, V. Kumar, *Biotech*. **2024**, *14*, 112. https://doi.org/10.1007/s13205-024-03958-z.
[131] J. V. Pham, M. A. Yilma, A. Feli, M. T. Majid, N. Maffetone, J. R. Walker, E. Kim, H. J. Cho, J. M. Reynolds, M. C. Song, S. R. Park, Y. J. Yoon, *Front. Microbiol*. **2019**, *10*, 1404. https://doi.org/10.3389/fmicb.2019.01404.
[132] S. Shao, W. Zhang, H. Hu, B. Xue, J. Qin, C. Sun, Y. Sun, W. Wei, Y. Sun, *Nucleic Acids Res*. **2016**, *44*, e86, https://doi.org/10.1093/nar/gkw066.
[133] D. J. Menn, S. Pradhan, S. Kiani, X. Wang, *ACS Synth. Biol*. **2018**, *7*, 1929, https://doi.org/10.1021/acssynbio.8b00153.
[134] K. Lee, V. A. Mackley, A. Rao, A. T. Chong, M. A. Dewitt, J. E. Corn, N. Murthy, *eLife* **2017**, *6*, e25312, https://doi.org/10.7554/eLife.25312.
[135] L. E. Dow, J. Fisher, K. P. O'Rourke, A. Muley, E. R. Kastenhuber, G. Livshits, D. F. Tschaharganeh, N. D. Socci, S. W. Lowe, *Nat. Biotechnol*. **2015**, *33*, 390, https://doi.org/10.1038/nbt.3155.
[136] W. Zhou, W. Brown, A. Bardhan, M. Delaney, A. S. Ilk, R. R. Rauen, S. I. Kahn, M. Tsang, A. Deiters, *Angew. Chem. Int. Ed.* **2020**, *59*, 8998, https://doi.org/10.1002/anie.201914575.
[137] X. Liu, W. Xiong, Q. Qi, Y. Zhang, H. Ji, S. Cui, J. An, X. Sun, H. Yin, T. Tian, X. Zhou, *Nucleic Acids Res*. **2022**, *50*, 4769, https://doi.org/10.1093/nar/gkac255.
[138] K. Zhang, W. Shen, Y. Zhao, X. Xu, X. Liu, Q. Qi, S. Huang, T. Tian, X. Zhou, *Nucleic Acids Res*. **2025**, *53*, https://doi.org/10.1093/nar/gkaf082.

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (https://chimia.ch/chimia/about).

The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2026.108