RSC Chemical Biology

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

Published on 16 October 2026
Licensed under CC-BY 4.0

Check for updates

Cite this: *RSC Chem. Biol.*, 2026, 7, 2183

Received 16th September 2026, Accepted 29th September 2026

DOI: 10.1039/d6cb00303f

rsc.li/rsc-chembio

ARCaDia-selected short DNA core positions a latent warhead for protein-gated covalent capture

Chaohui Li,[a] Naoyuki Miyashita,[ab] Masayasu Kuwahara,[c] Jay Yang*[ad] and Masumi Taki*[ae]

A single-round affinity/reactivity co-selection of aptamer (ARCaDia), combined with Fast String-Based Clustering (FSBC) analysis of the selection output, identified a compact 10-nucleotide DNA-warhead core motif for streptavidin-selective covalent capture, showing that protein-gated covalent chemistry can arise from a short sequence-dependent placement of a latent warhead in a biotin-sensitive reaction microenvironment. This demonstrates that functional warhead-positioning information is already present after a single selection round.

Targeted covalent inhibitors (TCIs) convert reversible molecular recognition into durable target engagement. Their central design problem is kinetic discrimination: bond formation should be accelerated at the target protein while reaction with solvent, serum proteins and off-target nucleophiles remains suppressed.[1,2] This problem becomes especially important for middle-to-macro biomolecular TCIs (bioTCIs), where sequence-defined peptide, protein and nucleic-acid scaffolds can encode both target recognition and warhead presentation.[3,4] Representative recent examples now span peptidic bioTCIs, including phage- and other display-selected covalent peptides,[5–7] protein bioTCIs based on genetically encoded proximity-enhanced reactivity and fast-acting covalent protein platforms,[8–11] and nucleotidic bioTCIs[12] developed through covalent selection[13–19] or post-selection covalent aptamer engineering.[20–26]

A common mechanistic requirement across these systems is that intrinsic warhead electrophilicity must be balanced against target-gated activation. Latent electrophiles, such as sulfur(VI) fluoride exchange (SuFEx) aryl fluorosulfates (AFS),[27] are attractive in this respect because their effective reactivity depends strongly on proximity, geometry, and the local protein microenvironment rather than on bulk electrophilicity alone.[28–32] Consistent with recent analyses of SuFEx-modified nucleic acids, productive covalent capture is not solely determined by affinity or warhead structure in isolation, but rather emerges from a matchmaking reaction microenvironment consisting of the coupled arrangements of the scaffold, target pocket, and latent electrophile.[33] Among nucleotidic bioTCIs, recent SuFEx-based combinatorial screening (*i.e.*, covalent SELEX) studies[16,25] are especially informative, because they use aptameric scaffolds not only as binding modules but also as sequence-encoded presenters/positioners of latent sulfur(VI) fluoride electrophiles. These studies have shown that aptameric scaffolds can support target-gated activation, and that binding affinity alone does not reliably predict covalent product formation. A complementary question, however, remains less explored: how much DNA scaffold sequence information is already present after a single reactivity-first selection round, before iterative enrichment or mature aptamer optimisation has occurred?

Building on our previous single-round affinity/reactivity co-selection (ARC) study, here we ask whether bioinformatic analysis of the selection output can identify a compact sequence-warhead arrangement sufficient to position a latent warhead within a target-gated protein reaction microenvironment. Our ARC framework was initially developed using a T7-phage peptide library modified with a latent AFS warhead, establishing a selection logic in which survival required both target engagement and activation of a weakly reactive warhead in the matchmaking microenvironment.[5] We subsequently extended this ARC logic to aptamer screening and established a single-round/non-SELEX system, namely ARCaDia, using a DNA library functionalised with one latent AFS warhead at a single defined position.[17] Using streptavidin (SA) as a structurally defined model target,[15] we now show that bioinformatic analysis of the single-round ARCaDia output identifies a

[a] *Department of Engineering Science, The Graduate School of Informatics and Engineering, The University of Electro-Communications (UEC), 1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585, Japan. E-mail: jyang75@wisc.edu, taki@pc.uec.ac.jp; Tel: +81-42-443-5980*

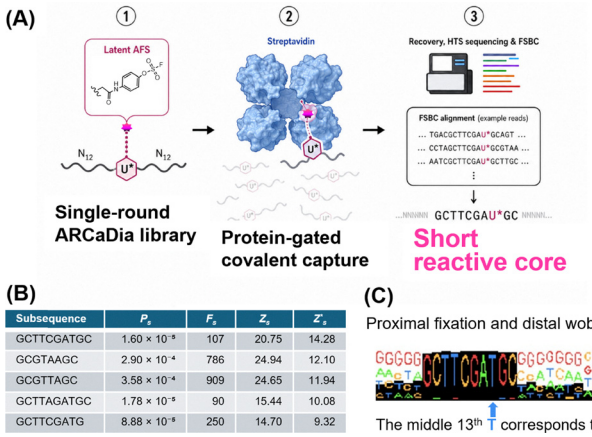
[b] *Department of Biological Systems Engineering, Graduate School of Biology-Oriented Science and Technology, KINDAI University, 930 Nishimitani, Kinokawa, Wakayama, 649-6493, Japan*

[c] *Graduate School of Integrated Basic Sciences, Nihon University, Setagaya-ku, Tokyo, 156-8550, Japan*

[d] *School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin, 53706, USA*

[e] *Institute for Advanced Science, UEC, Chofu, 182-8585, Japan*

© 2026 The Author(s). Published by the Royal Society of Chemistry



Subsequence	P_s	F_s	Z_s	Z'_s
GCTTCGATGC	1.60 × 10⁻⁵	107	20.75	14.28
GCGTAAGC	2.90 × 10⁻⁴	786	24.94	12.10
GCGTTAGC	3.58 × 10⁻⁴	909	24.65	11.94
GCTTAGATGC	1.78 × 10⁻⁵	90	15.44	10.08
GCTTCGATG	8.88 × 10⁻⁶	250	14.70	9.32

Fig. 1 Conceptual workflow of single-round ARCaDia followed by FSBC bioinformatics analysis of HTS reads. (A) A latent-warhead DNA library consisting of a central octadiynyl residue modified with an AFS warhead (U*) flanked by N_{12} random sequences on both sides was subjected to ARCaDia against streptavidin, and the recovered pool was analysed by FSBC to identify a warhead-proximal compact reactive core. The AFS warhead is shown in the dotted square. (B) Post-selection FSBC output of top-ranked over-represented subsequences (*i.e.*, strings). For each subsequence s, P_s, F_s, Z_s, and Z'_s denote the expected occurrence probability, observed read count, enrichment *Z*-score, and length-normalised ranking score, respectively. (C) Alignment of the 25-nucleotide (nt) read sequences containing the highest-scoring 10-nt core motif GCTTCGAU*GC (Core-AFS), characterised by proximal fixation and distal wobble. The U* position was assigned from the defined library design.

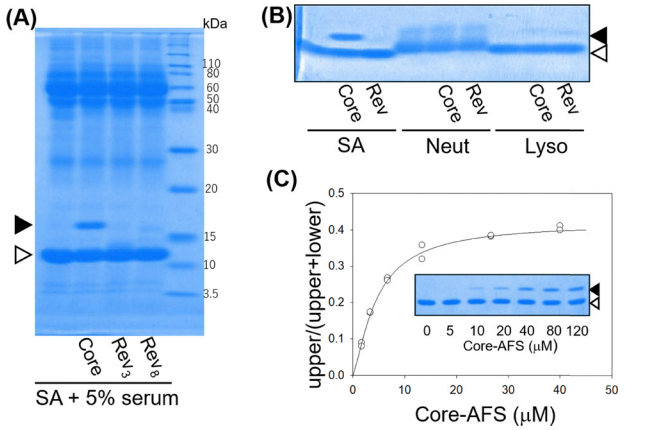


Fig. 2 AFS-conjugated Core selectively forms a covalent adduct with SA. (A) Core-AFS, but not the reverse-order sequence (Rev_3- or Rev_8-AFS; Table S1, SI), produced a clear mobility-shifted SA covalent adduct (closed arrow) in the presence of 5% human serum. The non-shifted unreacted SA was seen as a monomer (open arrow) in this reducing SDS-PAGE. All subsequent experiments used Rev_3 as the negative control. (B) Same as A, except that Core-AFS or Rev-AFS was reacted with NeutrAvidin (Neut) or lysozyme (Lyso). (C) Concentration-dependent covalent capture of SA by Core-AFS, shown by SDS-PAGE (inset) and gel quantitation of the fraction, upper/(upper + lower), plotted against Core-AFS concentration. The solid line is a Hill equation fit to the data (open circles), fraction shift = $A_{max}[C]^n/(EC_{50}{}^n + [C]^n)$, with A_{max} = 0.42, n = 1.3, EC_{50} = 4.57 ± 0.27 μM, all $P < 0.001$. Experiments in (A) and (B) were performed with SA or other target proteins at 20 μM and Core-AFS or Rev-AFS at 90 μM for 24 h at 37 °C; in (C), Core-AFS was varied as indicated.

compact 10-nucleotide (nt) DNA scaffold sufficient to position the latent warhead for SA-gated covalent capture.

Fig. 1A shows a conceptual workflow of the present experiments designed to identify a compact DNA scaffold capable of target-gated covalent capture after a single round of ARCaDia screening. We first performed ARCaDia against SA using the defined DNA library (Fig. S3), as previously described,[17] and then analysed the HTS readouts by Fast String-Based Clustering (FSBC)[34] rather than relying on conventional fixed-length *k*-mer[17,35] or full-length readout abundance counting (Fig. 1B). Simple *k*-mer analysis reports what is frequent (Table S2, SI), whereas FSBC asks what local sequence information is recurrent enough to define an experimentally testable candidate local sequence cluster/motif. This distinction was critical here: single-round ARCaDia was designed not to recover a mature full-length aptamer family, but to reveal a compact warhead-proximal DNA scaffold capable of positioning latent chemistry within a target-gated reaction microenvironment. FSBC identified 8–10-nt strings, and alignment of the full 25-nt readouts containing the highest-scoring short motif GCTTCGATGC revealed a conserved short core flanked by variable 5′ and 3′ regions (Fig. 1C). Because the AFS-modified U* was placed at the 13th nucleotide between two N_{12} random regions in the starting library, and this fixed position mapped predominantly to the 8th nucleotide of the aligned seed-containing reads (Table S3, SI), this motif was assigned as GCTTCGAU*GC and tested directly as Core-AFS. Core-AFS retained SA-dependent covalent capture in the presence of human serum, whereas reverse-order controls with the warhead in the 3rd or 8th position (Rev_3 or Rev_8, respectively), with matched length, nucleotide composition, and the same AFS-modified nucleotide, did not (Fig. 2A). Notably, Rev_8-AFS retained U* at the same position as Core-AFS yet showed negligible covalent capture, indicating that warhead position alone is insufficient and supporting a critical role for the local sequence arrangement in productive reactivity. This covalent capture by Core was SA-specific: the closely related NeutrAvidin (Neut) and the unrelated lysozyme (Lyso) showed negligible mobility shift on reducing SDS-PAGE (Fig. 2B). The Core-AFS reaction with SA was concentration-dependent and was well described by a Hill equation (Fig. 2C). Productive capture therefore cannot be explained simply by oligonucleotide charge alone,[36] electrophile exposure, or the presence of AFS. Rather, it requires sequence-dependent placement of the latent AFS warhead within a matchmaking microenvironment by this short DNA scaffold to enable covalent reaction with the SA protein. Consistent with previous observations,[22] Core-AFS covalently bound to SA demonstrated relative resistance to nuclease digestion (Fig. S4, SI).

Time-dependent formation of the mobility-shifted SA band was quantified by reducing SDS-PAGE. For each Core-AFS concentration, the shifted-band fraction increased with time and was empirically fitted to a mono-exponential approach to plateau, yielding an apparent first-order rate constant, k_{obs}. At 10 μM Core-AFS, k_{obs} was 0.15 h^{-1} (Fig. 3A). To interpret the concentration dependence of k_{obs}, we first considered the classical two-step irreversible-binding scheme shown in the

© 2026 The Author(s). Published by the Royal Society of Chemistry

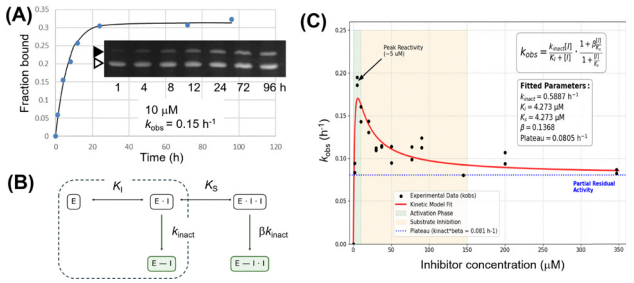


Fig. 3 Kinetic analysis of the Core-AFS covalent reaction with SA. (A) SDS-PAGE (Oriole staining) of Core-AFS (10 μM) reacted with SA (6.7 μM) for the indicated times (h), with gel quantification used to obtain the fraction bound (upper/(upper + lower)) fitted to a mono-exponential equation $f(t) = f_\infty(1-e^{-k_{obs}\cdot t})$. (B) Kinetic model used to describe the concentration dependence of k_{obs}. The scheme shows the kinetic model in which *I* denotes Core-AFS and *E* denotes SA; the boxed three-state model corresponds to the canonical reversible complex formation followed by an irreversible covalent conversion. Additional states representing partial substrate inhibition were included to reproduce the concentration-dependent suppression of k_{obs}. (C) Apparent k_{obs} obtained at various Core-AFS concentrations. Each dot indicates an independent experiment, with at least duplicate measurements for each concentration. The red curve shows the best fit to the phenomenological partial-substrate-inhibition model. Fitting used a weighting factor $W = (k_{obs} + 0.05)^2$.

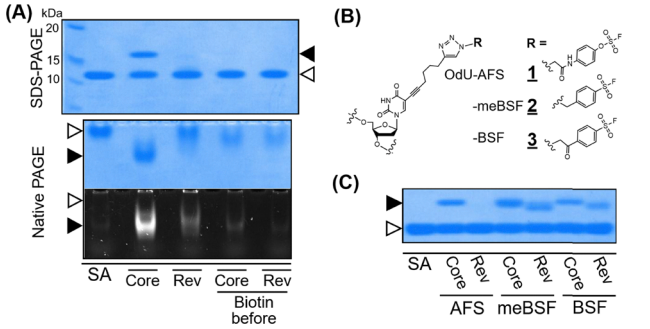


Fig. 4 The selected 10-nt Core positions the latent AFS warhead for biotin-sensitive, SA-gated covalent capture. (A) Core-AFS or Rev-AFS was reacted with SA in the absence of biotin, except for the lanes indicated as "Biotin before", in which SA was pre-incubated with a two-fold molar excess of biotin relative to SA monomer before addition of the oligonucleotide conjugate. Each reaction mixture was analysed by both denaturing SDS-PAGE and native PAGE. On SDS-PAGE, the lower band is the unreacted SA (open arrow) and the mobility-shifted upper band is the covalently modified SA monomer (closed arrow). On native PAGE, unmodified tetrameric SA is indicated by an open arrow and mobility-shifted tetrameric SA by a closed arrow; the two panels are the same gel stained with CBB for protein (upper) and SYBR Gold for DNA (lower). (B) Structures of the octadiynyl (OdU)-linked warheads installed by click conjugation: AFS (**1**), meBSF (**2**), and BSF (**3**). (C) Core and Rev modified with AFS, meBSF, or BSF were reacted with SA in the same manner and analysed by SDS-PAGE. AFS gave the clearest Core-over-Rev discrimination, whereas meBSF and BSF increased Rev-mediated background capture.

dashed box in Fig. 3B, in which reversible complex formation precedes covalent conversion.[37] However, the data did not follow the simple hyperbolic saturation expected for this model (Fig. 3B and C). Instead, k_{obs} reached a maximum at low micromolar Core-AFS concentrations and decreased thereafter, indicating that increasing Core-AFS does not simply increase the population of productive reaction complexes. This non-hyperbolic behaviour is consistent with the existence of one or more concentration-dependent less-reactive states, such as oligonucleotide self-association, secondary binding, or occupancy-dependent effects within tetrameric SA, which would reduce the apparent fraction of SuFEx-competent complexes at elevated Core-AFS concentrations. A phenomenological partial-substrate-inhibition model[38,39] reproduced the overall trend (Fig. 3C, red line), but multiple physical mechanisms (Fig. S5, SI) can account for the observed suppression. Because these alternatives cannot be distinguished by the present experiments, the fitted parameters should be regarded as operational kinetic descriptors rather than true microscopic kinetic parameters and do not confirm that partial-substrate inhibition is the underlying mechanism. Nevertheless, the kinetic data support the conclusion that the single-round ARCaDia-derived 10-nt Core is sufficient for sequence-dependent, target-gated covalent conversion. At the same time, the slow apparent conversion and non-hyperbolic concentration dependence indicate that this Core is not yet an optimised high-affinity covalent aptamer, but rather a compact warhead-positioning scaffold.

The biotin-binding architecture of tetrameric SA therefore provided a means to test whether Core-AFS capture was associated with the biotin-binding pockets. SA exists as a homotetramer and binds biotin at well-defined pockets deep within the β-barrel, with contributions from residues at the subunit interface.[40,41] A biotin competition experiment demonstrated that biotin suppressed Core-AFS capture, but only when applied before Core-AFS (Fig. 4A and Fig. S6, SI). Although these experiments do not provide residue-level assignments of the conjugated amino acids, they argue against non-specific protein labelling and support warhead placement near a biotin-sensitive region of SA.[42,43]

The Core-AFS-induced, biotin-sensitive mobility-shifted SA band on SDS-PAGE accounted for approximately half of the total SA signal (Fig. 2B and 4A, lane SA/Core), and we asked whether this reflected incomplete covalent reaction at the tetramer level or a monomer-level readout of tetrameric SA. We therefore analysed the same reaction mixtures in parallel by denaturing SDS-PAGE and native PAGE (Fig. 4A). In denaturing SDS-PAGE, Core-AFS reproducibly gave shifted and unshifted SA subunit bands at an approximately 1 : 1 ratio (Fig. 4A, top). In contrast, native PAGE, where the oligomeric SA is preserved, showed that Core-AFS induced a complete mobility shift of oligomeric SA, whereas Rev-AFS did not, and this shift was suppressed by biotin preincubation (Fig. 4A, middle). Thus, the 50% mobility-shifted fraction on SDS-PAGE does not reflect an incomplete tetramer-level capture but is consistent with the dissociation of Core-AFS-modified tetramers into conjugated and unconjugated monomers. Furthermore, the observed mobility shift of the monomer is consistent with a single Core-AFS modification per shifted SA monomer. Together, these data are consistent with an average of approximately two Core-AFS modifications per SA tetramer.

To further examine whether Core-dependent covalent capture required a matched warhead-latency window, rather than simply

© 2026 The Author(s). Published by the Royal Society of Chemistry

increasing with intrinsic electrophilicity,[5,28] we compared three sulfur(VI) fluoride variants[29–32] on the same Core and Rev scaffolds (Fig. 4B): the latent AFS,[5,17] a less latent methylene-linked benzenesulfonyl fluoride (meBSF),[23] and the most intrinsically reactive carbonyl-linked benzenesulfonyl fluoride (BSF).[20,22] AFS gave the clearest sequence discrimination, whereas both meBSF and BSF increased background capture with Rev, consistent with partial collapse of sequence-dependent gating under the tested conditions (Fig. 4C). Thus, within the Core–Rev scaffold pair, AFS provided the best balance between productive SA capture and suppression of sequence-independent background reactivity.

In summary, FSBC analysis of single-round ARCaDia output identified a compact 10-nt warhead-proximal DNA scaffold capable of SA-gated covalent capture. Sequence reversal, biotin competition, serum-containing reactions, non-target controls with lysozyme and NeutrAvidin, and warhead comparison together show that productive capture arises from sequence-dependent warhead placement/orientation in a biotin-sensitive reaction microenvironment, rather than from non-specific electrophile exposure. Thus, the single-round ARCaDia coupled with FSBC enabled extraction of local sequence-warhead information defining a compact scaffold that supports target-gated reactivity without constituting a mature high-affinity aptamer. If similar compact sequences against biologically important targets can be found, analyses of the interactions between such tractable compact scaffolds and their respective targets may provide a mechanistic understanding at the molecular level of the key determinants governing target-gated reactivity. In turn, this could enhance our ability to design high-affinity covalent aptameric drugs.

Author contributions

C. L., J. Y., and M. T.: conceptualization, investigation, and writing of the manuscript; N. M. and M. K.: review and editing. M. T.: funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Data availability

All data supporting the findings of this study are included in the article and supplementary information (SI). The SI includes the raw HTS data as FASTQ files, the locally used pFSBC_custom.py script, and the details of the web-server-based analysis procedures. Several references are cited only in the SI.[44–46] See DOI: https://doi.org/10.1039/d6cb00303f.

Acknowledgements

This work was mainly supported by JSPS KAKENHI grant (#26K08543), and the Takahashi Industrial Economic Research Foundation. We thank Dr Riku Katsuki (Keio University), a former member of our group, for early programming support for the pFSBC_custom.py script. Part of this work was also supported by "Advanced Research Infrastructure for Materials and Nanotechnology in Japan (ARIM)" of the MEXT (Proposal No. JPMXP1225UE0028), the Asahi Glass Foundation, and JKA through its promotion funds from KEIRIN RACE. We sincerely thank the anonymous reviewers for their valuable comments.

References

1 J. Singh, R. C. Petter, T. A. Baillie and A. Whitty, *Nat. Rev. Drug Discovery*, 2011, **10**, 307–317.
2 L. Boike, N. J. Henning and D. K. Nomura, *Nat. Rev. Drug Discovery*, 2022, **21**, 881–898.
3 J. Yang, Y. Tabuchi, R. Katsuki and M. Taki, *Int. J. Mol. Sci.*, 2023, **24**, 3525.
4 T. Tamura, M. Kawano and I. Hamachi, *Chem. Rev.*, 2025, **125**, 1191–1253.
5 Y. Tabuchi, T. Watanabe, R. Katsuki, Y. Ito and M. Taki, *Chem. Commun.*, 2021, **57**, 5378–5381.
6 S. Wang, F. F. Faucher, M. Bertolini, H. Kim, B. Yu, L. Cao, K. Roeltgen, S. Lovell, V. Shanker, S. D. Boyd, L. Wang, R. Bartenschlager and M. Bogyo, *J. Am. Chem. Soc.*, 2025, **147**, 7461–7475.
7 J. H. Walker, K. Yan and R. Derda, *Biochemistry*, 2026, **65**, 149–165.
8 B. Yu, S. Li, T. Tabata, N. Wang, L. Cao, G. R. Kumar, W. Sun, J. Liu, M. Ott and L. Wang, *Chemistry*, 2022, **8**, 2766–2783.
9 Z. Xiang, H. Ren, Y. S. Hu, I. Coin, J. Wei, H. Cang and L. Wang, *Nat. Methods*, 2013, **10**, 885–888.
10 Q. Li, Q. Chen, P. C. Klauser, M. Li, F. Zheng, N. Wang, X. Li, Q. Zhang, X. Fu, Q. Wang, Y. Xu and L. Wang, *Cell*, 2020, **182**, 85–97.
11 Q. Fan, J. Mei, T. Li, C. Zang, M. Li, J. Tang, Y. Xu, G. Yu, D. Liu, K. Chen, B. Yang, J. Huang, T. Zhou and B. Dang, *Science*, 2026, **392**, eadv3081.
12 S. Albright, J. Boette, M. Cacace and A. Deiters, *RSC Chem. Biol.*, 2026, **7**, 9–30.
13 D. Smith, G. P. Kirschenheuter, J. Charlton, D. M. Guidot and J. E. Repine, *Chem. Biol.*, 1995, **2**, 741–750.
14 J. Charlton, G. P. Kirschenheuter and D. Smith, *Biochemistry*, 1997, **36**, 3018–3026.
15 I. S. MacPherson and T. Liedl, *Aptamers*, 2018, **2**, 74–81.
16 Z. Qin, K. Zhang, P. He, X. Zhang, M. Xie, Y. Fu, C. Gu, Y. Zhu, A. Tong, H. Wei, C. Zhang and Y. Xiang, *Nat. Chem.*, 2023, **15**, 1705–1714.
17 M. Taki, M. Kuwahara, C. Li, N. Tomoda, N. Miyashita, T. Kan and J. Yang, *Chem. Commun.*, 2024, **60**, 14964–14967.
18 N. Soxpollard, S. Strauss, R. Jungmann and I. S. MacPherson, *Commun. Chem.*, 2024, **7**, 174.
19 S. Bian, Q. Qiu, C. Zhang, H. Cao, H. Xuan, X. Yan, J. Li, W. Chen, X. Liu, Y. Chen, J. Long, H. Su, J. Wang and L. Zhang, *J. Am. Chem. Soc.*, 2025, **147**, 46780–46786.
20 Y. Tabuchi, J. Yang and M. Taki, *Chem. Commun.*, 2021, **57**, 2483–2486.

© 2026 The Author(s). Published by the Royal Society of Chemistry

21 Y. Tivon, G. Falcone and A. Deiters, *Angew. Chem., Int. Ed.*, 2021, **60**, 15899–15904.
22 Y. Tabuchi, J. Yang and M. Taki, *Int. J. Mol. Sci.*, 2022, **23**, 7778.
23 D. Wang, J. Zhang, Z. Huang, Y. Yang, T. Fu, Y. Yang, Y. Lyu, J. Jiang, L. Qiu, Z. Cao, X. Zhang, Q. You, Y. Lin, Z. Zhao and W. Tan, *ACS Cent. Sci.*, 2023, **9**, 72–83.
24 H. Jo, S. Ju, M. Kim, J. Beon, S.-Y. Jang, S. P. Pack, C. Y. Son, J.-S. Kim and S. S. Oh, *J. Am. Chem. Soc.*, 2025, **147**, 9328–9340.
25 K. Zhang, J. Li, Y. Shi, S. Hou, Z. Qin, W. Shi, Y. Zhu, J. Zhang, A. Tong and Y. Xiang, *J. Am. Chem. Soc.*, 2025, **147**, 46112–46124.
26 Z. Zhang, H. Qian, R. Xu, Q. Duan, Y. Wang, T. Fu, X. Wu, C. Sha, Q. Wu, Y. Wang and W. Tan, *J. Am. Chem. Soc.*, 2025, **147**, 45951–45965.
27 J. Dong, L. Krasnova, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2014, **53**, 9430–9448.
28 T. V. Frankovich, H. M. McCann, K. S. Hoffman and A. F. Rullo, *ACS Cent. Sci.*, 2025, **11**, 2006–2017.
29 L. H. Jones, *Chem. Sci.*, 2025, **16**, 10119–10140.
30 L. H. Jones, *ACS Med. Chem. Lett.*, 2018, **9**, 584–586.
31 Q. Zheng, J. L. Woehl, S. Kitamura, D. Santos-Martins, C. J. Smedley, G. Li, S. Forli, J. E. Moses, D. W. Wolan and K. B. Sharpless, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 18808–18814.
32 L. Hillebrand, X. J. Liang, R. A. M. Serafim and M. Gehringer, *J. Med. Chem.*, 2024, **67**, 7668–7758.
33 M. Chromiński, M. Żmudziński and J. Jemielity, *RSC Chem. Biol.*, 2026, **7**, 1482–1504.
34 S. Kato, T. Ono, H. Minagawa, K. Horii, I. Shiratori, I. Waga, K. Ito and T. Aoki, *BMC Bioinf.*, 2020, **21**, 263.
35 S. Hoon, B. Zhou, K. D. Janda, S. Brenner and J. Scolnick, *Biotechniques*, 2011, **51**, 413–416.
36 G. Marson, M. Palumbo and C. Sissi, *Curr. Pharm. Des.*, 2012, **18**, 2027–2035.
37 R. Kitz and I. B. Wilson, *J. Biol. Chem.*, 1962, **237**, 3245–3249.
38 M. Yoshino and K. Murakami, *SpringerPlus*, 2015, **4**, 292.
39 Y. Lin, P. Lu, C. Tang, Q. Mei, G. Sandig, D. A. Rodrigues, T. H. Rushmore and M. Shou, *Drug Metab. Dispos.*, 2001, **29**, 368–374.
40 P. C. Weber, D. H. Ohlendorf, J. J. Wendoloski and F. R. Salemme, *Science*, 1989, **243**, 85–88.
41 W. A. Hendrickson, A. Pähler, J. L. Smith, Y. Satow, E. A. Merritt and R. P. Phizackerley, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 2190–2194.
42 T. Bing, X. Yang, H. Mei, Z. Cao and D. Shangguan, *Bioorg. Med. Chem.*, 2010, **18**, 1798–1805.
43 V. J. B. Ruigrok, E. van Duijn, A. Barendregt, K. Dyer, J. A. Tainer, R. Stoltenburg, B. Strehlitz, M. Levisson, H. Smidt and J. van der Oost, *ChemBioChem*, 2012, **13**, 829–836.
44 J. Hoinka, R. Backofen and T. M. Przytycka, *Mol. Ther. Nucleic Acids*, 2018, **11**, 515–517.
45 S. Kurtz, A. Narechania, J. C. Stein and D. Ware, *BMC Genomics*, 2008, **9**, 517.
46 S. Kato, T. Ono, M. Ito, K. Ito, H. Minagawa, K. Horii, I. Shiratori, I. Waga and T. Aoki, *EAI Endorsed Trans. Bioeng. Bioinform.*, 2021, **1**, e3.

© 2026 The Author(s). Published by the Royal Society of Chemistry