

Spatial barcoding reveals reaction radii and contact-dependent mechanism of proximity labeling

Received: 2 December 2024

Accepted: 31 October 2025

Published online: 24 December 2025

 Check for updates

Zhe Yang¹, Yu Zhang², Yuxin Fang³, Yuan Zhang¹, Jiasheng Du¹, Xiaowen Shen², Kecheng Zhang², Peng Zou^{2,3,4,5}✉ & Zhixing Chen^{1,2,6,7}✉

Proximity labeling techniques such as TurboID and APEX2 have become pivotal tools for studying protein interactions. However, the spatial patterns of labeling methods within the submicrometer range remain poorly understood. Here we used DNA nanostructure platforms to precisely measure the labeling radii of TurboID and APEX2 through in vitro assays. Our DNA nanoruler design enables the deployment of oligonucleotide-barcoded labeling targets with nanometer precision near the enzymes. By quantifying labeling yields using qPCR and mapping them against target distances, we uncovered surprising insights into the labeling mechanisms. Contrary to the prevailing diffusive labeling model, our results demonstrate that TurboID primarily operates through contact-dependent labeling. Similarly, APEX2 shows high labeling efficiency within its direct contact range. In parallel, it exhibits low-level diffusive labeling toward more distant phenols. These findings reframe our understanding in the mechanism of proximity labeling enzymes while highlighting the potential of DNA nanotechnology in spatially profiling reactive species.

Elucidating protein interaction networks is crucial for understanding fundamental biology and developing new therapeutics. Proximity labeling technology is an emerging yet increasingly popular tool for identifying protein–protein, protein–DNA and protein–RNA interactions^{1–4}. Essentially, proximity labeling involves the site-specific generation of reactive molecules that covalently label biomolecules within a certain distance range, followed by the identification of labeled biomolecules. Since the first conceptual prototypes, BioID⁵ and APEX⁶, the application of various biochemical reactions has led to numerous proximity labeling methods, such as PUP-IT⁷, CAP-seq⁸, μ Map⁹, FucolID¹⁰, PhoXCELL¹¹, QMID¹², CAT-S¹³, BAP-seq¹⁴, MultiMap¹⁵, APEX2 (ref. 16) and TurboID¹⁷,

with the latter two being the most widely used. These proximity labeling methods provide various options for exploring biomolecular interactions, as their mechanistical differences give rise to varying labeling reactivities and spatial preferences.

Specifically, the radius of proximity labeling is a key parameter determined by the intrinsic reactivity of the active species. Practically, measuring the labeling radius is essential for choosing the right proximity labeling system in various experimental contexts. Despite the widespread use of proximity labeling in protein interactomics, the accurate profiling the distance of proximity labeling has been disproportionately lagging behind because of a lack of molecular tools. There have been a

¹College of Future Technology, Institute of Molecular Medicine, National Biomedical Imaging Center, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Peking University, Beijing, China. ²Peking-Tsinghua Center for Life Science, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ³College of Chemistry and Molecular Engineering, Synthetic and Functional Biomolecules Center, Beijing National Laboratory for Molecular Sciences, Key Laboratory of Bioorganic Chemistry and Molecular Engineering of the Ministry of Education, PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China. ⁴PKU-IDG/McGovern Institute for Brain Research, Beijing, China. ⁵Chinese Institute for Brain Research, Beijing, China. ⁶State Key Laboratory of Membrane Biology, Peking University, Beijing, China. ⁷PKU-Nanjing Institute of Translational Medicine, Nanjing, China. ✉e-mail: zoupeng@pku.edu.cn; zhixingchen@pku.edu.cn

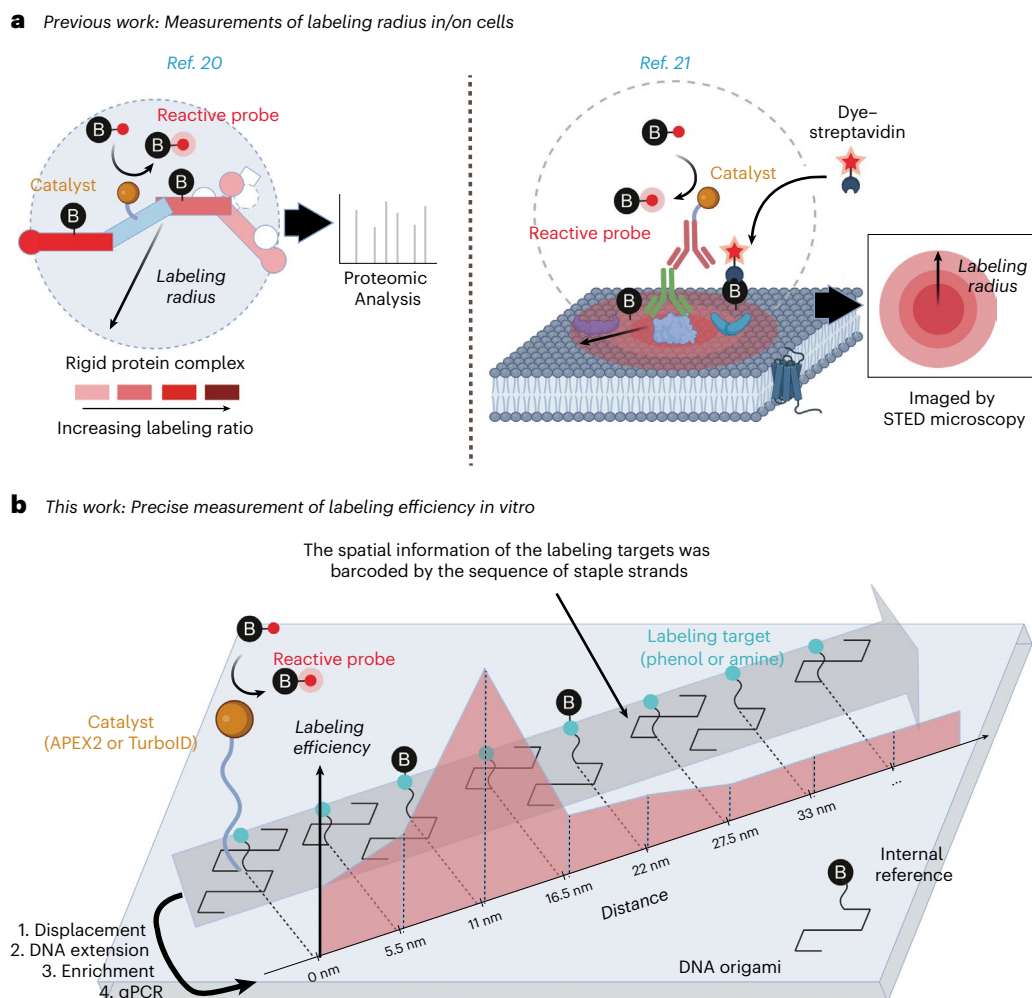


Fig. 1 | Selected methods for measuring the working distance of proximity labeling. **a**, Labeling radii were previously measured in the live-cell context using spatial proteomic analysis of a rigid protein complex or super-resolution imaging of membranous targets. **b**, In this work, proximity labeling with DNA nanostructure templates enables accurate profiling of spatial ranges in vitro.

few studies profiling the labeling radius of proximity enzymes^{18,19}. For example, the Roux group suggested that the labeling radius of BioID is around 10 nm in cells using a rigid nuclear pore protein complex as a molecular ruler²⁰. The MacMillan group developed a method to image the labeling spot of the proximity labeling using super-resolution microscopy²¹ (Fig. 1a), concluding that the labeling radius of horseradish peroxidase is 268 nm. While these methods provide valuable spatial information, the in cellulo measurements are intrinsically complicated by the cellular environments. Moreover, it is challenging to perform mechanistic studies and establish the full profile of the distance–labeling efficiency relationship without an in vitro reconstructed system of proximity labeling. Without such biochemical parameters on the distance profile, our understanding on the overarching mechanism of proximity labeling remains incomplete.

In this study, we devised DNA nanostructures that can accurately template proteins and labeling targets with designated addresses at nanometer accuracy, thereby reconstructing the proximity labeling systems in vitro. DNA nanostructures offer exquisite programmability and ideal distance ranges (1–100 nm) matching the working distances of proximity labeling. Currently, protein–DNA nanodevice systems are used to study various enzymatic cascade reactions by controlling the relative distance and orientation of enzyme components^{22–24}. DNA nanostructures can also serve as platforms for single-molecule chemical reactions^{25,26}, making them ideal candidates for probing the

spatial dynamics of proximity labeling reactions. Our system leverages rigid DNA nanostructures to control the distance between the enzyme catalyst that generates reactive species and the labeling targets, with spatial information of the labeling targets barcoded by the sequence of staple strands. Using pulldown and qPCR quantification to measure labeling efficiency at various distances, we were able to profile the full distance–labeling efficiency map of the APEX2 and TurboID proximity labeling reactions in a controlled microenvironment, which reveals new insights into the mechanism of these proximity labeling reactions.

Results

Spatially addressed enzymatic catalyst and labeling targets on DNA origami

We aimed to reconstruct proximity labeling systems in vitro to create a chemically defined environment without unknown factors. Considering the prior knowledge that proximity labeling mainly functions in the nanometer to micrometer range, we regarded DNA nanostructures as ideal molecular rulers for this purpose thanks to their structural rigidity and synthetic programmability. The catalyst, APEX2 or TurboID, can be conjugated to designer DNA oligos with a fusion HaloTag protein. The DNA–catalyst dyad can be hybridized to a specific address on a DNA origami (Fig. 1b). In this setup, the active center of protein catalyst is about 10 nm away from its branching spot on DNA origami. Moreover, selected staple strands were chemically functionalized at the 5' ends

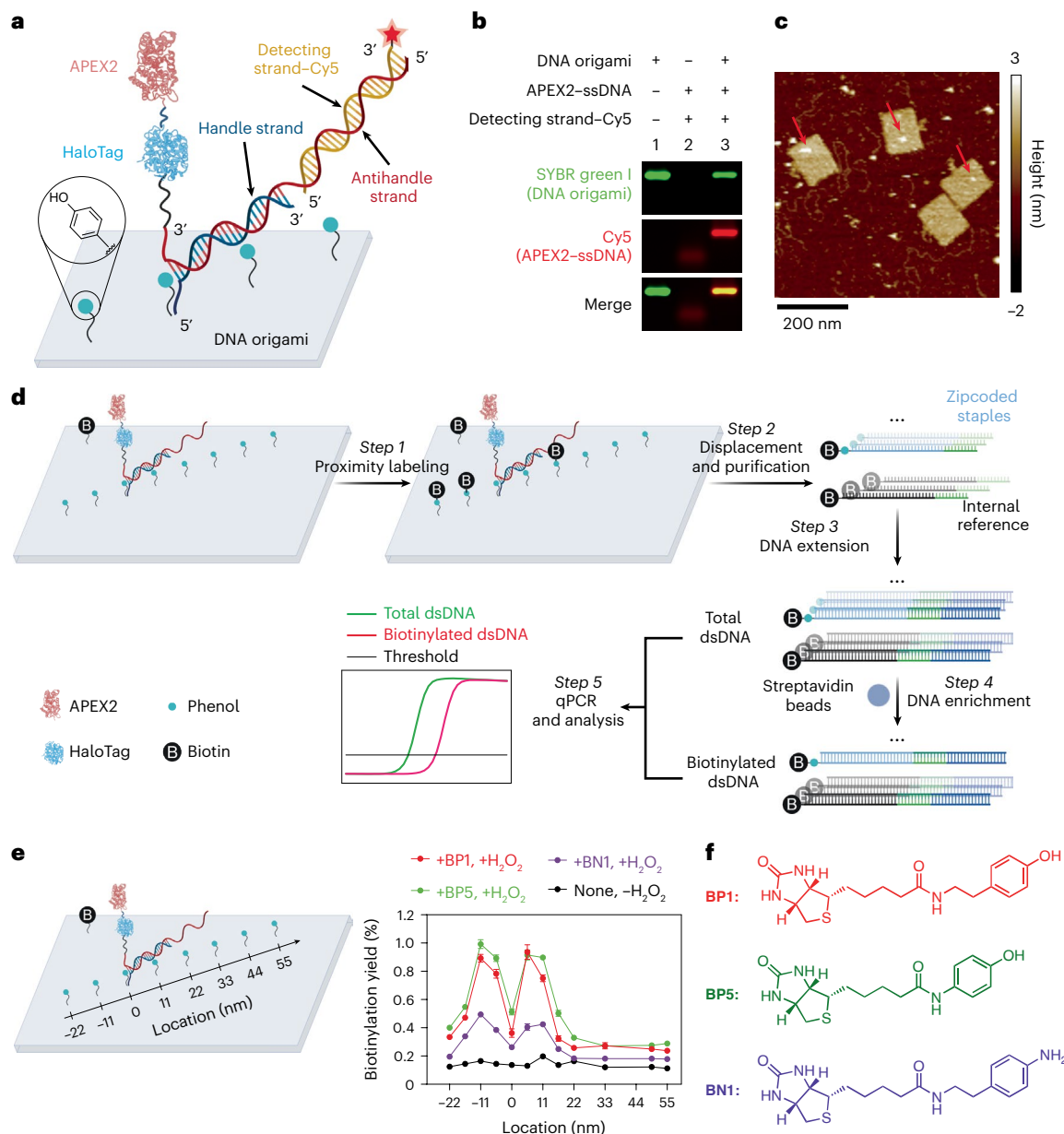


Fig. 2 | Measuring the labeling radius of APEX2 with DNA origami.

a, Schematic diagram of assembly of DNA origami and APEX2. **b**, Gel characterization of the assembly of DNA origami and APEX2. This experiment was independently repeated three times. **c**, AFM image of DNA origami-APEX2 assemblies. This experiment was independently repeated three times. **d**, Workflow of labeling efficiency measurements at destined addresses.

After performing the proximity labeling reaction, the DNA oligos are purified, elongated, enriched and subjected to qPCR quantifications. Benchmarking with an internal reference gives the labeling yields at different locations. **e**, Plot of biotinylation yield against phenol location. Data are presented as the mean values $\pm$ s.e.m. ($n = 3$ biological replicates for each group). **f**, Chemical structures of biotin-phenol/aniline substrates.

with labeling targets, such as phenols for APEX labeling or amines for TurboID. These functional staple strands were weaved into the DNA origami at designated addresses with known distance to the catalyst. In this way, an array of proximity labeling reactions could be programmed at various distances, whose labeling efficiencies could be then read using standard nucleic acid quantification methods.

To establish a fixed spatial arrangement for APEX2 labeling on DNA origami platform, we first designed modification sites for phenol groups at varying distances from the APEX2 site (Supplementary Fig. 1 and Supplementary Tables 2 and 3). Then, we installed the phenol groups on the selected staple strands by modifying the amino-modified staple strands with phenol-NHS, as confirmed by mass spectrometry (MS) (Supplementary Fig. 2). The phenol-modified staple strands were annealed onto the DNA origami using standard protocols (Methods).

Finally, the APEX2-ssDNA (single-stranded DNA) conjugate, prepared separately through the reaction of HaloTag ligand-ssDNA and purified recombinant APEX2-HaloTag fusion protein (Supplementary Fig. 3), was hybridized to the extended handle strand on the DNA origami with a known address (Fig. 2a). In this de novo designed molecular architecture, both the anchoring site of APEX2 and the phenol group are programmed on the same side of the DNA sheet²⁷. Notably, while APEX2 can label guanosine bases, the phenoxyl free radical shows substantially lower reactivity toward double-stranded DNA (dsDNA) compared to its preferred substrate, phenol^{28,29}.

To confirm the assembly of the APEX2-origami complex, a detecting strand modified with Cy5 at the 3' end was hybridized to the anti-handle strand covalently conjugated to the APEX2 fusion protein. Gel electrophoresis followed by in-gel fluorescence imaging indicated a

successful anchoring of APEX2 to the origami (Fig. 2b). Atomic force microscopy (AFM) images further confirmed that APEX2 was precisely positioned to the designated address on the origami (Fig. 2c and Supplementary Fig. 4) and the assembly efficiency of APEX2 on the DNA origami was approximately 60%.

Following the assembly of APEX2 onto the DNA origami, proximity labeling was initiated by the addition of biotin-phenol and hydrogen peroxide (Fig. 2d). After 30 min, the reaction was quenched by the addition of Trolox (10 mM) and sodium ascorbate (10 mM). The staple strands and scaffold strand in the origami were then peeled off by a strand-replacement reaction using the F-5.2 primer and Deep Vent polymerase (Fig. 2d and Supplementary Fig. 5a), followed by a purification step (Supplementary Fig. 5b,d and Methods). Notably, the incorporation efficiency of the modified staple strands into the DNA origami³⁰ and their subsequent release by strand displacement (Supplementary Fig. 8a,b) were largely consistent, providing a robust foundation for the quantitative comparison of labeling efficiencies across different positions. As the staple strands have similar base lengths (32 bp) to microRNAs, we used a protocol typically used for detecting miRNA to determine the biotinylation ratio of the modified staple strands³¹. The purified staple strands with poly(T)₂₅ were extended to an appropriate length (90 bp) using an adaptor strand containing poly(A) (Supplementary Fig. 5c,f). Subsequently, the extended staple strand, referred to as 'input', was enriched using streptavidin beads⁸, resulting in an enriched sample referred to as 'enrich' (Fig. 2d). Notably, we introduced a constitutively biotinylated staple strand as an internal reference in the very beginning to control the loss during the enrichment process.

Moving to the analysis step, these samples were subjected to qPCR using specific primers that did not interfere with each other (Supplementary Fig. 6). By comparing the cycle threshold (Ct) values of input and enrich, we calculated the biotinylation efficiency (F) of each phenol-group-modified staple strand using equation (1), where E_{staple} is the amplification efficiency of each staple (Supplementary Fig. 7 and Supplementary Table 1) and $Ct_{\text{staple, input}}$ and $Ct_{\text{staple, enrich}}$ are the Ct values before and after enrichment, respectively. Similarly, E_{IR} , $Ct_{\text{IR, input}}$ and $Ct_{\text{IR, enrich}}$ are the amplification efficiency of the internal reference and the Ct values before and after enrichment, respectively.

$$F = \frac{(E_{\text{staple}} + 1)^{-(Ct_{\text{staple, enrich}} - Ct_{\text{staple, input}})}}{(E_{\text{IR}} + 1)^{-(Ct_{\text{IR, enrich}} - Ct_{\text{IR, input}})}} \quad (1)$$

With the origami template and the DNA analysis protocol, the labeling efficiencies of phenols were plotted against their distance to APEX2 (Fig. 2e,f and Supplementary Fig. 8). The results indicate that notable labeling occurs within 55 nm, with the site within ± 11 nm showing the highest labeling. Screening of alternative substrates with different reactivities gave different labeling efficiencies yet similar distance profiles^{29,32}. The control group without substrate and hydrogen peroxide gave a low yet consistent signal ($<0.2\%$), indicating a relatively clean pulldown background. Notably, intermolecular labeling efficiency was also measured by the coaddition of an empty DNA origami with orthogonally barcoded phenols but without a catalyst (Supplementary Fig. 9a,b). At the origami concentration of 10 nM, the labeling at adjacent origami was very low, comparable to the pulldown background. To simulate the intracellular labeling of APEX2, cysteine was added to the buffer to create a reducing environment. Interestingly, the labeling efficiency of APEX2 is generally decreased but the overall trend over various distance remained the same (Supplementary Fig. 10). The labeling efficiency of APEX2 in 98% of heavy water showed a similar drop.

Overall, APEX labeling at DNA origami template gave two features. Expectedly, it was highly distance dependent, with intermolecular samples essentially unlabeled and strong labeling within 70 nm.

Paradoxically, labeling around 10 nm gave an abruptly high signal, which was even stronger than at the point of origin.

dsDNA is a fine ruler for measuring the distance of APEX2 labeling

We were intrigued by the APEX2 labeling behavior in the short range within 20 nm, where the maximal labeling happened at around 10 nm on origami. As it has become impractical to measure on origami with smaller step size, we turned to a dsDNA scaffold system to fine-tune the distance between the enzyme and the labeling targets with base-pair-level precision. Briefly, a phenol group was labeled at the 5' of ssDNA, while APEX2 was labeled at the 5' or a designated position in the middle of its complementary strand (Fig. 3a,b and Supplementary Fig. 11a,b). Because of the rigidity of dsDNA with a persistence length of 50 nm (refs. 33,34), when the two complementary DNAs hybridized to form a double strand, APEX2 and the phenol group could be well separated at designated distances. In this setup, the active center of APEX2 in the dsDNA system was about 7 nm away from its conjugation site on DNA. By adjusting the anchoring position of the APEX2 on the dsDNA structure, the distance to the phenol group could be modulated with a 1-nm step size (Fig. 3a).

Using this setup, the APEX2–dsDNA–phenol group systems with varying distances from 0 to 11 nm (Supplementary Tables 6 and 7) were pooled in the same tube. The labeling reaction was initiated by adding biotin-phenol and hydrogen peroxide. Following a 30-min incubation, the subsequent steps including double-strand extension, enrichment and qPCR were performed to calculate the labeling efficiency of APEX2 on the phenol group across different distances (Supplementary Fig. 12). The APEX2–dsDNA system gave consistent results compared to that of the APEX2–DNA origami system, demonstrating high labeling efficiency within 12 nm (Fig. 3c,d). In this setup, the highest labeling efficiency (8%) took place at 6 nm and 7 nm, which was approximately fourfold greater than that at other positions (2%). In intermolecular control, the labeling yield was the lowest (1%). Notably, there appeared to be sharp drops between the peaks at 6–7 nm and nearby spots at 4 or 8 nm. The dsDNA results corroborated the origami results, suggesting a labeling mechanism that deviates from simple diffusion of reactive phenol radicals, which should give a monotonic decay in labeling yield over distance.

APEX labeling has a proximity labeling portion and a contact labeling portion in parallel

It has been suggested that APEX2 labeling occurs through the interaction of the released phenoxy radicals with surrounding phenol groups. Supposedly, the labeling efficiency should be determined by the local concentration of phenoxy radicals diffused from the active center of APEX2. We used Brownian motion to simulate the distance-dependent diffusion of phenoxy radicals generated at the APEX2 active center. The concentration change of phenoxy radicals is described by equation (2), where $n(r,t)$ is the concentration of phenoxy radicals at a distance r from the APEX2 active center, D is the diffusion coefficient and t is the diffusion time³⁵.

$$n(r,t) = \frac{1}{(4\pi Dt)^{\frac{3}{2}}} \exp\left(-\frac{r^2}{4Dt}\right) \quad (2)$$

A prior study suggested a half-life of 0.1 ms for phenoxy radicals in the aqueous environment and in the absence of quenchers such as glutathione³⁶. Starting with a reported diffusion coefficient of phenol at $1,000 \mu\text{m}^2 \text{s}^{-1}$ (ref. 37), we inferred the diffusion coefficient of the biotin-phenoxy radical to be about $200 \mu\text{m}^2 \text{s}^{-1}$ on the basis of their molecular weights³⁸. Our simulation results revealed that the diffusion radius was within 500 nm (Supplementary Fig. 13a), aligned with labeling radius measured with stimulated emission depletion (STED) imaging (~ 268 nm)²¹. Our simulation also indicated that

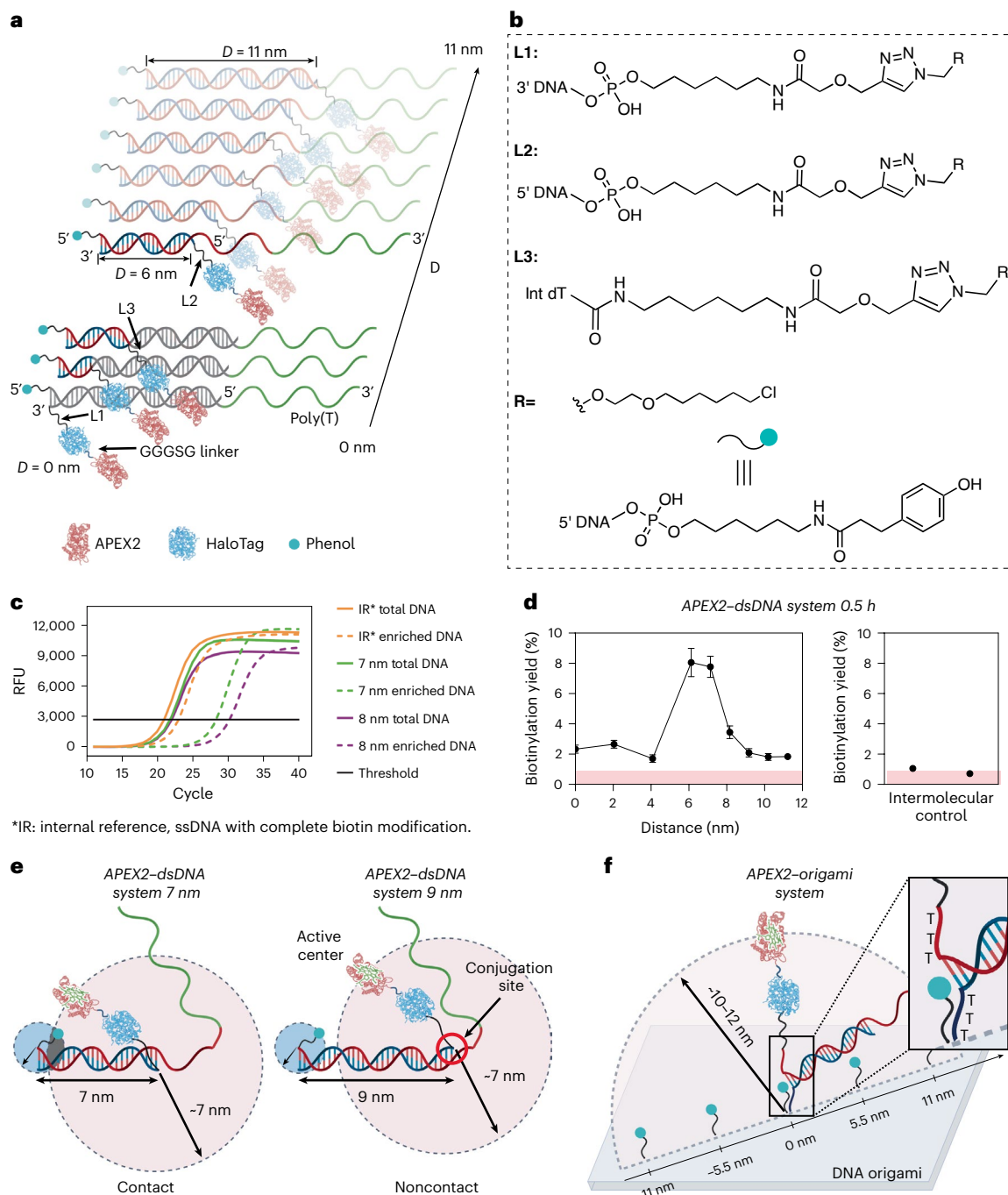


Fig. 3 | dsDNAs as finer nanorulers to accurately profile the labeling radius of APEX2. **a**, Schematic diagram of dsDNA-templated APEX2 with phenol groups at different distances. D refers to the distance between phenol and the conjugation sites of APEX2 on dsDNA. **b**, Chemical structures of linkers in **a**. **c**, Representative qPCR curves for quantifying labeling efficiencies. RFU, relative fluorescent units. **d**, Labeling efficiencies of APEX2 to phenol groups with different distances

spaced by dsDNA. Data are presented as the mean values $\pm$ s.e.m. ($n = 3$ biological replicates for each group). Pink areas represent the level of background intermolecular labeling. **e**, Schematic diagram of the contact-dependent labeling model (left) and the noncontact-dependent labeling model (right) of APEX2 in dsDNA system. **f**, Schematic diagram of APEX2 on DNA origami, showcasing a longer contact labeling radius attributed to longer triple T linkers.

the concentration of phenoxyl radicals decreased slightly (less than 6%) within 70 nm, suggesting that the labeling efficiency of APEX2 remained relatively constant within short distance, which aligns with our findings within 22–66 nm of the origami system and 9–11 nm in the dsDNA system (Figs. 2e and 3d and Supplementary Fig. 9). However, the high labeling efficiency observed at 5.5–11 nm in the origami system and 6–7 nm in the dsDNA system contradicted the diffusion model, indicating an alternative labeling mechanism beyond phenoxyl radical diffusion labeling.

Taking the size of APEX2 (Supplementary Fig. 14a)³⁹ and the linker length into account, the distance between the active center of APEX2 and the DNA conjugation site was approximately 10–12 nm in the APEX2–DNA origami system (Supplementary Fig. 14b). Consequently, the phenol group positioned at 5.5 nm and 11 nm away in the APEX2–DNA origami system could reach the APEX2 active center with a preferred conformation (Fig. 3f). Similarly, the phenol groups at 6 nm and 7 nm in the APEX2–dsDNA system were also capable of direct contact labeling with the APEX2 active center (Fig. 3e and

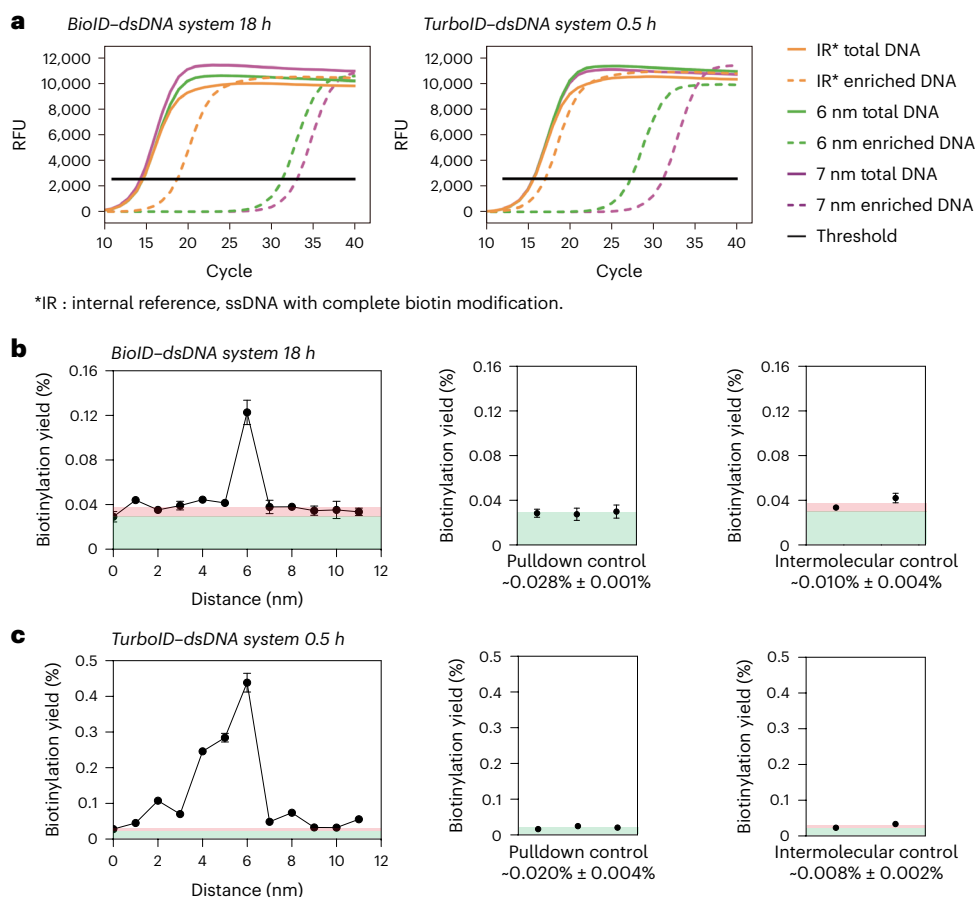


Fig. 4 | Measuring the labeling radius of BioID and TurboID by dsDNA-nanorulers. **a**, Representative qPCR curves for quantifying labeling efficiencies of BioID and TurboID to amines with different distances spaced by dsDNA. **b**, Biotinylation yield–distance plot of BioID. Data are presented as the mean

values ± s.e.m. ($n = 3$ biological replicates for each group). **c**, Biotinylation yield–distance plot of TurboID. Data are presented as the mean values ± s.e.m. ($n = 3$ biological replicates for each group). Pink areas represent the intermolecular labeling background and green areas represent the pulldown background.

Supplementary Fig. 14c). Therefore, we attributed the enhanced labeling efficiency at 5.5–11 nm in the origami system and 6–7 nm in the dsDNA system to the direct oxidation of phenol with the Fe^{IV} heme complex at the active center of APEX2, which is substantially faster than a relay oxidation from a heme-sensitized phenoxy radical. The resulting phenoxyl radical would then engage in diradical recombination with the biotin-phenoxyl radicals present in the solution⁴⁰, finishing the biotin labeling. Notably, while the APEX2–HaloTag protein remained the same, in the dsDNA system, the redundant triple T linkers in the origami system were removed (Fig. 3e) and the chemical linkers for click chemistry were further optimized (Supplementary Figs. 3 and 11), giving a shorter effective linker than in the origami. Such tuning in linkers matched exactly with the difference in labeling radius (6–7 nm on dsDNA versus 5–11 nm on origami), further supporting the contact mechanism. In summary, our findings inferred that APEX2 exhibits a dual labeling mechanism, featuring both a previously appreciated diffusive labeling process that covers the distal phenols within 500 nm and a newly unveiled strong contact labeling within 10 nm.

Measuring the distance of BioID and TurboID using dsDNA ruler

We further used the DNA nanoruler to investigate the labeling distance and mechanism of BioID and TurboID, both of which are engineered enzymes derived from the biotin ligase BirA. Previous studies have shown that BioID can label proteins within a distance of 10 nm in cells^{5,20}. Therefore, we chose the fine ruler of dsDNA system functioning within 11 nm to measure the labeling distance of these two enzymes.

TurboID can modify the lysine residues of proteins and aliphatic primary amines on oligonucleotides⁴¹. Therefore, we adapted the dsDNA scaffold from the previous setup, replacing phenol groups with amino groups to mimic lysine residues and substituted APEX2 with BioID or TurboID (Supplementary Fig. 15). The conjugation of BioID and TurboID to ssDNA was confirmed by SDS–PAGE (Supplementary Fig. 16). The reaction buffer was changed to sodium tetraborate buffer (0.1 M $\text{Na}_2\text{B}_4\text{O}_7$, pH 8.5) to deprotonate and enhance the reactivity of the amino groups. Proximity labeling was initiated by the addition of biotin and ATP. After the labeling reaction, Tris (10 mM) was added to terminate the reaction and the labeling efficiency was assessed using qPCR following the same workflow (Supplementary Fig. 12).

In the BioID–dsDNA system, there is a highest labeling efficiency of ~0.1% at 6 nm after an 18-h reaction (Fig. 4a,b), with undetectable labeling at shorter times (Supplementary Fig. 17). The pulldown background, which indicated the nonspecific absorption of unlabeled DNA by streptavidin beads, was only ~0.028% ± 0.001%. The intermolecular labeling of BioID to another amino-ssDNA (2 nM in solution) was even less (~0.010% ± 0.004%). The labeling efficiency at 6 nm was much higher than the pulldown background while the labeling at other distances was comparable to intermolecular labeling. In the TurboID–dsDNA system, we observed a high labeling efficiency within 6 nm after a reaction time of 0.5 h. However, the labeling efficiency dropped sharply beyond 6 nm, approaching background levels (Fig. 4a,c). Similar results were observed for extended labeling times (Supplementary Fig. 18). These results are consistent with the previous studies showing that TurboID has a higher labeling efficiency than

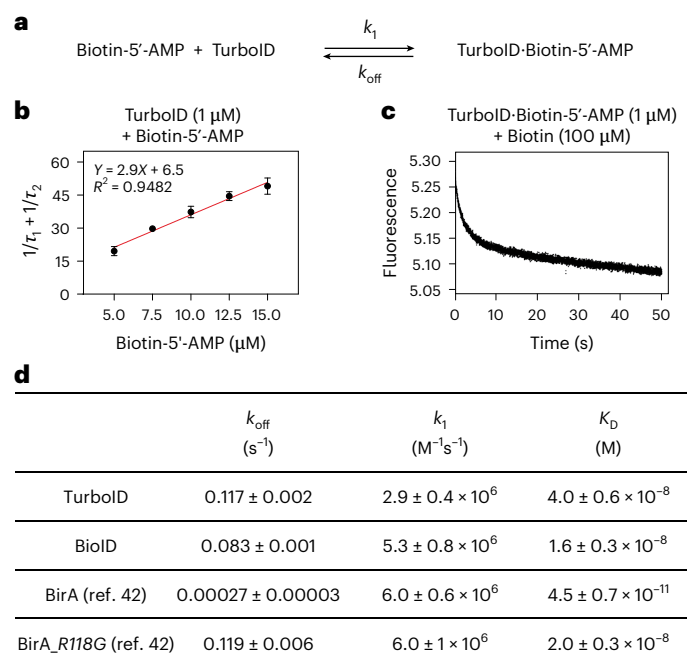


Fig. 5 | Enzymatic parameters of TurboID. **a**, Chemical equation of association and dissociation of TurboID with biotin-5'-AMP. **b**, Dependence of the sum of the apparent rates ($1/\tau_1$ and $1/\tau_2$) of the two kinetic phases on ligand concentration for TurboID, where τ_1 and τ_2 are the relaxation times for the first and second kinetic phases⁴², respectively (Methods). Data are presented as the mean values ± s.e.m. ($n = 3$ biological replicates for each group). **c**, Dissociation of the complexes of TurboID bound to biotin-5'-AMP measured by stopped-flow fluorescence. **d**, Kinetic and thermodynamic parameters of biotin-5'-AMP binding to TurboID, BirA and BioID.

BioID¹⁷. Interestingly, the labeling efficiency dropped substantially at 7 nm in both the BioID–dsDNA system and the TurboID–dsDNA system, which deviates from the previous perceived labeling mechanism. Moreover, labeling within 6 nm was paradoxically less efficient than at 6 nm. We speculate that the labeling of BioID and TurboID functions through a contact-dependent mechanism instead of the diffusion of active biotin-5'-AMP; at 6 nm, BioID and TurboID holding the biotin-5'-AMP could physically reach the amino group with optimal geometry (Supplementary Fig. 19a), resulting in a highest labeling efficiency. Labeling within 6 nm was sterically unfavored, whereas labeling beyond 6 nm was out of range, resulting in negligible labeling (Supplementary Fig. 19b).

Enzymology of TurboID supports a contact labeling mechanism

It was previously suggested that biotin-5'-AMP is first generated by TurboID using ATP and biotin as precursors, followed by the release of biotin-5'-AMP, which labels distal amines. Our data from the DNA ruler, however, suggested a strong preference for contact labeling. To further understand the labeling profile of TurboID, we measured its affinity with biotin-5'-AMP. Following an established protocol⁴², we chemically synthesized biotin-5'-AMP (Supplementary Fig. 20) and determined its association and dissociation rate constants with BioID and TurboID using stopped-flow spectroscopy (Fig. 5a–c and Supplementary Fig. 21). The association between TurboID and biotin-5'-AMP was as fast as that between BirA and BioID (Fig. 5d). Surprisingly, the dissociation kinetics (k_{off}) of BioID and TurboID were relatively close to each other. Consequently, the dissociation constant K_D for the BioID-biotin-5'-AMP complex was measured to be 1.6×10^{-8} M (Fig. 5d and Supplementary Figs. 21 and 22), whereas TurboID-biotin-5'-AMP exhibited a K_D of 4.0×10^{-8} M (Fig. 5d). The measured K_D values of

TurboID-biotin-5'-AMP and BioID-biotin-5'-AMP were comparable to that reported for BirA-R118G in a previous study⁴². Upon the generation of biotin-5'-AMP in the pockets of BioID and TurboID, the lifetime of the complex was on the order of 10 s before the dissociation.

Although TurboID was devised with drastically higher labeling efficiency than BioID, the underlying mechanism was elusive. One speculation from the original TurboID paper was that it offers faster dissociation with biotin-5'-AMP¹⁷. However, our data, showcasing that TurboID was only marginally more prone to biotin-5'-AMP release than BioID, contradicted this speculation. Looking at the distance profile, BioID and TurboID gave consistent labeling preference of reachable amines, suggesting that most of the labeling reactions between biotin-5'-AMP and amines happened within the BioID or TurboID pocket thanks to the relatively long dissociation lifetime (~10 s). Overall, the enzymology of TurboID revealed a tight binding with biotin-5'-AMP, supporting the contact labeling mechanism unveiled with DNA rulers.

Discussion

In this study, we developed a DNA nanostructure-based platform to precisely control the distance between enzymes and labeling targets, allowing the reconstruction of proximity labeling systems in vitro. This design allows for precise spatial control and ensures that all reactive groups are presented in a uniform and reproducible chemical environment—an experimental condition that is difficult to achieve on native proteins. By using qPCR to measure the relative labeling efficiency at different distances, we elucidated new aspects of the proximity labeling mechanism that challenges the current paradigm. Our results suggest that APEX2 labeling involves both contact and noncontact mechanisms (Fig. 6a), whereas BioID and TurboID labeling is primarily through a contact-based mechanism within the reach of ligase (Fig. 6b). Such universal contact-based labeling is substantiated with solid evidence using DNA rulers. From a proteomics perspective, our findings provide practical guide into selecting appropriate proximity labeling methods and insights for data interpretation. TurboID is recommended for more accurate identification of interacting proteins because of its strict contact-dependent labeling, whereas APEX2 is better suited for longer-distance labeling.

Mechanistically, this work addresses several long-standing controversies. Firstly, the half-life of biotin-5'-AMP at neutral pH should be minutes⁴³, which would confer BioID and TurboID with a long-range proximity labeling method if they function through diffusive labeling mechanism (Supplementary Fig. 23). However, multiple studies have suggested that BioID and TurboID function within 10 nm (refs. 1,20). The working model proposed in this work, where BioID and TurboID function through the direct labeling of adjacent amines with the BioID-biotin-5'-AMP or TurboID-biotin-5'-AMP complex, reconciles with previous data showing that BioID and TurboID labeling happens predominantly within 10 nm. Notably, the contact-based labeling mechanism proposed for BioID and TurboID in this study resembles the endogenous 'spray-like' post-translational modification mechanisms observed with writer enzymes such as p300/CBP and PARP1 (ref. 44). This similarity suggests that contact-dependent enzymatic modifications may be a broadly used and functionally essential mechanism in biological systems. Secondly, for APEX and peroxidase labeling, while electron microscopy suggested a labeling resolution of ~20 nm (refs. 6,18) and STED microscopy gave a labeling radius of ~270 nm (ref. 21), our data provide a unified model. The heavy signal within 20 nm can be attributed to the direct contact sensitization from heme, where the diffusive labeling of phenol radical gives a small portion of distal signal with a radius of ~270 nm. Overall, this work adds to the few approaches to measuring the distance parameter of proximity labeling. The in vitro labeling patterns herein unveil previously uncharted mechanistic insights that reconcile the existing evidence obtained using different approaches.

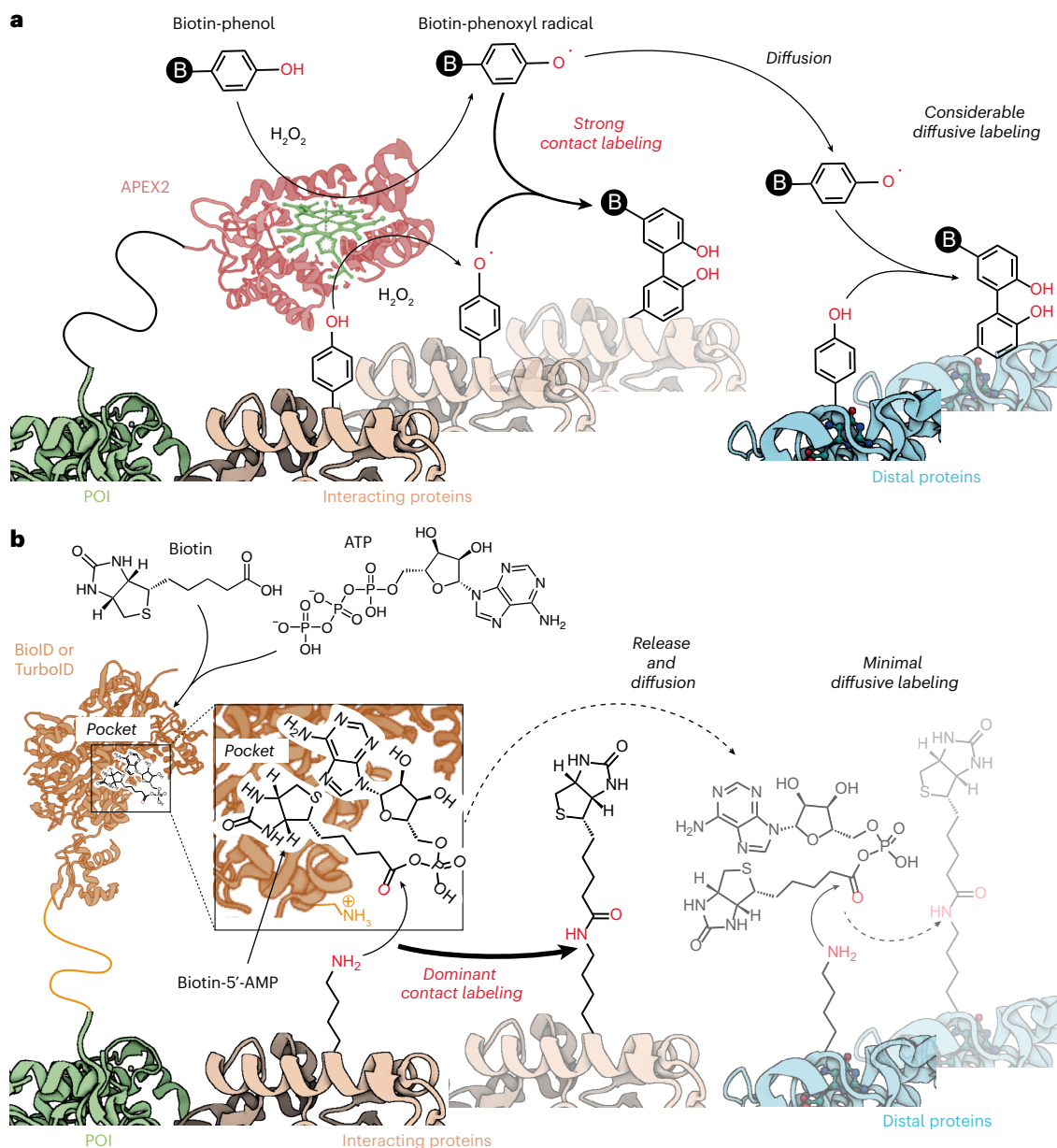


Fig. 6 | Proposed labeling profile of APEX2 and TurboID, whose distinctive mechanisms give rise to different spatial preference. a, APEX2 involves a combination of contact and diffusive labeling. In parallel to the widely appreciated phenol radical-mediated diffusive labeling, the proximity-induced oxidation of adjacent phenols directly from the heme is favored, giving strong

contact labeling that has been largely neglected in previous mechanistic studies. **b,** BioID and TurboID function primarily through contact-dependent labeling, where the generated biotin-5'-AMP is mainly consumed by adjacent amino groups from interacting proteins before release from the pocket, giving little diffusive labeling at distal proteins. POI, protein of interest.

From the view of chemical tools, the nanoruler platform we established through DNA nanostructures proved versatile for measuring the diffusion of active molecules at the nanoscale. Complementing the fluorescence resonance energy transfer dynamic ranges of 1–10 nm (ref. 45) and super-resolution imaging that offers precise measurements at >20 nm (refs. 46,47), this study exemplifies the power of DNA rulers for profiling microscopical patterns in the 1–100-nm window^{48,49}. Notably, this work integrates a qPCR-based DNA quantification scheme with the DNA origami platform, capitalizing on the precise addressability of DNA origami structures. This work opens new avenues for investigating biochemical reactions within microenvironments and provides fresh insights into detecting complex reaction intermediates. While earlier studies relied on AFM to image single-molecule reactions on DNA origami platforms^{25,26}, we advocate to introduce detection methods from modern molecular biology such as qPCR with boosted

sensitivity. Additionally, the use of DNA origami structures with internal cavities—such as rectangular origami with voids⁵⁰, U-shaped DNA origami⁵¹ or DNA origami clamps⁵²—could enable physical separation of enzymes and labeling targets to eliminate potential ‘relay’ effects of the origami scaffold on intermediate diffusion. This represents an important future direction for optimization and refinement. Overall, we foresee a broad potential for DNA nanotechnology to serve as a robust quantitative tool in biochemical research.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-025-02086-w>.

References

1. Qin, W., Cho, K. F., Cavanagh, P. E. & Ting, A. Y. Deciphering molecular interactions by proximity labeling. *Nat. Methods* **18**, 133–143 (2021).
2. Zhou, Y. & Zou, P. The evolving capabilities of enzyme-mediated proximity labeling. *Curr. Opin. Chem. Biol.* **60**, 30–38 (2021).
3. Zhang, S., Tang, Q., Zhang, X. & Chen, X. Proximitomics by reactive species. *ACS Cent. Sci.* **10**, 1135–1147 (2024).
4. Kang, M.-G. & Rhee, H.-W. Molecular spatiomics by proximity labeling. *Acc. Chem. Res.* **55**, 1411–1422 (2022).
5. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
6. Martell, J. D. et al. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).
7. Liu, Q. et al. A proximity-tagging system to identify membrane protein–protein interactions. *Nat. Methods* **15**, 715–722 (2018).
8. Wang, P. et al. Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling. *Nat. Chem. Biol.* **15**, 1110–1119 (2019).
9. Geri, J. B. et al. Microenvironment mapping via dexter energy transfer on immune cells. *Science* **367**, 1091–1097 (2020).
10. Liu, Z. et al. Detecting tumor antigen-specific T cells via interaction-dependent fucosyl-biotinylation. *Cell* **183**, 1117–1133 (2020).
11. Liu, H. et al. Antigen-specific T cell detection via photocatalytic proximity cell labeling (phoxcell). *J. Am. Chem. Soc.* **144**, 5517–5526 (2022).
12. Zhang, X. et al. Cellular-scale proximity labeling for recording cell spatial organization in mouse tissues. *Sci. Adv.* **9**, eadg6388 (2023).
13. Liu, Z. et al. Bioorthogonal photocatalytic proximity labeling in primary living samples. *Nat. Commun.* **15**, 2712 (2024).
14. Pani, S., Qiu, T., Kentala, K., Azizi, S.-A. & Dickinson, B. C. Bioorthogonal masked acylating agents for proximity-dependent RNA labelling. *Nat. Chem.* **16**, 717–726 (2024).
15. Lin, Z. et al. Multiscale photocatalytic proximity labeling reveals cell surface neighbors on and between cells. *Science* **385**, ead15763 (2024).
16. Lam, S. S. et al. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2015).
17. Branon, T. C. et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).
18. Mayer, G. & Bendayan, M. Amplification methods for the immunolocalization of rare molecules in cells and tissues. *Prog. Histochem. Cytochem.* **36**, 3–84 (2001).
19. Kotani, N. et al. Biochemical visualization of cell surface molecular clustering in living cells. *Proc. Natl Acad. Sci. USA* **105**, 7405–7409 (2008).
20. Kim, D. I. et al. Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl Acad. Sci. USA* **111**, E2453–E2461 (2014).
21. Oakley, J. V. et al. Radius measurement via super-resolution microscopy enables the development of a variable radii proximity labeling platform. *Proc. Natl Acad. Sci. USA* **119**, e2203027119 (2022).
22. Erkelenz, M., Kuo, C.-H. & Niemeyer, C. M. DNA-mediated assembly of cytochrome p450 bm3 subdomains. *J. Am. Chem. Soc.* **133**, 16111–16118 (2011).
23. Fu, J., Liu, M., Liu, Y., Woodbury, N. W. & Yan, H. Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures. *J. Am. Chem. Soc.* **134**, 5516–5519 (2012).
24. Fu, J. et al. Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.* **9**, 531–536 (2014).
25. Voigt, N. V. et al. Single-molecule chemical reactions on DNA origami. *Nat. Nanotechnol.* **5**, 200–203 (2010).
26. Helmig, S. et al. Single molecule atomic force microscopy studies of photosensitized singlet oxygen behavior on a DNA origami template. *ACS Nano* **4**, 7475–7480 (2010).
27. Castro, C. E. et al. A primer to scaffolded DNA origami. *Nat. Methods* **8**, 221–229 (2011).
28. Fazal, F. M. et al. Atlas of subcellular RNA localization revealed by APEX-seq. *Cell* **178**, 473–490 (2019).
29. Zhou, Y. et al. Expanding APEX2 substrates for proximity-dependent labeling of nucleic acids and proteins in living cells. *Angew. Chem. Int. Ed. Engl.* **58**, 11763–11767 (2019).
30. Strauss, M. T., Schueder, F., Haas, D., Nickels, P. C. & Jungmann, R. Quantifying absolute addressability in DNA origami with molecular resolution. *Nat. Commun.* **9**, 1600 (2018).
31. Shi, R. & Chiang, V. L. Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* **39**, 519–525 (2005).
32. Ke, M. et al. Spatiotemporal profiling of cytosolic signaling complexes in living cells by selective proximity proteomics. *Nat. Commun.* **12**, 71 (2021).
33. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
34. Manning, G. S. The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophys. J.* **91**, 3607–3616 (2006).
35. Pathria, R. K. & Beale, P. D. In *Statistical Mechanics* (eds Pathria, R. K. & Beale, P. D.) 599–657 (Elsevier, 2022).
36. Folkes, L. K., Trujillo, M., Bartesaghi, S., Radi, R. & Wardman, P. Kinetics of reduction of tyrosine phenoxyl radicals by glutathione. *Arch. Biochem. Biophys.* **506**, 242–249 (2011).
37. Winkelmann, J. in *Diffusion in Gases, Liquids and Electrolytes*, Vol 15B2 (ed. Lechner, M. D.) 2389–2390 (Springer, 2018).
38. Edward, J. T. Molecular volumes and the Stokes–Einstein equation. *J. Chem. Educ.* **47**, 261 (1970).
39. Sharp, K. H., Moody, P. C. E., Brown, K. A. & Raven, E. L. Crystal structure of the ascorbate peroxidase–salicylhydroxamic acid complex. *Biochemistry* **43**, 8644–8651 (2004).
40. Hope, T. O. et al. Targeted proximity-labelling of protein tyrosines via flavin-dependent photoredox catalysis with mechanistic evidence for a radical–radical recombination pathway. *Chem. Sci.* **14**, 7327–7333 (2023).
41. Wilbanks, B. et al. DNA modifications enabling proximity biotinylation. *Bioconjug. Chem.* **34**, 611–615 (2023).
42. Kwon, K. & Beckett, D. Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci.* **9**, 1530–1539 (2000).
43. DeMoss, J. A., Genuth, S. M. & Novelli, G. D. The enzymatic activation of amino acids via their acyl-adenylate derivatives. *Proc. Natl Acad. Sci. USA* **42**, 325–332 (1956).
44. Lee, Y.-B. & Rhee, H.-W. Spray-type modifications: an emerging paradigm in post-translational modifications. *Trends Biochem. Sci.* **49**, 208–223 (2024).
45. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl Acad. Sci. USA* **58**, 719–726 (1967).
46. Sigal, Y. M., Zhou, R. & Zhuang, X. Visualizing and discovering cellular structures with super-resolution microscopy. *Science* **361**, 880–887 (2018).
47. Sahl, S. J., Hell, S. W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat. Rev. Mol. Cell Biol.* **18**, 685–701 (2017).
48. Mao, X., Liu, M., Li, Q., Fan, C. & Zuo, X. DNA-based molecular machines. *JACS Au* **2**, 2381–2399 (2022).

49. Zhan, P. et al. Recent advances in DNA origami-engineered nano-materials and applications. *Chem. Rev.* **123**, 3976–4050 (2023).
50. Endo, M., Katsuda, Y., Hidaka, K. & Sugiyama, H. Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure. *J. Am. Chem. Soc.* **132**, 1592–1597 (2010).
51. Nickels, P. C. et al. Molecular force spectroscopy with a DNA origami-based nanoscopic force clamp. *Science* **354**, 305–307 (2016).
52. Funke, J. J. & Dietz, H. Placing molecules with Bohr radius resolution using DNA origami. *Nat. Nanotechnol.* **11**, 47–52 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2025

Methods

Construction of plasmids

For cloning, PCR fragments of HaloTag–APEX2, HaloTag–BioID, HaloTag–TurboID, BioID and TurboID with a 6×His tag at the C terminus were amplified using Phanta polymerase (Vazyme, P520-01). Then, the fragments were cloned into pET21a vectors by homemade Gibson assembly reagents.

Expression of recombinant proteins

Escherichia coli BL21 (DE3) cells were transformed with HaloTag–APEX2, HaloTag–BioID, HaloTag–TurboID, BioID and TurboID constructs. Then, these cells were cultured in 500 ml of LB medium containing 100 µg ml^{−1} of ampicillin at 37 °C for 4 h. Once the optical density at 600 nm reached 0.5, as measured by a Tecan plate reader, protein expression was induced by the addition of 0.5 mM IPTG. Subsequently, the cultures were incubated at 16 °C for 20 h. The cells were harvested by centrifugation at 5,000 rpm for 10 min. After resuspending in binding buffer (50 mM Tris and 300 mM NaCl, pH 7.5), the cells were lysed by sonication for 30 min. The lysate was then centrifuged at 11,000g for 30 min at 4 °C to collect the supernatant, which was then incubated with a gravity column containing 2 ml of Ni-NTA agarose beads. Excess liquid was removed by gravity flow. The beads were then washed sequentially with wash buffers, first with a buffer containing 10 mM imidazole, followed by one with 20 mM imidazole. The recombinant protein was eluted using an elution buffer (50 mM Tris, 300 mM NaCl and 200 mM imidazole, pH 7.5). The purified protein was concentrated using a 10-kDa ultrafiltration tube and the buffer was exchanged for PBS. Finally, the proteins were aliquoted and stored at −80 °C for further use.

Preparation of staple strands modified with phenol group

The amino-modified ssDNA was ordered from Sangon Biotech and dissolved in 0.1 M Na₂B₄O₇ pH 8.5 to a concentration of 100 µM. It was then mixed with 100 equivalents of phenol-NHS (CAS: 34071-95-9, dissolved in DMSO). The mixture was shaken at 1,000 rpm overnight at room temperature. The phenol-modified ssDNA was recovered using an ethanol precipitation method. Specifically, 0.1 volumes of 3 M sodium acetate (pH 5.2) was added to the reaction solution, followed by 2.5 volumes of anhydrous ethanol. The mixture was placed on ice for 30 min and subsequently centrifuged at 12,000 rpm for 30 min at 4 °C. The supernatant was carefully discarded and the pellet was washed twice with 70% ethanol. The tube cap was left open for 5 min to allow residual alcohol to evaporate. The pellet was resuspended in ultrapure water. Finally, the phenol-modified ssDNA was verified by MS (LTQ XL linear ion trap MS instrument, Thermo) and stored at −20 °C before use.

Preparation and purification of DNA origami

The rectangular DNA origami sequence was based on a previous study⁵³. The staple strands were purchased from Sangon Biotech. M13mp18 scaffold strand was produced in-house according to previous work^{54,55}. A 50-µl solution was prepared, containing M13mp18 scaffold strand (10 nM), phenol-modified staple strands (200 nM), APEX2-anchoring strand (200 nM), internal reference strand (200 nM) and other staple strands (100 nM), all in a TAE-Mg²⁺ buffer (40 mM Tris–acetic acid pH 8.0, 2 mM EDTA and 12.5 mM MgCl₂). The solution was annealed using the following program: 95 °C for 15 min, then cooled to 20 °C at a rate of 1 °C min^{−1}. The assembled origami was purified with PEG-8000 to remove the excess staple strand⁵⁶. Specifically, an equal volume (50 µl) of precipitation buffer (15% PEG-8000 (w/v), 5 mM Tris pH 8.0, 1 mM EDTA and 505 mM NaCl) was added. The mixture was then placed at 4 °C for 10 min and centrifuged at 5,000g for 20 min. The supernatant was gently discarded and the pellet was resuspended in 50 µl of TAE-Mg²⁺ buffer. The above process was repeated twice; the pellet was resuspended in 50 µl of TAE-Mg²⁺ buffer and stored at −20 °C.

Preparation of APEX2–ssDNA

First, we prepared ssDNA conjugated with a HaloTag ligand. Alkynyl-NHS (CAS: 1858242-47-3) and NH₂-HaloTag ligand (CAS: 1035373-85-3) were mixed in DMSO at a ratio of 1:10 and reacted at room temperature for 30 min. Then, ssDNA-azide (poly-LGT-N₃) was added to the mixture. A solution (50 µl) containing ssDNA-azide (100 µM), alkynyl-NHS (1 mM), NH₂-HaloTag ligand (10 mM), THPTA (2 mM), CuSO₄ (1 mM) and sodium ascorbate (5 mM) was reacted at room temperature for 1 h at 1,000 rpm. Ethanol precipitation was then performed according to the previous method to prepare the ssDNA–HaloTag ligand. The conjugate was resuspended in PBS and stored at −20 °C.

Subsequently, ssDNA–HaloTag ligand (100 µM) and APEX2–HaloTag (50 µM) were mixed and reacted at room temperature for 1 h. The mixture was purified using a Ni-NTA column to remove unreacted ssDNA–HaloTag ligand, followed by ion-exchange chromatography to separate out any unreacted proteins. For ion-exchange chromatography, the APEX2–ssDNA conjugate was concentrated using a 10-kDa ultrafiltration tube and the buffer was exchanged with buffer A (50 mM Tris pH 7.5). The mixture was then loaded onto an anion-exchange column in ÄKTA Pure chromatography system and eluted using gradient of buffer B (50 mM Tris and 1 M NaCl, pH 7.5), with the gradient increasing from 0% to 30% over 5 min, then 40% to 70% over 30 min and finally 70% to 100% over 5 min. The APEX2–ssDNA conjugate product was concentrated and buffer-exchanged to PBS using a 10-kDa ultrafiltration tube. The conjugate was stored at −80 °C before use.

Assembly and purification of APEX2–origami

The APEX2–ssDNA conjugate was incubated with DNA origami at room temperature for 30 min, with the conjugate at a fivefold excess. Unbound APEX2–ssDNA was removed using a 100-kDa ultrafiltration tube by centrifuging it five times at 4,000g for 10 min each. Finally, the buffer was exchanged with a reaction buffer (PBS and 12.5 mM MgCl₂) using an ÄKTA Pure chromatography system and the sample was stored at −20 °C before performing subsequent reactions.

AFM imaging

For topographic AFM imaging under liquid conditions, a 15 µl of 0.5 nM sample was deposited onto a freshly cleaved mica surface for 1 min followed by adding 1.5 µl of 100 mM NiCl₂ onto the mica for 30 s. Silicon nitride probe (ScanAsyst-Fluid+, Bruker) with a tip radius of 2 nm and a nominal spring constant of 0.7 N m^{−1} was typically used. Imaging was performed at room temperature in peak force tapping mode on BioScope Resolve AFM (Bruker) with a resolution of 256 × 256 pixels at a scan rate of 0.5 Hz. A peak force amplitude of 25 nm was used and the force imparted on the sample during imaging was 80 pN. All AFM images were processed and analyzed with NanoScope Analysis software (Bruker, version 1.8).

Verification of qPCR specificity

To verify the specificity of primer A for a specific dsDNA A, all functional strands were mixed and recorded as sample 1. dsDNA A was removed from sample 1, creating sample 2. Both sample 1 and sample 2 were subjected to qPCR using P141 and specific primer A to observe the amplification and melting curves.

Workflow

Proximity labeling using APEX2–DNA origami system (step 1). To initiate the reaction, 500 µM biotin-phenol and 1 mM hydrogen peroxide were added to the prepared APEX2–origami sample and allowed to react for 30 min at room temperature. The reaction was terminated by the addition of 10 mM Trolox and 10 mM sodium ascorbate. The mixture was then purified using PEG-8000 precipitation⁵⁶. The pellet was resuspended in 20 µl of TAE-Mg²⁺ buffer (40 mM Tris–acetic acid pH 8.0, 2 mM EDTA and 12.5 mM MgCl₂).

Displacement and purification (step 2). The staple strand was separated from the DNA origami by strand displacement. To the resuspended sample in step 1, 2 μl of primer F-5.2, 1 μl of Deep Vent DNA polymerase (New England Biolabs, M0259) and 5 μl of 10 $\times$ buffer were added, with water added to reach a total volume of 50 μl . The reaction proceeded at 72 °C for 1 h. The reaction sample was purified and separated using the DNA cleanup kit (New England Biolabs, T1030). Briefly, the mixture (50 μl) was combined with 200 μl of binding buffer, loaded onto a column (in DNA cleanup kit) and centrifuged for 1 min at 13,000 rpm to collect the filtrate. Then, 50 μl of water and 600 μl of anhydrous ethanol were added to the filtrate. The mixture was transferred to a new column and centrifuged for 1 min at 13,000 rpm. The column was washed twice with wash buffer and finally eluted with 20 μl of ultrapure water.

DNA extension (step 3). We first optimize the condition of staple extension. The functional staple strands were mixed and 2 μl of TL141-AA (10 μM) and 40 μl of 2 $\times$ Phanta Max (Vazyme, P520-01) were added to the mixture. The mixture was then extended for 1 min at each of the following temperatures: 35 °C, 40 °C, 50 °C, 60 °C and 72 °C. The sample was then purified using a DNA cleanup kit and verified by a 10% PAGE gel.

To the purified sample in step 2, 2 μl of TL141-AA (10 μM) and 40 μl of 2 $\times$ Phanta Max were added and then extended at 50 °C for 1 min. The mixture was purified using DNA cleanup kit and finally eluted with 20 μl of ultrapure water, recorded as Input.

DNA enrichment (step 4). Streptavidin beads were blocked with a blocking buffer (1 μg μl^{-1} random DNA and 1 μg μl^{-1} BSA) for 2 h. Then, 0.6 μl of blocked streptavidin beads and 14.4 μl of 2 $\times$ loading buffer (100 mM Tris pH 7.5, 1 M NaCl, 10 mM EDTA and 0.2% v/v Tween-20) were added to 15 μl of the input sample and then the mixture was incubated at 1,000 rpm for 45 min. After removing the supernatant, the beads were washed three times with washing buffer (100 mM Tris pH 7.5, 4 M sodium chloride, 10 mM EDTA and 0.2% v/v Tween-20) at room temperature, twice with washing buffer at 50 °C for 10 min and twice with PBS. The beads were resuspended in 20 μl of elution buffer (95% v/v formamide, 10 mM EDTA and 1.5 mM biotin, pH 8.0) and incubated at 95 °C for 10 min. The supernatant was purified using a DNA cleanup kit and recorded as enrich.

qPCR (step 5). Both the input and enrich samples were subjected to qPCR (instrument: CFX96, Bio-Rad) using primer P141 and specific primers with the Takara qPCR kit (RR820A). The Ct values of each staple and internal reference were recorded to calculate the relative labeling efficiency of each site using equation (1).

Determination and calculation of amplification efficiency of each staple strand

The Input sample was diluted with EASY Dilution (Takara, 9160); then, the diluted samples underwent qPCR with P141 and corresponding primers. The fitting curve of N and Ct values was drawn according to equation (3) and then the amplification efficiency E was calculated.

$$\lg(N) = -Ct \times \lg(1 + E) + C \quad (3)$$

Preparation and purification of scaffold strands–HaloTag–APEX2, scaffold strands–HaloTag–TurboID and scaffold strands–HaloTag–BioID

Scaffold strands modified with 3'-amino and 5'-amino groups were ordered from Sangon Biotech, dissolved in 0.1 M $\text{Na}_2\text{B}_4\text{O}_7$ pH 8.5 and mixed in equimolar amounts. To the mixture, 100 equivalents of alkynyl-NHS (CAS: 1858242-47-3, dissolved in DMSO) was added and reacted at room temperature overnight at 1,000 rpm. The modified

scaffold strands were purified by ethanol precipitation as previously described, with the pellet resuspended in PBS and stored at –20 °C until needed.

HaloTag ligand conjugation with scaffold strands was performed using a click reaction. A solution (50 μl) containing alkynyl-modified scaffold strands (100 μM), azide-HaloTag ligand (CAS: 2568146-55-2, 10 mM), THPTA (2 mM), CuSO_4 (1 mM) and sodium ascorbate (5 mM) was reacted at room temperature for 1 h at 1,000 rpm. The mixture was purified by ethanol precipitation as previously described. The conjugate was resuspended in PBS and stored at –20 °C.

Subsequently, scaffold strands conjugated with the HaloTag ligand (100 μM) were mixed with APEX2–HaloTag (50 μM), TurboID–HaloTag (50 μM) or BioID–HaloTag (50 μM) and reacted at room temperature for 1 h. Unreacted ssDNA–HaloTag ligand was removed using a Ni-NTA column, followed by further purification of the sample using ion-exchange chromatography as previously described. The conjugates were stored at –80 °C until needed.

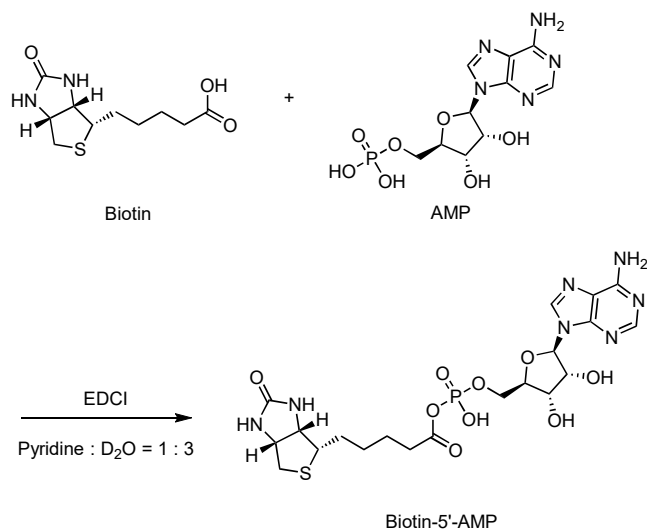
Measurement of TurboID and BioID labeling in dsDNA system

The amino-modified staple strands were mixed with three equivalents of either scaffold strands–HaloTag–TurboID or scaffold strands–HaloTag–BioID. The mixture was incubated at room temperature for 30 min, after which the buffer was replaced with 0.1 M $\text{Na}_2\text{B}_4\text{O}_7$ pH 8.5 and 12.5 mM Mg^{2+} using a desalting column in the ÄKTA Pure chromatography system. The final concentration of each amino-modified staple strand was 1 nM and the proximity labeling was initiated by adding 1 mM ATP and 1 mM biotin. The reaction was allowed to proceed for 30 min, 2 h and 18 h. The reaction was quenched with 10 mM Tris-HCl, followed by the addition of 300 μl of formamide to denature the dsDNA into ssDNA. Subsequently, the mixture was purified using a DNA cleanup kit and eluted with 20 μl of ultrapure water.

Next, 2 μl of TL141-AA (10 μM) and 40 μl of 2 $\times$ Phanta max were added to the purified sample and extended at 50 °C for 1 min. Then, 200 μl of lysis buffer (PBS and 300 mM NaCl) was added and scaffold strands–HaloTag–TurboID or scaffold strands–HaloTag–BioID were removed using a Ni-NTA column, with the filtrate collected. The filtrate was purified with a DNA cleanup kit and eluted with 20 μl of ultrapure water, recorded as input.

The pulldown and qPCR steps were the same as those used in the APEX2–origami system.

Synthesis of biotin-5'-AMP



Biotin (24.4 mg, 0.10 mmol, 1.0 eq) and AMP (38.2 mg, 0.11 mmol, 1.1 eq) were dissolved in 0.4 ml of a solution containing pyridine and D_2O in a 1:3 ratio at 0 °C. 1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide

hydrochloride (76.7 mg, 0.40 mmol, 4.0 eq) was then added to the mixture. The mixture was stirred at 0 °C for 3 h until the reaction was complete. The progress of the reaction was monitored by liquid chromatography (LC)–MS (4.6 mm × 150 mm, 3-μm C18 column; 2-μl injection; 5–100% CH₃CN/H₂O, linear gradient, with constant 0.1% v/v formic acid additive; 6.5-min run; 0.4 ml min⁻¹ flow; electrospray ionization (ESI); positive ion mode; ultraviolet detection with Acquity photodiode array). The reaction mixture was then purified by high-performance LC (eluent, a 45-min linear gradient, from 10% to 100% solvent B; flow rate, 10.0 ml min⁻¹; detection wavelength, 254 nm; eluent A, double-distilled H₂O containing 0.1% trifluoroacetic acid (v/v); eluent B, acetonitrile; column, Orienda Prep C18 column, FLA-2015-5-120A-BP, 120 Å, 5 μm, 20 × 150 mm). Biotin-5'-AMP (0.14 mmol, 69.8% yield) was finally obtained as a white solid upon freeze-drying. ESI-MS: *m/z* calculated for C₂₀H₂₉N₇O₉PS⁺, Biotin-5'-AMP (M + H)⁺ 574.2, found 574.4.

Determination of dissociation constant and binding constant by stopped-flow spectroscopy

The rates of association and dissociation were measured by stopped-flow spectroscopy using an SX20 stopped-flow instrument (Applied Photophysics) following the methods of a previous study⁴². In detail, the excitation wavelength was set at 280 nm for all measurements and fluorescence emission was monitored above 340 nm using a cutoff filter. For association of biotin-5'-AMP and TurboID or BioID, TurboID or BioID in standard buffer (10 mM Tris-HCl pH 7.5, 200 mM KCl and 2.5 mM MgCl₂) and a solution containing biotin-5'-AMP were rapidly mixed at a 1:1 ratio. Subsequently, the time-dependent change in the intrinsic protein fluorescence was measured. The final concentration of TurboID or BioID was 1 μM and the concentrations of biotin-5'-AMP were varied from 5 μM to 15 μM to satisfy pseudo-first-order conditions. Stopped-flow fluorescence traces obtained are well described by the following double exponential equation (3), where *F* is the measured fluorescence intensity and *A*₁ and *τ*₁ and *A*₂ and *τ*₂ are the amplitudes and relaxation times for the first and second kinetic phases, respectively. We obtained *τ*₁ and *τ*₂ at different concentrations of biotin-5'-AMP according to equation (4). We then calculated the dissociation constant *k*₁ using equation (5).

$$F = A_1 \exp\left(t \times \frac{1}{\tau_1}\right) + A_2 \exp\left(t \times \frac{1}{\tau_2}\right) + C \quad (4)$$

$$\frac{1}{\tau_1} + \frac{1}{\tau_2} = k_1[\text{ligand}] + k_{-1} + k_2 + k_{-2} \quad (5)$$

For dissociation of biotin-5'-AMP and TurboID or BioID, preformed complexes of TurboID-biotin-5'-AMP or BioID-biotin-5'-AMP in standard buffer were rapidly mixed with a solution containing 100 μM biotin in a 1:1 ratio. The subsequent time-dependent decrease in fluorescence was then measured. The resulting kinetic traces were analyzed using a single-exponential model (equation (6)) and equation (7) to obtain unimolecular dissociation constants *k*_{off}.

$$F = A \exp\left(\frac{t}{\tau}\right) + C \quad (6)$$

$$k_{\text{off}} = \frac{1}{\tau} \quad (7)$$

Finally, we calculated *K*_D using equation (8).

$$K_D = \frac{k_{\text{off}}}{k_1} \quad (8)$$

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data associated with this study are presented in the article and its Supplementary Information. Source data are provided with this paper.

References

- Rothmund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Douglas, S. M., Chou, J. J. & Shih, W. M. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc. Natl Acad. Sci. USA* **104**, 6644–6648 (2007).
- Li, S. et al. A DNA nanorobot functions as a cancer therapeutic in response to a molecular trigger in vivo. *Nat. Biotechnol.* **36**, 258–264 (2018).
- Stahl, E., Martin, T. G., Praetorius, F. & Dietz, H. Facile and scalable preparation of pure and dense DNA origami solutions. *Angew. Chem. Int. Ed. Engl.* **53**, 12735–12740 (2014).

Acknowledgements

This study was supported by the National Natural Science Foundation of China (32088101), Ministry of Science and Technology (2022YFA1304700), Li Ge-Zhao Ning Science Youth Research Foundation, Center for Life Sciences of Peking University and Beijing National Laboratory for Molecular Sciences (BNLMS-CXXM-202403). P.Z. is sponsored by a Bayer Investigator Award. We thank the National Center for Protein Sciences at Peking University, particularly S. Qin, for technical assistance with AFM. We thank the group of H. Cheng for support in qPCR instrumentation (CFX96, Bio-Rad). We thank A. Ting, B. Wang, J. Ling, T. Liu, Y. Li and X. Zhou for helpful discussions.

Author contributions

Z.C., P.Z. and Z.Y. designed the research. Z.Y. performed the research. Yu Zhang, Y.F., X.S. and K.Z. analyzed part of the data. Yuan Zhang and J.D. synthesized the biotin-5'-AMP. Z.Y. and Z.C. wrote the paper with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-025-02086-w>.

Correspondence and requests for materials should be addressed to Peng Zou or Zhixing Chen.

Peer review information *Nature Chemical Biology* thanks Chunhai Fan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Corresponding author(s): Dr. Peng Zou and Dr. Zhixing Chen

Last updated by author(s): Sep 14, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Bio-Rad CFX Manager 3.1.1517.0823

Data analysis GraphPad Prism 10.1.2, NanoScope Analysis 1.80, Bio-Rad CFX Manager 3.1.1517.0823,

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data presented in this study are available in the main text and supplementary materials.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All quantitative experiments were performed on replicate samples of fields of view as reported in Figure legends and supplementary materials.
Data exclusions	No data exclusion was performed.
Replication	The numbers of replicated experiments are indicated in the main text and figure legends.
Randomization	There is no bias when collecting data. All the samples were randomly allocated into experimental groups.
Blinding	The experiments of this study were not done blinded. Data were collected in parallel to minimize bias.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Non-participation

State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.

Randomization

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample

Describe the research sample (e.g. a group of tagged *Passer domesticus*, all *Stenocereus thurberi* within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection

Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work?

☐ Yes

☐ No

Field work, collection and transport

Field conditions

Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access & import/export

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance

Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

- n/a Involved in the study
- ☒ ☐ Antibodies
- ☒ ☐ Eukaryotic cell lines
- ☒ ☐ Palaeontology and archaeology
- ☒ ☐ Animals and other organisms
- ☒ ☐ Clinical data
- ☒ ☐ Dual use research of concern
- ☒ ☐ Plants

- n/a Involved in the study
- ☒ ☐ ChIP-seq
- ☒ ☐ Flow cytometry
- ☒ ☐ MRI-based neuroimaging

Antibodies

Antibodies used *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s) *State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*

Authentication *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See [ICLAC](#) register) *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*

Specimen deposition *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals *For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*

Wild animals *Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Reporting on sex *Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall*

numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | No | Yes | |
|--------------------------|--------------------------|----------------------------|
| ☐ | ☐ | Public health |
| ☐ | ☐ | National security |
| ☐ | ☐ | Crops and/or livestock |
| ☐ | ☐ | Ecosystems |
| ☐ | ☐ | Any other significant area |

Experiments of concern

Does the work involve any of these experiments of concern:

- | No | Yes | |
|--------------------------|--------------------------|---|
| ☐ | ☐ | Demonstrate how to render a vaccine ineffective |
| ☐ | ☐ | Confer resistance to therapeutically useful antibiotics or antiviral agents |
| ☐ | ☐ | Enhance the virulence of a pathogen or render a nonpathogen virulent |
| ☐ | ☐ | Increase transmissibility of a pathogen |
| ☐ | ☐ | Alter the host range of a pathogen |
| ☐ | ☐ | Enable evasion of diagnostic/detection modalities |
| ☐ | ☐ | Enable the weaponization of a biological agent or toxin |
| ☐ | ☐ | Any other potentially harmful combination of experiments and agents |

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A

ChIP-seq

Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication. *For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*

Files in database submission *Provide a list of all files available in the database submission.*

Genome browser session
(e.g. [UCSC](#)) *Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates *Describe the experimental replicates, specifying number, type and replicate agreement.*

Sequencing depth *Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*

Antibodies *Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Peak calling parameters *Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality *Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*

Software *Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation *Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*

Instrument *Identify the instrument used for data collection, specifying make and model number.*

Software *Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI

☐ Used

☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See [Eklund et al. 2016](#))

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involvement in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.