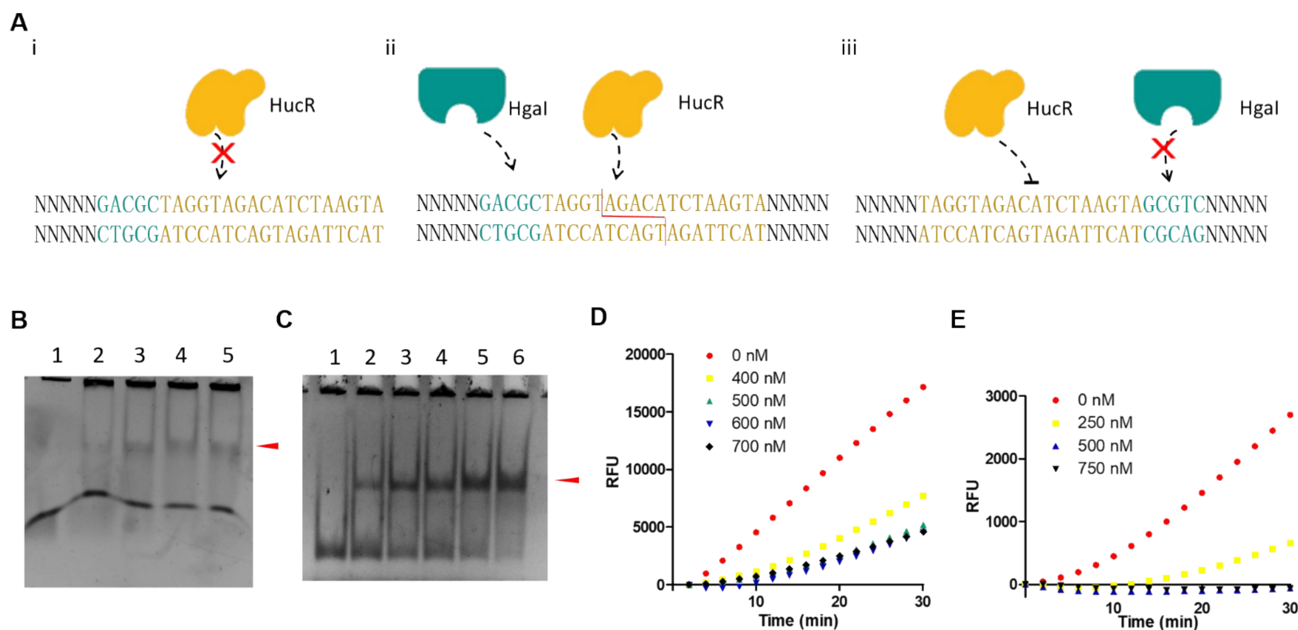
Electronic Supplementary Material (ESI) for Analyst.
This journal is © The Royal Society of Chemistry 2021

Supplementary Information

Rational design of allosterically regulated toehold mediated strand displacement circuit for sensitive and on-site detection of small molecule metabolites

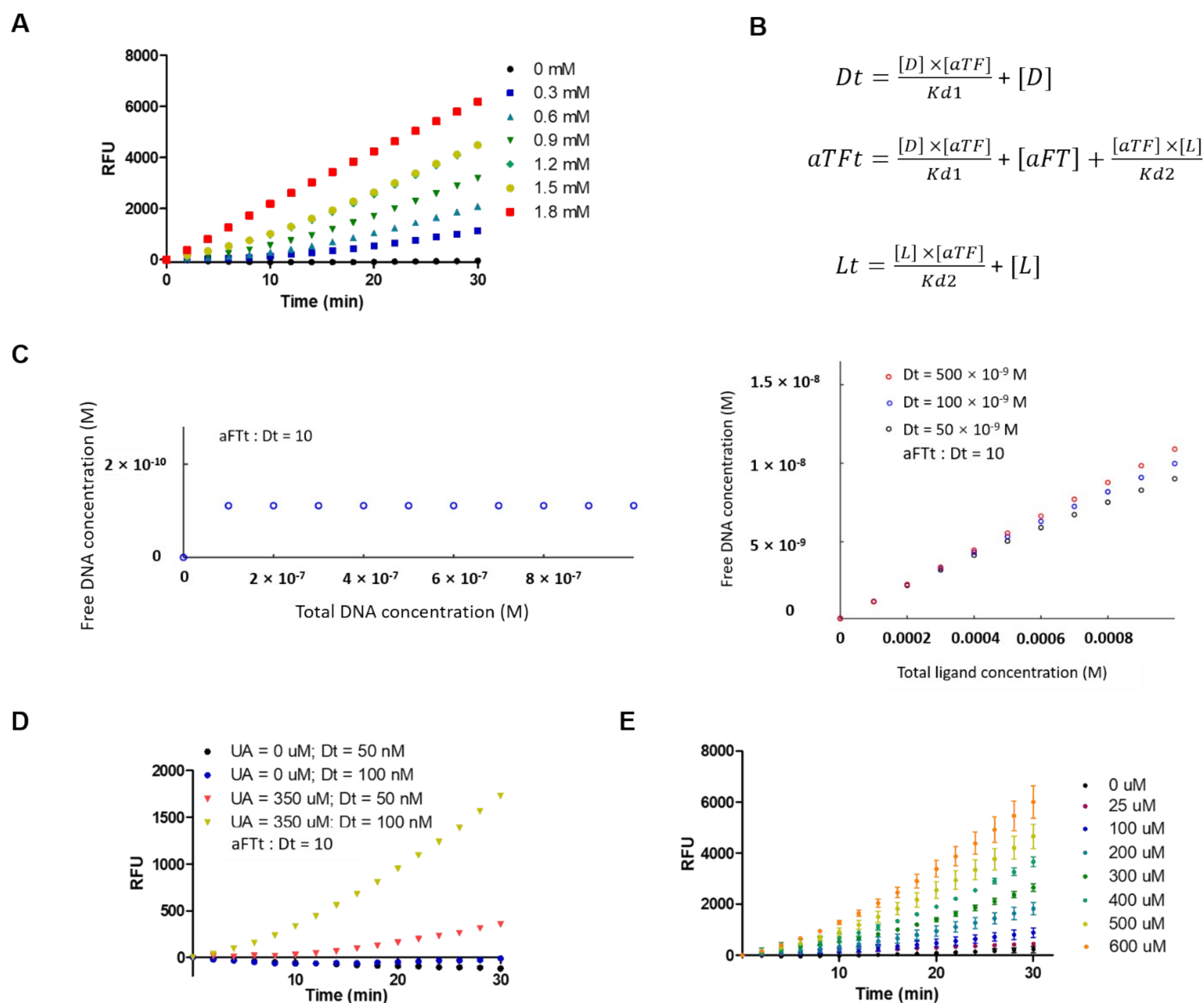
Haosi Lin, Alan F. Rodríguez-Serrano & I-Ming Hsing*

Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
* To whom correspondence should be addressed. Tel: (852) 2358 7122 E-mail: kehsing@ust.hk

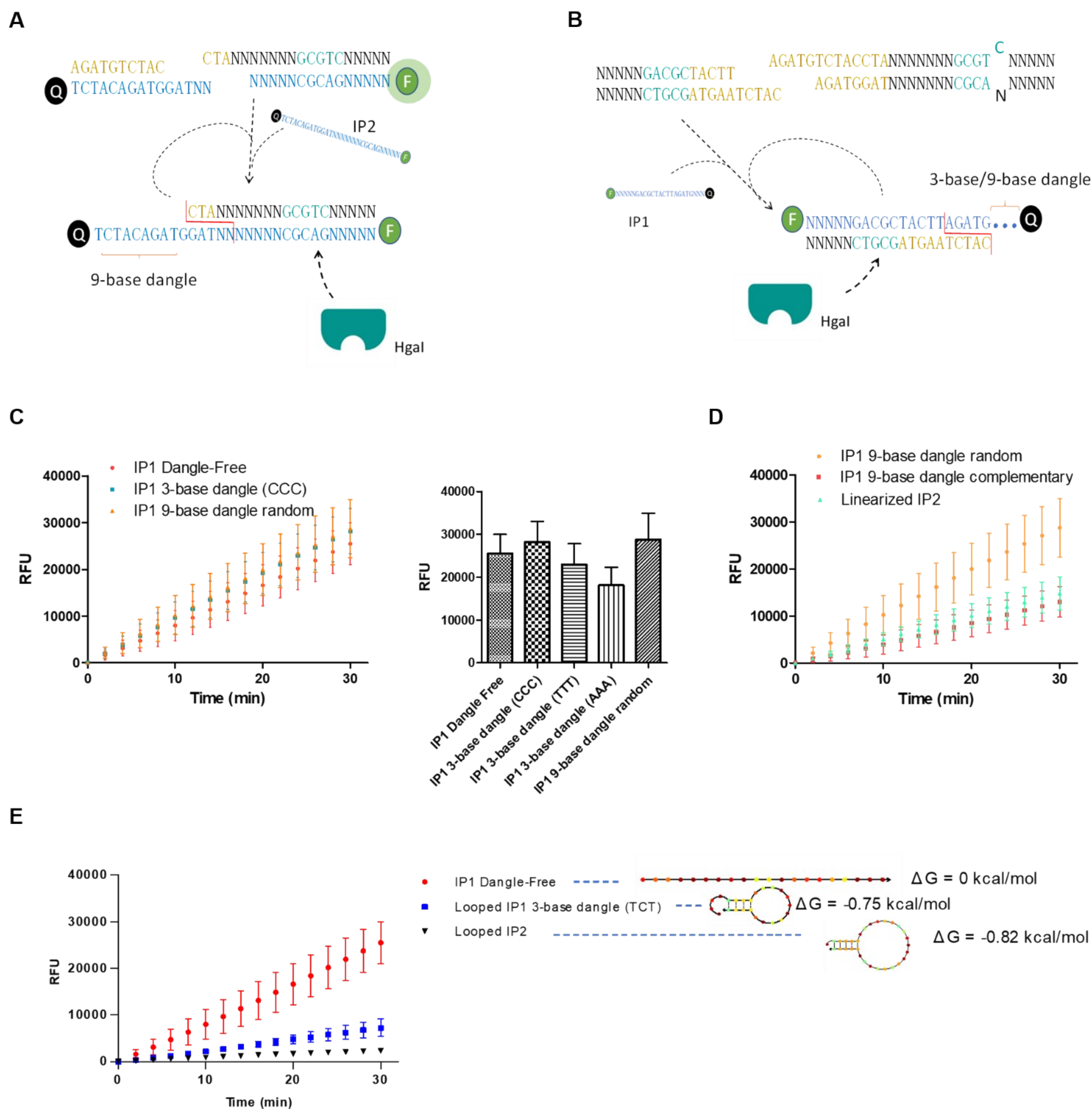


Supplementary Figure S1. Regulation of toehold mediated strand displacement (TMSD) by HucR. **(A)** Illustration of HucR-DNA binding and the protection from HgaI cleavage by HucR binding. Sequences in cyan indicate HgaI recognition site; sequences in gold indicate the HucO sequence; red line indicates the cleavage by HgaI. **(i)**: Upstream HgaI recognition site, no downstream sequence. **(ii)**: Upstream HgaI recognition site, random downstream sequence. **(iii)**: Random upstream sequence, downstream HgaI recognition site. Terms upstream and downstream are used in reference to HucO sequence. **(B)** EMSA for verifying HucR binding with HucO without extra sequences downstream as illustrated in **i** of panel **A**. 1: HucO = 50 nM; 2: HucO = 50 nM, HucR = 750 nM; 3: HucO = 50 nM, HucR = 1.25 μM; 4: HucO = 50 nM, HucR = 2.5 μM; 5: HucO = 50 nM, HucR = 5 μM; HucR-DNA complex are marked by red triangle. **(C)** EMSA showing HucR binding to a template with downstream extension as illustrated in **ii** of panel **A**. 1: HucO = 50 nM; 2: HucO = 50 nM, HucR = 50 nM; 3: HucO = 50 nM, HucR = 100 nM; 4: HucO = 50 nM, HucR = 150 nM; 5: HucO = 50 nM, HucR = 200 nM; 6: HucO = 50 nM, HucR = 250 nM; HucR-DNA complexed are marked by red triangles. **(D)** HucR repression of TSMD circuit with substrate containing upstream HgaI recognition site and downstream extra sequence, as illustrated in **ii** of panel **A**; DNA substrate = 50 nM; invading probe = 1.25 μM. **(E)** Repression of TSMD circuit with downstream HgaI recognition site by HucR, as illustrated in **iii** of panel **A**; DNA substrate = 50 nM; invading probe = 1.25 μM. RFU in panels **D** and **E** indicates relative fluorescence units; n = 1.

This journal is © The Royal Society of Chemistry 20xx

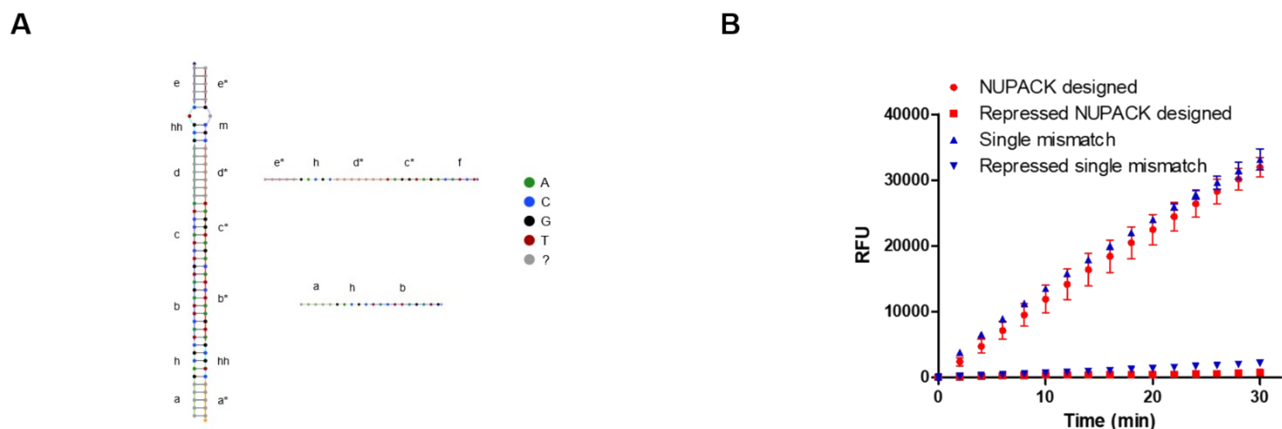


Supplementary Figure S2. Detection of uric acid with allosteric transcription factor (aTF)-regulated TMSD circuit. **(A)** Detection of uric acid (UA) with the one-cycle TMSD circuit. DNA substrate = 50 nM; HucR = 500 nM; IP=1.25 µM; n=1. **(B)** Equations describing the equilibrium of DNA-aTF-DNA interaction. Dt: Total DNA concentration; $aTFt$: Total aTF concentration; Lt: Total ligand concentration; $[D]$: Free-DNA concentration; $[aTF]$: Free aTF concentration; $[L]$: Free ligand concentration; K_{d1}: dissociation constant of aTF-DNA complex; K_{d2}: dissociation constant of ligand-aTF complex. **(C)** Simulation of equilibrium concentration of free-DNA at different total DNA (left) and total ligand (right) concentrations by solving the equations in panel **B**; Dt: total DNA concentration; aTFt: total aTF concentration. **(D)** UA induction of repressed TMSD circuit; UA: uric acid sample concentration; Dt: total DNA concentration; aTFt: total aTF concentration; n=1. **(E)** UA detection with optimized TMSD circuit; error bars indicate mean ± S.D., n = 3. RFU in panels **A, D** and **E** indicates relative fluorescence units.

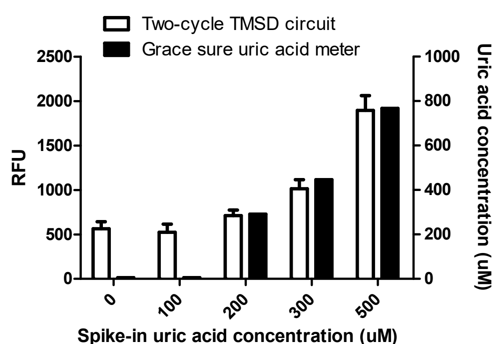


Supplementary Figure S3. Impact of 3' end dangles and 3' end secondary structure on TMSD-based signal amplification. **(A)** Generation of nine-base 3' dangle in cycle 2. IP2: cycle 2 invading probe. **(B)** Generation of 3' dangle in cycle 1. IP1: cycle 1 invading probe. **(C)** Signal amplification of TMSD circuit with IP1 containing dangles of different lengths and sequences. Data in the histogram was taken after 30-minute reaction. DNA template = 100 nM; IP1 = 1.25 µM **(D)** Signal amplification of TMSD circuit with IPs containing nine-base dangles. IP2 contains inherent 9-base complementary dangle. DNA template = 100 nM; IPs = 1.25 µM. **(E)** Signal amplification of TMSD with invading probe containing secondary structure involving the 3' end. DNA template = 100 nM; HucR = 1 µM; IPs = 1.25 µM. Template structure prediction and calculation of free energy were conducted with NUPACK (material: DNA; temperature: 37 °C; Mg^{2+} = 0.01 M; Na^{+} = 0.05 M). Error bars in panels **C-E** indicate mean ± S.D., n=3. RFU in panels **C-E** indicates relative fluorescence units.

This journal is © The Royal Society of Chemistry 20xx



Supplementary Figure S4. *In-silico* design of two-cycle TMSD circuit **(A)** The in-put structure of DNA template and invading probes for NUPACK design. Labels a, b, c, d, e, f, h, hh, and m representing different domains; a*, b*, c*, d*, e*, f* representing the complementary sequences of a, b, c, d, e, f; h and hh correspond to the HgaI recognition site and its complementary sequence repectively. **(B)** Signal amplification and HucR repression of the NUPACK designed system. Error bars in panel **B** indicate mean ± S.D. n=3. RFU in panels **B** indicates relative fluorescence units. Error bars in panels **B** indicate mean ± S.D., n=3.



Supplementary Figure S5. UA detection in human serum with *in-silico* design of two-cycle TMSD circuit in comparison with a commercial UA meter. Error bars indicate mean ± S.D., n=3. For detection by commercial UA meter n=1. RFU indicates relative fluorescence units; UA concentration in the right axis indicates the UA concentration read by the UA meter.

Supplementary Note S1.

The NUPACK-based *in silico* design was done with the following script:

material = dna

magnesium = 0.01

temperature[C] = 37.0

trials = 2

sodium[M] = 0.05

structure probes = U20 + U30

structure substrate = D38 (U1 D6 + U1)

domain a = H N4

domain h = GACGC

domain hh = GCGTC

domain b = tacttAGATG

domain c = TCTACcta

domain d = N7

domain m = N5

domain e = N4 D

domain f = catct

strand A = a h b c d hh e

strand B = e* m d* c* b* hh a*

probes.seq = a h b e* h d* c* f

substrate.seq = A B

prevent = AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY

The result of the design containing sequences of the two probes and two strands of the substrate. The known sequences in the script were the sequences of HucO and HgaI recognition site. These sequences could be changed for application of different aTFs.

This journal is © The Royal Society of Chemistry 20xx

Supplementary Table 1. Protein, DNA, and oligonucleotides used in the experiment.

No.	Name	Sequences
1	6 His-TEV-HucR	HHHHHHSSGLVPRGSHMENLYFQSMSARMDNDTAALLERIRSDWARLNHGQ GPDSDGLTPSAGPMLTLLLLERLHAALGREIERTYAASGLNAAGWDLLLTLYRS APPEGLRPTELSALAAISGPSTSNRIVRLLEKGLIERREDERDRRSASIRLTPQGR ALVTHLLPAHLATTQRVLAPLSAQEQRTLEELAGRMLAGLEQGV
2	HucO	TAGGTAGACATCTAAGTA
3	One-cycle substrate	ATAAAGACGCTAGGTAGACATCTAAGTA
4	One-cycle substrate'	TACTTAGATGTCTACCTAGCGTCTTTAT
5	One-cycle substrate - extended	ATAAAGACGCTAGGTAGACATCTAAGTAAATGACCTGGATACATGGCTCAGT
6	One-cycle substrate – extended'	ACTGAGCCATGTATCCAGGTCATTTACTTAGATGTCTACCTAGCGTCTTTAT
7	IP – one-cycle extended	5'6-FAM-ATAAAGACGCTAGGTAGACATCT-3'BHQ1
8	One-cycle substrate – downstream HgaI site	ATAAAGTTACTAGGTAGACATCTAAGTAGCGTCTCCTG
9	One-cycle substrate – downstream HgaI site'	CAGGAGACGCTACTTAGATGTCTACCTAGTAACTTTAT
10	IP – downstream HgaI site – TCT dangle	5'6-FAM-CAGGAGACGCTACTTAGATGTCT-3'BHQ1
11	Two-cycle substrate – looped IP2	AAGATGACGCTACTTAGATGTCTACCTAATAAGTTGCGtcATCTT
12	Two-cycle substrate – looped IP2'	AAGATacCGCAACTTATTAGGTAGACATCTAAGTAGCGTCATCTT
13	Looped IP2	5'6-FAM-AAGATGACGCAACTTAT TAGGTAGACATCT-3'BHQ1
14	Two-cycle substrate – double mismatch	AAGATGACGCTACTTAGATGTCTACCTAATAAGTTGCGtcTATAG
15	Two-cycle substrate – double mismatch'	CTATAacCGCAACTTATTAGGTAGACATCTAAGTAGCGTCATCTT
16	Two-cycle substrate – single mismatch	AAGATGACGCTACTTAGATGTCTACCTAATAAGTTGCGtCTATAG
17	Two-cycle substrate – single mismatch'	CTATAGcCGCAACTTATTAGGTAGACATCTAAGTAGCGTCATCTT
18	IP1 – dangle free	5'6-FAM-AAGATGACGCTACTTAGATG-3'BHQ1
19	IP1 – CCC dangle	5'6-FAM-AAGATGACGCTACTTAGATGCCC-3'BHQ1
20	IP1 – AAA dangle	5'6-FAM-AAGATGACGCTACTTAGATGAAA-3'BHQ1
21	IP1 – TTT dangle	5'6-FAM-AAGATGACGCTACTTAGATGTTT-3'BHQ1
22	IP1 – 9-base random dangle	5'6-FAM-AAGATGACGCTACTTAGATGCCCTATATAA-3'BHQ1
23	IP1 – 9-base complementary dangle	5'6-FAM-AAGATGACGCTACTTAGATGTCTACCTAAT-3'BHQ1
24	IP2 – optimized	5'6-FAM-CTATAGACGCAACTTATTAGGTAGACATCT-3'BHQ1
25	Two-cycle substrate – *in silico* designed	CCATTGtCGCGATTAGTTAGGTAGACATCTAAGTAGCGTCGTGGG
26	Two-cycle substrate – *in silico* designed'	CCCACGACGCTACTTAGATGTCTACCTAACTAATCGCGtCAATGG
27	IP1 –*in silico* designed	5'6-FAM-CCCACGACGCTACTTAGATG-3'BHQ1
28	IP2 –*in silico* designed	5'6-FAM-CCATTGACGCGATTAGTTAGGTAGACATCT-3'BHQ1
Sequences underlined indicate HucO sequence. Sequences in red indicate HgaI recognition site. Lower case letters indicate mismatch site.		

Supplementary Table 2. Comparison among different aTF-based UA biosensor and a commercial UA meter.

	aTF-based UA biosensors				Commercial uric acid meter
	Two-cycle TMSD	CaT-SMelor	*Hin*dIII-HF-qPCR/ Alpha	aTF-NAST	
Detection range	25 – 600 μM	25 – 500 nM	5 – 300 nM/ 0.1 – 10 μM [a]	10 – 600 nM/ 10 – 200 nM 0.0005 – 10 nM [b]	179 – 1190 μM
Turnaround	ca. 15 min	15 – 25 min	70 – 110 min [c]	70 – 170 min [d]	30 s

time					
Operation steps	1	3 [e]	4 [c]	3 [d]	1
Modularity [f]	aTF responding to targeted molecule	Cellulose binding domain fused aTF	Operator sequence constraint [g]	aTF responding to targeted molecule	Target-specific electrode
Sample type	Saliva	Whole blood/ serum	Serum	Serum	Whole blood/ Serum
Equipment [h]	Incubator, fluorescence reader	Incubator, centrifuge, fluorescence reader	Real-time PCR system/ microplate reader.	Incubator, fluorescence reader	Portable electrochemical device.
Source	This work	[1]	[2]	[3]	Grace Medical

[a.] *Hin*dIII-HF-qPCR: 5 – 300 nM; *Hin*dIII-HF-Alpha: 0.1 – 10 μM.[2]
[b.] aTF – NAST (qPCR): 10 – 600 nM; aTF – NAST (RCA): 10 – 200 nM; aTF – NAST (RPA): 0.0005 – 10 nM.[3]
[c.] 20 – 25 min HucR binding, 15 – 20 min *Hin*dIII reaction, 10-min inactivation plus qPCR (40-cycle qPCR where each cycle contains 95 °C 5 s, 55 °C 10 s, and 72 °C 15 s) or Alpha luminescence (slow cool-down and 60-min incubation in dark).[2]
[d.] 20-min aTF-DNA binding, 30-min ligation and signal amplification range from 20 min (RPA) to 2 hours (RCA).[3]
[e.] Step 1: Setup; step 2: incubation; step 3: detection and data analysis.[1]
[f.] Requirements for applying the system to another target small molecule.
[g.] This design utilized a modified HucO sequence overlapping with *Hin*dIII recognition site.[2]

Reference

1. M. Liang, Z. Li, W. Wang, J. Liu, L. Liu, G. Zhu, L. Karthik, M. Wang, K. F. Wang, Z. Wang, J. Yu, Y. Shuai, J. Yu, L. Zhang, Z. Yang, C. Li, Q. Zhang, T. Shi, L. Zhou, F. Xie, H. Dai, X. Liu, J. Zhang, G. Liu, Y. Zhuo, B. Zhang, C. Liu, S. Li, X. Xia, Y. Tong, Y. Liu, G. Alterovitz, G. Y. Tan and L. X. Zhang, *Nat Commun*, 2019, **10**, 3672.
2. Y. Yao, S. Li, J. Cao, W. Liu, F. Qi, W. Xiang, K. Yang, W. Wang and L. Zhang, *Applied microbiology and biotechnology*, 2018, **102**, 7489-7497.
3. J. Cao, Y. Yao, K. Fan, G. Tan, W. Xiang, X. Xia, S. Li, W. Wang and L. Zhang, *Science advances*, 2018, **4**, eaau4602.

This journal is © The Royal Society of Chemistry 20xx