Supporting Information

3-aminophthalic acid: A new cereblon ligand for targeted protein degradation by O'PROTAC

Yuqian Yan[1,5], Jingwei Shao[2,5], Donglin Ding[1], Yunqian Pan[1], Phuc Tran[2], Wei Yan[2], Zhengyu Wang[2], Hong-yu Li[2]* and Haojie Huang[1,3,4]*

[1]Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA

[2]Department of Pharmaceutical Sciences, College of Pharmacy, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

[3]Department of Urology, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA

[4]Mayo Clinic Cancer Center, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA.

[5]These authors contributed equally.

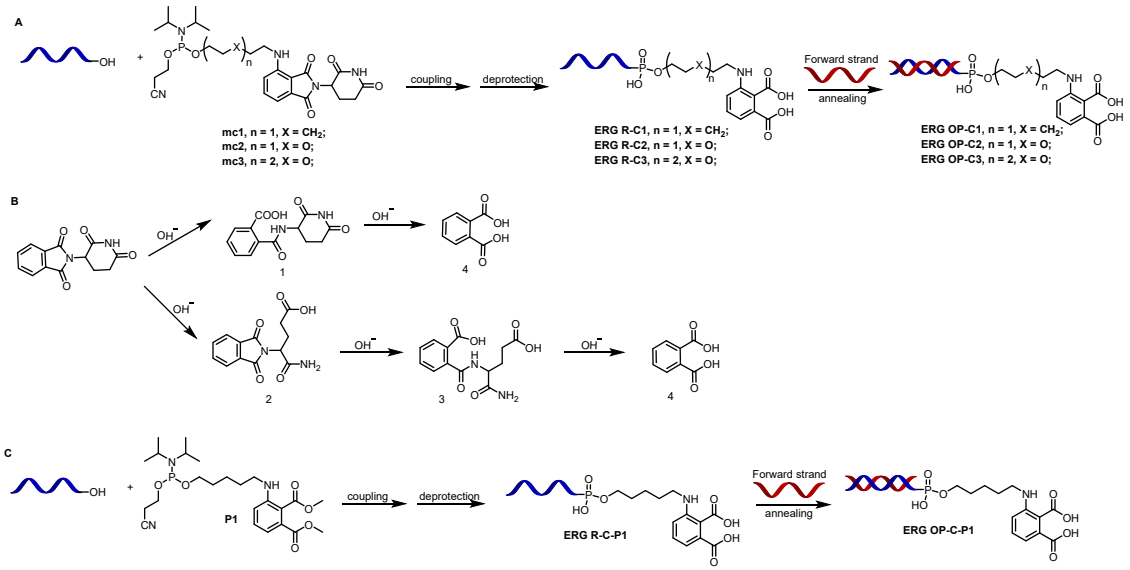
*** Corresponding authors**

Hong-yu Li (HLi2@uams.edu) and Haojie Huang (huang.haojie@mayo.edu)

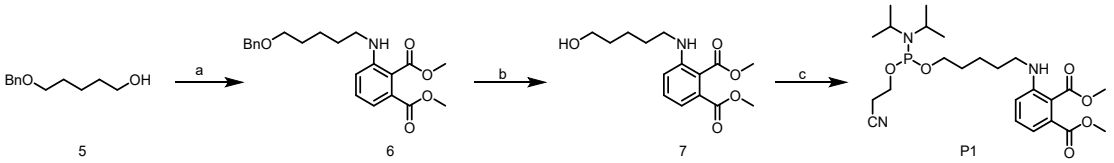
EXPERIMENTAL METHODS

General Chemistry Methods

Scheme S1



Scheme S2[a]:



[a]Reagents and conditions: a) 1. Dess-Martin periodinane (DMP), methylene chloride (DCM); 2. dimethyl 3-aminophthalate, $NaBH(OAc)_3$, AcOH, DCM; b) Pd/C, H_2, MeOH; c) Cl-POCEN^{i}Pr$_2$, N,N-Diisopropylethylamine (DIPEA), DCM, 2 h, room temperature (RT).

Procedure:

Dimethyl 3-((5-(benzyloxy)pentyl)amino)phthalate (6): Compound 5(1.94 g, 10 mmol) was dissolved in DCM (30 ml), then DMP(5.5 g, 13 mmol) was added. The mixture was stirred at RT for 2 h. The white solid was filtered off and washed with EA. The filtrate was concentrated. The residue was dissolved in Et_2O and washed with water. The organic phase was dried with Na_2SO_4, filtered and concentrated. The residue was dissolved in DCM (30 ml), then dimethyl 3-

aminophthalate (836 mg, 4 mmol) and 3 drops of AcOH were added. The mixture was stirred at RT for 30 min, then NaBH $(OAc)_3$ (1.22 g, 6 mol) was added. The reaction was stirred at RT overnight. After completion, the reaction solution was diluted with DCM, and washed with water. The organic phase was dried with Na_2SO_4, filtered and concentrated. The residue was purified with flash chromatography (Hexane:EA =100:0 to 80:20, V/V), giving product as yellow oil (915 mg, 59.4%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.35 – 7.30 (m, 6H), 6.80 (t, *J* = 1.1 Hz, 1H), 6.79 – 6.77 (m, 1H), 4.50 (s, 2H), 3.86 (s, 3H), 3.82 (s, 3H), 3.49 (t, *J* = 7.3, 2H), 3.16 (t, *J* = 7.1 Hz, 2H), 1.71 – 1.63 (m, 4H), 1.53 – 1.47 (m, 2H).

Dimethyl 3-((5-hydroxypentyl)amino)phthalate (7): Compound 6 (900 mg, 2.33 mmol) was dissolved in MeOH (15 ml), then Pd/C (180 mg, 20% wt) was added. The mixture was stirred at RT under H_2 atmosphere overnight. Pd/C was filtered off and washed with MeOH. The filtrate was concentrated and purified with flash chromatography (Hexane:EA =100:0 to 65:35, V/V), giving product as yellow oil (530 mg, 77%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 – 7.27 (m, 1H), 6.77 (t, *J* = 1.5 Hz, 1H), 6.75 (m, 1H), 3.85 – 3.82 (s, 3H), 3.81 (s, 3H), 3.65 (t, *J* = 7.8, 2H), 3.16 (t, *J* = 7.0 Hz, 2H), 1.67 (dd, *J* = 14.6, 7.2 Hz, 2H), 1.63 – 1.56 (m, 2H), 1.51 – 1.42 (m, 2H).

Dimethyl 3-((5-(((2-cyanoethoxy)(diisopropylamino)phosphaneyl)oxy)pentyl) amino)phthalate (P1): Compound 7 (130 mg, 0.44 mmol) was dissolved in anhydrous DCM (5 ml), DIPEA (218 µl, 1.32 mmol) and Cl-POCEN^{i}Pr$_2$ (147 µl, 0.66 mmol) was added. The mixture was stirred at RT for 1 h. Solvent was removed, and the residue was purified with flash chromatography (Hexane:Actone (5%TEA)=100:0 to75:25, V/V), giving product as colorless oil (135 mg, 62%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.31 (t, *J* = 8.0 Hz, 1H), 6.78 (s, 1H), 6.76 (t, *J* = 2.8 Hz, 1H), 3.88 – 3.83 (m, 4H), 3.83 – 3.77 (m, 4H), 3.71 – 3.55 (m, 4H), 3.17 (dd, *J* = 12.3, 6.9 Hz, 2H), 2.63 (t, *J* = 6.5 Hz, 2H), 1.66 (m, 4H), 1.54 – 1.46 (m, 2H), 0.92 – 0.83 (m, 12H).

Synthesis of oligonucleotides and annealing reaction

All oligonucleotides used in this study were synthesized by ExonanoRNA (Columbus, OH). The sequences of oligonucleotides and compositions of O'PROTACs are listed in **Table S1 and S2, respectively**. For oligo annealing reaction, single-stranded forward and reverse oligonucleotides were mixed in an assembly buffer (10 mM Tris-HCl [pH7.5], 100 mM NaCl, 1 mM EDTA), and heated to 90 ^{o}C for 5 min, then slowly cooled down to 37 ^{o}C within 1 h. Double-stranded O'PROTACs were mixed well, aliquoted and stored at -20 ^{o}C for the future use.

Plasmids and reagents

The siRNA constructs (siNS and siCRBN) were purchased from GE Dharmacon and the sequence information is listed in **Table S3**. The mammalian expression vector for HA-Ub was purchased from Addgene while pMCV-HA-ERG was constructed using cDNA of VCaP cells as a template. Cycloheximide (CHX) and MG132 were purchased from Sigma Aldrich. The antibodies used were: HA (Cat# MMS-101R) from Covance; Flag (M2) (Cat# F-3165) from Sigma; ERK2 (sc-1647) from Santa Cruz; CRBN (Cat#71810S) from Cell Signaling Technology; ERG from Biocare Medical (Cat#901-421-101520). For western blots, all the antibodies were diluted 1:1,000 with 5% BSA in TBST.

Cell lines, cell culture and transfection

The immortalized human embryonic kidney cell line 293T and two prostate cancer cell lines (VCaP and 22Rv1) were purchased from ATCC (Manassas, VA). The 293T and VCaP cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% of FBS

(Thermo Fisher Scientific). The 22Rv1 cells were cultured in RPMI 1640 medium supplemented with 10% of FBS. The cells were maintained in a 37°C humidified incubator supplied with 5% CO_2. Transient transfection was performed by Lipofectamine 2000 (Cat# 11668500, Thermo Fisher Scientific) according to the manufacturer's instruction.

Protein extraction and western blot

The cells were washed with PBS once before being lysed into lysis buffer containing 25 mM Tris–HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40, and 5% glycerol for 30 mins on ice. The lysate was centrifuged at 13,000 rpm for 15 mins and the supernatant containing 50 μg of total protein was applied to SDS-PAGE gel. The protein gel was transferred to the nitrocellulose membrane, which was blocked by 5% skim milk for 1 h, followed by incubation with primary antibody at 4 °C overnight and secondary antibody at RT for 1 h. The protein signal was developed with Pierce™ ECL Western Blotting Substrate (Cat#32106, Thermo Fisher Scientific). All the raw images are shown in Fig. S6 and Fig. S7.

RNA extraction and RT-qPCR

Total RNA was extracted and reversely transcribed into cDNA as previously described,[1] followed by quantitative PCR using iQ SYBR Green Supermix (Cat# 1708880, Bio-Rad). The ΔCT was calculated by normalizing the threshold difference of a certain gene with glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Primers used for RT-qPCR are listed in **Table S4**.

Nuclear extraction and electrophoretic mobility shift assay (EMSA)

The VCaP cell nuclear protein was extracted using NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (Cat# 78833, Thermo Fisher Scientific). EMSA was performed with LightShift™ Chemiluminescent EMSA Kit (Cat# 20148, Thermo Fisher Scientific) according to the manufacturer's instruction. Briefly, ERG OP-C-P1 containing the potential ERG binding motif was incubated with VCaP nuclear protein for 30 mins at RT, followed by separation with 6% acrylamide DNA gel. The biotin-labeled probe was incubated with 0.5 or 1 μg of ERG antibody for 1 h before loading into 6% of Polyacrylamide DNA gel. All the raw images are shown in Fig. S7.

Three-dimension (3D) sphere

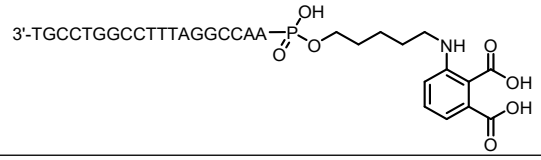
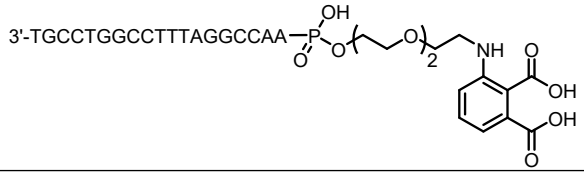
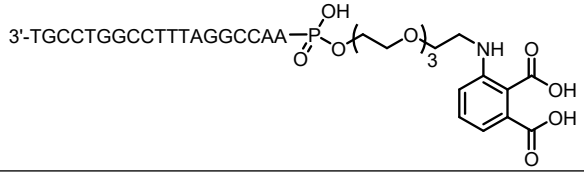
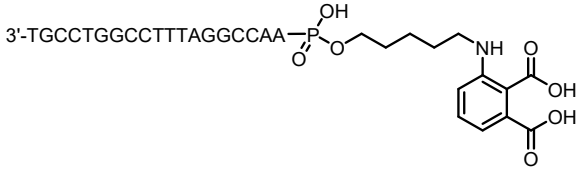
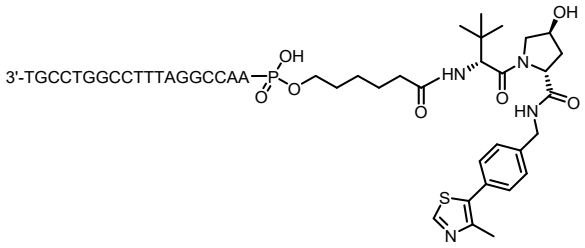
The procedure followed a previous study.[2] Briefly, ~120 μl of matrigel matrix (Cat# 354234, BD Bioscience) was pre-coated onto the bottom of the wells of 24-well plate at 37 °C for 30 mins. Approximately 20,000 VCaP cells transfected with ERG OP-C-P1 (200 nM) were resuspended in 250 μl of DMEM/F12 medium containing 10% FBS and seeded on the top of matrigel pre-coated wells. After 30 min, when the cells were settled down, they were covered with another layer of 10% matrigel diluted with DMEM/F12 medium. The medium was changed every 2–3 days.

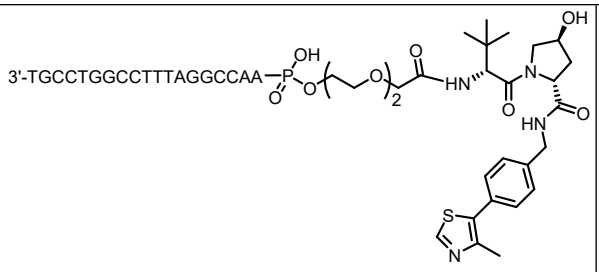
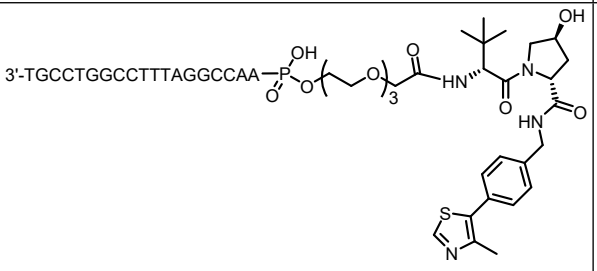
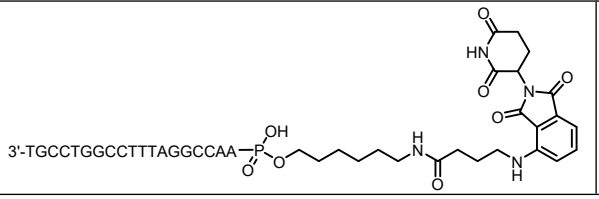
Cell invasion

The 22Rv1 cells were transfected with 100 nM of OP-C-P1 and 0.5 μg of pCMV-HA-ERG. Approximately 50,000 transfected 22Rv1 cells were re-suspended with 200 μl of serum-free RPMI-1640 medium and seeded onto matrigel invasion chamber (Cat#354480, Corning). The

chambers were then placed into the wells filled with 800 µl of RPMI-1640 medium supplemented with 10% FBS.

Table S1. Sequences of control and ERG-bound DNA oligos in O'PROTACs

Name[a)]	DNA sequence	Mass calc.	Mass obs.
ERG F	5'-ACGGACCGGAAATCCGGTT-3'	5837.8	5838.0
ERG FITC-F	5'-FITC-ACGGACCGGAAATCCGGTT-3'	6404.3	6405.8
ERG Biotin-F	5'-Biotin-ACGGACCGGAAATCCGGTT-3'	6273.3	6274.0
ERG R-C1[b)]		6108.9	6109.6
ERG R-C2[b)]		6154.9	ND[c)]
ERG R-C3[b)]		6198.9	ND[c)]
ERG R-C-P1		6108.9	6108.6
ERG R-V1		6386.0	6386.2

ERG R-V2		6418.0	6418.4
ERG R-V3		6462.0	6462.5
CTRL F	5’-TGTGCTAGCTGATGTGCTA-3’	5849.9	5850.3
CTRL R-C-N1 *d)*		6286.0	6286.6

a) F, forward; R, reverse; b) not purified. c) ND, not determined. d) CTRL R-C-N1 was synthesized through NHS-ester modification as previously described.[2]

Table S2. Composition of ERG O’PROTACs

Name	Forward strand	Reverse strand
ERG OP-C1	ERG F	ERG R-C1
ERG OP-C2	ERG F	ERG R-C2
ERG OP-C3	ERG F	ERG R-C3
ERG OP-C-P1	ERG F	ERG R-C-P1
ERG OP-V1	ERG F	ERG R-V1

ERG OP-V2	ERG F	ERG R-V2
ERG OP-V3	ERG F	ERG R-V3
ERG FITC-OP-C1	ERG FITC-F	ERG R-C1
ERG FITC-OP-C-P1	ERG FITC-F	ERG R-C-P1
ERG Biotin-OP-C-P1	ERG Biotin-F	ERG R-C-P1

Table S3. Sequences of siRNA oligonucleotides

Name	**Species**	**Catalog No.**	**Sequence (5'-3')**
non-targeting siRNA (siNS)	Human	M-021086-01-0005	UAGCGACUAAACACAUCAAUU
siCRBN Pool	Human	M-021086-01-0005	
siCRBN-1	Human	D-021086-06	GAAUAAAUGUACUUCCCUU
siCRBN-2	Human	D-021086-07	GUGCUGAUAUGGAAGAAUU
siCRBN-3	Human	D-021086-08	CCAGCAAGCUAAAGUGCAA
siCRBN-4	Human	D-021086-09	AGACAAAGGUUCAAAGUCC

Table S4. Sequences of primers for RT-qPCR

Gene	**Species**	**usage**	**Forward (5'-3')**	**Reverse (5'-3')**
GAPDH	Human	RT-qPCR	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC
ADAM19	Human	RT-qPCR	GCCTATGCCCCCTGAGAGTG	GCTTGAGTTGGCCTAGTTTGTTGTTC
MMP3	Human	RT-qPCR	TTCATTTTGGCCATCTCTTCCTTCAG	TATCCAGCTCGTACCTCATTTCCTCT
MMP9	Human	RT-qPCR	TGCCCGGACCAAGGATACAGT	AGCGCGTGGCCGAACTCAT
PLAT	Human	RT-qPCR	CACTGGGCCTGGGCAAACATA	CACGTCAGCCTGCGGTTCTTC
PLAU	Human	RT-qPCR	TACGGCTCTGAAGTCA	CCCCAGCTCACAATTCC

			CCACCAAAAT	AGTCAA
ERG FL	Human	RT-qPCR	GAGAGTGTGCAAGAGATC	GCTTTTGGTCAACACGG
ERG T1/E4	Human	RT-qPCR	CGCCCGGAGGTGAAAGCG	GCTTTTGGTCAACACGG

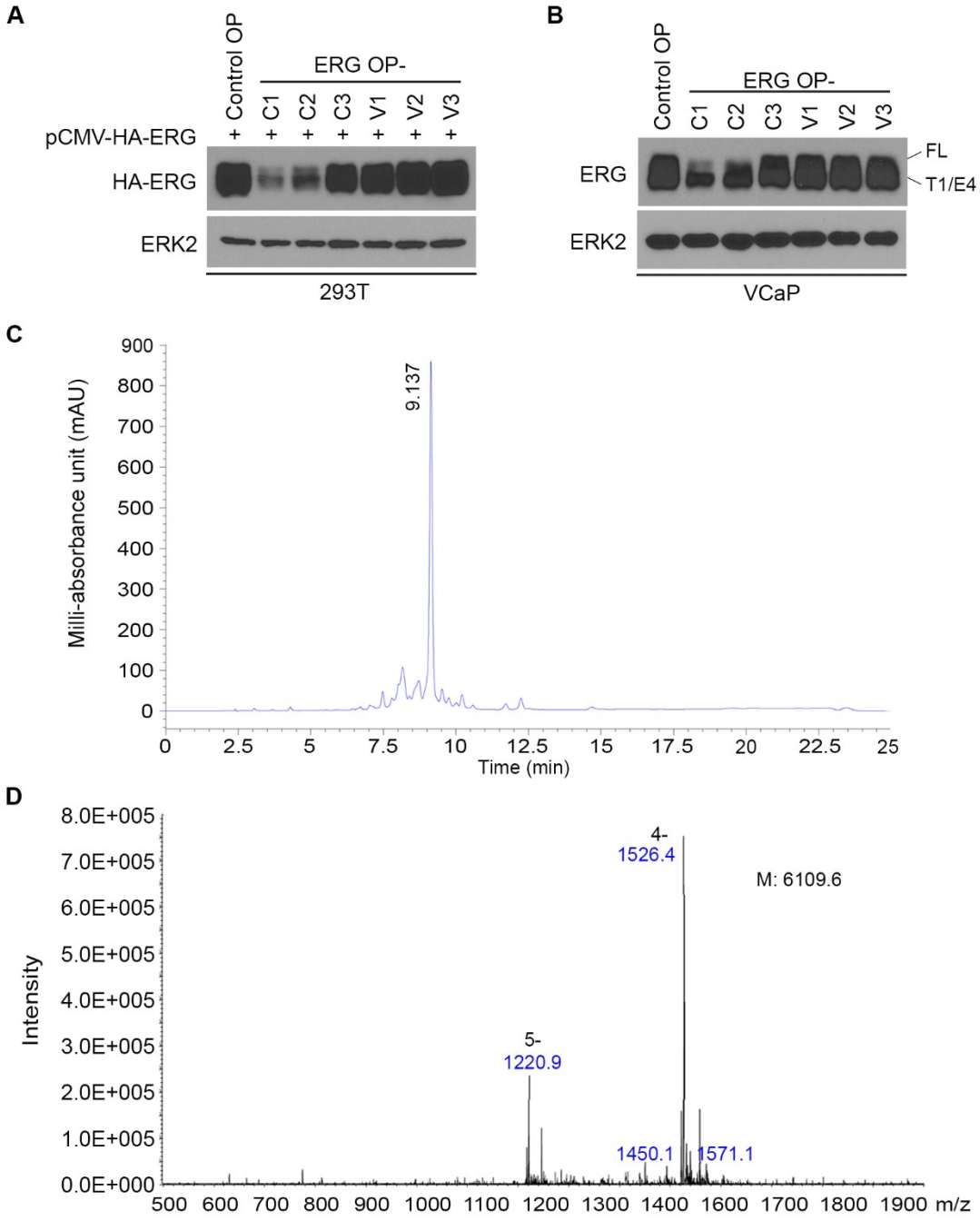


Fig. S1 Protein degradation efficacy of ERG O'PROTACs and HPLC and mass spectrometry data of ERG R-C1. (A) The 293T cells were transfected with pCMV-HA-ERG plasmid and control or six indicated ERG O'PROTACs (100 nM) and harvested for western blot analysis 48 h post-transfection. ERK2 was used as a loading control. C stands for CRBN-based OPs while V is for VHL. (B) The VCaP cells were transfected with control or six indicated ERG

O’PROTACs (100 nM) and harvested for western blot analysis 48 h post-transfection. Both endogenous full-length (FL) (wild-type) and truncated TMPRSS2-ERG (T1/E4) were detected. (C) The HPLC spectrum of ERG R-C1, detected at 260 nm. (D) The mass spectrum of ERG R-C1, the deconvoluted mass is shown in the upper right corner.

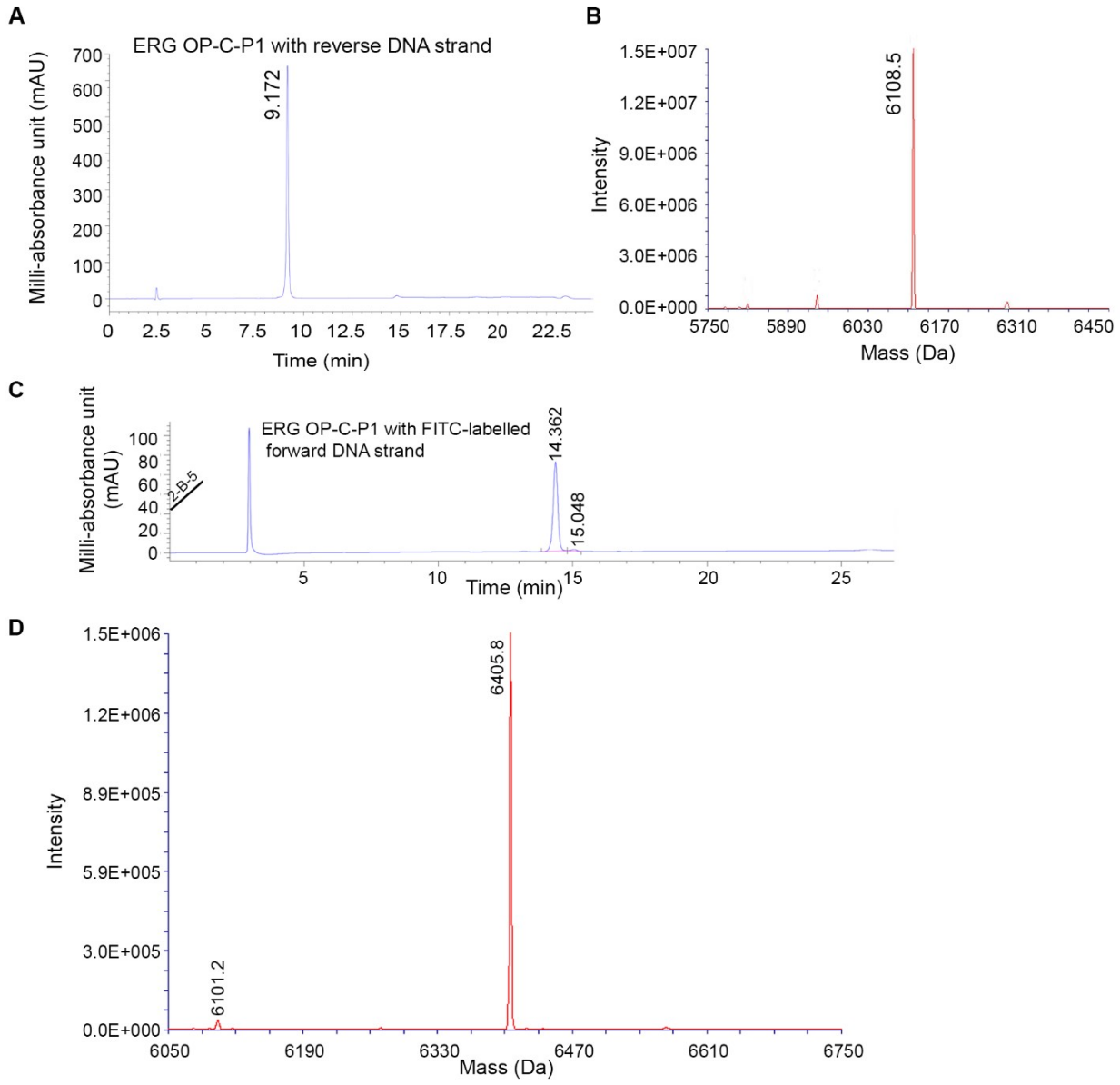


Fig. S2 HPLC and mass spectrum of oligonucleotides of ERG OP-C-P1. (A) The HPLC spectrum of ERG OP-C-P1 reverse strand, detected at 260 nm. (B) The deconvoluted mass spectrum of ERG OP-C-P1 reverse strand. (C) The HPLC spectrum of FITC-labelled ERG OP-C-P1 forward strand, detected at 260 nm. (D) The deconvoluted mass spectrum of FITC-labelled ERG OP-C-P1 forward strand.

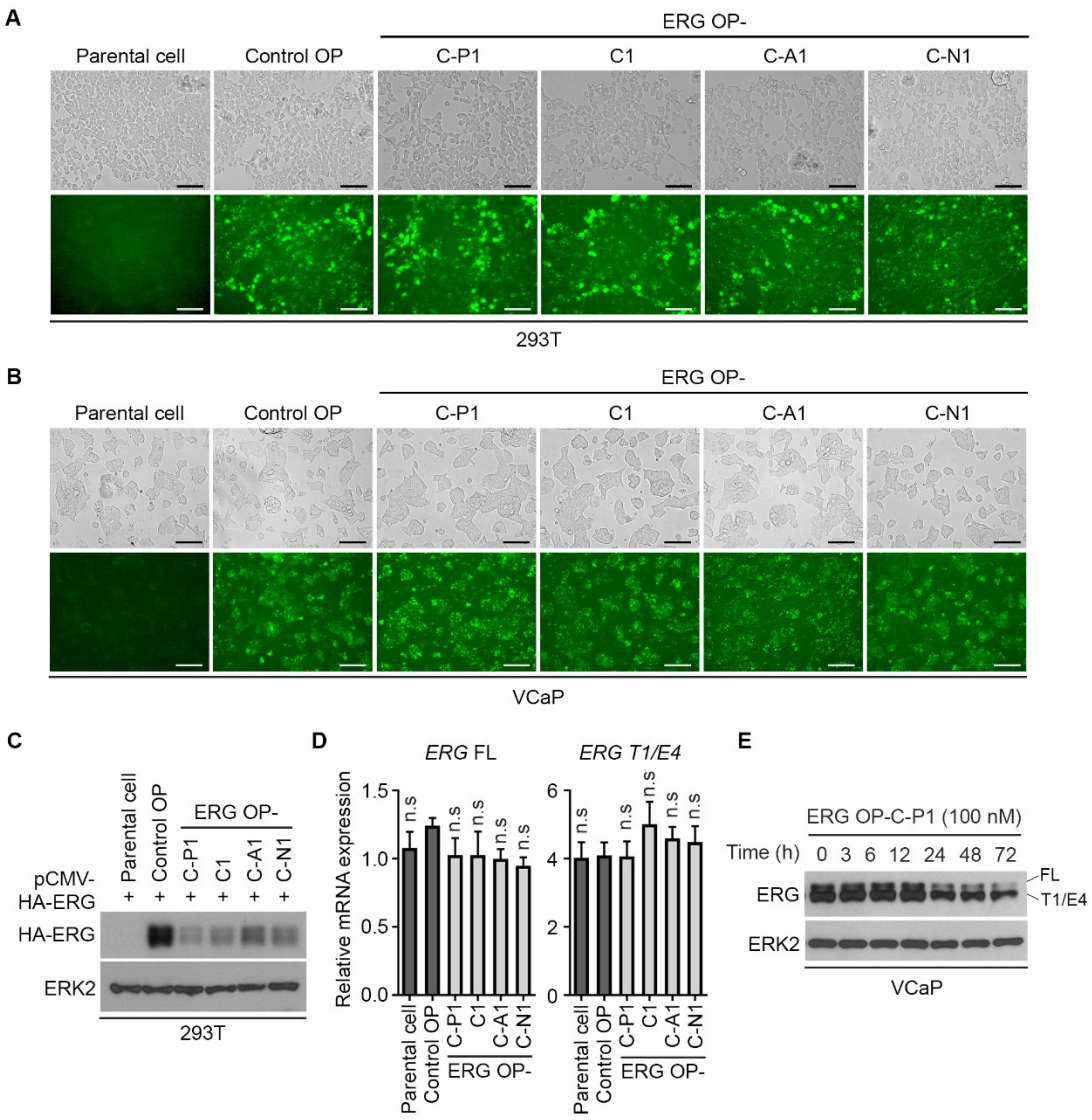


Fig. S3 Phthalic acid-based ERG O'PROTAC degrades ERG oncoprotein. (A and B) FITC-labeled ERG O'PROTACs, including ERG OP-C-P1 and OP-C1, and two previously published ERG O'PROTACs OP-C-A1 and C-N1 as positive controls [2], were individually transfected into 293T (A) and VCaP cells (B) at a final concentration of 100 nM with Lipofectamine 2000. Parental cells without any transfection were used as a negative control. Representative images of bright (upper) and fluorescent (lower) fields for indicated O'PROTACs are shown. Scale bar: 50 μm.

(C) The 293T cells were transfected with control or four indicated ERG O'PROTACs at a final concentration of 100 nm and harvested for western blot analysis. (D) The VCaP cells were transfected with control or four indicated ERG O'PROTACs at a final concentration of 100 nm and harvested for RT-qPCR to detect the mRNA level of ERG FL and T1/E4 after 48 h transfection. Data represents means ± SD (n = 3). The P values were determined using the unpaired two-tailed Student's t-test. The n.s. represents not significant by comparing the values in ERG O'PROTAC-treated groups to the control OP-treated group. (E) The VCaP cells were transfected with ERG OP-C-P1 at a final concentration of 100 nM and harvested at different time points, followed by western blot analysis of ERG protein expression.

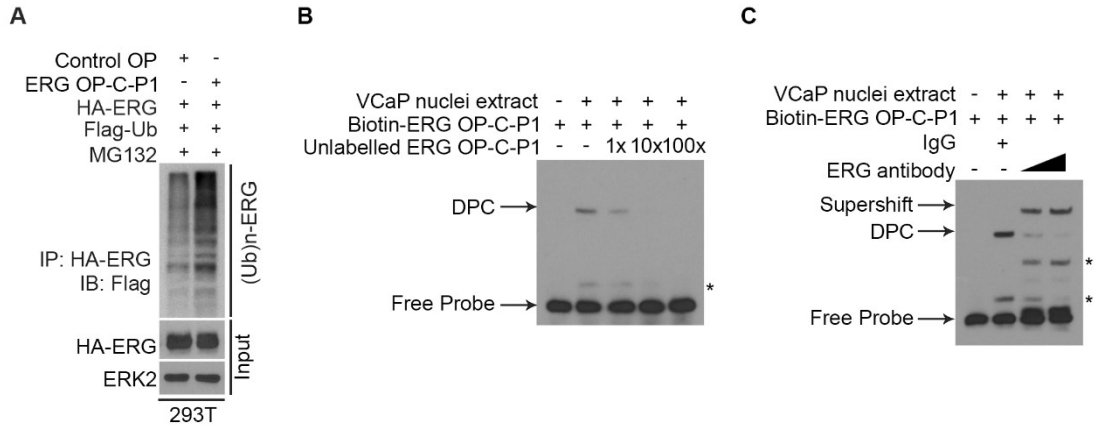


Fig. S4 Phthalic acid-based ERG OP degrades ERG via CRBN and the proteasome pathway. (A) The 293T were transfected with the indicated plasmids and ERG OP-C-P1 at a final concentration of 100 nM for 36 h and treated with the proteasome inhibitor MG132 (20 μM) for 12 h before harvested for protein extraction. ERG protein was immunoprecipitated with HA by protein A/G beads to detect its ubiquitination level by western blot analysis. (B) Biotin-labeled ERG OP-C-P1 (100 nm) was incubated with VCaP nuclear extract in the presence of an increasing amount of the non-biotin-labeled counterparts (1-, 10-, and 100-fold higher than the

concentration of biotin-labeled probe) followed by electrophoretic mobility shift assay (EMSA). DPC stands for DNA-protein complex. (C) Biotin-labeled ERG OP-C-P1 was incubated with VCaP nuclear extract in the presence of increasing amounts of ERG antibody (0.5 and 1 μg), followed by EMSA.

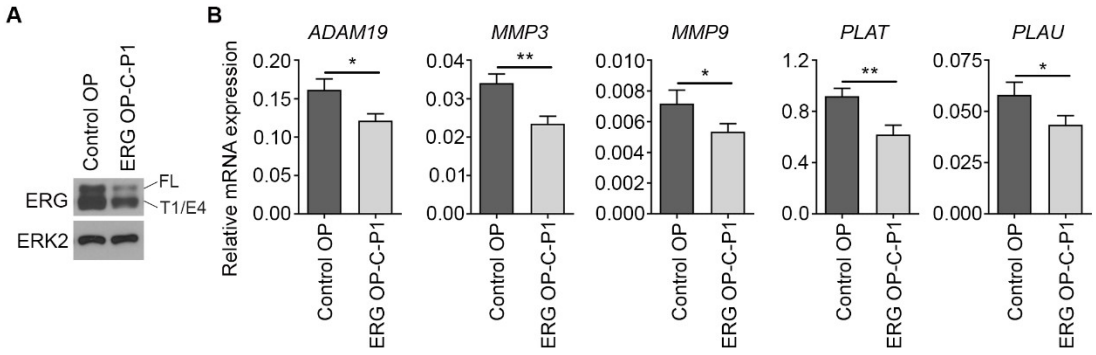


Fig. S5 Phthalic acid-based ERG OP inhibits ERG target gene expression.

(A and B) The VCaP cells were transfected with control OP or ERG OP-C-P1 at a final concentration of 100 nM for 48 h and harvested for western blot analysis (A) and RT-qPCR for the indicated ERG target genes (B). Data represents means ± SD (n = 3). P values were determined using the unpaired two-tailed Student's t-test; * $P < 0.05$; ** $P < 0.01$.

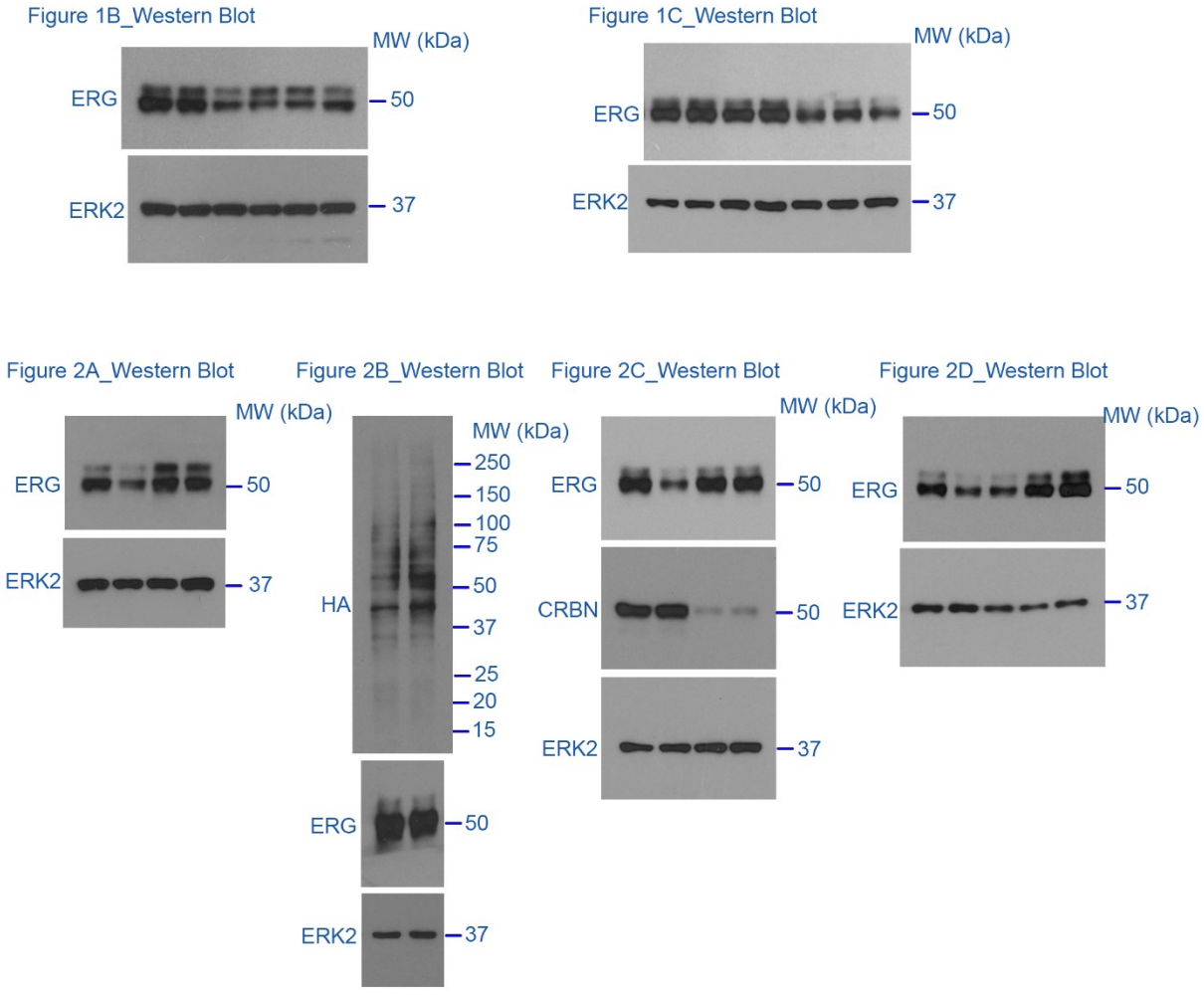


Fig. S6 Raw Western blot images for main figures (Figure 1B, 1C, Figure 2A, 2B, 2C and 2D).

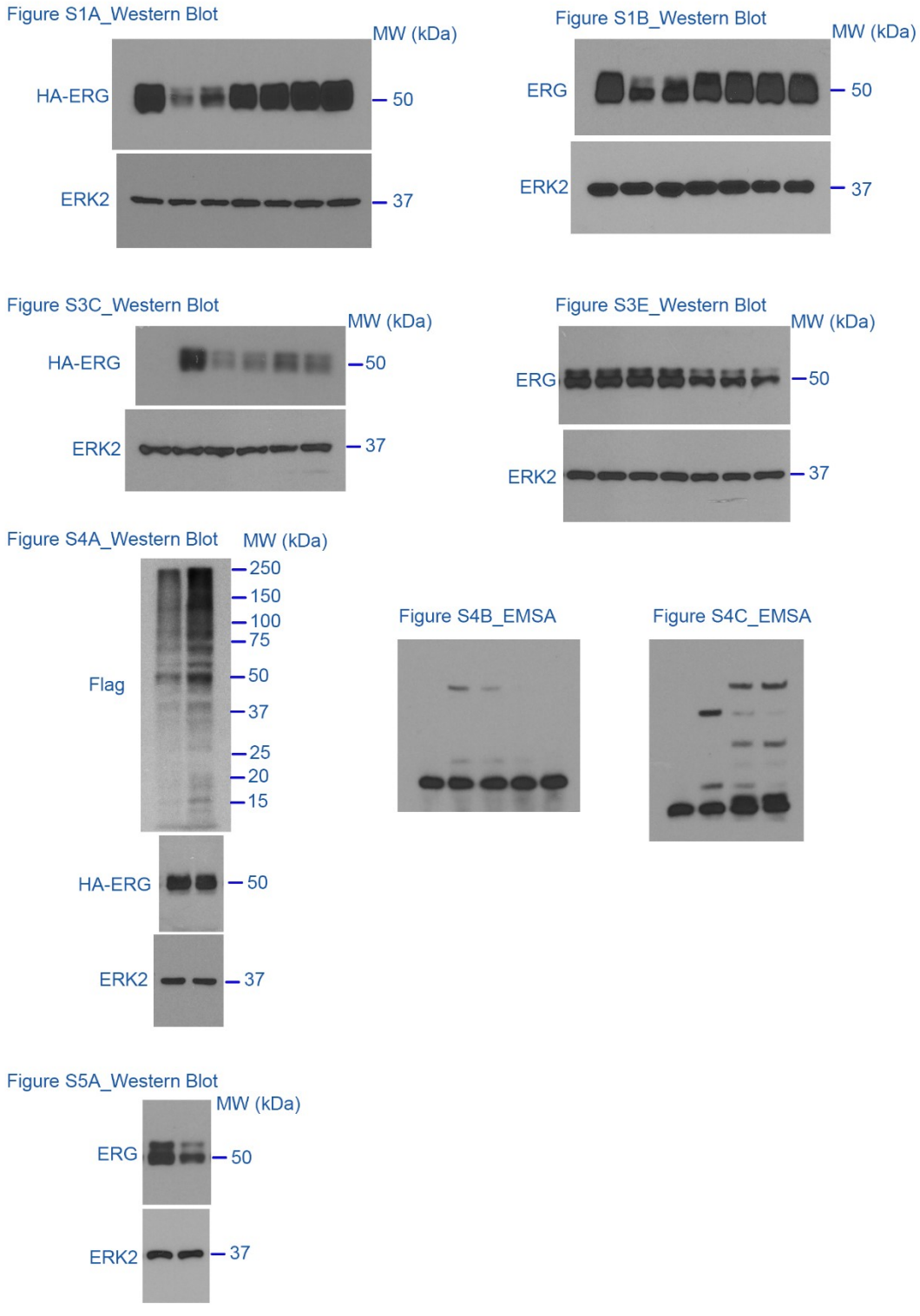


Fig. S7 Raw Western blot images for supplementary figures (Figure S1A, S1B, Figure S3C, S3E, Figure S4A, S4B, S4C, Figure S5A).

REFERENCES

1.Y. Yan, J. An, Y. Yang, D. Wu, Y. Bai, W. Cao, L. Ma, J. Chen, Z. Yu, Y. He, X. Jin, Y. Pan, T. Ma, S. Wang, X. Hou, S. J. Weroha, R. J. Karnes, J. Zhang, J. J. Westendorf, L. Wang, Y. Chen, W. Xu, R. Zhu, D. Wang and H. Huang, *EMBO Mol Med*, 2018, **10**.

2.J. Shao, Y. Yan, D. Ding, D. Wang, Y. He, Y. Pan, W. Yan, A. Kharbanda, H.-y. Li and H. Huang, *Advanced Science*, 2021, **n/a**, 2102555.