nature chemical biology

BRIEF COMMUNICATION

PUBLISHED ONLINE: 6 APRIL 2015 | DOI: 10.1038/NCHEMBIO.1793

Small molecule–triggered Cas9 protein with improved genome-editing specificity

Kevin M Davis[1,2], Vikram Pattanayak[3], David B Thompson[1,2], John A Zuris[1,2] & David R Liu[1,2]*

Directly modulating the activity of genome-editing proteins has the potential to increase their specificity by reducing activity following target locus modification. We developed Cas9 nucleases that are activated by the presence of a cell-permeable small molecule by inserting an evolved 4-hydroxytamoxifen–responsive intein at specific positions in Cas9. In human cells, conditionally active Cas9s modify target genomic sites with up to 25-fold higher specificity than wild-type Cas9.

The RNA-guided endonuclease Cas9 from the type II CRISPR-Cas system enables simple and efficient genome editing in a wide variety of organisms. Virtually any target DNA locus can be cleaved by programming Cas9 with a single guide RNA (sgRNA) that contains a stretch of ~20 nucleotides complementary to the target sequence[1–4]. Owing to its simplicity and robustness, the Cas9 system has been widely adopted for biological research and therapeutic development. The DNA cleavage specificity of Cas9 is imperfect[5–9], however, raising concerns over off-target genome modification that may limit its usefulness in therapeutic or research applications. Researchers have reduced Cas9 off-target activity through protein[10–13] and sgRNA[14] engineering, and by direct delivery of Cas9–sgRNA protein–RNA complexes into cells[15–17].

A complementary, underexplored strategy to improve Cas9 specificity is to reduce its activity once it has had sufficient opportunity to modify the target DNA locus. Indeed, higher concentrations of Cas9 in cells have been observed to degrade specificity[5–7] (defined as the ratio of on-target to off-target DNA cleavage activity), presumably because any Cas9 protein present after the target locus has been modified can only process off-target substrates. Unfortunately, wild-type Cas9 nucleases are not known to be regulated by other molecules and therefore are used in their constitutively active form. Although Cas9 can be regulated at the transcriptional level through the use of inducible promoters[18,19], transcriptional control cannot limit activity to the short temporal windows that may be necessary to maximize genome-editing specificity[17,20], in contrast with the high temporal resolution of post-translational strategies that directly control protein activity.

To enable post-translational control over Cas9 in cells, we sought to engineer variants of Cas9 that can be controlled with a readily available, cell-permeable small molecule. We previously evolved inteins that undergo protein splicing only in the presence of 4-hydroxytamoxifen (4-HT)[21]. These inteins were developed by inserting the human estrogen receptor ligand-binding domain into the *M. tuberculosis* RecA intein and evolving the resulting inactive fusion protein into a conditionally active intein that requires the presence of 4-HT[21–23]. Subsequent evolution at 37 °C yielded a second-generation intein, 37R3-2, with improved splicing properties in mammalian cells[23]. We envisioned that inserting the 37R3-2 intein into Cas9 at a location that disrupts Cas9 activity until protein splicing has taken place could result in conditionally active Cas9 nucleases that are active only in the presence of 4-HT (**Fig. 1a**).

We genetically inserted the 4-HT–dependent intein at each of 15 positions in Cas9 (Cys80, Ala127, Thr146, Ser219, Thr333, Thr519, Cys574, Thr622, Ser701, Ala728, Thr995, Ser1006, Ser1154, Ser1159 and Ser1274), chosen to distribute the location of the intein across the

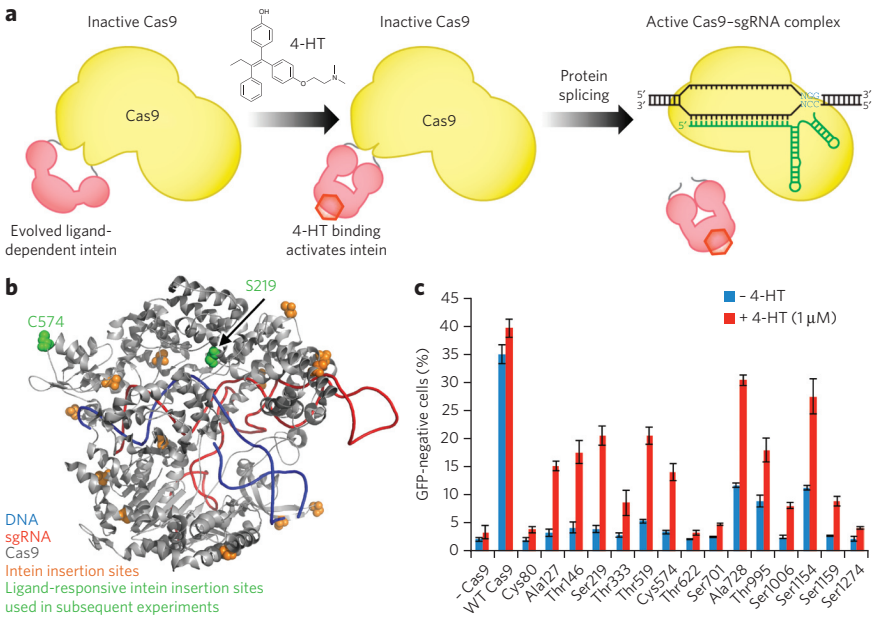

Figure 1 | Insertion of an evolved ligand-dependent intein enables small-molecule control of Cas9. (**a**) Intein insertion renders Cas9 inactive. Upon 4-HT binding, the intein undergoes conformational changes that trigger protein splicing and restore Cas9 activity. (**b**) The evolved intein was inserted to replace each of the colored residues. Intein-inserted Cas9 variants at Ser219 and Cys574 (green) were used in subsequent experiments. (**c**) Genomic *EGFP* disruption activity of wild-type Cas9 and intein-Cas9 variants in the absence or presence of 4-HT. Intein-Cas9 variants are identified by the residue replaced by the intein. Error bars reflect the s.d. of three biological replicates.

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA. [2]Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts, USA. [3]Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA. *e-mail: drliu@fas.harvard.edu

structural domains of Cas9 (ref. 24) (**Fig. 1b**). Because intein splicing leaves behind a single cysteine residue, the intein was inserted in place of one Cas9 amino acid in each of the 15 candidate constructs. In addition to replacing natural cysteine amino acids, we also favored replacing alanine, serine or threonine residues to minimize the likelihood that the resulting cysteine point mutation resulting from protein splicing would disrupt Cas9 activity. The 15 intein-Cas9 candidates were expressed in HEK293-GFP cells together with a sgRNA that targets the genomic *EGFP* locus in these cells. Twelve hours after transfection, cells were treated with or without 1 μM 4-HT. Five days after transfection, cells were analyzed on a flow cytometer for loss of GFP expression from Cas9-mediated *EGFP* cleavage and subsequent nonhomologous end joining.

Eight of the candidates, corresponding to intein insertion at Ala127, Thr146, Ser219, Thr333, Thr519, Cys574, Ser1006 and Ser1159, demonstrated 4-HT–dependent loss of GFP expression, consistent with 4-HT–triggered Cas9 activity (**Fig. 1c**). Notably, three intein-Cas9 variants (insertion at Ala728, Thr995 and Ser1154) showed high DNA modification rates both in the presence and absence of 4-HT, suggesting that large protein insertions at these positions do not substantially inhibit nuclease activity or that the intein lost its 4-HT dependence owing to context-dependent conformational perturbations. We speculate that it may be possible to engineer split Cas9 variants by dividing the protein at these locations, given their tolerance of a 413-residue insertion. The lack of nuclease activity of the remaining four intein-Cas9 variants (insertion at Cys80, Thr622, Ser701 and Ser1274) in the presence or absence of 4-HT could result from the inability of the intein to splice in those contexts, the inability of Cas9 to refold properly following splicing or intolerance of replacement of native threonine or serine residues with cysteines. We pursued two intein-Cas9 variants corresponding to insertion at Ser219 and Cys574 (**Fig. 1b**). These two variants combined high activity in the presence of 4-HT and low activity in the absence of 4-HT.

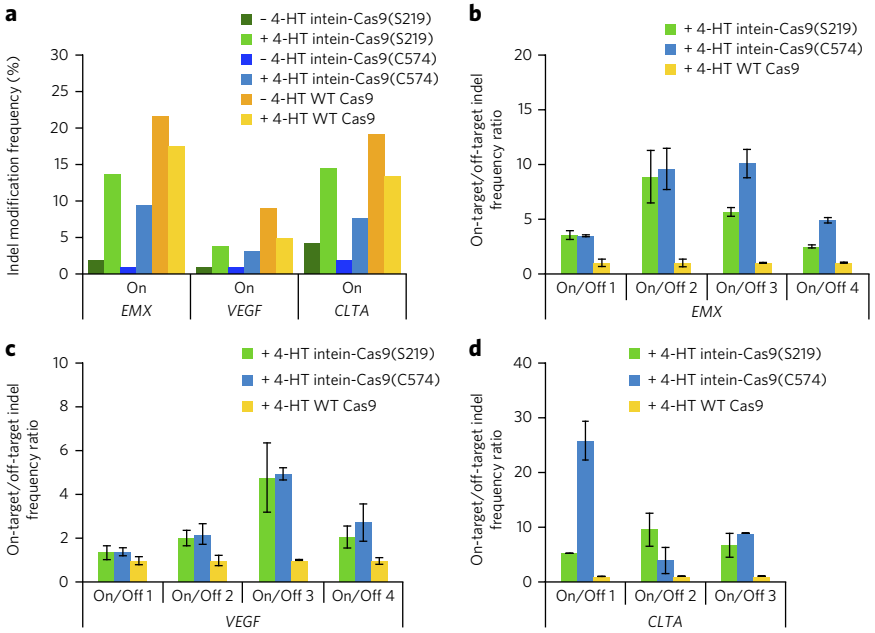

Figure 2 | Genomic DNA modification by intein-Cas9(S219), intein-Cas9(C574) and wild-type Cas9. (**a**) Indel frequency from high-throughput DNA sequencing of amplified genomic on-target sites in the absence or presence of 4-HT. Note that a substantial number of indels were observed at the *CLTA* on-target site even in the absence of a targeting sgRNA (**Supplementary Table 7**). (**b**–**d**) DNA modification specificity, defined as on-target/off-target indel frequency ratio[5–7], normalized to wild-type Cas9. Cells were transfected with 500 ng of the Cas9 expression plasmid and incubated at transfection with medium containing 1 μM 4-HT. After 12 h, the medium was replaced with medium lacking 4-HT. *P* values are $<10^{-15}$ for the Fisher exact test (one-sided, up) on comparisons of indel modification frequency in the presence versus the absence of 4-HT for intein-Cas9(S219) and intein-Cas9(C574). *P* values were adjusted for multiple comparisons using the Benjamini-Hochberg method and are listed in **Supplementary Table 3**. Error bars reflect the range of two independent experiments conducted on different days.

To evaluate the genome modification specificity of conditionally active Cas9 variants, we expressed intein-Cas9(S219), intein-Cas9(C574) and wild-type Cas9 in HEK293-GFP cells together with each of three previously described[12] sgRNAs that target the well-studied *EMX*, *VEGF* and *CLTA* genomic loci. We assayed these Cas9–sgRNA combinations in human cells for their ability to modify the 3 on-target loci as well as 11 known off-target genomic sites (**Supplementary Results**, **Supplementary Table 1**)[5,6,11,14]. Cells were treated with or without 1 μM 4-HT during transfection, and after 12 h the medium was replaced with fresh medium lacking 4-HT. We observed no cellular toxicity arising from 12 h or 60 h of treatment with 1 μM 4-HT in untransfected or transfected HEK293 cells (**Supplementary Fig. 1**). Genomic DNA was isolated 60 h after transfection and analyzed by high-throughput DNA sequencing.

Overall on-target genome modification frequency of intein-Cas9(S219) and intein-Cas9(C574) in the presence of 1 μM 4-HT was similar to that of wild-type Cas9 (**Fig. 2a**, **Supplementary Tables 2** and **3** and **Supplementary Data Set 1**). On-target modification frequency in the presence of 4-HT was 3.4- to 7.3-fold higher for intein-Cas9(S219) and 3.6- to 9.6-fold higher for intein-Cas9(C574) than in the absence of 4-HT, whereas modification efficiency for wild-type Cas9 was 1.2- to 1.8-fold lower in the presence of 4-HT (**Fig. 2a**). Both intein-Cas9 variants exhibited a low level of background activity in the absence of 4-HT, consistent with previous reports[21–23]. Western blot analysis of intein-Cas9(S219) from transfected HEK293 cells confirmed the presence of spliced product at the earliest assayed time point (4 h) following 4-HT treatment; no spliced product was detected in the absence of 4-HT (**Supplementary Fig. 2**). Together, these results indicate that intein-Cas9(S219) and intein-Cas9(C574) are slightly less active than wild-type Cas9 in the presence of 4-HT, most likely owing to incomplete splicing (**Supplementary Fig. 2**), but are much less active in the absence of 4-HT.

High-throughput sequencing of 11 previously described off-target sites that are modified by wild-type Cas9–sgRNA complexes targeting the *EMX*, *VEGF* and *CLTA* loci revealed that both intein-Cas9 variants, when treated with 4-HT for 12 h, exhibit substantially improved specificity compared to that of wild-type Cas9 (**Supplementary Fig. 3**, **Supplementary Tables 2**, **4** and **5** and **Supplementary Data Sets 1** and **2**). On-target/off-target indel modification ratios for both intein-Cas9 variants were on average sixfold higher and as much as 25-fold higher than that of wild-type Cas9 (**Fig. 2b**–**d**). In the absence of 4-HT, the genome modification specificity of both intein-Cas9 variants was on average 14-fold higher than that of wild-type Cas9 in the absence of 4-HT (**Supplementary Fig. 4**), presumably resulting from the much lower activity of the intein-Cas9 variants in the absence of 4-HT[5–7].

As intein-Cas9s can result in slightly lower on-target modification rates compared to wild-type Cas9 (**Fig. 2a**), we sought to verify that the improvements in specificity among the intein-Cas9s were not simply a result of reduced activity. Both on- and off-target activity

of Cas9 has been shown to be dependent on the amount of Cas9 expression plasmid transfected[5–7]. By transfecting lower amounts of the wild-type Cas9 expression plasmid, we compared intein-Cas9s with wild-type Cas9 under conditions that result in very similar levels of on-target modification. To minimize potential differences in transfection efficiency, we supplemented with a plasmid that does not express Cas9 so that the same total amount of plasmid DNA was transfected into each sample. High-throughput sequencing revealed that wild-type Cas9 shows slightly improved specificity, as expected, as the on-target cleavage rate is reduced. The intein-Cas9 variants, however, remain substantially more specific than wild-type Cas9 at similar on-target DNA cleavage rates (**Supplementary Figs. 5–7**, **Supplementary Tables 4** and **6** and **Supplementary Data Set 2**). For example, intein-Cas9(C574) and wild-type Cas9 (80 ng) have virtually identical on-target DNA cleavage rates (both 6.4%) at the *EMX* locus, but all four off-target sites are modified at, on average, fourfold lower frequencies ($P < 1 \times 10^{-13}$) by intein-Cas9(C574) than by wild-type Cas9 (80 ng). These findings indicate that specificity improvements of intein-Cas9 variants do not simply arise from differences in overall genome editing activity.

Intein 37R3-2 can be activated by other estrogen receptor modulators. To enable intein-Cas9 applications in which endogenous β-estradiol is present, we inserted into the estrogen receptor ligand-binding domain a point mutation (G521R) that renders the domain more specific for 4-HT[25]. This mutation slightly reduces affinity for 4-HT but almost abolishes affinity for β-estradiol. The addition of this mutation to intein-Cas9(S219) eliminates the ability of β-estradiol to trigger Cas9 activity (**Supplementary Fig. 8**).

The intein-Cas9 variants developed here enable small-molecule control of Cas9 function, thereby enhancing genome modification specificity. Together with our recent observations that direct delivery of transient Cas9 protein into cells can also improve its specificity[17], these findings strongly suggest that limiting the opportunity of genome-editing agents to modify off-target loci following a period of activity sufficient to induce desired levels of on-target modification results in substantial specificity improvements. Our findings complement those of a recent study[26] describing small molecule–dimerized split Cas9 halves with improved specificity, published while this study was in review. Because the intein-Cas9s, unlike split Cas9 proteins, are expressed as single polypeptides that, upon splicing, generate Cas9 proteins with an identical (or nearly identical) amino acid sequence as wild-type Cas9, these two approaches offer complementary strengths. We anticipate that the use of ligand-dependent Cas9 variants will provide greater control over genomic modification efficiencies and specificities than what is currently achievable with constitutively active or transcriptionally regulated genome editing. In principle, this approach could synergize with other specificity-augmenting strategies such as using truncated guide RNAs[14], paired Cas9 nickases[10,11] or FokI-dCas9 fusions[12,13]. This approach could also be applied to other genome engineering proteins to enable, for example, small-molecule control of TALE-based or Cas9-mediated transcriptional regulators.

Received 12 January 2015; accepted 9 March 2015;
published online 6 April 2015

Methods

Methods and any associated references are available in the online version of the paper.

References

1. Jinek, M. *et al. Science* **337**, 816–821 (2012).
2. Mali, P. *et al. Science* **339**, 823–826 (2013).
3. Cong, L. *et al. Science* **339**, 819–823 (2013).
4. Jinek, M. *et al. eLife* **2**, e00471 (2013).
5. Pattanayak, V. *et al. Nat. Biotechnol.* **31**, 839–843 (2013).
6. Fu, Y. *et al. Nat. Biotechnol.* **31**, 822–826 (2013).
7. Hsu, P.D. *et al. Nat. Biotechnol.* **31**, 827–832 (2013).
8. Cho, S.W. *et al. Genome Res.* **24**, 132–141 (2014).
9. Tsai, S.Q. *et al. Nat. Biotechnol.* **33**, 187–197 (2015).
10. Mali, P. *et al. Nat. Biotechnol.* **31**, 833–838 (2013).
11. Ran, F.A. *et al. Cell* **154**, 1380–1389 (2013).
12. Guilinger, J.P., Thompson, D.B. & Liu, D.R. *Nat. Biotechnol.* **32**, 577–582 (2014).
13. Tsai, S.Q. *et al. Nat. Biotechnol.* **32**, 569–576 (2014).
14. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. *Nat. Biotechnol.* **32**, 279–284 (2014).
15. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. *Genome Res.* **24**, 1012–1019 (2014).
16. Ramakrishna, S. *et al. Genome Res.* **24**, 1020–1027 (2014).
17. Zuris, J.A. *et al. Nat. Biotechnol.* **33**, 73–80 (2015).
18. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. *Science* **343**, 80–84 (2014).
19. González, F. *et al. Cell Stem Cell* **15**, 215–226 (2014).
20. Pruett-Miller, S.M., Reading, D.W., Porter, S.N. & Porteus, M.H. *PLoS Genet.* **5**, e1000376 (2009).
21. Buskirk, A.R., Ong, Y.C., Gartner, Z.J. & Liu, D.R. *Proc. Natl. Acad. Sci. USA* **101**, 10505–10510 (2004).
22. Yuen, C.M., Rodda, S.J., Vokes, S.A., McMahon, A.P. & Liu, D.R. *J. Am. Chem. Soc.* **128**, 8939–8946 (2006).
23. Peck, S.H., Chen, I. & Liu, D.R. *Chem. Biol.* **18**, 619–630 (2011).
24. Jinek, M. *et al. Science* **343**, 1247997 (2014).
25. Danielian, P.S., White, R., Hoare, S.A., Fawell, S.E. & Parker, M.G. *Mol. Endocrinol.* **7**, 232–240 (1993).
26. Zetsche, B., Volz, S.E. & Zhang, F. *Nat. Biotechnol.* **33**, 139–142 (2015).

Acknowledgments

This work was supported by National Institutes of Health R01 GM095501, Defense Advanced Research Projects Agency HR0011-11-2-0003 and N66001-12-C-4207 and the Howard Hughes Medical Institute. K.M.D. acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC) for a Postgraduate Scholarship–Doctoral (PGS D) award. J.A.Z. is a Ruth L. Kirchstein National Research Service Awards Postdoctoral Fellow (F32 GM 106601-2). We are grateful to J. Doudna, S. Sternberg, D. Taylor, M. Jinek and F. Jiang for providing the structural coordinates of Cas9 and to J. Guilinger, Y. Kim, M. Li and A. Badran for helpful discussions.

Author contributions

K.M.D., V.P., D.B.T. and D.R.L. designed the research. K.M.D. performed the experiments, and J.A.Z. assisted with high-throughput sequencing. K.M.D., V.P., D.B.T. and J.A.Z. analyzed the data. D.R.L. supervised the research. All authors wrote the manuscript.

Competing financial interests

The authors declare competing financial interests: details accompany the online version of the paper.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to D.R.L.

ONLINE METHODS

Cas9, intein-Cas9 and sgRNA expression plasmids. A plasmid encoding the human codon-optimized *Streptococcus pyogenes* Cas9 nuclease with an NLS and 3× Flag tag (Addgene plasmid 43861)[6] was used as the wild-type Cas9 expression plasmid. Intein 37R3-2 was subcloned at the described positions into the wild-type Cas9 expression plasmid using USER (NEB M5505) cloning. sgRNA expression plasmids used in this study have been described previously[12]. Plasmid constructs generated in this work will be deposited with Addgene.

Modification of genomic GFP. HEK293-GFP stable cells (GenTarget), which constitutively express Emerald GFP, served as the reporter cell line. Cells were maintained in 'full serum medium': Dulbecco's Modified Eagle's Medium plus GlutaMax (Life Technologies) with 10% (v/v) FBS and penicillin/streptomycin (1×, Amresco). 5×10^4 cells were plated on 48-well collagen-coated Biocoat plates (Becton Dickinson). 16–18 h after plating, cells were transfected with Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Briefly, 1.5 µl of Lipofectamine 2000 was used to transfect 650 ng of total plasmid: 500 ng Cas9 expression plasmid, 125 ng sgRNA expression plasmid and 25 ng near-infrared iRFP670 expressing plasmid (Addgene plasmid 45457)[27]. Twelve hours after transfection, the medium was replaced with full serum medium, with or without 4-HT (1 µM, Sigma-Aldrich T176). The medium was replaced again 3–4 d after transfection. Five days after transfection, cells were trypsinized and resuspended in full serum medium and analyzed on a C6 flow cytometer (Accuri) with a 488-nm laser excitation and 520-nm filter with a 20-nm band pass. Transfections and flow cytometry measurements were performed in triplicate.

High-throughput DNA sequencing of genome modifications. HEK293-GFP stable cells were transfected with plasmids expressing Cas9 (500 ng) and sgRNA (125 ng) as described above. For treatments in which a reduced amount of wild-type Cas9 expression plasmid was transfected, pUC19 plasmid was used to bring the total amount of plasmid to 500 ng. 4-HT (1 µM final), where appropriate, was added during transfection. 12 h after transfection, the medium was replaced with full serum medium without 4-HT. Genomic DNA was isolated and pooled from three biological replicates 60 h after transfection using a previously reported[12] protocol with a DNAdvance Kit (Agencourt). 150 ng or 200 ng of genomic DNA was used as a template to amplify by PCR the on-target and off-target genomic sites with flanking HTS primer pairs described previously[12]. PCR products were purified using RapidTips (Diffinity Genomics) and quantified using the PicoGreen dsDNA Assay Kit (Invitrogen). Purified DNA was PCR amplified with primers containing sequencing adaptors and purified with the MinElute PCR Purification Kit (Qiagen) and AMPure XP PCR Purification (Agencourt). Samples were sequenced on a MiSeq high-throughput DNA sequencer (Illumina), and sequencing data were analyzed as described previously[5].

Western blot analysis of intein splicing. HEK293-GFP stable cells were transfected with 500 ng Cas9 expression plasmid and 125 ng sgRNA expression plasmid. 12 h after transfection, the medium was replaced with full serum medium, with or without 4-HT (1 µM). Cells were lysed and pooled from three technical replicates 4 h, 8 h, 12 h or 24 h after 4-HT treatment. Samples were run on a Bolt 4–12% Bis-Tris gel (Life Technologies). An anti-Flag antibody (Sigma-Aldrich F1804) and an anti-mouse 800CW IRDye (LI-COR) were used to visualize the gel on an Odyssey IR imager.

Statistical analysis. Statistical tests were performed as described in the figure captions. All *P* values were calculated with the R software package. *P* values for the Fisher exact test were calculated using the fisher.test function, with a one-sided alternative hypothesis (alternative = 'greater' or alternative = 'less', as appropriate). Upper bounds on *P* values that are close to zero were determined manually. The Benjamini-Hochberg adjustment was performed using the R function p.adjust (method = 'fdr').

27. Shcherbakova, D.M. & Verkhusha, V.V. *Nat. Methods* **10**, 751–754 (2013).

doi:10.1038/nchembio.1793