



Published in final edited form as:

Nat Biotechnol. 2019 October ; 37(10): 1209–1216. doi:10.1038/s41587-019-0242-8.

Multi-input chemical control of protein dimerization for programming graded cellular responses

Glenna Wink Foight^{1,2,†}, Zhizhi Wang^{2,3,†}, Cindy T. Wei¹, Per Greisen Jr^{2,4,6}, Katrina M. Warner¹, Daniel Cunningham-Bryant¹, Keunwan Park^{2,4,7}, T. J. Brunette^{2,4}, William Sheffler^{2,4}, David Baker^{2,4,5}, Dustin J. Maly^{1,4,*}

¹Department of Chemistry, University of Washington, Seattle, Washington, USA.

²Institute for Protein Design, University of Washington, Seattle, Washington, USA.

³Department of Biological Structure, University of Washington, Seattle, Washington, USA.

⁴Department of Biochemistry, University of Washington, Seattle, Washington, USA.

⁵Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA.

⁶Current addresses: Global Research, Novo Nordisk A/S, Måløv, Denmark.

⁷Current addresses: Systems Biotechnology Research Center, Korea Institute of Science and Technology, Gangneung, Republic of Korea.

Abstract

Chemical and optogenetic methods for post-translationally controlling protein function have enabled new discoveries and the engineering of synthetic cellular functions. However, most of these methods only confer single-input, single-output control. To increase the diversity of post-translational behaviors that can be programmed we built a system based on a single protein receiver that can integrate multiple drug inputs, including approved therapeutics. Our system translates drug inputs into diverse outputs with engineered reader proteins that provide variable dimerization states of the receiver protein. We show that our single receiver protein architecture can be used to program diverse cellular responses, including graded and proportional dual-output control of transcription and mammalian cell signaling. We apply our tools to titrate the competing activities of the Rac and Rho GTPases to control cell morphology. Our receiver protein and suite

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding author: djmaly@uw.edu.

†These authors contributed equally to this work

Author contributions

G.W.F., Z.W., D.B., and D.J.M. conceived of the work. G.W.F. and D.J.M. wrote the manuscript. G.W.F. designed and validated the grazoprevir/NS3a reader, performed all improvement and validation of the danoprevir/NS3a reader, and completed and analyzed all mammalian cell experiments. Z.W. performed the danoprevir/NS3a reader crystallography. Z.W., P.J.G., and K.P. wrote the design script and contributed the danoprevir/NS3a reader design. C.T.W. designed and screened the grazoprevir/NS3a reader library. K.W. performed the CD95 titration. D.C.B. established the ANR/NS3a system. T.B. contributed the DHR scaffold set. W.S. created RIFdock.

Competing interests

The authors G.W.F, Z.W., C.T.W., P.G., D.C.B., K.P., D.B., and D.J.M. are inventors on US Patent application 62/775,171, "Reagents and Methods for Controlling Protein Function and Interaction" submitted by University of Washington.

of reader proteins provides researchers with a versatile toolset to post-translationally program mammalian cellular processes and to engineer cell therapies.

Cells exhibit proportional, graded, digital and temporal behaviors in sensing and responding to multiple environmental or autologous inputs.^{1–3} Biologists seeking to reproduce natural functions, or create new ones, need tools that can program a similar range of behaviors. Most reported synthetic biology tools are based on transcriptional circuits that can enable a wide variety of quantitative control modes.^{4,5} However, methods for rapid, protein-level manipulation of cellular processes have lagged behind due to the difficulty of engineering complex post-translational control schemes.

For mammalian synthetic biology applications, post-translational control systems that use small molecules as extrinsic inputs are desirable for many applications because they are easy to use *in vitro*, *ex vivo* and *in vivo* and confer temporal modulation.⁶ Chemically-controlled proteases and degradation domains have been applied for post-translational control.^{7–9} Two recently-developed, chemically-controlled systems that use catalytically-active hepatitis C virus (HCV) protease NS3a as a cleavage-based modulator of mammalian cellular processes are particularly attractive because they use orally-available, clinically approved drugs that are orthogonal to mammalian systems as extrinsic inputs.^{10,11} Chemically-induced dimerization (CID) systems, which modulate cellular processes through small molecule-induced protein proximity, are advantageous for applications that require more rapid cellular responses, like cellular signaling, than protease- or degradation-based systems.^{12–14} Although there has been recent success in expanding the diversity of small molecules that can be used in CID systems, no system that uses a clinically-approved drug that lacks an endogenous mammalian target has been described to date.¹⁵

A limitation of current chemically-controlled systems is that they rely on single small molecule inputs that are translated into single outputs, which limits the types of cellular responses that can be programmed. There has been success in combining orthogonal CID systems to achieve digital logic control of cell signaling and transcription.^{14,16} In addition, combining composable, single-input/single-output protease-based systems has allowed the assembly of a diversity of digital circuits.¹⁷ While digital logic is useful, current post-translational control systems lack robust analog outputs, such as graded and proportional control, that are needed to fully mimic natural cellular processes.

Here, we present a new post-translational control system that utilizes the NS3a protease as a central receiver protein that is targeted by multiple clinically-approved drug inputs. To translate different drug-bound states of NS3a into diverse outputs, we engineer computationally-designed “reader” proteins that recognize specific inhibitor-bound states of NS3a and use a genetically-encoded peptide that selectively recognizes the *apo* form of this protease (Fig. 1a). Our system, called Pleiotropic Response Outputs from a Chemically-Inducible Single Receiver (PROCISiR), can be used to program diverse cellular responses owing to its single receiver protein architecture.

Results

Computational design of NS3a readers

Rosetta interface design allowed us to develop protein readers that selectively recognize a binding surface centered on NS3a-bound inhibitors (Fig. 1b).¹⁸ First, we used a set of stable, *de novo*-designed proteins as scaffolds on which to design an interface with the danoprevir:NS3a complex, including leucine-rich repeat proteins (LRRs), designed helical repeat proteins (DHRs), ferredoxins, and helical bundles.^{19–21} As a starting point, PatchDock was used to center each scaffold over NS3a-bound danoprevir, followed by RosettaDesign on the scaffold surface that forms the binding interface.²² DHR design D5, one of 31 designs selected for testing via yeast surface display based on interface scoring metrics, showed modest, drug-dependent binding to NS3a, which required designed interface residues (Fig. 1c). To improve D5's affinity for the NS3a:danoprevir complex, we used two sequential yeast surface display libraries (Supplementary Fig. 1, Supplementary Notes). Our final variant of the danoprevir:NS3 complex reader (DNCR), DNCR2 had an apparent affinity for the NS3a:danoprevir complex of 36 pM, no detectable binding to apo NS3a, and >20,000-fold selectivity over NS3a bound to the drugs grazoprevir or asunaprevir (Supplementary Table 1, Supplementary Fig. 2a). Further biochemical analysis confirmed that DNCR2 does not bind substantially to free danoprevir and that DNCR2:danoprevir:NS3a form a 1:1:1 complex (Supplementary Fig. 2b,e, Supplementary Notes). A 2.3 Å resolution structure of the DNCR2:danoprevir:NS3a complex revealed a modest shift for DNCR2 relative to the D5 model, but with both interfaces sharing an overlapping set of positions (Fig. 1d,e, Supplementary Fig. 2c,d, Supplementary Table 2). The structural basis for the selective binding of DNCR2 to the NS3a:danoprevir complex is clearly apparent when structures of asunaprevir- or grazoprevir-bound NS3a are aligned to the DNCR2:danoprevir:NS3a complex (Supplementary Fig. 2f).^{23,24}

The high specificity of DNCR2 provided confidence that we could design additional readers that selectively recognize other NS3a:drug complexes. We computationally designed a reader of the grazoprevir:NS3a complex by applying a similar methodology, this time focusing exclusively on DHR scaffolds due to the success of D5 as a reader of the danoprevir/NS3a complex (see Supplementary Notes for more design details). We also employed a new reverse rotamer docking protocol, RIFdock, in addition to the more coarse-grained PatchDock method.²⁵ One design of the 29 tested, G3, showed modest, grazoprevir-dependent binding, which was not observed for the original scaffold or interface mutants (Fig. 2a). Screening a single library for improved affinity yielded grazoprevir:NS3a complex reader 1 (GNCR1) (Supplementary Fig. 3a,b, Supplementary Notes). GNCR1 demonstrated an apparent affinity for the grazoprevir:NS3a complex of 140 nM and little-to-no affinity for apo, danoprevir-, or asunaprevir-bound NS3a (Supplementary Fig. 3c, Supplementary Table 1). To provide a third reader, we used a previously described, genetically-encoded peptide that binds to the active site of apo NS3a, here called apo NS3a reader (ANR).^{26,27} ANR forms a basal complex with NS3a with an affinity of 10 nM that is disrupted by NS3a-targeting drugs (Fig. 1a).

Validation of PROCISiR in mammalian cells

We next sought to evaluate the unique, multi-input/multi-output behaviors that PROCISiR could achieve in mammalian cells. First, we verified that danoprevir was capable of rapidly colocalizing DNCR2 with membrane-targeted NS3a ($t_{1/2}$ of 76 ± 27 sec (mean, standard deviation)), which was capable of activating PI3K-Akt signaling when DNCR2 was fused to the inter-SH2 domain from the p85 regulatory subunit of PI3K (Supplementary Fig. 4).²⁸ The drug specificity of DNCR2 was maintained in cells, as neither grazoprevir nor asunaprevir promoted DNCR2-EGFP colocalization with mitochondrially-targeted NS3a (Fig. 2b). We then combined DNCR2 with GNCR1 or ANR to divergently localize mCherry-NS3a to different subcellular spaces. We observed that grazoprevir exclusively colocalized NS3a-mCherry with membrane-targeted GNCR1, while only danoprevir led to colocalization with mitochondrially-targeted DNCR2 (Fig. 2c, Supplementary Fig. 5a). Likewise, membrane-targeted ANR pre-localized NS3a-mCherry to the plasma membrane, and danoprevir treatment recruited NS3a to the nucleus with nuclear-targeted DNCR2 (Fig. 2d, Supplementary Fig. 5b). These and additional colocalization experiments (Supplementary Figs. 6,7, Supplementary Notes) validated that the DNCR2, GNCR1, and ANR readers are selective for their targeted state of NS3a and can be used in concert.

PROCISiR enables programmable transcriptional control

The ability of our readers to discriminate between different states of NS3a allows complex control modes to be achieved by combining inputs and/or readers, a capability not shared by chemically inducible systems for which there is only one input and one subsequent protein complex formed. First, we used danoprevir as an agonist and grazoprevir as an antagonist to temporally and proportionally control transcription of one endogenous gene in HEK293 cells using DNCR2-VPR (a transcriptional activator) and an NS3a-dCas9 fusion (*Streptococcus pyogenes*).^{29,30} By chasing with grazoprevir after pulsing transcriptional activation with DNCR2-VPR and danoprevir, we were able to rapidly reverse CXCR4 expression (Fig. 3a). Next, to achieve graded control of the steady-state transcriptional output from DNCR2-VPR, we co-treated cells with danoprevir and increasing grazoprevir proportions, producing a linear output from endogenous CXCR4 and CD95 promoters over three orders of magnitude of danoprevir input (Fig. 3b,c). We found that this precise titration of gene expression on a single-cell level could also be produced from an exogenous promoter (Supplementary Fig. 8a). Commonly used methods for mammalian gene induction, such as highly cooperative TetR-based systems, are ineffective at achieving intermediate levels of gene expression.³¹

We also combined two inputs with two readers to achieve proportional and graded control of two transcriptional outputs at once. Titration of each drug alone in the two-gene transcriptional activation system shown in Figure 3d, which utilizes orthogonal RNA hairpin/RNA-binding proteins to target two different promoters, demonstrated that observed EC_{50} values, 0.16 ± 0.03 nM and 0.79 ± 0.15 nM (mean \pm standard deviation) for the grazoprevir:NS3a and danoprevir:NS3a readers, respectively, are in close agreement with NS3a's K_i value for each drug (Supplementary Fig. 8c,d).^{23,32} Hence, the transcriptional output from these systems depends on the binding affinity of the drugs to NS3a, not the affinity of the readers for their drug/NS3a complex. The dependence of DNCR2- and

GNCR1-promoted transcription levels on their inducer's K_i value allowed us to model the steady-state output from each reader:drug:NS3a complex at mixed danoprevir and grazoprevir concentrations (Fig. 3e, Supplementary Fig. 9, Supplementary Notes). Proportional expression output of CXCR4 and GFP across a matrix of grazoprevir and danoprevir concentrations demonstrated close concordance with predicted NS3a:drug complex levels (Fig. 3e,f). See Supplementary Notes and Supplementary Fig. 10 for a description of additional transcriptional control modes that can be achieved, including 3-gene control and switchable repression/overexpression. The responsive nature of the PROCISiR architecture enables diverse modes of temporal, graded, or proportional, and multi-state transcriptional control, all realized at the post-translational level.

Proportional control of cell signaling with PROCISiR

Finally, we used PROCISiR to interrogate signaling pathway crosstalk by colocalizing signaling effector domains fused to DNCR2 or GNCR1 with membrane-localized NS3a in HeLa and NIH3T3 cells (Fig. 4a). We predicted danoprevir and grazoprevir concentrations that would provide graded and proportional regimes of DNCR2 and GNCR1 colocalization with NS3a by modeling NS3a:drug complex levels (Supplementary Fig. 9, Supplementary Notes). Confocal microscopy of EGFP-DNCR2 and BFP-GNCR1 with membrane-localized NS3a validated that these drug combinations provided the desired colocalization regimes (Fig. 4b, Supplementary Fig. 11c). These datasets based on fluorescent protein colocalization at the membrane provide a useful model for the competition between DNCR2 and GNCR1 for binding to membrane-bound NS3a at different danoprevir and grazoprevir concentrations.

We next used these same danoprevir/grazoprevir drug combinations with DNCR2-TIAM (Rac1 GEF) and GNCR1-LARG (RhoA/B/C GEF) fusions to provide proportional and graded control over the activation of Rac and Rho GTPases, which have opposing effects on the actin cytoskeleton and cell morphology.³³ In cells co-expressing DNCR2-TIAM and GNCR1-LARG with membrane-localized NS3a, with F-actin visualized by Lifeact-mCherry, danoprevir treatment resulted in cell expansion due to the activity of Rac in promoting actin polymerization and formation of lamellipodia (Fig. 4c, left panel). Grazoprevir treatment caused cell contraction and stress fibers (visualized by bright actin staining along the cell periphery), which is consistent with Rho activation (Fig. 4c, right panel).

We then used time-lapse, widefield microscopy to track the morphological changes of hundreds of cells co-expressing DNCR2-TIAM and GNCR1-LARG (dual effector) under the drug treatment conditions shown in Fig. 4b, which were compared to responses in cells expressing a single effector, DNCR2-TIAM or GNCR1-LARG (Fig. 4d–g, Supplementary Fig. 12). Changes to cell size measured by total cell area (Fig. 4d,e) and perimeter (Supplementary Fig. 12a,b) show that Rac activity induced by danoprevir treatment caused modest increases in cell size in cell lines with DNCR2-TIAM, while grazoprevir-mediated Rho activity caused cells to contract in cells with GNCR1-LARG. We also tracked changes to cell shape measured by solidity (defined as the cell's area divided by the area of an ellipse drawn around the cell, such that a cell with more membrane protrusions will have lower

solidity) and circularity (0–1, where 1 is a perfect circle, and 0 a line) (Fig. 4f,g; Supplementary Fig. 12c,d). Rho activity promoted by grazoprevir treatment in cells with GNCR1-LARG caused cells to increase in solidity and circularity, indicative of contraction of protrusions and cell length.

A notable difference between dual-effector and single-effector lines was that titration of a single effector resulted in only a few gradations in morphology change, which plateaued at a maximum with at most one intermediate level (blue and teal mean values in 4e,g, Supplemental Figure 12b,d). In contrast, dual-effector cells showed more graded population-level changes in cell morphology as a result of different ratios of Rac and Rho activity. However, concurrent variation in Rho and Rac activity does not result in an average phenotype, but one closer to the Rho-only (GNCR1-LARG line) condition, indicating that for this pair of opposing signaling effectors, Rho has the dominant effect on cell morphology. Single-cell distributions of the changes in cell morphology statistics show that the majority of cells in each treatment population responded to GTPase activation (Supplementary Fig. 12 e–g). See the Supplementary Notes for more discussion of single-cell level responses. Coactivation of Rac and Rho also resulted in graded population-level solidity and circularity changes in NIH3T3 cells co-expressing DNCR2-TIAM and GNCR1-LARG (Supplementary Fig. 11, Supplementary Notes). However, the increase in area and perimeter from danoprevir treatment was lower compared to that observed in HeLa cells, likely due to general morphological differences between the more adherent and flat, fibroblast NIH3T3 cells and the more rounded, epithelial HeLa cell lines. Thus, PROCISiR's ability to proportionally co-activate multiple signaling pathways provides basic insight into competing cellular processes and the achievement of finer gradations in phenotypic output than is observable with the titration of single effectors.

Discussion

We report here that PROCISiR can control switching behavior and provide graded, proportional control over two cellular outputs at once. The control modalities in PROCISiR can be used to manipulate mammalian cellular processes and potentially engineer drug-regulated cell therapies. The main challenge in developing PROCISiR was the generation of two new CID systems, which we accomplished through the *de novo* design of protein readers that selectively recognize distinct inhibitor-bound states of NS3a. To address the challenge of generating proteins that selectively recognize protein-small molecule complexes, we applied Rosetta protein design to highly stable *de novo* scaffolds. Importantly, the affinity and specificity achieved did not require any special inputs of known binding motifs or directed interactions with inhibitors or NS3a, in contrast to many recent *de novo* protein-protein and protein-small molecule interface design applications.^{34–37} The specificity achieved by the design and engineering process relied only on the chemical differences between inhibitors, as there is little-to-no rearrangement of NS3a in different inhibitor complexes.²³ Thus, our design strategy for generating protein-based readers should be scalable to other clinically-approved NS3a inhibitors, allowing the development of additional orthogonal PROCISiR components.³⁸ Our computational approach for generating protein-based readers is complementary to a phage display-based selection strategy that was recently used to identify an antibody fragment that selectively binds to a complex of the anti-

apoptotic protein Bcl-x_L bound to the small molecule ABT-737.¹⁵ While both strategies are capable of generating high affinity readers of protein-small molecule complexes, the highly-stable designed helical repeat proteins used in our computational approach are better suited than antibody fragments for intracellular synthetic biology applications.²⁰

CID systems have been used in numerous applications and are the first post-translational control module to enter the clinic in the form of inducible caspase 9 suicide switches for cell therapies.^{39,40} All current CID systems use small molecules that either have off-target effects in mammalian systems or do not behave well *in vivo*.^{12–15,41} Thus, the ability to modulate PROCISiR with clinically-approved drugs that are orthogonal to mammalian systems makes it a promising replacement for single-input/single-output CID systems that are currently being explored for *ex vivo* and *in vivo* cellular therapies, including improving the safety of therapeutic CAR T-cells. One limitation of PROCISiR and any CID system is that the constitutive presence of a regulated protein activity in the cell may lead to higher background than desired for some applications. The availability of ANR to basally localize a protein activity fused to NS3a away from a site of action, or to act as an intramolecular regulatory switch, provides an appealing option for reducing background, but protease-based NS3a control systems offer the possibility of obtaining the complete absence of a protein activity in the cell prior to the addition of a small molecule input.^{10,11,27,42} However, the reduced background that can be obtained with protease-based systems comes at the expense of speed, as protein accumulation is required for activation. Furthermore, we have found that catalytically-dead NS3a is an equally effective component of PROCISiR as the active protease, which is advantageous for applications in which off-target proteolytic activity may be a concern.

The emergent properties arising from the multi-input/multi-output architecture of PROCISiR enables new types of synthetic cellular behaviors including graded and proportional control of two-or-more outputs in a temporally-regulated manner. Using single or combined drug dosing, we can tune our output response from the switch-like response of bimolecular binding, to a more graded, linear response. Binary, switch-like responses allow digital logic operations to be performed, while graded, analog responses are useful for the fine-tuning of gene expression in metabolic engineering applications or studies of small changes in gene expression or signaling outputs.^{4,43} Combination of multiple chemical inputs with our readers enabled multi-state control that could be used to create dynamic signaling or transcriptional programs centered around the NS3a receiver hub. We demonstrated combinatorial outputs in the form of proportional transcriptional and signaling control, a form of ratiometric computation that cells have been shown to perform in responding to competitive ligands such as galactose/glucose and BMP ligands.^{1,44} When paired with single-cell transcriptomics or high-throughput microscopy, the fine-scale manipulations achieved with PROCISiR are well-suited for addressing questions related to single-cell decision making. The ability to either orthogonally or proportionally control two outputs, as we demonstrate with Rac and Rho signaling, makes PROCISiR a versatile system for investigating and creating new regimes of signaling pathway crosstalk. We envision that the expanded post-translational control behaviors that can be achieved with PROCISiR will facilitate new investigations into basic cellular mechanisms and the engineering of unique cellular behaviors. Furthermore, PROCISiR is well suited for integration with engineered

synthetic systems that respond to intrinsic inputs like synNotch, a system that enables transcriptional control in response to recognition of extracellular ligands.⁴⁵ The combination of small molecule-based, extrinsic control offered by PROCISiR with intrinsic, cell-autonomous decision making is an attractive approach for enabling safer and more effective cell therapies. The robust, rapid, multi-input/multi-output control offered by PROCISiR makes it a useful addition to the growing set of tools for post-translational control of *in vitro* and *in vivo* mammalian synthetic biology.

ONLINE METHODS

Protein design

Target preparation—PDB IDs 3SUD and 3M5L were used as the target NS3a:grazoprevir and NS3a:danoprevir complexes, respectively, for docking and design. Inhibitor parameter files were generated with OpenBabel, starting from the SDF file downloaded from the 3SUD or 3M5L PDB entry. Example commands for grazoprevir (SUE):

```
~directorypath/openbabel/bin/babel SUE_ideal.sdf SUE_ideal.mol2
```

```
~directorypath/Rosetta/main/source/scripts/python/public/molfile_to_params.py -c -n SUE  
SUE_ideal.mol2
```

The full-atom inhibitor pdb thus generated was concatenated with the NS3a coordinates, from which all non-bonded atoms had been removed. Each complex was then prepared with one round of minimization.

Scaffold set and docking—The scaffold sets for the danoprevir reader design included: 31 repeat protein crystal structures (PDB IDs 4GMR, 4GPM, 4HB5, 4HQD, 4HXT, 4PQ8, 4PSJ, 4R58, 4R5C, 4R5D, 4R6F, 4R6G, 4R6J, 4RV1, 4RZP, 5CWB, 5CWC, 5CWD, 5CWF, 5CWG, 5CWH, 5CWI, 5CWJ, 5CWK, 5CWL, 5CWM, 5CWN, 5CWO, 5CWP, 5CWQ), and 49 LRR models (models available upon request). PatchDock was run with standard settings to dock scaffolds on the prepared structure, directing the docking at the inhibitor. The top 50 PatchDock poses were used as inputs for design for each scaffold.

For the grazoprevir reader design, three scaffold sets were used: the 31 repeat protein crystal structures, 761 designed helix repeat (DHR) protein models, 30 DHR junction models (models available upon request). The top 10–100 PatchDock poses were used as inputs for design for each scaffold. Chains in the outputted PatchDock pdbs were reordered as follows: chain B (scaffold), chain A (NS3a), chain X (inhibitor).

A second Rosetta-based docking protocol, rotamer interaction field docking, RIFdock, was also used to dock the same scaffold sets on NS3a:grazoprevir. RIFdock is a module in Rosetta that defines a rotamer interaction field (RIF) on a designated target surface, and then matches scaffold backbones to the sidechains of the RIF. The target surface of the NS3a:grazoprevir complex was selected visually in PyMOL (PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) to include residues: 50, 52–55, 68–74, 85, 88–93, 95, 121, 132–135, 144, 145, 148, 149, 167, 169–174, 177–183, 186, 192, where residue 192

was grazoprevir, which was made part of chain A. The rotamer interaction field (RIF) was generated with the following command (see Code section for rifgen_grazo.flags):

```
~directorypath/rif_v4.1/build/apps/rosetta/rif_dock_test @rifgen_grazo.flags
```

The scaffold sets were then docked to the rotamer field using RIFdock. An example command is shown below and the file rifdock_grazo.flags is included in the Code section).

```
~directorypath/release/rif_v4.1/build/apps/rosetta/rif_dock_test @rifdock_grazo.flags - scaffolds
```

```
~directorypath/Scaffolds/scaffold.pdb
```

This generated ~500 starting poses each from the 31 repeat protein crystal structures and 30 DHR junction models, and ~3000 starting poses from the 761 DHR models, all of which were used for design.

Design—The RosettaScripts design protocol cid_design.xml was run on each PatchDock starting pdb, generating one design per input pose. For the grazoprevir reader design a cutoff of solvent accessible surface area of $>1100 \text{ \AA}^2$ was imposed, based on the empirical observation that designs with sasa scores below this often had significant empty space between the scaffold and NS3a:grazoprevir. The ROSETTA scoring function talaris_2014 was used for the danoprevir READER designs, and beta_November2016 was used for the grazoprevir reader design. An example command line is given:

```
~directorypath/rosetta_scripts.default.linuxgccrelease -database
```

```
~directorypath/Rosetta/main/database -parser:protocol ~directorypath/cid_design.xml -s
```

```
~directorypath/3SUD_scaffold.pdb -extra_res_fa ~directorypath/SUE.fa.params - holes:dalphaball ~directorypath/DAlphaBall.gcc
```

We generated 27,840 designs for the danoprevir reader, from which we picked the designs to order as described below.

For the grazoprevir reader, this first round of design generated 2,280 structures from PatchDock-generated poses, and 3,200 structures from RIFDock-generated poses. We narrowed the PatchDock designs by visually examining designs that passed the score cutoffs of ddg (full atom attract) per 1000 sasa < -25 (i.e., Rosetta G score per 1000 \AA of solvent accessible surface area) and shape complementarity (sc) > 0.25 and hand-picked 72 structures to redesign. We hand-picked 30 structures from the RIFDock-generated designs that passed the score cutoffs of ddg < -17 , sc > 0.50 , interface buried sasa > 1700 . We ran cid_roll_design.xml on these structures with -nstruct 100 to generate 100 designs per input, resulting in about 10,000 more designs. The script cid_roll_design.xml is same as the original cid_design.xml, with the addition of a random rotation or translation of the scaffold relative to the NS3a:inhibitor complex performed prior to design in order to access slightly different binding modes.

Filtering—For the danoprevir reader, we filtered designs based on solvent accessible surface area, ligand interface energy, shape complementarity, ddg, packstat (a packing metric), and buried unsaturated H-bonds. A complete definition of all Rosetta filtering metrics is available at https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/Filters/Filters-RosettaScripts. 200 designs passing these filters were visually examined in PyMOL. We chose 31 designs to order: 10 with the interface on the concave side of the LRRs, 10 with the interface on the convex side of the LRRs, and 11 from other de novo scaffold folds. Additionally, for the designs that we ordered, we reverted all residues to native that did not improve the binding energy with the inhibitor.

For the grazoprevir reader, we filtered all ~15,000 designs output from both the PatchDock and RIFDock starting inputs. We evaluated designs that passed the cutoffs: ddg: < -17, interface shape complementarity > 0.50, interface buried sasa > 1700. RIFDock redesigns were required to pass the more stringent cutoffs: ddg: < -20, interface shape complementarity > 0.55. 152 designs passing these criteria were manually examined in the Rosetta graphical user interface Foldit standalone. The best redesigns (by ddg) were compared to the original design, and the one design chosen from each redesign set based on optimal scores and hydrogen bonding. For the ordered design G3, which resulted in the best grazoprevir reader, a favored mutation from the redesigns was inserted into the original design (A112M).

The 152 best grazoprevir reader designs were also evaluated for their potential to clash with danoprevir, as specificity for grazoprevir was a design goal. The NS3a:danoprevir structure (3M5L) was aligned to 3SUD in each design and visually examined for clashes.

Based on the visual examinations, 29 grazoprevir reader designs were ordered with the following provenance: 6 from the DHR crystal structure set (2 PatchDock, 4 RIFDock), 16 from DHR model set (5 PatchDock, 11 RIFDock), and 7 from the DHR junction model set (1 PatchDock, 6 RIFDock). The best design that showed modest binding, G3, was generated using RIFDock on the DHR18 crystal structure.

Constructs

Note that there were three variants of the NS3a protein sequence used in this study. A solubility optimized NS3a/4a (either catalytically active or catalytically dead, S139A) derived from HCV genotype 1a was used for the majority of the work with the designed readers.⁴⁶ Genotype 1a NS3a/4a does not interact with the peptide ANR, which was selected to interact with genotype 1b NS3a; therefore, we engineered a hybrid NS3a/4a, NS3aH1, which is the solubility optimized NS3a/4a with four mutations needed for interaction with ANR: A7S, E13L, I35V, and T42S.²⁷ NS3aH1 (catalytically active) was used for the majority of the microscopy colocalization and transcription-control constructs. NS3a/4a solubility optimized S139A was used for membrane signaling constructs with DNCR2 and GNCR1. The NS3a/4a fusion is referred to as NS3a throughout the paper. The NS3a variant used is described for each experiment below and in Supplementary Table 5.

Bacterial expression constructs: Biotinylated proteins were expressed from the pDW363 vector, which encodes a bi-cistronic BirA biotin ligase.⁴⁷ Proteins were N-terminally tagged

with the biotin acceptor peptide, followed by a His₆ tag. Constructs were cloned into pDW363 via PCR-linearization of the vector, followed by Gibson assembly with the gene insert. Untagged proteins were expressed from the pCDB24 vector (gift of Christopher Bahl, Baker lab), which encodes proteins with an N-terminal His₁₀-Smt3 tag, which is scarlessly removed by ULP1. Linear gene inserts with overhangs and a stop codon added were inserted via Gibson assembly into pCDB24 that had been linearized with XhoI (New England Biolabs).

Yeast surface expression constructs: Danoprevir:NS3a reader designs were synthesized as linear genes by Gen9. All yeast constructs were cloned by homologous recombination in yeast with linearized pETCON vector (NdeI/XhoI-cut, New England BioLabs). pETCON encodes Aga-2, the inserted gene, and a C-terminal c-myc tag for expression detection.⁴⁸ Grazoprevir:NS3a reader designs were synthesized and constructed in complete pETCON plasmids by Genscript.

Mammalian expression constructs: All constructs were made in pcDNA5/FRT/TO (Thermo Fisher Scientific) unless otherwise noted. pcDNA5/FRT/TO was either linearized via PCR, or cut by BamHI and EcoRV, and inserts and vector were assembled by Gibson assembly. Dual expression constructs of DNCR2-VPR/KRAB and NS3aH1-dCas9 were made in PiggyBac vectors (pSLQ2818 pPB: CAG-PYL1-KRAB-IRES-Puro-WPRE-SV40PA-PGK-ABI-tagBFP-SpdCas9 and pSLQ2817 pPB: CAG-PYL1-VPR-IRES-Puro-WPRE-SV40PA-PGK-ABI-tagBFP-SpdCas9, gifts from Stanley Qi (Addgene plasmids 84241 and 84239)). The PiggyBac vectors were linearized by restriction enzyme digest, and PCR amplified inserts and digested vector were assembled by Gibson assembly. pcDNA5/FRT/TO-MCP-NS3a-P2a-DNCR2-KRAB-MeCP2-P2a-GNCR1-VPR-IRES-BFP was assembled with fragments PCR amplified from the following sources: MCP from pJZC34 (see below), KRAB-MeCP2 was a gift from Alejandro Chavez & George Church (Addgene 110821), VPR from one of the above-mentioned pPB vectors, and DNCR2, GNCR1, and NS3a (solubility optimized S139A) from gBlocks.

Single-guide RNAs (CXCR4, CD95, TRE3G) were cloned into the gRNA Cloning Vector, a gift from George Church (Addgene plasmid 41824). DNA corresponding to the guide target was ordered as a single stranded oligo with overlap to the vector and assembled with AflIII-digested gRNA vector by Gibson Assembly. Scaffold RNAs (targeting CXCR4, CD95, or TRE3G with com, PP7, or MS2, respectively) were cloned into dual insert vectors derived from pSico, expressing the scaffold RNA under a U6 promoter and the protein inserts under a CMV promoter: pJZC33 or 34 (MS2/MCP), pJZC43 (PP7/PCP), pJZC48 (com/com), gifts from Jesse Zalatan.³² All RNA-binding protein-reader fusions were expressed with P2a-tagBFP in place of the IRES-mCherry in the original vectors. This vector was also the basis of the scRNA-only vectors, which were used when all readers/RBPs were expressed separately. These vectors expressed only a tagBFP downstream of the CMV, and the guide plus 2x MS2 (wt + f6 sequences) under the U6 promoter.

pcDNA5/FRT/TO-Lifeact-mCherry was created from mCherry-Lifeact-7, a gift from Michael Davidson (Addgene plasmid 54491). pEF5-FRT-mCherry-NS3a-CAAX-IRES-EGFP-DNCR2-P2a-BFP-GNCR1 was created by assembling readers and fluorescent

proteins from other constructs in a pEF5-FRT backbone obtained by digestion of Addgene plasmid 61684, a gift from Maxence Nachury. pPB-NS3a-CAAX-IRES-EGFP-DNCR2-TIAM-P2a-BFP-GNCR1-LARG and the pEF5 version of the same construct, as well as NS3a-CAAX-IRES-EGFP-DNCR2-TIAM-P2a-BFP-GNCR1 and NS3a-CAAX-IRES-EGFP-DNCR2-P2a-BFP-GNCR1-LARG and were assembled with NS3a, reader, and fluorescent protein fragments from the previously mentioned construct, with addition of signaling effector domains from the following sources: human TIAM DH-domain residues 1033–1240 from a Maly lab source, and LARG DH-domain was a gift from Michael Glotzer (Addgene plasmid 80408). The PiggyBac vector used for these two constructs was linearized by digesting the multiple cloning site of PB501B (Systems Biosciences) with EcoRI and BamHI (New England BioLabs).

pLenti-UAS-minCMV-mCherry/CMV-Gal4DBD-NS3a-P2a-DNCR2-VPR was based on a pLenti-UAS-minCMV-mCherry/CMV-Gal4DBD-ERT2VP16 vector, a gift from Kenneth Matreyek, (from which the Gal4-UAS-minCMV was from Addgene plasmid 79130, a gift from Wendell Lim) which was digested with BamHI-HF and SexA1 to insert the NS3a-P2a-DNCR2-VPR fragment.

All cloning PCR reactions were performed with Q5 polymerase (New England BioLabs), and all Gibson assembly reactions were performed with NEBuilder HiFi Assembly Master Mix (New England BioLabs). Oligonucleotides and gBlocks were synthesized by Integrated DNA Technologies. Presence of all components of each construct was verified by sequencing (Genewiz). Select mammalian expression vectors constructed in this study are available on Addgene, and bacterial or yeast expression vectors are available upon request. See Supplementary Table 5 for all sequences.

Inhibitor sources

Grazoprevir was purchased from MedChem Express (MK-5172, product number HY-15298). Asunaprevir (BMS-650032, product number A3195) and danoprevir (RG7227, product number A4024) were purchased from ApexBio.

Protein expression and purification

Proteins were expressed in BL21 (DE3) *E. coli* at 37 °C to an O.D.₆₀₀ of 0.5–1.0, then moved to 18 °C and induced to 0.5 mM IPTG overnight. For biotinylated constructs, 12.5 mg D(+)-biotin/L culture was added upon inoculation with overnight culture. After 16–20 hours of overnight growth, cultures were harvested, and cell pellets frozen at –80 °C. Cell pellets were resuspended in 20 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole, 1 mM DTT, 0.1% v/v Tween-20. All buffers for NS3a purifications additionally included 10% v/v glycerol. Cells were lysed by sonication, and supernatant was incubated with NiNTA resin (Qiagen) for at least 1 h at 4 °C. Resin was washed with 20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, and proteins were eluted with 20 mM Tris pH 8.0, 500 mM NaCl, 300 mM imidazole. Biotinylated constructs were then further purified by size exclusion chromatography on a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, 10% v/v glycerol. Proteins were stored in this buffer at –80 °C. For proteins tagged with His₁₀-Smt3, the tag was removed by overnight cleavage at

room temperature using His-tagged ULP1 protease (purified in-house) at a ratio of 1 mg ULP1: 250 mg protein. Cleavage was performed concurrent with dialysis (3.5 kDa mwco Slide-A-Lyzer dialysis cassettes, Thermo Scientific) in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT, 10% v/v glycerol. Cleaved protein was then put through a second NiNTA purification, with the desired protein collected in the flowthrough and wash (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 10% v/v glycerol). NS3a S139A and DNCR2 for crystallization were further purified via ion exchange chromatography on a HiTrap SP column (GE Healthcare) and HiTrap Q column (GE Healthcare), respectively, followed by size exclusion chromatography on a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM Tris pH 8.0, 100 mM NaCl, 2 mM DTT. 60 μ M NS3a and 100 μ M DNCR2 were mixed with 500 μ M danoprevir and incubated at 4 °C overnight. The NS3a S139A/DNCR2/danoprevir complex was further purified by size exclusion chromatography on a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM Tris pH 8.0, 50 mM NaCl, 2 mM DTT. The protein complex peak fractions were pooled and subsequently concentrated to 7 mg/mL for crystallization.

Crystallization of the DNCR2, NS3a and danoprevir

Crystals were obtained using the hanging drop method by adding 1 μ l of the above NS3a:DNCR2:danoprevir complex to 1 μ l of a well solution containing 100 mM Bis-Tris, pH 6.5, 200 mM LiSO₄ and 22% w/v PEG 3350. Crystals formed in 24–36 h at room temperature. Crystals were flash-frozen with liquid nitrogen in a cryoprotectant with 20% v/v glycerol.

X-ray data collection and structure determination

Data collection was performed at the ALS beamlines 8.2.1 and 8.2.2. The diffraction data was processed by the HKL2000 package in the space group $P2_1$.⁴⁹ The structure was determined, at 2.3 Å resolution, using one data set collected at a wavelength of 1.00 Å, which was also used for refinement (Supplementary Table 2). The initial phases were determined by molecular replacement with the program Phaser, using the crystal structure of NS3a (PDB code: 3M5L) as the initial search model.⁵⁰ Two NS3/4a were found in one asymmetric unit, and the experimental electron density map clearly showed the presence of two molecules of DNCR2 with two molecules of danoprevir in one asymmetric unit. The complex model was improved using iterative cycles of manual rebuilding with the program COOT and refinement with Refmac5 of the CCP4 program suite.^{51,52} There were no Ramachandran outliers (98.3% most favored, 1.7% allowed).

Analytical size exclusion chromatography

5 nmoles of each protein or drug were mixed in 300 μ L total volume (16.7 μ M final concentration), in a buffer of 20 mM Tris pH 8.0, 300 mM NaCl, 10% glycerol, 1 mM DTT. Complexes were incubated on ice for 1 h before injection of 250 μ L into a 500 μ L loop and onto a Superdex-75 10/300 GL column (GE Healthcare) at 4 °C. Untagged NS3a S139A (solubility optimized) and untagged DNCR2 were used for SEC.

Combinatorial library design

Library design to improve the affinity of the original designs proceeded through three stages: 1) Redesign of the D5 or G3 interface using RosettaDesign, 2) selection of positions to vary in the library, and 3) optimization of degenerate codon choices to encode the library using a previously described integer linear programming approach.^{53–55}

Redesign of the interfaces was done using the RosettaScript `cid_roll_design.xml` (Code section). ~1000 redesigns were generated for D5/G3. Unique sequences from designs that had a ROSETTA ddg score below that of the original design (700–800 sequences) were used to assemble a position specific scoring matrix (PSSM).

To select positions to vary in the library, this PSSM was visually examined with reference to the original design and the redesign models. Positions with significant changes in the redesigns that were proximal to the interface were chosen to vary in the library. Additionally, to enable construction of each library from two oligonucleotides, the positions varied were constrained to two helices (helices 5 and 7 for D5, and helices 2 and 4 for G3).

The library design scripts require two inputs: a short list of residues required to be varied in the library, and a longer list of preferred residues and/or a PSSM.⁵⁵ Required residue lists generally included the original residue from the design, with a further hand-selected set of residues highly preferred in the redesigns. Preferred residue lists included all amino acids occurring in the redesigns. The D5 library was designed by optimizing degenerate codon choice to encode as many preferred residues as possible within a DNA library size constraint of 10^7 . The resulting library encoded 4.1×10^6 protein variants and was encoded by the following degenerate codons (original D5 residue/degenerate codon/amino acids encoded, where Z=stop codon): T122/WAY/ILST, V131/GYC/AV, N132/VWM/DEHIKLNQV, N173/VAW/DEHKNQ, I174/MDC/HILNRS, W177/WKG/LMRW, L181/HKT/CFILRS, Q184/DRS/CDEGKNRSWYZ, V187/KMW/ADESYZ). The G3 library was designed by optimizing the sum of the PSSM scores from the redesigns within a DNA library size constraint of 10^7 . The resulting G3 library encoded 7.1×10^6 protein variants and was encoded by the following degenerate codons: D38/KWC/DFVY, N41/WDK/CFIKLMNRSWYZ, E42/VWA/EIKLQV, E44/DDK/CDEFGIKLMNRSVWYZ, A49/DYG/ALMSTV, S52/MKT/ILRS, R105/MKT/ILRS, I108/WTK/FILM, M112/RYG/AMTV.

DNCR1 combinatorial library design used the same library optimization approach as above, but used experimentally determined mutational preferences as the input, rather than design-determined preferences. The enrichment values from the DNCR1 SSM library (see below) were standardized (Z-value) for each positive sort (performed at 50 nM or 500 nM NS3a). The Z-values for the two sorts were then averaged. These average, standardized enrichment values were used as a PSSM input to the library design script. Positions to vary were hand-chosen based on their proximity to the designed interface (based on the original D5 model), as well as the presence of multiple enriched mutations in the SSM results. The mutations that were required to be included in the library design were also hand-picked from the most enriched mutations (top 10% of enrichment values), while the inclusion of additional mutations was optimized by maximizing the sum of the enrichment scores. Some large codon choices were removed to enforce a modest number of mutations at each position.

Additionally, chemical diversity classes were defined to prioritize inclusion of certain classes of residues. The library DNA size was constrained to be $<10^8$ variants, and final size in protein sequences was 2.76×10^7 . The final DNCR1 combinatorial library design was as follows: L121/YWT/FHLY, I125/HWS/FHIKLMNQYZ, E128/RWA/EIKV, V131/RKA/GIRV, H132/HWS/FHIKLMNQYZ, E171/VWK/DEHIKLMNQV, K178/ADG/KMR, V180/RKK/GIMRSV, E185/VWG/EKLMQV, S189/MRC/HNRS.

Yeast display library construction

Combinatorial libraries were assembled from two ultramer oligonucleotides (Integrated DNA Technologies), which contained a short, overlapping region corresponding to part of the constant helix between the two varied helices (helix 6 for the D5 libraries, and helix 3 for the G3 library). Linear, double stranded fragments were generated in the first PCR by pairing each varied primer with a constant primer that annealed 5' or 3' to the end of the full gene. These fragments were excised and extracted from an agarose gel. A second round of PCR was performed to overlap these fragments, with further amplification by addition of the outside primers in the 10th cycle (out of 35). The correct-sized product was gel extracted and used as the template for 1–2 more rounds of PCR with the outside primers to yield sufficient DNA. The DNCR1 SSM library was assembled using a pair of primers (Integrated DNA Technologies) for each of the 75 protein positions varied, where the forward primer contained the NNK site in a central position, and the reverse primer overlapped with the 5' end of the forward primer.⁵⁶ Linear fragments corresponding to each primer pair were overlapped in a second round of PCR to yield the full gene insert. Combinatorial library PCRs were performed with Q5 polymerase (New England BioLabs), and the SSM library PCRs were performed with Phusion polymerase (Thermo Fisher Scientific). For all libraries, the linear library DNA was combined with NdeI- and XhoI-digested pETCON at a ratio of 4 μ g insert:1 μ g vector and electroporated into freshly-prepared electrocompetent EBY100 *S. cerevisiae*.

Yeast surface display analysis and sorting

Yeast were grown overnight at 30 °C in yeast minimal media (-ura for strain selection, -trp for pETCON selection) supplemented with 2% w/v glucose. Overnights were used to inoculate SGCAA cultures (2% w/v galactose, 0.67% w/v yeast nitrogen base, 0.5% w/v casamino acids, and 0.1 M sodium phosphate, pH 6.6) to an O.D. 600 of 1.0–2.0 and protein expression was induced overnight at 30 °C. Before sorting or analysis, cells were pelleted and resuspended in PBS supplemented with 0.5% w/v bovine serum albumin (PBSA). Protein solutions of biotinylated NS3a with danoprevir or grazoprevir were made in PBSA and incubated with the yeast for 30 min–1 h at 22 °C. For analysis and sorting of initial, low-affinity designs, NS3a was pre-tetramerized by incubation with streptavidin-phycoerythrin (SAPE, Invitrogen) at a molar ratio of 1 SAPE:4 NS3a for at least 10 minutes prior to incubation with yeast; these sorts are denoted as “avid” below. Cells were washed in cold PBSA and incubated for 15 min on ice with SAPE and fluorescein isothiocyanate-conjugated chicken anti-c-myc (Immunology Consultants Laboratory), both diluted 1:100 in PBSA. After the labeling incubation, cells were washed again in cold PBSA and analyzed on a C6 flow cytometer (Accuri) or a FACSCanto cytometer (BD Biosciences), or sorted on a SH800 (Sony Biotechnology) cell sorter or a FACSAria III (BD Biosciences) cell sorter.

All FACS data were analyzed using FlowJo (v.10.1). See Supplementary Fig. 13 for yeast gating strategies. Sorted yeast recovered for 1–2 days at 30 °C in yeast minimal media plus 2% w/v glucose.

Titration curves for NS3a:drug complexes on yeast-displayed designs used construct NS3a_3 (solubility-optimized, catalytically active). Drug concentrations were at a fixed molar ratio of 10 drug:1 NS3a, with the exception of the DNCR2-danoprevir titration, for which a fixed concentration of 50 nM danoprevir was used for all points to stay above the NS3a:danoprevir K_i . Curves were fit using Graphpad Prism 5 to a one-site specific binding model with Hill coefficient. See Supplementary Table 1 for K_D values and Supplementary Fig. 14 for binding curves.

For the first D5 library, the following sequential sorts were performed using catalytically active NS3a (NS3a_3): 1 μ M NS3a/10 μ M danoprevir, 0.5 μ M NS3a avid/5 μ M danoprevir, 0.5 μ M NS3a avid/5 μ M danoprevir, 0.25 μ M NS3a avid/2.5 μ M danoprevir, 2 μ M NS3a/20 μ M danoprevir, 20 nM NS3a/200 nM danoprevir. The highest 1–3% PE/FITC-positive events were collected for each sort, with the gate set along the binding/expression diagonal. For the DNCR1 combinatorial library, the following sequential sorts were performed using catalytically inactive NS3a (NS3a_2): 100 nM NS3a/1 μ M danoprevir, 100 nM NS3a/1 μ M danoprevir, 50 nM NS3a/500 nM danoprevir, 5 nM NS3a/50 nM danoprevir, 500 pM NS3a/50 nM danoprevir, 20 pM NS3a/50 nM danoprevir. The top 0.5–9% were collected in each sort. For the G3 library, the following sequential sorts were performed using catalytically inactive NS3a (NS3a_2): 500 nM NS3a avid/5 μ M grazoprevir, 50 nM NS3a avid/500 nM grazoprevir, 500 nM NS3a/5 μ M grazoprevir, 500 nM NS3a/5 μ M grazoprevir, 250 nM NS3a/2.5 μ M grazoprevir, 100 nM NS3a/1 μ M grazoprevir, 30 nM NS3a/300 nM grazoprevir.

The most-enriched clones were assessed by colony PCR and sequencing (Genewiz) of ~50 colonies from the final 2–3 pools of each library. Titrations of NS3a:drug were performed on several of the most enriched clones to verify that the most-enriched clones (DNCR1 and GNCR1) exhibited the tightest binding. DNCR2 was selected from multiple indistinguishably-high affinity clones based on its superior expression on yeast.

For the DNCR1 site saturation mutagenesis (SSM) library, two sorts were performed on the same day at 50 nM NS3a (NS3a_2)/500 nM danoprevir and 500 nM NS3a (NS3a_2)/5 μ M danoprevir. For both conditions, a positive-sort gate was set to collect the top 1% of binders, and a negative-sort gate was set to collect the bottom 6% of binders. All gates were set along the binding/expression diagonal (Supplementary Fig. 13). The naïve population for sequencing analysis was saved from the same day of growth.

DNCR1 SSM library sequencing

At least 20 million cells were harvested for each selected library pool and the naïve library, and DNA was extracted and prepared for Illumina sequencing as in Berger et al.³⁵ The first round of qPCR, to amplify the 150 bp varied region, was performed for 25–35 cycles using Phusion polymerase. After gel extraction, a second round of PCR was performed to add on barcodes and Illumina adaptors. Sequencing was performed with a 600-cycle reagent kit

(Illumina) on a MiSeq sequencer (Illumina). Enrich was used to align and filter the paired-end reads.⁵⁷ An average quality for each read was required to be greater than 20, no N's were allowed, and the maximum number of nucleotide mutations allowed per sequence was 3. The sequence counts output by Enrich were processed by an in-house Python script to calculate the enrichment value (enrichment ratio for each mutant, normalized by the wild-type enrichment ratio): $\log_2 (F_{v,sel}/F_{v,inp})/(F_{wt,sel}/F_{wt,inp})$, where F_v is the frequency of the variant in the selected or input (naïve library) pool, and F_{wt} is the frequency of the wild-type residue. Only single mutants that had at least 15 counts in the naïve library were included in the analysis.

Mammalian cell culture

All cells were cultured in high-glucose DMEM, 4 mM L-glutamine, 10% fetal bovine serum (FBS, Life Technologies) at 37 °C, 5% CO₂. Cells were tested and found free of mycoplasma at least every 6 months.

Confocal microscopy for fluorescent protein colocalization analysis

A Leica SP8X system was used for confocal microscopy. A UV laser at 405 nm was used to excite tagBFP. White light lasers of 488 and 587 nm were used for EGFP and mCherry, respectively. TagBFP emission was recorded on a PMT detector, and EGFP and mCherry were detected by separate HyD detectors. All images were taken using a 63x objective with oil, at 512×512 resolution.

All fluorescent protein colocalization experiments were performed in NIH3T3 cells (Flp-In-3T3, Thermo Fisher Scientific), with the exception of the DNCR2/GNCR1/NS3a colocalization in HeLa cells in Fig. 4b and Supplementary Fig. 12a,b (described below). For fixed-cell experiments, cells were plated at 3×10^4 cells/mL on sterile glass coverslips placed in 12-well culture plates. Cells were transfected 24 hours after plating with Lipofectamine 2000 or 3000 (Thermo Fisher Scientific) at a ratio of 3 μ L reagent: 1 μ g DNA, according to manufacturer's instructions. 3-vector transfections were performed with 0.3 μ g NS3a and 0.35 μ g each ANR/DNCR2/GNBP vectors, while 2-vector transfections were performed with 0.3 μ g free component and 0.7 μ g of the immobilized component. One day after transfection, cells were treated with drug or DMSO and fixed. Drug additions were performed by exchanging the media for DMEM + 10% v/v FBS plus drug. To fix, cells were washed once with DPBS (Thermo Fisher Scientific), then incubated with 4% v/v paraformaldehyde in DPBS for 15 minutes. After washing twice with DPBS, coverslips with cells were removed from the plate and mounted on glass slides using Fluoromount-G (SouthernBiotech).

For the live cell experiment assaying DNCR2 membrane association time, cells were plated at 3×10^4 cells/mL in 35 mm glass-bottomed dishes (Matek), that were coated with poly-D-lysine. Experiments were performed in FluorBrite DMEM (Thermo Fisher Scientific) media supplemented with GlutaMax (Thermo Fisher Scientific) and 10% v/v FBS. Cells were imaged with dishes open on a heated stage (~55 °C, which resulted in the media at the center of the plate remaining at ~30 °C). 5 μ M drug additions were performed by removing 1 mL media from the dish, mixing with drug, and returning to the dish after 2 minutes of imaging.

All cells were imaged within 30 minutes of removal from incubator, and no environmental controls were used beyond heating. The constructs used for live cell membrane localization kinetics were myristoyl-tag-mCherry-NS3a and DNCR2-EGFP.

Colocalization of NS3a and DNCR1 at the plasma membrane, nucleus, mitochondria and Golgi was performed with two sets of constructs, with either NS3a or DNCR1 as the immobilized component. mCherry-NS3a was used with Tom20-DNCR1-EGFP, DNCR1-EGFP-Giantin, and 3xNLS-DNCR1-EGFP. DNCR1-EGFP was used with Tom20-mCherry-NS3a, mCherry-NS3a-Giantin, 3xNLS-mCherry-NS3a, and myristoyl-tag-mCherry-NS3a. Drug specificity of DNCR1 was analyzed with mCherry-NS3a and Tom20-DNCR1-EGFP or DNCR1-EGFP-Giantin, and drug specificity of DNCR2 and NS3a with DNCR2-EGFP and Tom20-mCherry-NS3a. Colocalization was analyzed after 1 h of 10 μ M drug or equal volume DMSO treatment.

Colocalization of NS3a, ANR, and DNCR2 was performed with NS3aH1-mCherry in combination with 2 separate vectors encoding 3xNLS-DNCR2-EGFP and ANR-ANR-BFP-CAAX (0.3 μ g, 0.35 μ g, 0.35 μ g, respectively) or one vector encoding Tom20-BFP-ANR-ANR-P2a-DNCR2-EGFP-CAAX (0.3 μ g NS3a, 0.75 μ g ANR/DNCR2). Colocalization of NS3a, DNCR2 and GPCR1 was performed with NS3aH1-mCherry, Tom20-DNCR2-EGFP, and GPCR1-BFP-CAAX (2-location; 0.3 μ g, 0.35 μ g, 0.35 μ g, respectively), or with DNCR2-EGFP, GPCR1-BFP, and NS3aH1-mCherry-CAAX (1-location; 0.25 μ g, 0.25 μ g, 0.5 μ g, respectively). For all 3-color experiments, 15-minute drug treatments with 5 μ M danoprevir or grazoprevir or equal volume DMSO were performed prior to fixing.

For the colocalization experiment shown in Supplementary Fig. 12c, a single pEF5 vector expressing mCherry-NS3a(S139A)-CAAX-IRES-EGFP-DNCR2-P2a-BFP-GPCR1 was transiently transfected into NIH3T3 cells as described above. The same construct was used for the equivalent experiment in Fig. 4b, which was performed in transiently transfected TReX-HeLa cells (ThermoFisher Scientific). HeLa cells were transiently transfected using HeLaFect (OZ Biosciences) according to the manufacturer's instructions. Cells were treated with combinations of danoprevir and grazoprevir or equal volume DMSO for 1 hour before fixing.

All images were analyzed using ImageJ. Pearson's r values reported are Rcolocalization values generated using an automatic thresholding program (Colocalization Threshold plugin).⁵⁸ For DNCR2 membrane associate kinetics analysis, a square ROI was set to include only cytoplasm. EGFP fluorescence was quantified in the ROI over the timecourse. 15 min timecourses (2 min pre-drug addition, 13 min post-drug) were collected for 18 cells from 4 independent plates. The cytoplasmic fluorescence was normalized to the value in the first and last frame for each cell. Because the cells were imaged at different time points (every ~20–30 seconds), we used an in-house Python script to fit a 1-D interpolation to each timecourse and plotted the average and standard deviation value of the 1-D functions at 20 second intervals. Time points after drug addition were fit to an exponential decay model to calculate a $t_{1/2}$ using Graphpad Prism 5 ($y=(y_0 - b)*e^{-kx} + b$, where b was constrained to 0, but y_0 was left unconstrained to account for minor variability in drug addition and mixing times).

Widefield microscopy for signaling phenotype analysis

Widefield images for cell morphology experiments were collected in an environmental chamber with humidity control, 37 °C, and 5% CO₂ on a Leica DMI8 automated fluorescence microscope with a 20x objective with 2x pixel binning (Fig. 4c (left panel, danoprevir cell) Supplementary Fig. 11d images were collected with a 40x objective). Cells were plated on glass-bottomed 96-well plates (Cellvis). Plates were treated with 10 µg/mL bovine fibronectin (Sigma Aldrich) for 1 hour and washed once with PBS.

The cell line used was TREx-HeLa (ThermoFisher Scientific), into which Lifeact-mCherry was stably integrated into the doxycycline-regulated Flp-In site by co-transfection of the pCDNA5-FRT/TO-Lifeact-mCherry vector with the Flp recombinase plasmid pOG44 (ThermoFisher Scientific) according to manufacturer's protocols. The PB-NS3a-CAAX-IRES-EGFP-DNCR2-TIAM-P2a-BFP-GNCR1-LARG and equivalent TIAM-only or LARG-only constructs were incorporated stably into the TREx-Lifeact-HeLa line using the Super Piggy Bac transposase (Systems Biosciences). After ~5 days of selection in 3 µg/mL puromycin, the top 1% of GFP-expressing cells were collected using a BD FACSARIA III (BD Biosciences). These cells were grown out to serve as the stable lines for the TIAM/LARG timecourse experiments in HeLas. One day prior to these experiments, 3000 cells were plated in each well of a fibronectin-coated 96-well plate in DMEM, 10% v/v FBS, supplemented with 1 µg/mL doxycycline. The following day, media was aspirated, cells were washed once with PBS, and cells were serum starved for 3–8 hours before imaging in 100 µL FluorBrite DMEM (Thermo Fisher Scientific) media supplemented with GlutaMax (Thermo Fisher Scientific) ("imaging media") plus 1 µg/mL doxycycline. Images collected for the mCherry (Lifeact) and EGFP (DNCR2-TIAM) channels. Cells were imaged for 10 minutes prior to drug addition, and drug was added by pipetting 100 µL 2x drug in prewarmed imaging media, after which cells were imaged for a further 60 minutes, at 2-minute intervals.

For TIAM/LARG experiments in NIH3T3 cells, pCDNA5-FRT/TO-Lifeact-mCherry and pEF5-NS3a-CAAX-IRES-EGFP-DNCR2-TIAM-P2a-BFP-GNCR1-LARG were transiently transfected in at a ratio of 9 pEF5:1 pCDNA5. The next day, cells were sorted for GFP and mCherry positive cells using a BD FACSARIA III (BD Biosciences), with 4000 cells collected in each well of a fibronectin-coated 96-well plate used for imaging. Widefield microscopy experiments were performed the next day, as described above for HeLa cells.

Analysis of widefield TIAM/LARG experiments for cell morphology parameters

Segmentation and computation of cell morphology parameters was performed in ImageJ using the ADAPT plugin.⁵⁹ Segmentation of the cell boundary was performed using the mCherry (Lifeact) signal as the reference channel, and the GFP signal distribution (EGFP-DNCR2-TIAM) was output for later use in filtering for expression-positive cells. ADAPT running parameters for the HeLa experiments were: gray level threshold 0.90, frames per minute 0.5, minimum trajectory length 34, and minimum object size 500. ADAPT running parameters for the NIH3T3 experiments were: gray level threshold 0.95, frames per minute 0.5, minimum trajectory length 27, and minimum object size 50. Area, perimeter, solidity and circularity and roundness were computed according to standard ImageJ morphology

parameter methods. The output from ADAPT was processed and filtered by in-house Python scripts. The mean GFP fluorescence of each cell was computed from the first frame fluorescence distribution output by ADAPT. For the HeLa experiments, cells were excluded from analysis if they fell below the 10th percentile of GFP fluorescence, if they fell above the 95th percentile of area or perimeter in any frame, or if they fell above the 95th percentile of solidity, circularity and roundness in the first frame. For the NIH3T3 experiments, cells were excluded from analysis if they fell below the 50th percentile of GFP fluorescence, if they fell above the 95th percentile of area or perimeter in any frame, if they fell below the 5th percentile of area or perimeter in any frame, or if they fell above the 95th percentile of solidity, circularity and roundness in the first frame. Percentiles were calculated separately for each different construct/cell line. These filtering cutoffs were intended to remove cells that did not express the effector construct substantially (GFP), were debris or mis-segmented (area and perimeter) or were dead or dividing at the start of the experiment (solidity/circularity/roundness). Additionally, cells were required to be present in all frames of the experiment.

AKT Western blots

COS-7 cells (ATCC), were plated in 24-well plates at 2×10^5 cells/mL (0.5 mL volume). One day later, cells were transfected using TurboFectin 8.0 (OriGene) according to the manufacturer's instructions with 0.75 μ g myristoyl-tag-mCherry-NS3a and 0.25 μ g DNCR1- or DNCR2-iSH2 vectors. One day after transfection, cells were washed once with DPBS, and media was replaced with serum-free DMEM. After serum-starving for 22 hours, cells were exposed to a 15-min drug treatment of 10 μ M danoprevir or DMSO. After drug treatment, cells were washed once in DPBS, then lysed in 50 μ L modified RIPA buffer (50 mM Tris-HCl, pH 7.8, 1% v/v IGEPAL CA-630, 150 mM NaCl, 1 mM EDTA, 1x Pierce Protease Inhibitor Tablet) for 30 minutes on ice. Cell debris was cleared by centrifugation at 17 kg for 10 min at 4 °C. Lysate was mixed with protein loading dye and denatured at 95 °C for 7 minutes then run on an SDS-PAGE gel (Criterion, Bio-Rad) and transferred to nitrocellulose. Blocking and primary antibody incubations were done in a 1:1 mix of TBS plus 0.1% v/v Tween-20 (TBST) and blocking buffer (Odyssey). Primary antibodies used were pSER473 AKT (1:2000, Cell Signaling Technologies #4060), pan-AKT (1:2000, Cell Signaling Technologies #2920), and GAPDH (1:2000, Cell Signaling Technologies #2118). Blots were washed with TBST, then incubated with secondary antibodies diluted 1:10,000 in TBST (goat anti-rabbit-IRDye 800 CW (926–32211) and goat anti-mouse-IRDye 680LT (926–68020), LI-COR), washed, and imaged on a LI-COR Odyssey scanner. See Supplementary Fig. 15 for full Western blots.

dCas9 transcription control

CXCR4 and CD95 induction experiments with DNCR2-VPR and NS3aH1-dCas9 were performed in HEK293T cells (293T/17, ATCC) following the protocol and using the same materials as detailed in Gao *et al.*¹⁶ Antibodies used were: APC anti-human CD184 (CXCR4) [12G5] (BioLegend 306510), PE anti-human CD95 (Fas) [DX2] (BioLegend 305607), PE Mouse IgG1, κ Isotype Ctrl [MOPC-21] (BioLegend 400111), APC Mouse IgG2b, κ Isotype Ctrl [MPC-11] (BioLegend 400322). No binding of isotype controls was observed to HEK293T cells; therefore, no background adjustments were made for isotype

binding. Briefly, cells were plated in 12-well plates at 6×10^4 cells/mL on day 1 and transfected with TurboFectin 8.0 (OriGene) according to the manufacturer's instructions on day 2. 1 μg total DNA was transfected per well (0.5 μg pB-DNCR2-VPR/NS3a-dCas9, 0.5 μg equal mix of 3 CD95 or CXCR4 guide RNA vectors (or unrelated guide for "No guide" controls)). 10 μM danoprevir was added on day 3, and cells were harvested on day 5 (VPR), incubated with antibodies for 1 hr, and analyzed on a FACSLSR II (BD Biosciences). For gene repression experiments with KRAB, cells were passaged on day 5, incubated with fresh drug, and analyzed on day 7. For all mammalian FACS experiments (unless otherwise noted), 10,000 single cell events were collected for each sample, and the median fluorescence signal of cells with BFP signals greater than that of untransfected cells were reported. All FACS data were analyzed using FlowJo (v.10.1). See Supplementary Fig.13 for mammalian cell gating strategies.

Danoprevir/grazoprevir titrations to linearize CXCR4 or CD95 expression were performed with DNCR2-VPR and NS3a-dCas9 following the protocol detailed above for gene induction with VPR, but in 24-well plates with 0.5 μg total DNA. Danoprevir was titrated in 12 concentrations in 2.5-fold dilutions starting from 1000 nM. Grazoprevir dilutions were added to the danoprevir titration, all starting from 10 nM grazoprevir, and decreasing across 12 concentration points in 2-, 1.5-, or 1.25-fold dilutions. The exact drug concentrations used are shown in Supplementary Table 5. Data were fit to four-parameter log dose-response curves (fitting EC_{50} , upper and lower baselines, and Hill coefficient) in Graphpad Prism 5.

Induction and reversion timecourses of CXCR4 expression that were analyzed by qPCR were performed in a similar manner, with 10 μM danoprevir replaced by 10 μM grazoprevir or equal DMSO after 24 hours of danoprevir treatment. Wells (in triplicate for each condition) were harvested at each time point by aspirating, washing with 1 mL DPBS, adding 300 μL Versene (ThermoFisher Scientific) and incubating for 5 minutes at 37 $^{\circ}\text{C}$, then pelleting at 3.5 krpm for 2 minutes at 4 $^{\circ}\text{C}$, aspirating, and freezing the pellets at -80°C .

GFP expression experiments were performed in a HEK293T cell line with GFP stably integrated in a single tetracycline-inducible landing pad (7xTRE3G operator with rTA) created in a similar manner as a previously published TetBxb1BFP-rTA HEK293T cell line (gift from Doug Fowler).⁶⁰ Combined CXCR4 and GFP induction was performed in this line transfected with 0.3 μg pCDNA5-FRT/TO-dCas9, 0.3 μg pCDNA5-FRT/TO-NS3aH1-VPR, 0.2 μg CXCR4-2xMS2/MCP-GNCR1-P2a-BFP (equal mix of 3 scRNAs), and 0.2 μg TRE3G-2xPP7/PCP-DNCR2-P2a-BFP. Drug treatment (48 hours) with 10 μM danoprevir or 10 μM grazoprevir or danoprevir/grazoprevir matrix, harvesting, CXCR4 antibody incubation and FACS analysis were performed as described above for immunofluorescence analysis.

The 3-gene experiment was performed in the GFP reporter HEK293T cell line transfected with 0.25 μg pCDNA5-FRT/TO-dCas9, 0.25 μg pCDNA5-FRT/TO NS3aH1-VPR, 0.166 μg TRE3G-2xMS2(wt+f6)/MCP-ANR-ANR-P2a-BFP, 0.166 μg CXCR4-com/com-GNCR1-P2a-BFP (equal mix of 3 scRNAs), and 0.166 μg CD95-2xPP7/PCP-DNCR2-P2a-BFP (equal mix of 3 scRNAs). Cells were plated in 12-well plates at 6×10^4 cells/mL on day 1

and transfected with TurboFectin 8.0 (OriGene) according to the manufacturer's instructions on day 2 and 1 μM or 10 μM drug was added on day 3. Cells were harvested on day 5 as described above for other samples to be analyzed to qPCR.

For RT-qPCR analysis, RNA was extracted with the Arum Total RNA Mini Kit (Bio-Rad). Integrity of the total RNA was confirmed by running on an agarose gel. Reverse transcription was performed on 1 μg total RNA using the iScript Reverse Transcription Kit (Bio-Rad), according to manufacturer's instructions. A no-RT control was performed on several samples per experiment to confirm that there was no significant genomic DNA contamination. qPCR was performed on 50 ng cDNA (1 μL of RT reaction) in a 10 μL reaction volume using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). For each biological sample, technical duplicates of the qPCR were performed and averaged. qPCR primers for GAPDH (reference gene), CXCR4, CD95, and GFP are listed in Supplementary Table 2. CXCR4 and GAPDH primers are from Zalatan et al., and CD95 and GFP primers were designed to amplify a 94 bp product using Primer3 (v. 0.4.0).^{32,61} A thermocycle of 95 °C for 2 min, (95 °C 10 sec, 58 °C 30 sec)x40 cycles, 65 °C–95°C at 0.5 °C increments 5 sec/step was performed on a Bio-Rad CFX Connect Real-Time System. For the CXCR4 reversibility experiment, fold-change in CXCR4 expression was calculated relative to a 0 hr timepoint using the $2^{-\text{CT}}$ method.⁶² For analysis of the 3-gene experiment, fold-change was calculated relative to untransfected TRE3G-GFP HEK293Ts.

The switchable gene expression/repression experiment on CXCR4 and CD95 was performed in TREx-HEK293 cell (ThermoFisher Scientific), into which Sp dCas9 was stably integrated using vector pCDNA5/FRT/TO-nFLAG-dCas9 and the Flp recombinase vector pOG44, according to manufacturer's protocols. This experiment followed our general dCas9 transcription experiment workflow described above. Briefly, cells were plated on day 1, transfected and induced with doxycycline on day 2, had 100 nM danoprevir or grazoprevir or equal volume DMSO added on day 3, and harvested for FACS analysis on day 5. All readers were transfected in via one plasmid, pCDNA5/FRT/TO-MCP-NS3a-P2a-DNCR2-KRAB-MeCP2-P2a-GNCR1-VPR-IRES-BFP. A mix of 3 guides each for CXCR4 or CD95 were transfected, or a gal4-4 control guide, all in a pU6-guide-2xMS2(wt+f6)/CMV-BFP vector. 0.5 μg reader and 0.5 μg guide plasmids were co-transfected in each well. Cells were incubated with antibodies and analyzed as described above, with 20,000 single-cell events collected per sample, and the median fluorescence plotted for cells with the top ~30% BFP expression signal.

Inducible Gal4 transcription factor

HEK293T/17 cells (ATCC) were plated at 7×10^4 cells/mL in 0.5 mL in 24-well plates. One day later, they were transfected with 0.35 μg pLenti-UAS-mCherry/CMV-Gal4DBD-NS3a-P2a-DNCR2-VPR and 0.15 μg of a BFP-expressing vector to use for gating on transfection-positive cells. The next day, a 12-point dilution series of danoprevir was added with 2.5-fold dilutions starting at 100 nM danoprevir. Two days later, cells were removed from the plate with Versene (Gibco) and analyzed for mCherry and BFP fluorescence on an FACSLSR11 (BD Biosciences). 20,000 single-cell events were collected, and median mCherry fluorescence was reported for the cells with the top ~50% of BFP signal for each sample.

Statistics

P values for experiments with multiple comparisons are from one-way ANOVA tests with significance for each comparison ($P < 0.05$) determined by Tukey's multiple comparisons test, and other P values are from unpaired, two-sided t-tests, computed using Graphpad Prism 5.

Reporting Summary

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

Data availability

The atomic coordinates and experimental data of the DNCR2/danoprevir/NS3a crystal structure have been deposited in the Protein Data Bank (ID 6N4N). Illumina data is deposited in the Sequence Read Archive under accession number SRP126450. Key DNA constructs are deposited in Addgene. All plasmid sequences are available in this Benchling project: <https://benchling.com/s/seq-OJeDzb2IUfv5dpTIXfro>.

Code availability

ROSETTA software can be downloaded from www.rosettacommons.org and is available free to academic users. Online documentation can be found at: http://www.rosettacommons.org/manuals/archive/rosetta3.5_user_guide/index.html. Instructions for RosettaScripts syntax is available at: http://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/RosettaScripts. A comprehensive list of command line options for ROSETTA can be found at: www.rosettacommons.org/docs/latest/full-options-list. All RosettaScripts written for this study are included in the Code section. Python scripts written for data processing are available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the Institute for Protein Design Protein Production Core for production of proteins. Thanks to D. Fowler and J. Zalatan and their labs for helpful discussions and materials. Thanks to G. Butterfield and J. Nelson for assistance with Illumina sequencing. Thanks to E. Dieter (UW, Seattle) and J. Rose (UW, Seattle) for construction of stable cell line resources. This work was supported by the NIH (R01GM086858 (D.J.M.) and S10 OD016240 (UW W.M. Keck Center)). D.B. acknowledges support from HHMI. Design calculations were facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. We are grateful to the staff at Advanced Light Source (ALS) beamlines BL 8.2.1 and 8.2.2 for assistance with synchrotron data collection. G.W.F. and Z.W. were funded by Washington Research Foundation Innovation Postdoctoral Fellowships.

References

1. Antebi YE et al. Combinatorial Signal Perception in the BMP Pathway. *Cell* 170, 1184–1196.e24 (2017). [PubMed: 28886385]

2. Giorgetti L et al. Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Mol Cell* 37, 418–428 (2010). [PubMed: 20159560]
3. Freed DM et al. EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. *Cell* 171, 683–695.e18 (2017). [PubMed: 28988771]
4. Daniel R, Rubens JR, Sarpeshkar R & Lu TK Synthetic analog computation in living cells. *Nature* 497, 619–623 (2013). [PubMed: 23676681]
5. Weinberg BH et al. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat Biotechnol* 35, 453–462 (2017). [PubMed: 28346402]
6. Stanton BZ, Chory EJ & Crabtree GR Chemically induced proximity in biology and medicine. *Science* 359, eaao5902 (2018). [PubMed: 29590011]
7. Banaszynski LA, Chen L-C, Maynard-Smith LA, Ooi AGL & Wandless TJ A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* 126, 995–1004 (2006). [PubMed: 16959577]
8. Chung HK et al. Tunable and reversible drug control of protein production via a self-excising degenon. *Nat Chem Biol* 11, 713–720 (2015). [PubMed: 26214256]
9. Lin MZ & Tsien RY TimeSTAMP tagging of newly synthesized proteins. *Curr Protoc Protein Sci* Chapter 26, Unit 26.5–26.5.11 (2010).
10. Tague EP, Dotson HL, Tunney SN, Sloas DC & Ngo JT Chemogenetic control of gene expression and cell signaling with antiviral drugs. *Nat Methods* 15, 519–522 (2018). [PubMed: 29967495]
11. Jacobs CL, Badiie RK & Lin MZ StaPLs: versatile genetically encoded modules for engineering drug-inducible proteins. *Nat Methods* 15, 523–526 (2018). [PubMed: 29967496]
12. Spencer DM, Wandless TJ, Schreiber SL & Crabtree GR Controlling signal transduction with synthetic ligands. *Science* 262, 1019–1024 (1993). [PubMed: 7694365]
13. Liang F-S, Ho WQ & Crabtree GR Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci Signal* 4, rs2–rs2 (2011). [PubMed: 21406691]
14. Miyamoto T et al. Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. *Nat Chem Biol* 8, 465–470 (2012). [PubMed: 22446836]
15. Hill ZB, Martinko AJ, Nguyen DP & Wells JA Human antibody-based chemically induced dimerizers for cell therapeutic applications. *Nat Chem Biol* 14, 112–117 (2018). [PubMed: 29200207]
16. Gao Y et al. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* 13, 1043 EP–1049 (2016). [PubMed: 27776111]
17. Gao XJ, Chong LS, Kim MS & Elowitz MB Programmable protein circuits in living cells. *Science* 361, 1252–1258 (2018). [PubMed: 30237357]
18. Fleishman SJ et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS ONE* 6, e20161 (2011). [PubMed: 21731610]
19. Park K et al. Control of repeat-protein curvature by computational protein design. *Nat Struct Mol Biol* 22, 167–174 (2015). [PubMed: 25580576]
20. Brunette TJ et al. Exploring the repeat protein universe through computational protein design. *Nature* 528, 580–584 (2015). [PubMed: 26675729]
21. King IC et al. Precise assembly of complex beta sheet topologies from de novo designed building blocks. *Elife* 4, e11012 (2015). [PubMed: 26650357]
22. Schneidman-Duhovny D, Inbar Y, Nussinov R & Wolfson HJ PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33, 7 (2005).
23. Romano KP et al. The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. *PLoS Pathog* 8, e1002832 (2012). [PubMed: 22910833]
24. Soumana DI, Ali A & Schiffer CA Structural analysis of asunaprevir resistance in HCV NS3/4A protease. *ACS Chem Biol* 9, 2485–2490 (2014). [PubMed: 25243902]
25. Dou J et al. De novo design of a fluorescence-activating β -barrel. *Nature* 561, 485–491 (2018). [PubMed: 30209393]
26. Kügler J et al. High affinity peptide inhibitors of the hepatitis C virus NS3–4A protease refractory to common resistant mutants. *J Biol Chem* 287, 39224–39232 (2012). [PubMed: 22965230]

27. Cunningham-Bryant D et al. A Chemically Disrupted Proximity System for Controlling Dynamic Cellular Processes. *J Am Chem Soc* 141, 3352–3355 (2019). [PubMed: 30735038]
28. Suh B-C, Inoue T, Meyer T & Hille B Rapid Chemically Induced Changes of PtdIns(4,5)P2 Gate KCNQ Ion Channels. *Science* 314, 1454–1457 (2006). [PubMed: 16990515]
29. Chavez A et al. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12, 326–328 (2015). [PubMed: 25730490]
30. Qi LS et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183 (2013). [PubMed: 23452860]
31. Loew R, Heinz N, Hampf M, Bujard H & Gossen M Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.* 10, (2010).
32. Zalatan JG et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350 (2015). [PubMed: 25533786]
33. MacKay JL & Kumar S Simultaneous and independent tuning of RhoA and Rac1 activity with orthogonally inducible promoters. *Integrative Biology* 6, 885–894 (2014). [PubMed: 25044255]
34. Chevalier A et al. Massively parallel de novo protein design for targeted therapeutics. *Nature Publishing Group* (2017). doi:10.1038/nature23912
35. Berger S et al. Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. *Elife* 5, e20352 (2016). [PubMed: 27805565]
36. Bick MJ et al. Computational design of environmental sensors for the potent opioid fentanyl. *Elife* 6, e28909 (2017). [PubMed: 28925919]
37. Tinberg CE et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501, 212–216 (2013). [PubMed: 24005320]
38. de Leuw P & Stephan C Protease inhibitor therapy for hepatitis C virus-infection. *Expert Opinion on Pharmacotherapy* 19, 577–587 (2018). [PubMed: 29595065]
39. Di Stasi A et al. Inducible Apoptosis as a Safety Switch for Adoptive Cell Therapy. *N Engl J Med* 365, 1673–1683 (2011). [PubMed: 22047558]
40. Stavrou M et al. A Rapamycin-Activated Caspase 9-Based Suicide Gene. *Mol Ther* 26, 1266–1276 (2018). [PubMed: 29661681]
41. Wu C-Y, Roybal KT, Puchner EM, Onuffer J & Lim WA Remote control of therapeutic T cells through a small molecule-gated chimeric receptor. *Science* 350, aab4077–aab4077 (2015). [PubMed: 26405231]
42. Rose JC et al. A computationally engineered RAS rheostat reveals RAS-ERK signaling dynamics. *Nat Chem Biol* 13, 119–126 (2017). [PubMed: 27870838]
43. Deaner M & Alper HS Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in *Saccharomyces cerevisiae*. *Metabolic Engineering* 40, 14–22 (2017). [PubMed: 28212815]
44. Escalante-Chong R et al. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proc Natl Acad Sci USA* 112, 1636–1641 (2015). [PubMed: 25605920]
45. Morsut L et al. Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* 164, 780–791 (2016). [PubMed: 26830878]
46. Wittekind M, Weinheimer S & Zhang Y Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes. *US Patent* 6800456 (2004).
47. Tsao K-L, Debarbieri B, Michel H & Waugh DS A versatile plasmid expression vector for the production of biotinylated proteins by site-specific, enzymatic modification in *Escherichia coli*. *Gene* 169, 59–64 (1996). [PubMed: 8635750]
48. Fleishman SJ et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332, 816–821 (2011). [PubMed: 21566186]
49. Otwinowski Z & Minor W Processing of X-ray diffraction data collected in oscillation mode. *Meth Enzymol* 276, 307–326 (1997).
50. Romano KP, Ali A, Royer WE & Schiffer CA Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. *Proc Natl Acad Sci USA* 107, 20986–20991 (2010). [PubMed: 21084633]

51. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486–501 (2010). [PubMed: 20383002]
52. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760–763 (1994). [PubMed: 15299374]
53. Chen TS, Palacios H & Keating AE Structure-based redesign of the binding specificity of anti-apoptotic Bcl-x(L). *J Mol Biol* 425, 171–185 (2013). [PubMed: 23154169]
54. Dutta S, Chen TS & Keating AE Peptide ligands for pro-survival protein Bfl-1 from computationally guided library screening. *ACS Chem Biol* 8, 778–788 (2013). [PubMed: 23363053]
55. Foight GW, Chen TS, Richman D & Keating AE Enriching Peptide Libraries for Binding Affinity and Specificity Through Computationally Directed Library Design. *Methods Mol Biol* 1561, 213–232 (2017). [PubMed: 28236241]
56. Procko E et al. Computational design of a protein-based enzyme inhibitor. *J Mol Biol* 425, 3563–3575 (2013). [PubMed: 23827138]
57. Fowler DM, Araya CL, Gerard W & Fields S Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* 27, 3430–3431 (2011). [PubMed: 22006916]
58. Costes SV et al. Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. *Biophys J* 86, 3993–4003 (2004). [PubMed: 15189895]
59. Barry DJ, Durkin CH, Abella JV & Way M Open source software for quantification of cell migration, protrusions, and fluorescence intensities. *J Cell Biol* 209, 163–180 (2015). [PubMed: 25847537]
60. Matreyek KA, Stephany JJ & Fowler DM A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res* 45, e102 (2017). [PubMed: 28335006]
61. Untergasser A et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40, e115–e115 (2012). [PubMed: 22730293]
62. Livak KJ & Schmittgen TD Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2⁻CT Method. *Methods* 25, 402–408 (2001). [PubMed: 11846609]

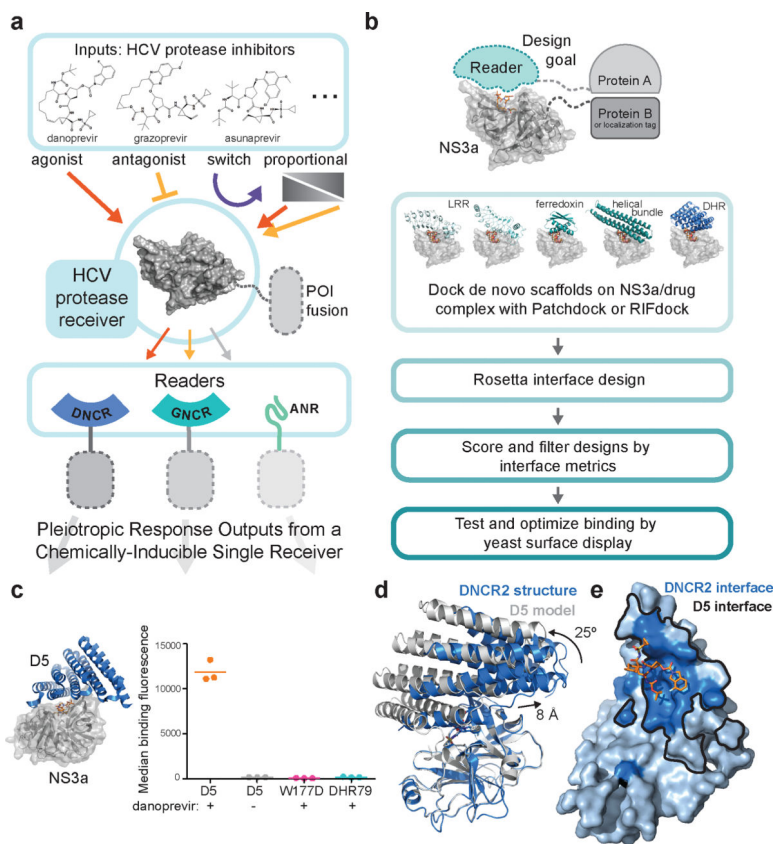


Fig. 1 | Design of a danoprevir:NS3a complex reader.

a, Schematic of the PROCISiR system. Multiple NS3a-targeting drugs are used as inputs that are interpreted by designed readers to generate multiple outputs. **b**, Goal and process for designing and optimizing drug:NS3a complex readers, starting from docking of several scaffold classes on a drug/NS3a complex, Rosetta design of the reader interface, filtering based on Rosetta interface scoring metrics, and finally testing and optimization via yeast surface display. **c**, Rosetta model for D5 (left) and binding of 1 μ M NS3a with avidity to yeast-displayed D5 in the presence or absence of 10 μ M danoprevir. A point mutant of the D5 interface, W177D, and the original DHR79 scaffold show no binding. Technical triplicates and means from one experiment. **d**, A co-crystal structure of the DNCR2:danoprevir:NS3a complex aligned with the D5:danoprevir:NS3a model via NS3a. Measurements indicate shifts of DNCR2 relative to D5. **e**, Residues within 4 Å of NS3a:danoprevir are highlighted in dark blue on the surface of DNCR2. Residues at the interface in the D5 model are outlined in black.

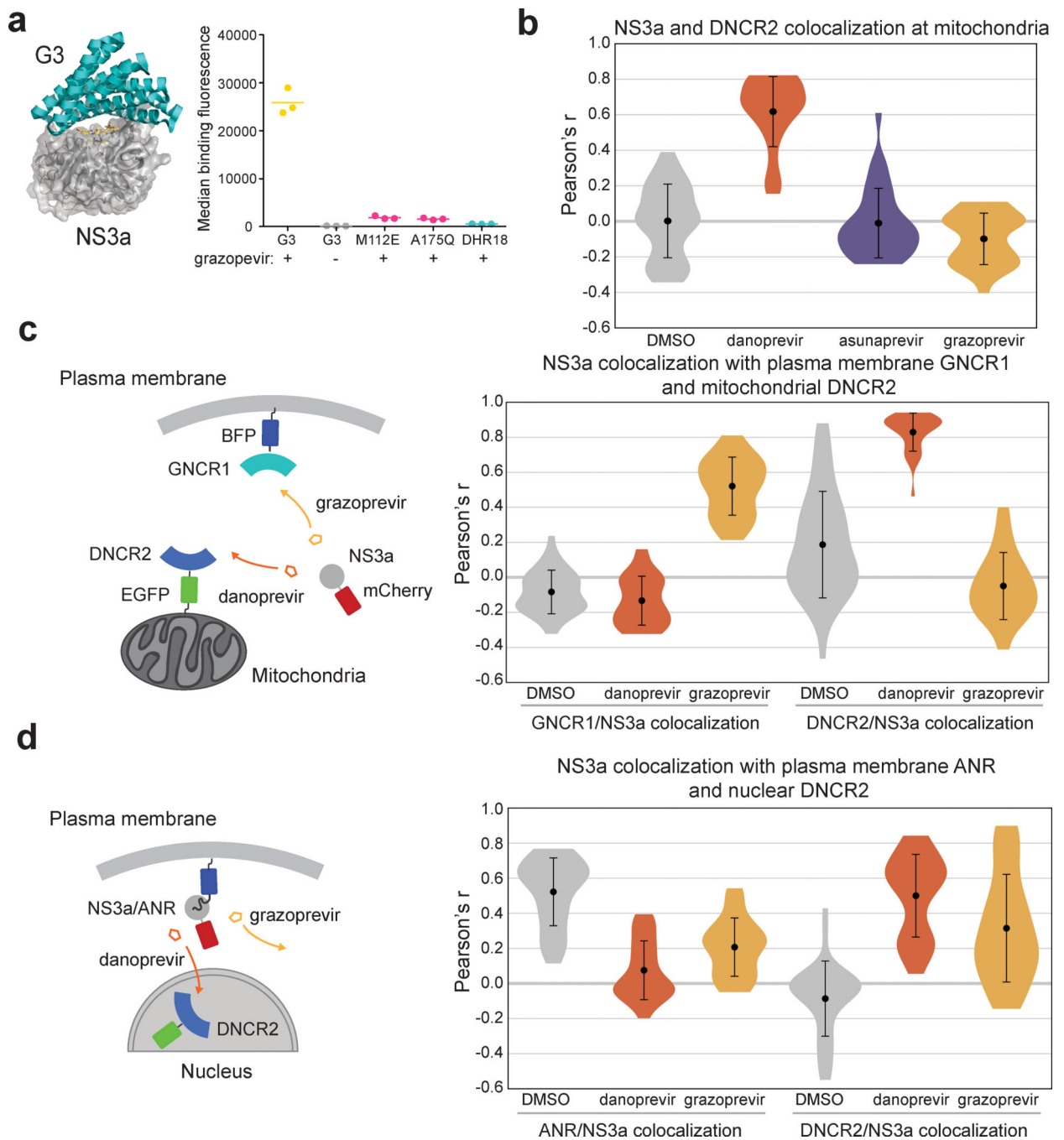


Fig. 2 | Design and testing of a grazoprevir:NS3a complex reader.

a, Rosetta model (left) and binding (right) of 1 μ M NS3a with avidity to yeast-displayed G3 in the presence or absence of 10 μ M grazoprevir. Point mutants at the G3 interface, M112E and A175Q, and the original DHR18 scaffold show no binding. Technical triplicates and means from one experiment. **b**, Colocalization of DNCR2-EGFP with Tom20-mCherry-NS3a after treatment with the drug indicated or DMSO. **c**, Colocalization of NS3a-mCherry with GPCR1-BFP-CAAX or Tom20-DNCR2-EGFP after treatment with danoprevir, grazoprevir, or DMSO. **d**, Colocalization of NS3a-mCherry with ANR-BFP-CAAX or NLS-

DNCR2-EGFP after treatment with danoprevir, grazoprevir, or DMSO. The mean (marked by dot) and standard deviation (error bars) of the Pearson's r of red/blue or red/green pixel intensities for the number of cells stated in Supplementary Table 3 is given for each condition in (b-d), along with the distributions of Pearson's r . See Supplementary Table 3 for sample sizes and P values.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

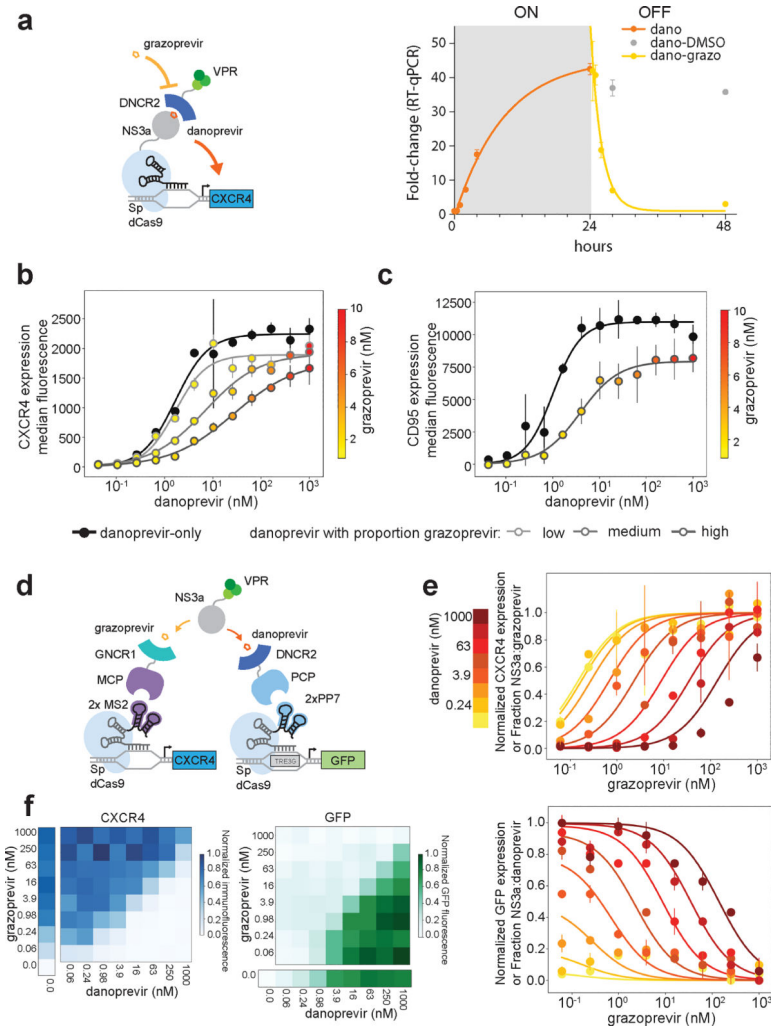


Fig. 3 | Temporal, graded, and proportional transcriptional control using PROCISIR.
a, Reversibility of CXCR4 induction from danoprevir-promoted recruitment of DNCR2-VPR to NS3a-dCas9. “OFF” conditions indicate replacement of danoprevir-containing media with DMSO- (gray) or grazoprevir-containing media (yellow). **b,c**, Co-titrating grazoprevir as a competitor in the presence of a uniform titration of danoprevir inducer in cells expressing the constructs shown in (a) extends the linear range of the CXCR4 (b) or CD95 (c) expression response resulting from DNCR2-VPR recruitment. The curves with higher proportions of the competitive inhibitor grazoprevir (darker gray) were created by performing lower fold-dilutions of grazoprevir (“low” (light gray): 2-fold, “medium” (medium gray): 1.5-fold, and “high” (dark gray): 1.25-fold serial dilutions). See Supplemental Table 4 for the exact drug concentrations used for each condition. (a,b,c) Mean and standard deviation of three biological replicates relative to DMSO-baseline subtracted values. **d**, Schematic of the transcriptional activation system used in (e,f) to simultaneously modulate expression of CXCR4 and GFP in cells co-expressing an MS2 scRNA targeting CXCR4, a PP7 scRNA targeting a GFP reporter, GNCR1-MCP, DNCR2-PCP, NS3a-VPR, and dCas9. **e**, Data from (f) (points, mean, standard deviation) overlaid with modeling (lines) of the fraction of NS3a bound to grazoprevir (top) or danoprevir

(bottom) at the drug concentrations used in (f), as described in the Supplementary Notes. **f**, Expression of CXCR4 and GFP after co-treatment with danoprevir and grazoprevir. Matrices are means of two biological replicates; single-drug titrations are means of three biological replicates. Raw median fluorescence values for (e,f) are shown in Fig. S8e,f.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

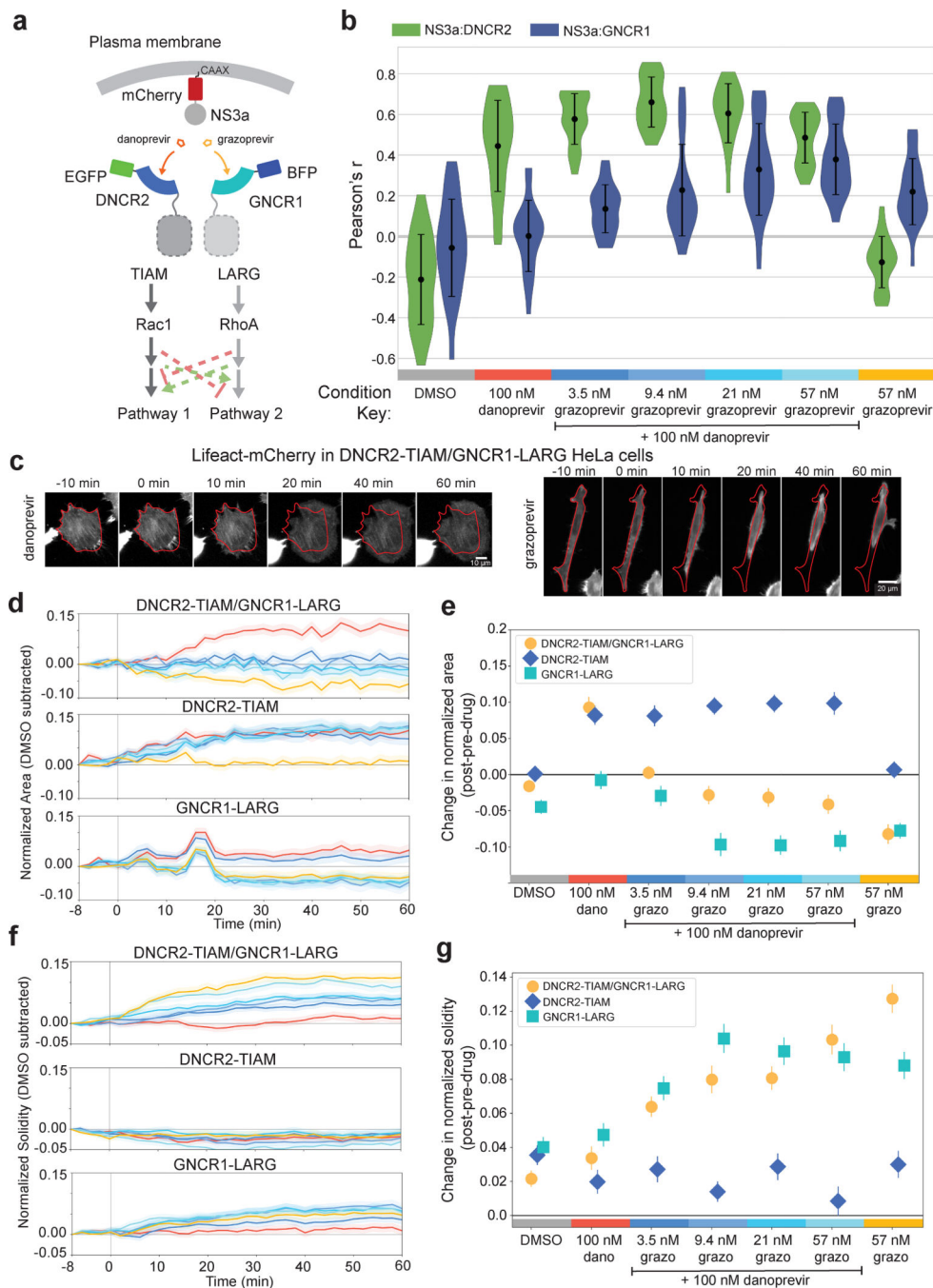


Fig. 4 |. Graded and proportional control of GTPase-driven signaling pathways.

a, Schematic of the colocalization system used in (c-g) for GTPase-driven signaling activation. Combinations of danoprevir and grazoprevir were used to control the proportions of DNCR2 and GNCR1 colocalizing with NS3a at the plasma membrane. **b**, Colocalization of EGFP-DNCR2 (green) or BFP-GNCR1 (blue) with mCherry-NS3a-CAAX quantified by Pearson's *r* from confocal images of HeLa cells. The mean (marked by dot) and standard deviation (error bars) of the Pearson's *r* of red/blue or red/green pixel intensities for the number of cells stated in Supplementary Table 3 is given for each condition, along with the

distributions of Pearson's r . **c**, Representative images from two experiments of HeLa cells co-expressing EGFP-DNCR2-TIAM, BFP-GNCR1-LARG, NS3a-CAAX, and Lifeact-mCherry treated with danoprevir (left) or grazoprevir (right) for the times indicated. Fluorescent signal for Lifeact-mCherry, which stains F-actin, is shown. A red outline illustrating the cell boundary at -10 min is overlaid on each panel to illustrate the change in cell size. **d,f**, Change in normalized area (d) or solidity (f) (DMSO baseline subtracted) over time (drug addition at 0 min) in HeLa cells expressing NS3a-CAAX with either DNCR2-TIAM and GNCR1-LARG (top), DNCR2-TIAM alone (middle), or GNCR1-LARG alone (bottom). Line colors correspond to the drug conditions in (b). **e,g**, Change in normalized area (e) or solidity (g) (average last 10 min-first 10 min). (d-g) Mean and s.e.m. of the number cells per condition listed in Supplementary Table 3 from four independent wells.