

REVIEW ARTICLE OPEN



Nanoscale programming of cellular and physiological phenotypes: inorganic meets organic programming

Nikolay V. Dokholyan ^{1,2,3,4} ✉

The advent of protein design in recent years has brought us within reach of developing a “nanoscale programming language,” in which molecules serve as operands with their conformational states functioning as logic gates. Combining these operands into a set of operations will result in a functional program, which is executed using nanoscale computing agents (NCAs). These agents would respond to any given input and return the desired output signal. The ability to utilize natural evolutionary processes would allow code to “evolve” in the course of computation, thus enabling radically new algorithmic developments. NCAs will revolutionize the studies of biological systems, enable a deeper understanding of human biology and disease, and facilitate the development of in situ precision therapeutics. Since NCAs can be extended to novel reactions and processes not seen in biological systems, the growth of this field will spark the growth of biotechnological applications with wide-ranging impacts, including fields not typically considered relevant to biology. Unlike traditional approaches in synthetic biology that are based on the rewiring of signaling pathways in cells, NCAs are autonomous vehicles based on single-chain proteins. In this perspective, I will introduce and discuss this new field of biological computing, as well as challenges and the future of the NCA. Addressing these challenges will provide a significant leap in technology for programming living cells.

npj Systems Biology and Applications (2021)7:15; <https://doi.org/10.1038/s41540-021-00176-8>

The history of programming dates back to ninth century when brothers Abū Jaʿfar, Muḥammad ibn Mūsā ibn Shākir, Abū al-Qāsim, Aḥmad ibn Mūsā ibn Shākir, and Al-Ḥasan ibn Mūsā ibn Shākir, who have first described an automated flute in their Book of Ingenious Devices¹. Since then, many inventions that automated instructions to perform a particular task were implemented on numerous platforms, including biological materials. Perhaps the most notable example was the experiment by Luigi Galvani in XVIII century, who controlled the contraction of detached frog legs using an electric current². Since then the electric control of live matter moved to tissue level with such notable applications as cardiac pacemakers³, brain⁴, and vagus nerve⁵ stimulators. Most recently, the emergence of the computer-brain interface is enabling “read and write” brain signals in a desirable fashion, thereby enabling control over arterial blood pressure⁶, restoration of the motor functions after stroke⁷, and conscious brain-to brain communication in humans⁸. The emergence of the field of synthetic biology^{9–11} moved the control to a single cell level. The programming, as we know it today, has undergone radical evolution in nineteenth century with the invention of the silicon-based computers. Bioprogramming, like silicon-based coding, is a set of instructions aimed to achieve a particular task, but unlike silicon-based programming, these instructions are operated on biological molecules, such as DNA, RNA and proteins, and aimed at manipulation of specific phenotypic output in living cells.

We are on the threshold of creating nanoscale cellular computers using biological molecules for bioprogramming cellular phenotypes. The revolution in the field of protein design^{12–14} has allowed us to establish rational control of proteins in living cells. With this progress, we are within reach of developing a “nanoscale programming language”, in which molecules serve as operands,

and their conformational states function as logic gates. Combining these operands through protein engineering into larger molecules and molecular complexes will allow us to write and execute “code” using NCAs. As with other computer languages, these agents would respond to input and return output signals. While the speed of the “computation” would be significantly slower than that of inorganic silicon-based computers, one cell could contain more computational agents than the number of CPUs in a supercomputer. Furthermore, the ability to utilize natural evolutionary processes would allow code to “evolve” in the course of computation, thus enabling radically new algorithmic developments.

While this vision may sound like science fiction, a number of elements of this technology already exist, and several laboratories have executed some of these programs, fueling the emergence of the field of synthetic biology. These elements include approaches to sense and control proteins in living cells. Streamlined nanoscale biological computation, *bioprogramming*, will allow direct interrogation of biological systems, enable a deeper understanding of human biology and disease, and introduce possibilities for precision therapeutics. Furthermore, since bioprogramming can be extended to novel reactions and processes not typically seen in biological systems, growth of this field will spark the development of biotechnological applications with impact outside of biological fields. Unlike traditional approaches in synthetic biology that are based on rewiring/hijacking signaling pathways in cells^{15–21}, NCAs are autonomous vehicles based on single-chain proteins or, plausibly in the future, RNA molecules. Although NCAs are susceptible to expression variability, they present a single expression variable compared to that of multicomponent circuit rewiring, done in classical synthetic biology approaches. NCAs offer a complementary mean for controlling cellular phenotypes. Importantly, the size of the “program” (i.e., DNA code) based on NCA

¹Departments of Pharmacology, Penn State College of Medicine, Hershey, PA 17033-0850, USA. ²Departments of Biochemistry & Molecular Biology, Penn State College of Medicine, Hershey, PA 17033-0850, USA. ³Departments of Chemistry, and Biomedical Engineering, Penn State University, University Park, PA 16802, USA. ⁴Departments of Biomedical Engineering, Penn State University, University Park, PA 16802, USA. ✉email: dokh@psu.edu

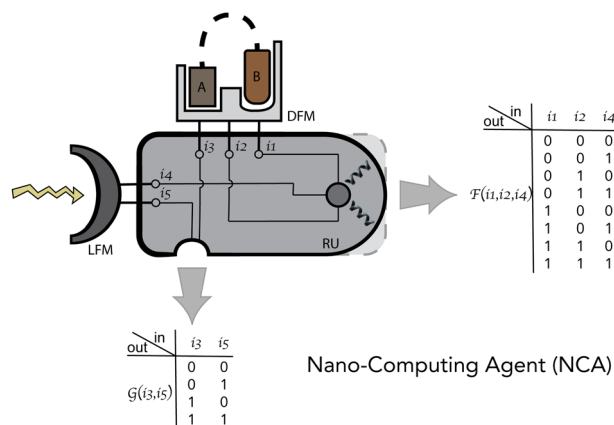


Fig. 1 A conceptual diagram of the NCAs. The response unit (RU) is controlled by light- or drug-sensitive regulatory units (LFM, DFM). The control is established either via allosteric networks within proteins or through direct steric interactions. The inputs $\{i_1, \dots, i_5\}$ and the outputs F and G are represented as binary functions for simplicity.

is drastically smaller than that of programs utilizing synthetic biology approaches. Such code “compression” is possible due to direct design of a protein function rather than indirect control of it through protein expression, as it is done in synthetic biology approaches.

The main component of the NCA is the response unit (RU) – a protein whose output is a biological signal (Fig. 1). RUs are akin to computer motherboards with attached outputs. The input to the RU can be provided by a number of functional modulators (FMs), such as light- or drug-sensitive functional modulators (LFMs and DFMs), as well as other specialized units, such as pH-sensitive, temperature-sensitive, and/or RNA sensitive units. The input domains can be combined to produce a complex response by RUs. In Fig. 1, two DFMs are combined in one so that the input from them generates a signal of higher complexity than that produced by one DFM: this combined DFM can respond to ligands A, B, and A and B (e.g., A and B could be brought separately or as one when connected via a (potentially cleavable) linker). In addition, LFM is “wired” through steric or allosteric networks to influence both output functions $F(x)$ and $G(x)$ of the RU (here, x is the input vector). Examples of the output can be catalysis, (de) activation, homo- or hetero-dimerization, oligomerization, localization, translocation and many other desired functions performed by proteins in cells. The output is generated either through conformational changes in the RU unit or changes in the dynamics of the RU’s active site. For example, in Fig. 1 function $F(x)$ depends on the surface of the RU that interfaces other binding partners, while function $G(x)$ regulates the active site dynamically without changing the conformation of the RU. While the output can be conceived as a binary response, this response can be fine-tuned to adapt to a desired dynamic range.

There are two principal requirements for using NCAs in cells. First, they must be “stealthy”²², meaning that the NCAs do not affect the cellular phenotype without activation of LFMs and DFMs, and that RU behaves as if it did not have regulatory units attached. To address this requirement, we can utilize protein allostery to regulate protein function^{22,23}. In this way, we will be able to avoid functionally important protein surfaces. Second, for simplicity and consistency of operation, the NCAs must be genetically encoded and introduced to cells either via transient transfection or generating a stable cell line.

The conceptual architecture of an NCA is as follows: the main unit RU is a protein or a protein domain, capable of multiple responses, such as alterations of (i) surfaces used to bind other proteins, (ii) active site structure, (iii) dynamics of the active site, (iv) post-translational modifications, and (v) other conformational

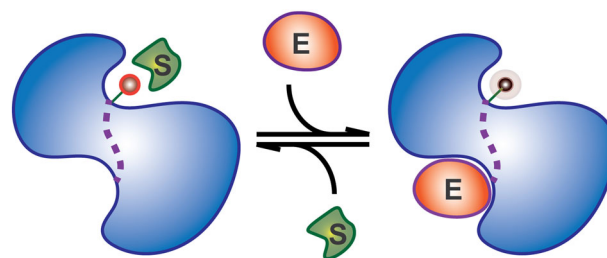


Fig. 2 Schematic diagram of dynamic allostery. An effector, E, binding to an enzyme results in increased dynamics of the active site residues, reducing the likelihood of interaction with the substrate, S, and thus reducing the enzymatic activity of the protein.

changes that result in altered function of this protein. This RU is controlled by functional modulators responding to light, drug, pH, temperature, RNA, or any other user-defined input. The wiring of these functional modulators is performed through allosteric networks^{24,25} or direct steric gating²⁶. In the latter, FM can be used to sterically interfere with the activity of the RUs. In the former, it is possible to utilize dynamic allostery²⁷ (Fig. 2); whereby, upon ligand binding, the active site exhibits altered dynamics thus affecting the RU’s function. In the process of ligand binding, RUs maintain structural equivalence of active versus inactive states of the unmodified RU, thereby limiting interference of our FMs with the RU’s endogenous interaction partners, and maintaining stealthy control over the RUs’ active sites^{22,24}.

Some of the established modes of controlling RUs are photo/chemo-allosteric activation and inhibition (Fig. 3). Other modes, such as steric gating²⁶ and the controlled split protein reassembly method (SPELL)²⁸ (Fig. 3D), offer additional methods of bioprogramming RUs. The proof of concept of simultaneous, multiplexed control of proteins in cells by modulating several RUs at the same time, was recently demonstrated by Dagliyan et al.²⁹

The other critical component of NCAs is a set of FMs or sensors (Fig. 4). Several groups have already developed and utilized light^{30–39} and drug-based sensors^{40–46}. Two or more FMs can be combined to regulate complex RUs: multidomain proteins provide a rich platform for functionalization with multiple MFs. RUs themselves can be also combined to create an even richer platform. Perception of external conditions, such as temperature and pH, are critical to all species. Nature has adapted many hierarchical mechanisms for sensing these conditions: from molecules that undergo conformational change, or shape change⁴⁷, to signaling within and between cells and organs. The sensing of conditions, as well as of molecules, is an important and critical step for developing a versatile palette of NCAs. Following our strategy of allosteric modulation of the RUs, we require that for designed FMs: (i) the C- and N- termini of FMs must be within 7–12 Å distance²⁹, and (ii) the pH, temperature, or binding to other molecules must not destabilize the RUs. It is possible to utilize natural proteins that respond to pH, temperature and binding to molecules, as insertable scaffolds.

Construction of NCAs will offer a novel direction in our ability to interrogate cellular and organismal life, and build novel pharmaceutical strategies. Among many future applications, we could pursue therapeutic interventions using NCAs by exploiting innate evolutionary pressure. For example, NCA may be designed to target kinases whose hyperactivity contributes to cancer. Failure of chemotherapy treatments often happens due to evolutionary adaptation of these kinases to drugs (e.g., via mutations in the drug-binding site). Adaptive changes of the designed NCA may be able to counter changes in kinases to keep them inactive. This example signifies radical new possibilities for establishing perpetual and autonomous regulation of proteins in living cells using NCAs.

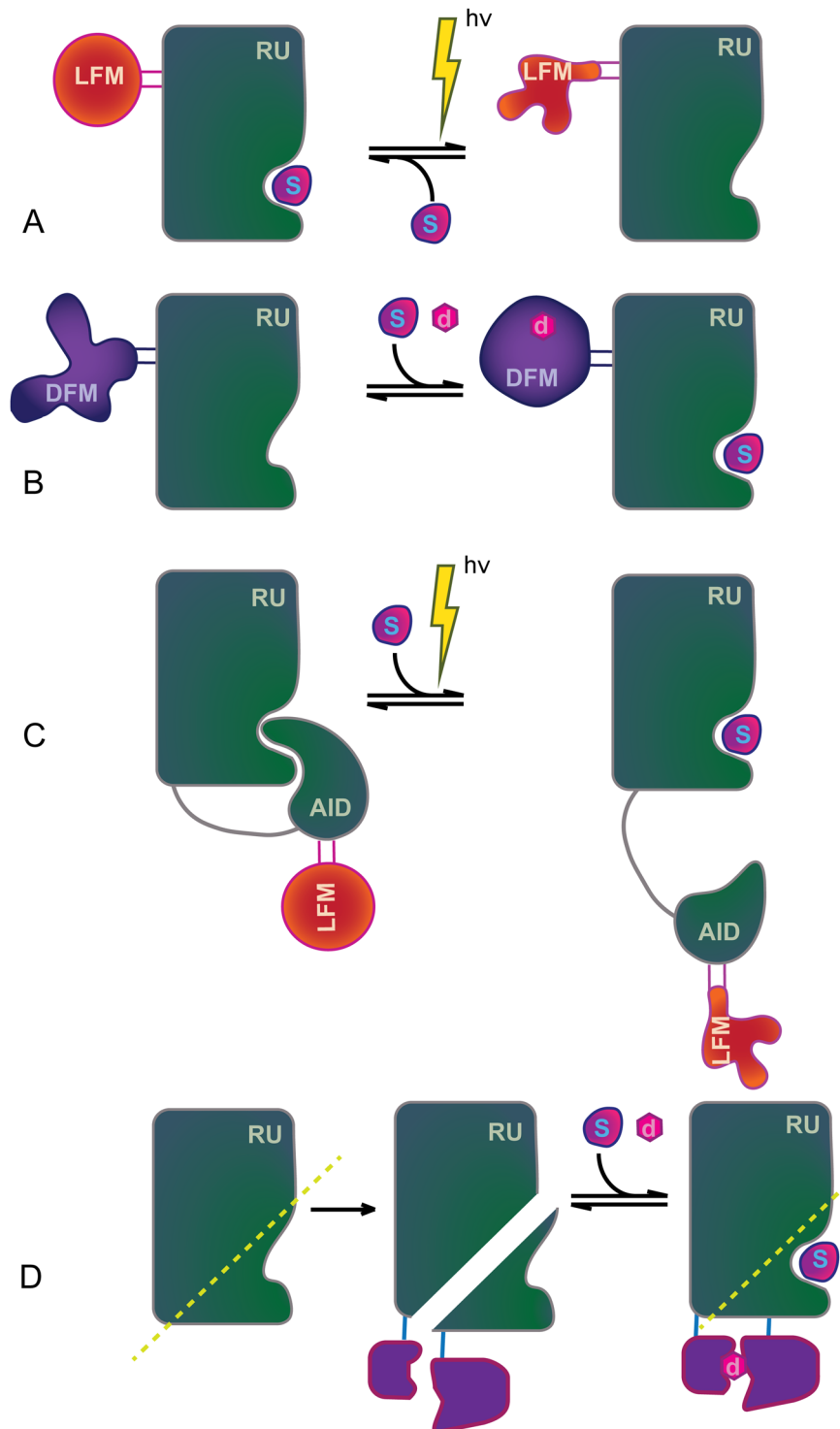


Fig. 3 Some of the established modes of protein control^{26,27}. **A** Photo-allosteric inhibition (e.g., using LOV2): upon irradiation with light, disorder induced in LOV2 (LFM) allosterically induces increased fluctuations in the active site, thereby inhibiting interaction with the substrate, S. **B** Chemo-allosteric inhibition: an engineered DFM (e.g., uniRapR) undergoes disorder-order transition upon addition of a small molecule (rapamycin); thereby, allosterically promoting interaction with the substrate. **C** Photo-allosteric activation: similar to **A** but LFM is inserted into autoinhibitory domains (AID). Upon irradiation by light, AID dissociates thereby activating the RU. Similarly, chemo-inhibition can be accomplished by targeting autoinhibitory domains (AID) by the drug-controlled domain uniRapR. **D** Controlling protein function via split reassembly²⁸. An RU split into N and C parts that are functionalized with two conditionally dimerizing proteins (e.g., iFKBP and FRB). iFKBP is a designed⁴⁴ FKBP variant that is predominantly disordered⁴⁷, thereby disallowing spontaneous reassembly of N and C parts. Upon addition of rapamycin, iFKBP and FRB dimerize, stabilize iFKBP and bring N and C parts to form functional RU.

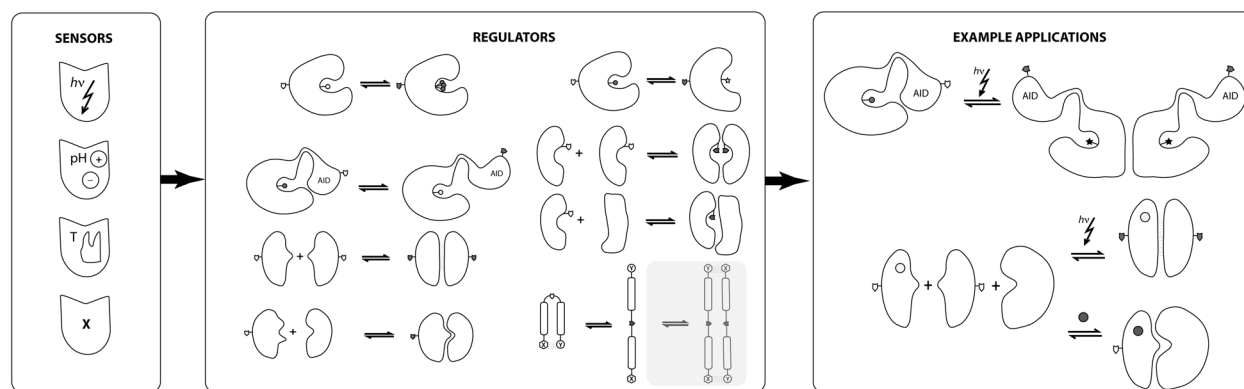


Fig. 4 Bioprogramming in action: A suite of disparate sensors (FMs) can be functionalized to RUs which can respond upon FM stimulations by functional activation/inactivation, hetero/homo-dimerization, conformational switching and many other potential responses. By combining RUs and FMs, we can assemble combinatorial number of possible applications with desired functions.

CHALLENGES

To fully enable nanoscale biological computation – *bioprogramming* – we need to: (i) expand the repertoire of inputs; (ii) include other biological molecules (e.g., RNA, lipids, DNA, macrocycles, metabolites) to aid or perform computation; and (iii) expand the portfolio of approaches for “writing” algorithms at the nanoscale level. Mapping allosteric communications within proteins has been a focus of many laboratories. A number of methods have been to accurately map allosteric pathways^{24,48,49} and even had success in disrupting allosteric connections within proteins^{50,51}. Designing of specific allosteric communications within proteins is a critical next challenge. While within reach, the technology to “rewire” allosteric networks in proteins has yet to be developed. Addressing these challenges will provide a significant leap in technology for programming living cells, and create a new direction in bioprogramming.

Received: 8 October 2020; Accepted: 12 February 2021;

Published online: 11 March 2021

REFERENCES

- Koetsier, T. On the prehistory of programmable machines: musical automata, looms, calculators. *Mech. Mach. Theory* **36**, 589–603 (2001).
- Rivnay, J., Owens, R. M. & Malliaras, G. G. The rise of organic bioelectronics. *Chem. Mater.* **26**, 679–685 (2014).
- McWilliam, J. A. Electrical stimulation of the heart in man. *Br. Med. J.* **1**, 348 (1889).
- Simon, D. T., Gabrielson, E. O., Tybrandt, K. & Berggren, M. Organic bioelectronics: bridging the signaling gap between biology and technology. *Chem. Rev.* **116**, 13009–13041 (2016).
- Koopman, F. A., Schuurman, P. R., Vervoordeldonk, M. J. & Tak, P. P. Vagus nerve stimulation: a new bioelectronics approach to treat rheumatoid arthritis? *Best. Pract. Res. Clin. Rheumatol.* **28**, 625–635 (2014).
- Gotoh, T. M., Tanaka, K. & Morita, H. Controlling arterial blood pressure using a computer–brain interface. *Neuroreport* **16**, 343–347 (2005).
- Naros, G. & Gharabaghi, A. Reinforcement learning of self-regulated β -oscillations for motor restoration in chronic stroke. *Front. Hum. Neurosci.* **9**, 391 (2015).
- Grau, C. et al. Conscious brain-to-brain communication in humans using non-invasive technologies. *PLoS ONE* **9**, e105225 (2014).
- Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802 (2003).
- Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl Acad. Sci. USA.* **101**, 17669–17674 (2004).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Dahiyat, B. I. & Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
- Gordon, D. B., Marshall, S. A. & Mayo, S. L. Energy functions for protein design. *Curr. Opin. Struct. Biol.* **9**, 509–513 (1999).
- Kuhlman, B. et al. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
- Bashor, C. J. & Collins, J. J. Understanding biological regulation through synthetic biology. *Annu. Rev. Biophys.* **47**, 399–423 (2018).
- Toda, S., Blauch, L. R., Tang, S. K. Y., Morsut, L. & Lim, W. A. Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science* **361**, 156–162 (2018).
- Andrews, L. B., Nielsen, A. A. K. & Voigt, C. A. Cellular checkpoint control using programmable sequential logic. *Science* **361**, eaap8987 (2018).
- Kong, W., Meldgin, D. R., Collins, J. J. & Lu, T. Designing microbial consortia with defined social interactions. *Nat. Chem. Biol.* **14**, 821–829 (2018).
- Roybal, K. T. & Lim, W. A. Synthetic immunology: hacking immune cells to expand their therapeutic capabilities. *Annu. Rev. Immunol.* **35**, 229–253 (2017).
- Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A. & Weiss, R. Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell Biol.* **15**, 289 (2014).
- Lim, W. A. Designing customized cell signalling circuits. *Nat. Rev. Mol. Cell Biol.* **11**, 393–403 (2010).
- Dokholyan, N. V. & Hahn, K. M. Stealthy control of proteins and cellular networks in live cells. *Cell Syst.* **4**, 3–6, <https://doi.org/10.1016/j.cels.2017.01.006> (2017).
- Dokholyan, N. V. Experimentally-driven protein structure modeling. *J. Proteom.* **220**, 103777 (2020).
- Dokholyan, N. V. Controlling Allosteric Networks in Proteins. *Chem. Rev.* **116**, 6463–6487 (2016).
- Wodak, S. J. et al. Allostericity in its many disguises: from theory to applications. *Structure* **27**, 566–578 (2019).
- Wu, Y. I. et al. A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461**, 104 (2009).
- Cooper, A. & Dryden, D. T. F. Allostericity without conformational change. *Eur. Biophys. J.* **11**, 103–109 (1984).
- Dagliyan, O. et al. Computational design of chemogenetic and optogenetic split proteins. *Nat. Commun.* **9**, 4042 (2018).
- Dagliyan, O. et al. Engineering extrinsic disorder to control protein activity in living cells. *Science* **354**, 1441 LP–1444 (2016).
- Yumerefendi, H. et al. Light-induced nuclear export reveals rapid dynamics of epigenetic modifications. *Nat. Chem. Biol.* **12**, 399–401 (2016).
- Guntas, G. et al. Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proc. Natl Acad. Sci. USA.* **112**, 112–117 (2015).
- Roth, B. L. DREADDs for neuroscientists. *Neuron* **89**, 683–694 (2016).
- Conklin, B. R. et al. Engineering GPCR signaling pathways with RASSLs. *Nat. Methods* **5**, 673–678 (2008).
- Berglund, K. et al. Combined optogenetic and chemogenetic control of neurons. *Optogenetics*, 207–225, https://doi.org/10.1007/978-1-4939-3512-3_14 (2016).
- Redchuk, T. A., Omelina, E. S., Chernov, K. G. & Verkhusha, V. V. Near-infrared optogenetic pair for protein regulation and spectral multiplexing. *Nat. Chem. Biol.* **13**, 633 (2017).
- Wang, H. et al. LOVTRAP: an optogenetic system for photoinduced protein dissociation. *Nat. Methods* **13**, 755–758 (2016).
- Smart, A. D. et al. Engineering a light-activated caspase-3 for precise ablation of neurons in vivo. *Proc. Natl Acad. Sci. USA.* **114**, E8174 LP–E8183 (2017).

AUTHOR CONTRIBUTIONS

N.V.D. has designed the idea of NCAs and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to N.V.D.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

38. Goglia, A. G. & Toettcher, J. E. A bright future: optogenetics to dissect the spatiotemporal control of cell behavior. *Curr. Opin. Chem. Biol.* **48**, 106–113 (2019).
39. Lerner, A. M., Yumerefendi, H., Goudy, O. J., Strahl, B. D. & Kuhlman, B. Engineering improved photoswitches for the control of nucleocytoplasmic distribution. *ACS Synth. Biol.* **7**, 2898–2907 (2018).
40. Kummer, L. et al. Knowledge-based design of a biosensor to quantify localized ERK activation in living cells. *Chem. Biol.* **20**, 847–856 (2013).
41. Dagliyan, O. et al. Engineering Pak1 allosteric switches. *ACS Synth. Biol.* **6**, 1257–1262 (2017).
42. Wu, H. D. et al. Rational design and implementation of a chemically inducible heterotrimerization system. *Nat. Methods* **17**, 928–936 (2020).
43. Dagliyan, O. et al. Rational design of a ligand-controlled protein conformational switch. *Proc. Natl Acad. Sci. USA.* **110**, 6800–6804 (2013).
44. Karginov, A. V., Ding, F., Kota, P., Dokholyan, N. V. & Hahn, K. M. Engineered allosteric activation of kinases in living cells. *Nat. Biotechnol.* **28**, 743–747 (2010).
45. Bi, S., Pollard, A. M., Yang, Y., Jin, F. & Sourjik, V. Engineering hybrid chemotaxis receptors in bacteria. *ACS Synth. Biol.* **5**, 989–1001 (2016).
46. Park, J. S. et al. Synthetic control of mammalian-cell motility by engineering chemotaxis to an orthogonal bioinert chemical signal. *Proc. Natl Acad. Sci. USA.* **111**, 5896 LP–5901 (2014).
47. Ding, F., Jha, R. K. & Dokholyan, N. V. Scaling behavior and structure of denatured proteins. *Structure* **13**, 1047–1054 (2005).
48. Wang, J. et al. Mapping allosteric communications within individual proteins. *Nat. Commun.* **11**, 1–13 (2020).
49. Proctor, E. A. et al. Rational coupled dynamics network manipulation rescues disease-relevant mutant cystic fibrosis transmembrane conductance regulator. *Chem. Sci.* **6**, 1237–1246 (2015).
50. Aleksandrov, A. A. et al. Regulatory insertion removal restores maturation, stability and function of DeltaF508 CFTR. *J. Mol. Biol.* **401**, 194–210 (2010).
51. He, L. et al. Correctors of $\Delta F508$ CFTR restore global conformational maturation without thermally stabilizing the mutant protein. *FASEB. J. Publ. Fed. Am. Soc. Exp. Biol.* **27**, 536–545 (2012).

ACKNOWLEDGEMENTS

We acknowledge support from the National Institutes for Health (1R35 GM134864) and the Passan Foundation.