Methods and Applications of Computational Design in Multiple States

Quinton M. Dowling

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading Committee:
David Baker, Chair
Neil King
Patrick Stayton

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Methods and Applications of Computational Design in Multiple States

Quinton Dowling

Chair of the Supervisory Committee
Dr. David Baker, Biochemistry
Biochemistry

The balance between flexibility and stability is a key property of proteins, contributing to the vast area of structures and functions observed in nature. Flexibility arises from the complex energy landscape that determines the available conformational states of a protein. Over the past two decades, our ability to design proteins has advanced significantly, however, our understanding of and ability to design multistate proteins remains an area of significant challenge. The multistate design challenge hinders our ability to engineer functional proteins like enzymes, and limits the size and complexity of designed protein materials. The focus of my doctoral research has been developing multistate design approaches and applying those methods to practical design problems; a peudosymmetric trimer used as a building block for assembling large, closed, cage-like nanostructures; and re-engineering a natural enzyme to remove allosteric dependence on double-stranded DNA for activity. In both cases I started by applying simple computational methods to identify mutations that are likely to shift the equilibrium from an undesired state to a desired state. I then incorporated bioinformatics data to improve the design pipeline. This approach was applied to convert a naturally occurring, symmetric, homotrimeric protein, 1WA3, into a pseudosymmetric heterotrimer. I used the resulting pseudosymmetric trimer as a building block for designing cage-like protein assemblies. Because of the choice of trimer as starting material I was able to efficiently build cage-like structures containing 240, 540, or 960 protein chains, significantly larger than any previous computationally designed, bounded, protein nanostructure. I also applied this multistate design

approach to a naturally occurring enzyme, cyclic GMP-AMP synthase (cGAS). Under normal circumstances cGAS adopts an enzymatically active conformation only when bound to double-stranded DNA (dsDNA). By applying multistate design I was able to engineer constitutively active variants of cGAS, which adopt the active conformation independent of dsDNA. We then showed the utility of CA-cGAS in an *in vivo* model. The methods developed here are generally applicable to multistate design problems in naturally occurring building blocks and also highlight the practical utility of the proteins engineered using these approaches.

This dissertation is based in part on the previously published articles listed below.

Dowling, Q. M. Park, Y. Gerstenmaier, N. Yang, E. C. Wargacki, A. Hsia, Y. Fries C. N. Ravichandran, R. Walkey, C. Burrell, A. Veesler, D. Baker, D. King N. P. (2023). Hierarchical design of pseudosymmetric protein nanoparticles.
Manuscript in preparation.

Dowling, Q. M. Volkman, H. E. Gray, E. E. Ovchinnikov, S. Cambier, S. Bera, A. K. Sankaran, B. Johnson, M. R. Bick, M. J. Kang, A. Stetson, D. B. King, N. P. (2023). Computational Design of Constitutively Active cGAS. Nature Structural & Molecular Biology 30 (1): 72–80.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge the mentors who have supported me throughout my journey. Neil King for the opportunity to work in his group and for his support and advice throughout the years, for always putting science and integrity first, and for creating a thoughtful and inclusive work environment. David Baker for his direction and support, and for fostering a supportive and inspiring community of protein scientists. I would also like to thank Brooke Fiala, Daniel Ellis, Cassandra Ogahara, Julia Burrows, Kate DaPron, Karla-Luise Herpoldt, Adam Wargacki for mentoring me as I started in the lab. I would also like to acknowledge the many members of the Rosetta community who have inspired me throughout the years, who have taken time out of their day to teach me and advise me, and who have taken time to criticize my work so that it can be better.

From that community I would like to acknowledge a few individuals in particular. First, Neil Gerstenmaier, for his willingness to run yet another experiment testing yet another crazy idea, Anika Burrell for running yet another gel, Adam Wargacki for his timely thermodynamics modeling when all of my results looked bleak, which gave me the confidence to continue, Yang Hsia for all of our conversations about symmetry, and for seeing that even a partial success could be made into something great, Chelsea Fries for taking up the mantle of resident symmetry obsessive, Will Sheffler for his programming expertise and writing rpxDock, Erin Yang for her leadership and insight, and Sergey Ovchinnikov for mentoring me in bioinformatics. I would also like to acknowledge the staff of the Institute for Protein Design Core Facilities for their technical support, conversations, and encouragement, in particular, Lauren Carter, Rashmi Ravichandran, and Andrew Borst.

I would also like to thank my fantastic collaborators, Dan Stetson, Elizabeth Gray, and Hannah Volkman, whose excellent work and guidance made Constitutively Active cGAS a reality, and David Veesler and Young-Jun Park for their excellent electron microscopy work.

I would also like to thank my doctoral advisory committee, Susie Pun, Patrick Stayton, and Dan Stetson, for their support, understanding, and insightful questions. Thanks also to Mark Gunderson, Scott Truksa, and Arthur Ayers, who inspired me to put down my fly fishing rod, if only for a minute, and pick up a pipette, starting me on my journey as a scientist. I would also like to thank Christopher Fox and Ryan Kramer for encouraging me to continue my development as a scientist.

Finally, I would like to thank my wife, Ying Yu, for her constant support, always encouraging me to take on challenges both scientific and personal. Your love, honesty, compassion, and commitment to growth inspire me. To my children, Lucas and Olivia, who started their journey in life as I started graduate school, thank-you for joining me in this adventure. Thank-you to my brother, who is always willing to have a deeply technical conversation. Finally, thank-you to my parents, without whom none of this would have been possible. Thanks for all of the childcare so I could spend late nights in the lab taking pictures of strange and beautiful proteins.

INTRODUCTION

Flexibility of form is a fundamental property that drives protein function. This phenomenon is observed from the allosteric regulation of enzymes to the assembly of large protein complexes in viruses and microcompartments [1]. The information driving protein form is encoded in their amino acid sequences, and functionality derives from the energy landscape accessible to the protein form. Flexibility therefore emerges from the energetic landscape of a protein, constrained by evolutionary history, biological needs, and a finite amino acid palette. For an enzyme, protein flexibility, defined here as the ability for a protein to adopt multiple conformations, emerges from the requirements of ligand binding and catalysis, from any allosteric controls, and from the thermodynamic requirements of function [2]. Proteins in protein complexes may also adopt multiple states. In the trivial case, two proteins in a complex can exist in a bound and unbound state, however much more sophisticated combinations of states often exist in naturally occurring protein complexes. Quasi-equivalence, for example, is a phenomenon where one protein occupies two or more unique chemical and symmetrical states, commonly observed in icosahedral virus capsids [3–7]. The unique properties of proteins; flexibility, functionality, genetic encoding, make them an appealing target for engineering and broad application. Indeed, methods for engineering proteins have been an active area of research for decades. Massive advances have been made in directed evolution, rational, and computational design of proteins. Numerous computational approaches have been developed to engineer proteins that adopt multiple conformations. Some of these methods are for specific applications, [8–10], while other methods are generalized protocols [11,12] that have been used successfully for multistate protein-protein interface design [13,14]. However, low design success rate, and computational intensiveness, limit the application of these methods. An improved understanding of how to engineer multistate-proteins will further expand the function and utility of designer proteins.

For example, self-assembling protein complexes are ubiquitous structures foundational to living systems. These structures span length and complexity scales from a few protein molecules in nanometer sized complexes to micron-sized viral capsids with thousands of protein subunits. These structures perform a wide variety of structural and biochemical functions [15–17]. As a result of their unique properties, self-assembling proteins have been exploited for applications in drug delivery, enzyme encapsulation, and vaccines [18–21]. Numerous designed assemblies have been engineered in recent years due to advances in the methods used to control how protein subunits interact, particularly those with finite, point-group symmetries (i.e., oligomers, nanoparticles, and capsids) [22]. Engineered nanoparticles and capsids have been generated using many different methods, for example computational protein design [23–31], genetic fusion [32–43], orthogonal chemical interactions [44–47], and directed evolution [4,5]. These designed protein nanoparticles have been utilized for encapsulation and cargo packaging and scaffolding of functional moieties including to enhance receptor-mediated signaling and multivalent antigen presentation [28,48–53]. Antigen presentation has been particularly successful, resulting in multiple vaccines currently in clinical development [51,54] or licensed for use in humans [55,56]. Computational design has been a particularly effective method for engineering novel protein nanoparticles.

One limitation of all design methods reported to date is a reliance on strict symmetry. This approach constrains the number, size, and complexity of designable architectures. Structures are either unbounded, with little control over the final assembly size and number of constituent parts [57], or bounded and limited by symmetry [23,24,27,31]. Of the computationally designed, bounded structures, the largest and most complex consists of 120 subunits arranged in strict icosahedral symmetry [26–28]. The possible bounded, strictly-symmetric structures have either cyclic, dihedral, tetrahedral, octahedral, or icosahedral symmetry. Multistate design provides one approach to increasing the size and complexity of computationally designed, bounded, protein nanomaterials.

Multistate design is a particularly appealing approach to breaking the bounds on finite protein nanomaterials because it harnesses principles already observed in natural systems. As mentioned earlier, quasi-equivalence is one principle by which viral capsids efficiently assemble into complex structures with hundreds-to-thousands of protein subunits. Quasi-equivalent proteins will adopt subtly different backbone conformations depending on their local environment. Engineering proteins capable of quasi-equivalence is likely still beyond the capabilities of current computational methods, although progress has been made [12]. Pseudosymmetry is another principle commonly observed in viral capsid structures. Pseudosymmetry is where genetically distinct subunits or domains adopt roughly symmetric orientations within oligomeric capsomers [58]. This can either take the form of three subunits containing two similar, but distinct domains roughly approximating a six-fold symmetric hexagonal structure [59–61], or two or more genetically distinct but similarly shaped subunits that likewise approximating a six-fold symmetric hexagonal structure [58]. Computationally designed pseudosymmetric oligomers could be generated by designing multiple, orthogonal interfaces simultaneously. Indeed, this method has been successfully implemented [62]. An alternative approach, presented here, is to introduce the minimal set of mutations necessary to convert a symmetric protein complex to a pseudosymmetric complex. By identifying mutations that are disruptive of the homomeric state, but not the heteromeric state, using a combination of computational and bioinformatics methods, I succeeded in accurately preserving the symmetry of the parent homooligomer in a novel heterooligomer, although with some loss in specificity. I then demonstrate the utility of this novel material by designing cage-like assemblies that far surpass the size and complexity of previously designed assemblies. Finally I discuss the implications of breaking strict symmetry on the energy landscape of computationally designed protein-protein interfaces.

I applied a similar combination of computational and bioinformatics-guided methods to the computational design of a constitutively active cyclic GMP-AMP Synthase (cGAS) enzyme.

As with our material design method, we posited that identifying the minimal set of mutations necessary to shift the enzyme from an undesired state to a desired state would be more likely to preserve enzymatic activity, compared to methods used to design multi-state protein switches [8–10], and generalized protocols and used for multistate protein-protein interface design [13,14], although recent advances may overcome some of these limitations [12]. Cyclic GMP-AMP Synthase (cGAS) is a pattern recognition receptor involved in the innate immune response to foreign DNA and DNA damage [63]. cGAS is normally present in an inactive resting state in the cytosol, or tightly bound in the inactive state in the nucleus [64]. cGAS is activated allosterically by binding double-stranded DNA (dsDNA), which induces a series of structural rearrangements leading to catalytic activity [65,66] production of the second messenger cyclic GMP-AMP (cGAMP)[67,68] and induction of a potent type I interferon response through STING activation [69]. Engineering a constitutively active cGAS (CA-cGAS) could be a useful alternative to modified STING-agonist therapeutics which cause tumor regression when injected intratumorally [70–72], or as research reagent.

Here I also present a combined computational and bioinformatics approach to generating CA-cGAS. We hypothesized that mutations that stabilized the active conformation, or destabilized the inactive conformation, would shift the equilibrium towards the active conformation. At the same time I compared the amino acid conservation of a structurally similar protein with low sequence homology, that only adopts a single conservation, reasoning that the amino acids that are conserved in one family but not the other likely contribute to the single or multiple-state nature of the protein. Applying this method to cGAS resulted in multiple CA-cGAS variants that are highly active *in vitro* and demonstrate tumor regression in a mouse model.

These two examples of multi-state protein design are quite different. However, the philosophical approach to the engineering problem and overarching computational methods are quite similar. Specifically, both methods seek to identify the minimum number of mutations necessary to drive the desired effect, reasoning that by minimizing the number of changes made to the starting protein will preserve other desirable features of the protein. In the case of pseudo-symmetric protein nanoparticles our minimal approach enabled a hierarchical approach which resulted in a huge increase in the size and complexity of the designed nanoparticles, with highly efficient use of designed interfaces. Both methods also use computational methods to identify mutations that shift the energy landscape towards a desired state, augmented with bioinformatics. Incorporating bioinformatics was critical to increasing the information content of the designs because predicting the effect of single mutations is notoriously difficult [73]. These general principles should be particularly useful when coupled with recent breakthroughs in accurate structure prediction and design [12,31,74,75].

Chapter 1. Hierarchical Design of Pseudosymmetric Protein Nanoparticles

There are numerous materials applications for multistate protein design. Flexible materials, materials sensitive to environmental conditions, and increased size and complexity are all possible applications. Computational design of closed structures with point-like architectures is limited by a reliance on strictly symmetric building blocks. Here I present a method to break local symmetry using multistate design. In this case the symmetric assembly is one state and the heteromeric assembly the second state. Introducing mutations against the symmetric state enables the formation of a pseudosymmetric heterotrimer, which I then used to build the largest known, bounded, computationally designed protein assemblies. What follows is a manuscript in preparation, describing in detail this work, performed primarily by me with collaborative assistance collecting and processing electron microscopy data from Young-Jun Park [76].

1.1 Introduction to pseudosymmetry and and protein nanoparticles

Self-assembling protein complexes are ubiquitous structures foundational to living systems. These structures span length scales from a few nanometers to micron-sized viral capsids and perform a wide variety of structural and biochemical functions [15–17]. The information driving assembly of these complexes is encoded in their amino acid sequences and functionally takes the form of the structures of individual protein subunits and the interactions between them. The unique properties of self-assembling proteins have been exploited for applications in drug delivery, enzyme encapsulation, and vaccines [18–21]. However, relying on naturally occurring assemblies constrains the engineer to existing sizes, shapes, and levels of complexity. Methods for generating new self-assembling proteins render additional classes of structures and functions accessible, enabling these properties to be tailored to specific applications [77].

Advances in methods for controlling or designing the way protein subunits interact has led to an explosion of new designed assemblies in recent years, particularly those with finite, point-group symmetries (i.e., oligomers, nanoparticles, and capsids) [22]. Engineered nanoparticles and capsids have been generated by computational protein design [23–30], rational design [78], genetic fusion and domain swapping [32–43], orthogonal chemical interactions such as metal coordination [44–47], and laboratory evolution [4,5]. Each of these methods has a characteristic level of precision and predictive capacity. Computational docking and protein-protein interface design stands out for its ability to consistently create new protein complexes with atomic-level accuracy, although with a relatively modest success rate due to the unique challenge posed by each interface design problem. Nevertheless, computationally designed protein nanoparticles have been engineered to encapsulate small molecules, nucleic acids, and other polymers [79–81]; evolved for improved cargo packaging and extended *in vivo* half-life [79]; applied to enhance receptor-mediated signaling and virus neutralization [26,29,82]; and used as scaffolds for structure determination [83–85], multi-enzyme co-localization [86], and multivalent antigen presentation [28,48–53], including in multiple vaccines currently in clinical development [51,54] or licensed for use in humans [55,56]. Further computational methods development will give rise to designed protein nanomaterials of continually increasing sophistication, leading to improved performance in these applications and making additional applications possible.

Design methods reported to date have relied on the use of strict symmetry and pre-existing oligomeric building blocks to reduce the number of new interfaces that must be designed [87,88]. Although this approach yields access to a handful of finite (i.e., bounded) symmetric architectures that require only a single designed interface [89], it nevertheless places a severe constraint on the architectures accessible to design and their size and complexity. The largest and most complex structures designed using this approach comprise 120 subunits and have strict icosahedral symmetry, featuring a single copy of each of two subunits in the icosahedral asymmetric unit [26–28]. Developing methods for breaking the symmetry of computationally designed protein assemblies is a key next step in developing more sophisticated self-assembling proteins.

Four routes to larger and more sophisticated protein assemblies exist, each of which is observed in naturally occurring self-assembling proteins. First, larger protein subunits could be used as building blocks, with titin providing an extreme example [90]. However, this approach is untenable as a general solution, as limits in protein translation, folding, stability, and flexibility are quickly encountered [91]. Second, the number of different kinds of subunits in the assembly (or its asymmetric unit; asu) could be increased by designing new asymmetric interactions between them, as observed in molecular machines such as RNA polymerases [92] or NADH-quinone oxidoreductase [93]. Although ultimately we expect this approach to become possible, it is currently impractical as it would compound the low success rates of existing interface design methods. Third, principles of quasi-equivalence could be used to design large assemblies from protein subunits that adopt subtly different conformations depending on their local symmetry environment, a phenomenon commonly found in icosahedral virus capsids [3–7]. However, current computational protein design methods lack the precision required to reliably encode in a single amino acid sequence the multiple subtly different backbone conformations required to implement this approach. Finally, pseudosymmetry could be used to enable asymmetric functionalization of oligomeric building blocks, opening up new routes to the design of larger assemblies. Pseudosymmetry is also frequently observed in icosahedral virus capsids, where genetically distinct subunits or domains adopt roughly symmetric orientations within oligomeric capsomers [58]. For example, pseudosymmetric trimers in virus capsids may comprise three subunits, each containing two related but slightly distinct domains that result in an (A-B)-(A-B)-(A-B) arrangement with roughly six-fold symmetry at the backbone level [59–61]. Such trimers can be arranged in hexagonal lattices that form the facets of very large icosahedral assemblies; the A and B domains form the distinct sets of contacts that are necessary to form non-porous lattices. Although designing pseudosymmetric assemblies requires the creation of multiple new protein-protein interfaces, a hierarchical approach in which pseudosymmetric oligomers are designed first and subsequently used as the building blocks of larger pseudosymmetric assemblies would allow the distinct interfaces to be designed and validated individually. This approach avoids compounding the relatively high failure rate of interface design and, as we show, permits the design of novel cage-like protein nanomaterials that far exceed the size and complexity of previously designed assemblies.

1.2 Pseudosymmetric heterotrimer design

We started our pseudosymmetric design with a homotrimeric aldolase from the hyperthermophilic bacterium *Thermotoga maritima* that is remarkably stable and tolerant of modification (PDB ID 1WA3; [94]). This trimer has previously been used to design multiple one- and two-component protein assemblies [23,27,95], which as we show below makes possible the re-use of these previously designed interfaces in the creation of large pseudosymmetric assemblies. We set out to identify the minimum set of mutations necessary to drive formation of a pseudosymmetric heterotrimer. We used two methods to identify individual mutations predicted to disrupt—as well as compensatory mutations predicted to restore—homotrimer stability, reasoning that combining sets of such mutations across three variants of the trimer subunit could yield pseudosymmetric heterooligomers (**Fig. 1A**). First, the energetic effects of all possible single and pairwise mutations in 98 contacting residue pairs over 36 positions in the 1WA3 homotrimer interface were evaluated using Rosetta. The Rosetta score and predicted homotrimerization energy (ddG) of each individual and double mutant were normalized to the wild-type (WT) score and ddG by simple subtraction. 96 unique individual mutations had normalized ddG or Rosetta scores >+100, suggesting they may disrupt the wild-type homotrimeric interface (**Fig. S1, A and B**). However, many of these had no compensatory mutation that brought the normalized total score or normalized ddG close to zero. Double mutants that did have favorable normalized score or ddG values were considered further (**red boxes, Fig. 1, B and C**). Second, because 1WA3 is a naturally occurring protein, we also used bioinformatics to guide our mutant and double mutant selection. Using GREMLIN [96,97] we inspected the coupling matrices of highly co-evolving residues at the trimer interface to identify low-frequency single mutations (e.g., H91I; PDB ID 1WA3 numbering) with high-frequency compensatory mutations (e.g., V118Y) (**Fig. 1D**). As expected, many of the predicted disrupting mutations identified by computational modeling or bioinformatics were mutations to bulky hydrophobic residues (**Fig. 1E**). Models of those single mutants were visually inspected and then paired with the best-scoring double mutant. In total, mutations from 76 mutant pairs were selected for experimental screening (**Table S1**).

First, the ability of each single mutation to disrupt trimer formation was screened in a lysate-based assay by evaluating its effect on the assembly of I53-50, a previously reported two-component nanoparticle comprising a trimeric component (I53-50A) derived from the 1WA3 trimer [27]. Clarified lysates from *E. coli* expressing the I53-50A mutants were mixed with purified I53-50B pentamer at three different pentamer concentrations and analyzed by native (i.e., non-denaturing) PAGE (**Fig. 1F**). Four of the 82 single mutants tested failed to yield a band corresponding to the assembled I53-50 nanoparticle, indicating that these either prevented soluble expression of the trimer variant or altered its geometry so that it was no longer assembly-competent. Mutations that prevented nanoparticle formation were then combined with their compensatory mutation to determine if the combination restored the ability to form I53-50 nanoparticles (**Fig. S2**). Through these analyses, three pairs of disrupting and compensatory mutations were identified: H91I/V118Y, P90F/P147A, and P114F/F131V.

We initially set out to generate an “ABC” heterotrimer, in which each subunit has a different amino acid sequence, by combining the three mutant pairs in a tricistronic expression construct (all novel amino acid sequences provided in **Table S2**). The A chain contained the V118Y mutation, the B chain contained F131V and P90F, and the C chain contained H91I, P114F and P147A. However, when we purified the expressed proteins by immobilized metal affinity chromatography (IMAC) and StrepTrap chromatography, analysis by SDS-PAGE suggested the presence of trimers comprising predominantly a mixture of the A and B subunits, with little C subunit (**Fig. S3, A and B**). We therefore expressed a bicistronic “AB” gene, which resulted in a mixture of two distinct heterooligomers we separated by IMAC and size exclusion chromatography (SEC) and identified by native mass spectrometry: a trimer comprising one copy of the A chain and two copies of the B chain (“ABB”), as well as a trimer comprising two copies of the A chain and one copy of the B chain (“AAB”) (**Fig. 1G** and **Fig. S3C**). As we describe below, such heterooligomers provide a simple route to designing large pseudosymmetric materials. To confirm that symmetry was preserved at the backbone level—a prerequisite for our hierarchical design approach—we determined if the heterotrimer mixture was assembly-competent by purifying and incubating it in a 1:1 molar ratio with I53-50B pentamer. Assemblies were purified by SEC and the presence of nanoparticles with the known I53-50 morphology [27] were observed by negative stain electron microscopy (nsEM; **Fig. 1H**).

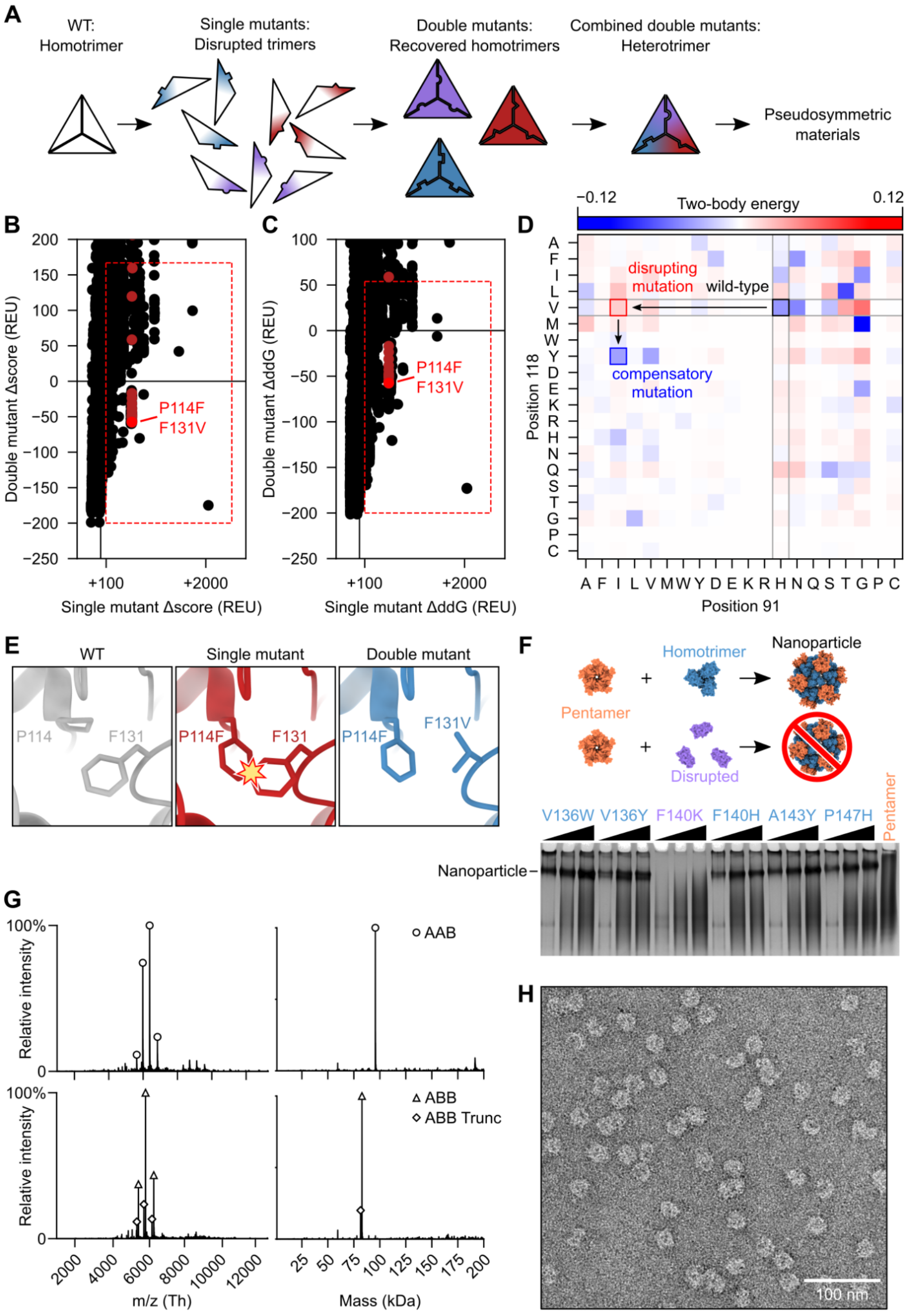


Fig. 1: Design and characterization of a pseudosymmetric heterotrimer. (**A**) The design protocol starts with a homotrimer to which single trimer-disrupting mutations are introduced, followed by compensatory mutations that rescue trimer assembly. Sets of orthogonal mutations (depicted as red, purple, and blue) are combined to generate a heterotrimer that can then be used as a component in pseudosymmetric materials. (**B**) Calculated changes in Rosetta score

(Δscore) and (**C**) predicted trimerization energy (ΔddG) upon mutation were used to evaluate single mutants (horizontal axis) and double mutants (vertical axis). The red boxes enclose mutants that met selection criteria for further evaluation, and mutant pairs containing P114F, shown in Fig. 1E, are highlighted in red. (**D**) Possible mutations were also evaluated by their co-evolution coupling matrix. Desirable mutations are those where the single mutant/WT pair are observed less often than expected (red; H91I/V118) and the double mutant pair more often than expected (blue; H91I/V118Y). (**E**) An example of a productive mutant pair where the WT residue F131 clashes with the mutant residue P114F and the second mutation F131V resolves the clash. (**F**) Disruption of trimer geometry was assayed by assembling mutant I53-50A trimers in clarified *E. coli* lysates with purified I53-50B pentamer and evaluating the presence or absence of I53-50 nanoparticles by native PAGE. Here, F140K was identified as a disrupting single mutation. Black wedges indicate increasing pentamer concentration in the assembly reaction. (**G**) Native mass spectrometry of (*top*) AAB-enriched and (*bottom*) ABB-enriched heterotrimer fractions purified by IMAC and SEC. ABB Trunc refers to a truncation product of the A chain in which the N-terminal 10 residues of the protein were missing. (**H**) Assembly of I53-50-like nanoparticles using an AAB/AAB mixture of I53-50A heterotrimer was verified by negative-stain EM.

1.3 Design and characterization of a 240-subunit, T=4 nanoparticle

We then used the pseudosymmetric heterotrimers to design large, pseudosymmetric assemblies with icosahedral symmetry. We had previously used the 1WA3 homotrimer to generate a single-component nanoparticle with icosahedral symmetry, I3-01, by designing a novel protein-protein interface with two-fold symmetry between the subunits of adjacent trimers [23]. The existence of this interface allowed us to generate a fifteen-subunit “pentasymmetron” [98] comprising five trimers by simply including the I3-01 mutations on the A chains of the AAB heterotrimer (**Fig. 2A**). Docking this pentasymmetron against C3-symmetric homotrimers (“CCC”) and designing novel sequences that create favorable interfaces between the B and C chains yielded models of 240-subunit nanoparticles with icosahedral symmetry (**Fig. 2, B-D**). The Caspar-Klug triangulation (T) number notation [3] is useful for describing these pseudosymmetric nanoparticles, although the assignment of subunits to geometric elements is different than in traditional use of the T number in structural virology. In our pseudosymmetric nanoparticles, trimeric building blocks form wireframe-like structures surrounding roughly pentagonal and hexagonal pores, each subunit interacting with exactly one other subunit from a different trimer. The original I3-01 nanoparticle can be thought of as T=1, with one (A) subunit in the asu, while the pentasymmetron-containing 240-subunit nanoparticles are T=4, with four (2×A, 1×B, 1×C) subunits in the asu. In these assemblies $k=0$, so the T number is equal to h^2, where h is a positive integer representing the number of steps required to traverse from one pentasymmetron to another, each step moving to the next pentagonal or hexagonal pore. Because this is one of the set of equations used to define class I Goldberg polyhedra [99], we refer to these nanoparticles using the naming convention GI_T-X, where G stands for Goldberg, I for icosahedral symmetry, T is used to denote the triangulation number of a particular architecture, and X is a unique identifier for each design. We expressed three initial designs in *E. coli* as tricistronic genes with a 6×His tag on only the C chain, and found that Ni^{2+} beads co-precipitated all three subunits of two of the designs, suggesting assembly (**Fig. S4**). We moved forward with the better expressing and more soluble of the two, GI_4-F7. To scale up expression of the AAB heterotrimer so that we could explore assembly of GI_4-F7 *in vitro* from purified components [100], we re-cloned it as a bicistronic AB construct with a 6×His tag on the A chain. Upon gradient elution during IMAC, we observed three peaks corresponding to an ABB-rich fraction, an AAB-rich fraction, and off-target AAA homotrimers that assembled to an I3-01-like nanoparticle (**Fig. S5A**). We polished the AAB peak by size exclusion chromatography

(SEC), discarding the I3-01-like nanoparticle fraction (**Fig. S5B-D**). In parallel, we purified 6×His-tagged CCC homotrimer—which was also derived from the 1WA3 trimer—by IMAC and SEC.

We mixed the AAB heterotrimer with an excess of the CCC homotrimer in the presence of detergent and initiated assembly by dialyzing overnight into Tris-buffered saline (TBS). The major assembly product was purified by SEC (**Fig. S6**), and images obtained by cryo-electron microscopy (cryoEM) of vitrified specimens revealed wireframe structures with large hexagonal pores that strikingly resembled the design model (**Fig. 2E**). We determined a single-particle reconstruction of GI_4-F7 at 4.4 Å resolution applying icosahedral symmetry and a 3.1 Å resolution structure of the four chains of the asu (cryoEM processing details in **Fig. S7** and **Table S3**). The cryoEM structure agrees well with the design model, with a Cα root mean square deviation (RMSD) of 9.3 Å across all 240 subunits and 3.0 Å within the asu (**Fig. 2F and G**, **Fig. S8**). The differences between the cryoEM structure and design model are mostly accounted for by slight rigid body deviations allowed by the limited degrees of freedom (DoFs) of the oligomeric building blocks in this symmetric architecture (**Fig. S8A**). The main rigid body deviation is a 5.9° clockwise rotation of the pentasymmetron, accompanied by a 5.8 Å translation away from the origin (**Fig. 2H**). The CCC homotrimer compensates by rotating 12.4° and translating 4.0 Å, resulting in only slight local shifts relative to the design model (2.1 Å across the B:C subunits; **Fig. 2I**). Within the pentasymmetron, the DoFs of the AAB heterotrimer are no longer restricted by the strict icosahedral symmetry of I3-01, resulting in a slight deviation from perfect two-fold symmetry between neighboring A chains (1.4 Å Cα RMSD; **Fig. 2J** and **Table S4**). The effects of these deviations on the atomic contacts at the A:A, A:B, and B:C interfaces are highlighted in **Fig. S8**. There were minor structural deviations within each subunit, primarily in the B:C interface (**Fig. S9**). Overall, the diameter of GI_4-F7 observed by cryoEM is within 2% of the design model, establishing that our method is capable of accurately designing pseudosymmetric protein nanomaterials comprising hundreds of subunits.

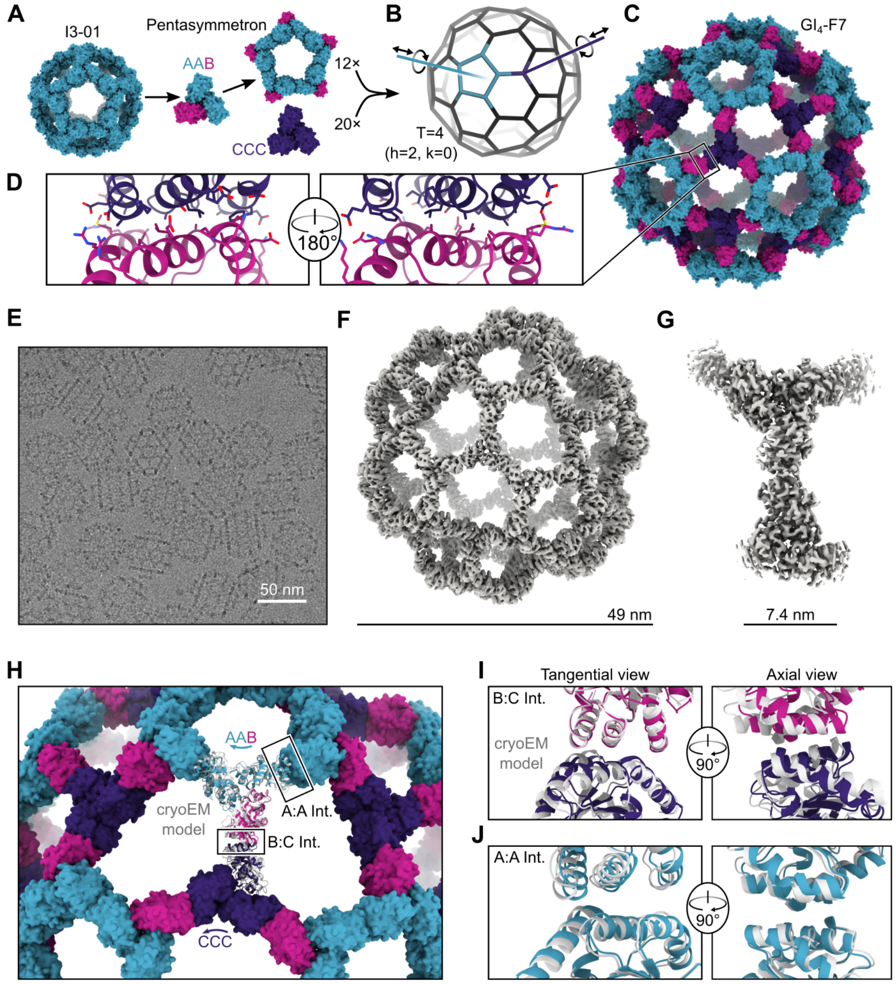


Fig. 2. Design and characterization of the 240-subunit GI$_4$-F7 nanoparticle. (**A**) Schematic of pentasymmetron generation from I3-01 and the AAB heterotrimer. The A (cyan) subunits in the pentasymmetron retain the two-fold symmetric I3-01 nanoparticle interface, while the B (magenta) subunits are available for docking. (**B**) Docking the pentasymmetron as a rigid body against CCC homotrimers (purple) yields 240-subunit, T=4 assemblies. Translational and rotational DOFs for the pentasymmetron and homotrimer components are indicated. (**C**) A design model of GI$_4$-F7. (**D**) Detail of the computationally designed interface between the B and C subunits of GI$_4$-F7 design model. (**E**) CryoEM micrograph of assembled GI$_4$-F7 nanoparticles embedded in vitreous ice. (**F**) 4.4 Å resolution density map of the entire GI$_4$-F7 nanoparticle. (**G**) 3.1 Å resolution density map from an asymmetric unit obtained via symmetry-expansion

and local refinement. (**H**) Comparison of the cryoEM structure derived from local refinement (gray ribbon) with the computational design model (colored ribbons), aligned using a single copy of the asu. Arrows indicate rigid body deviations of the pentasymmetron (cyan) and CCC homotrimer (purple). (**I**) Detail of the rigid body deviations from the design model at the B-C interface and (**J**) the A-A (I3-01) interface. In panel J, two neighboring copies of the AAB heterotrimer from the full nanoparticle reconstruction and the design model were aligned.

1.4 Observation and structure of a 540-subunit, T=9 nanoparticle

Unexpectedly, in a number of the GI_4-F7 micrographs we also observed a 71 nm nanoparticle with a similar wireframe morphology and hexagonal pores (**Fig. 3A**). By counting the hexagonal pores we found that *h*=3; thus the nanoparticle is T=9 and we refer to it as GI_9-F7. GI_9-F7 can be explained by the presence of small amounts of ABB heterotrimer in AAB heterotrimer preparations. Analogous to the AAB heterotrimer, which forms a pentasymmetron through five roughly two-fold-symmetric A:A interfaces inherited from I3-01, the ABB heterotrimer forms a two-trimer “disymmetron” structure held together by the same A:A interaction (**Fig. 3B**). In GI_9-F7 this disymmetron occupies the icosahedral two-fold symmetry axes, providing the edges that connect three-fold-symmetric facets containing three ABB heterotrimers and three CCC homotrimers. As a result, GI_9-F7 is quasisymmetric in addition to being pseudosymmetric: the A, B, and C subunits each occupy multiple, distinct environments in the assembly. We expanded GI_4-F7 to generate a design model for GI_9-F7 containing 12 pentasymmetrons constructed from AAB heterotrimers, 30 disymmetrons comprising ABB heterotrimers, and 60 CCC homotrimers (**Fig. 3B**). The asu of GI_9-F7 therefore comprises one AAB trimer, one ABB trimer, and one CCC trimer. To generate more homogenous preparations of GI_9-F7, we separately polished the AAB and ABB heterotrimer fractions from IMAC (see **Methods**) and assembled them with CCC homotrimer at a 1:1:1 ratio. Micrographs of SEC-purified GI_9-F7 assemblies revealed enrichment of the target assembly (**Fig. S10**), and we determined a cryo-EM structure of the nanoparticle to 6.7 Å resolution applying icosahedral symmetry, as well as a 4.0 Å resolution structure of the asu (**Fig. 3C**, **Fig. S7**, and **Table S3**). Consistent with the accuracy with which we designed GI_4-F7, the GI_9-F7 cryoEM structure deviates from the design model by only 11.5 Å Cα RMSD across all 540 subunits, or 1.6% of the nanoparticle diameter, and superimposition of the designed asu with the structure yields a Cα RMSD of 4.6 Å across all nine chains.

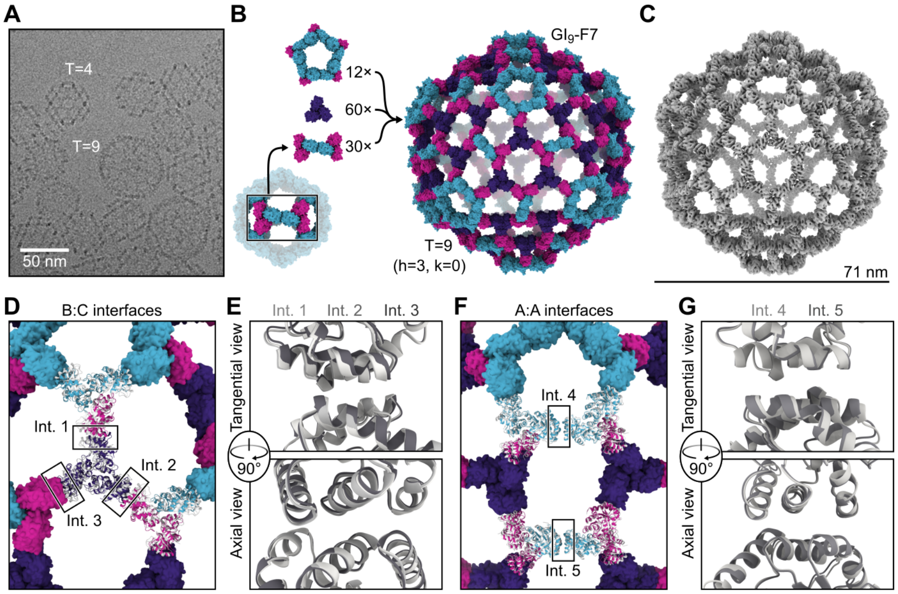


Fig. 3. Discovery and characterization of the 540-subunit GI$_9$-F7 nanoparticle. (**A**) A cryo-electron micrograph showing GI$_4$-F7 and GI$_9$-F7 nanoparticles in the same preparation. (**B**) GI$_9$-F7 comprises, from top to bottom, 12 pentasymmetrons consisting of five copies of the AAB heterotrimer, 60 copies of the CCC homotrimer, and 30 copies of the disymmetron consisting of two copies of the ABB heterotrimer. A subunits are cyan, B subunits are magenta, and C subunits are purple. (**C**) CryoEM map of GI$_9$-F7 at 6.7 Å resolution. (**D**) Comparison of a model derived from the cryoEM map (gray) with the computational design model (colors), aligned using a single asu (shown in cartoon). The three independent copies of the B:C interface in the asu are indicated. (**E**) Alignment of Int. 1. (light gray), Int. 2 (medium gray), and Int. 3 (dark gray) from the cryoEM model. (**F**) Alignment of two neighboring copies of AAB heterotrimers from the cryoEM model to the design model. The two independent copies of the A:A (I3-01) interface, located in the pentasymmetron (Int. 4) and the disymmetron (Int. 5), are indicated. (**G**) Alignment of the two A:A (I3-01) interfaces from the cryoEM model, Int. 4 (light gray) and Int. 5 (medium gray).

Although the pentasymmetron, disymmetron, and three homotrimers in GI_9-F7 are respectively constrained by the five-fold, two-fold, and three-fold icosahedral symmetry axes, no single trimer occupies a position constrained by icosahedral symmetry—the icosahedral three-fold instead passes through a large pore. Each trimer can therefore deviate from the design model along all six rigid body DoFs. As a result, the two designed nanoparticle interfaces (B:C and A:A) occupy five quasi-equivalent positions in GI_9-F7. Two of the B:C interfaces are located within the icosahedral asu, between the CCC homotrimer and the B chain of a neighboring pentasymmetron (interface 1) or disymmetron (interface 2) (**Fig. 3D**). The third B:C interface is between the CCC homotrimer and the B chain of a disymmetron in a neighboring asu (interface 3). Despite being unconstrained by symmetry, interfaces 1-3 fit well to the density, with a very small deviation from the design model comprising only a small rotation with very little radial translation (**Fig. 3E**, **Fig. S11A-D**, and **Table S5**). Interfaces 4 and 5 are the A:A (i.e., I3-01) interfaces in the pentasymmetron and disymmetron, respectively (**Fig. 3F**). Unlike interfaces 1-3, interfaces 4 and 5 appear to differ, with a Cα RMSD of 1.3 Å to each other and Cα RMSDs of 2.0 and 2.4 Å to the GI_9-F7 design model, respectively (**Fig. 3G** and **Fig. S11E-G**). This difference arises because the pentasymmetron interface (interface 4) is not symmetrically constrained, while the disymmetron interface (interface 5) is constrained by the icosahedral two-fold symmetry axis. We propose that the lowest-energy state of the A:A interface is not perfectly symmetric, but that the symmetry requirements for nanoparticle assembly force it to adopt a higher-energy, two-fold-symmetric configuration where appropriate.

1.5 Generation of pseudosymmetric nanoparticles with extensible hexagonal lattice facets

The geometries of I3-01, GI_4-F7, and GI_9-F7 are analogous to the first three instances in the infinite series of class I Goldberg polyhedra [3,99,101]. The larger instances in this series are effectively constructed by folding 20 roughly triangular two-dimensional (2D) hexagonal lattices into icosahedron-like shapes through the introduction of curvature at their edges and vertices. Theoretically, the next nanoparticle in the series would be GI_{16}-F7. As in GI_4-F7 and GI_9-F7, curvature in this structure would be provided by disymmetrons and pentasymmetrons. However, GI_{16}-F7 would have a C3-symmetric component centered on the icosahedral three-fold symmetry axis, as opposed to the pore-centered three-fold of GI_9-F7. Extrapolating from GI_4-F7 and GI_9-F7, this component must be a homotrimer of the B chain (“BBB”), and it must be coplanar with the six surrounding CCC homotrimers (i.e., their three-fold axes must be parallel; **Fig. 4A**). Nanoparticles beyond GI_{16}-F7 simply add more copies of the BBB and CCC homotrimers (and ABB disymmetrons) to form larger two-dimensional hexagonal arrays (**Fig. 4B**). Thus, obtaining GI_{16}-F7 and the larger nanoparticles in the series does not require new interface design, only production of BBB homotrimer. Indeed, analyzing an equimolar mixture of purified BBB and CCC homotrimers by nsEM yielded a 2D array with a characteristic hexagonal lattice diffraction pattern (**Fig. S12A-C**). The dimensions of the array agree well with a design model derived from the GI_9-F7 nanoparticle (**Fig. S12, D and E**).

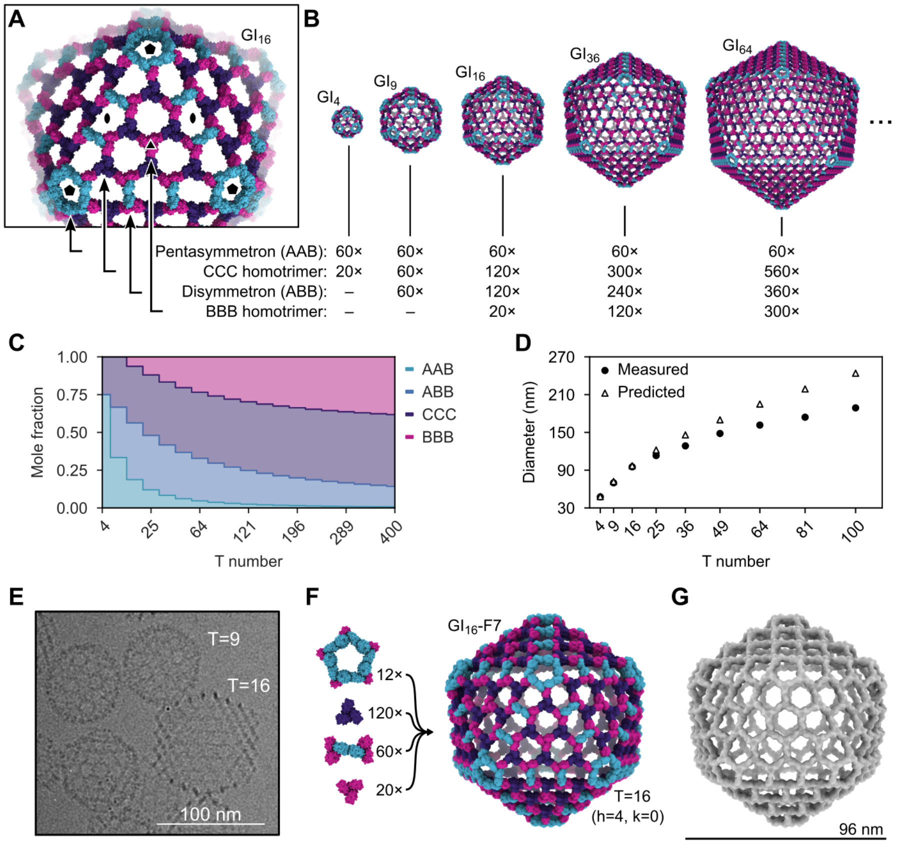


Fig. 4. Generation of pseudosymmetric nanoparticles with extensible hexagonal lattice facets. (**A**) The four types of trimers required to generate pseudosymmetric nanoparticles with T numbers ≥16, viewed in the context of the GI$_{16}$-F7 design model. Icosahedral five-, two-, and three-fold symmetry axes are indicated. (**B**) Design models and corresponding building block stoichiometries of GI$_4$-F7, GI$_9$-F7, GI$_{16}$-F7, GI$_{36}$-F7, and GI$_{64}$-F7 nanoparticles. The stoichiometries listed indicate the number of each kind of trimeric component. (**C**) Graphical depiction of trimer stoichiometry as a function of T number. (**D**) Theoretical nanoparticle diameters and Z-average hydrodynamic diameters measured by DLS as a function of trimer stoichiometry used during *in vitro* assembly (indicated by T number). (**E**) Cryo-electron micrograph of a sample containing GI$_9$-F7 and GI$_{16}$-F7

nanoparticles. (**F**) Composition and design model of the GI_{16}-F7 nanoparticle. (**G**) 14.9 Å resolution cryoEM map of GI_{16}-F7.

In the absence of other control mechanisms, the inclusion of BBB homotrimer in assembly reactions should yield distributions of large T-number assemblies rather than monodisperse preparations of a single species. However, the relative stoichiometries of these components in each assembly vary as a function of T number (**Fig. 4, B and C**), providing a potential mechanism for controlling assembly size. We prepared assembly reactions containing the four components at the stoichiometries corresponding to T=4, 9, 16, 25, 36, 49, 64, 81, and 100 nanoparticles. Consistent with our predictions, the Z-average hydrodynamic diameter measured by DLS increased with increasing target T number (from 47.5 ± 0.4 nm to 188 ± 1.1 nm), though the observed hydrodynamic diameter deviated from the predicted diameter at higher T numbers (**Fig. 4D**, **Table S6**). GI_{16}-F7 was readily observed by cryoEM in assembly reactions prepared at the T=16 stoichiometry (**Fig. 4E**). GI_{16}-F7 is predicted to have a diameter of 96 nm and contains 12 pentasymmetrons, 120 CCC homotrimers, 60 disymmetrons, and 20 BBB homotrimers for a total of 960 subunits (**Fig. 4F**). We determined a 14.9 Å resolution cryoEM map of GI_{16}-F7 and found that it closely matches the expected geometry of the design (**Fig. 4G**). This assembly has an internal volume that is roughly 90-fold larger than our previously designed nanoparticles with strict icosahedral symmetry [23,27] and adeno-associated viruses, commonly used vectors for gene therapy [102].

1.6 Conclusions and future directions

Here we show that designing pseudosymmetric protein building blocks, in which symmetry is broken at the sequence level while backbone symmetry is maintained, enables the construction of very large pseudosymmetric protein nanoparticles. By breaking the reliance on strict symmetry, this work represents an important leap forward relative to established methods for accurately designing novel self-assembling proteins [24,25,27,28,35,42,43,103]. Although those methods are general with respect to the choice of building block and can therefore give rise to a rich variety of potential assemblies, the space of strictly symmetric structures is nevertheless infinitesimally small compared to the space of asymmetric architectures. Continued development of computational methods for designing new proteins and protein assemblies, docking existing protein building blocks, and designing protein-protein interfaces will enable the generation of increasingly asymmetric protein nanomaterials. The technological impact of such computationally designed protein nanomaterials is highlighted by the recent regulatory approval and commercialization of a designed protein nanoparticle vaccine for COVID-19 [55].

Although some small viruses make purely pseudosymmetric capsids, many larger capsids are constructed by combining pseudosymmetry with quasisymmetry. Analogously, while GI_4-F7 and the assemblies reported in the accompanying manuscript (Lee & Kibler, et al.) are pseudosymmetric, with each distinct subunit in a single chemical environment, GI_9-F7 and larger structures are also quasisymmetric, with genetically identical subunits in more than one chemical environment. The A subunit occupies an asymmetric position in the pentasymmetron and either an asymmetric position in the disymmetron (for even T numbers) or both asymmetric and two-fold symmetric positions in disymmetrons (for odd T numbers >9). Likewise, the B and C subunits occupy different chemical environments depending on their locations in the assembly. Quasisymmetry is enabled by the use of two-component heterotrimers (ABB and AAB), which provides for economy in coding for larger assemblies. The T=4 structure requires only three distinct chains, compared to four chains for the more conceptually straightforward approach of a strictly pseudosymmetric ABC heterotrimer and DDD homotrimer. For larger particles the economy is even greater: for example, only three unique chains are required to make T=9 nanoparticles, but seven would be needed for the strictly pseudosymmetric approach. The tradeoff to this economy is a reduction in precision compared to the strictly pseudosymmetric approach we describe in the accompanying manuscript (Lee & Kibler, et al.), although as we have shown this can be partially overcome by controlling the stoichiometry of the assembly reaction.

We used a hierarchical design strategy to fulfill the requirement for multiple designed interfaces in our pseudosymmetric nanomaterials. By constructing pseudosymmetric heterotrimers first and combining these with an existing designed interface to generate pentasymmetrons, we produced 240-subunit and larger pseudosymmetric assemblies in a single dock-and-design step. Similar hierarchical and modular design strategies are widespread in reticular chemistry [104,105] and DNA nanotechnology [106–109], and should become increasingly powerful in protein nanomaterials design as the number and kinds of modular protein building blocks continue to increase [31,110–114].

1.7 Methods for pseudosymmetric protein nanoparticle design

1.7.1 Pseudosymmetric trimer design

To identify mutations for altering trimer assembly specificity, we first identified all pairs of interacting residues in the trimer interface. Contacts were defined as any residue with a heavy (i.e., non-hydrogen) atom within 4 Å of a heavy atom in a residue across the interface. We then used Rosetta to calculate the total score of poses containing all possible pairs of mutations, as well as the difference in score between the trimeric and monomeric states using the ddG filter. Individual mutations were evaluated by comparing their ddG and total scores to those of the WT interface according to Equation 1. The total scores and ddG values of the paired mutations were similarly normalized according to Equation 2.

Eq. 1 $Single\ mutant\ \Delta score = score_{mutant} - score_{WT};$

$$Single\ mutant\ \Delta ddG = ddG_{mutant} - ddG_{WT}$$

Eq. 2 $Double\ mutant\ \Delta score\ = score_{mutant} - score_{WT};$

$$Double\ mutant\ \Delta ddG\ = ddG_{mutant} - ddG_{WT}$$

Ideal mutant pairs were those where one or both single mutations increased the energy of the trimer relative to the WT (i.e., normalized scores > 0) while the double mutation had no effect or stabilized the trimer (i.e., normalized scores ≤0). We also identified likely positions for design using coevolutionary analysis [96,97]. Strongly co-evolving residues at the protein-protein interface were identified using GREMLIN. We then identified mutations that were negatively correlated with the WT pair for testing experimentally.

1.7.2 Mutant protein expression

Mutant I53-50A trimers were expressed at three scales. Small-scale expressions were performed at 1 mL culture volume in 96-well plates with 2 mL well volume. Medium-scale expressions were performed at 50 mL culture volume in 250 mL baffled shake flasks. Large-scale expressions were performed at 500 mL culture volumes in 2 L baffled shake flasks. All proteins were expressed in T7 competent *E. coli* in TB media, with IPTG induction for 3 hours at 37 °C. Cells were pelleted and frozen at -20 °C until lysis. Prior to lysis cells were defrosted on ice in lysis buffer (50 mM Tris pH 8.0, 250 mM NaCl, 20 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM dithiothreitol (DTT), 0.1 mg/mL DNase, and 0.1 μM RNase, unless otherwise noted). Small-scale expressions were lysed with a plate sonicator (QSonica), medium-scale expressions were lysed with a probe sonicator, and large-scale expressions by microfluidization (18,000 psi, one pass). Lysates from small-scale expressions were clarified by centrifugation in a swinging bucket rotor at 4,000 rcf. Lysates from medium- and large-scale expression lysates were clarified by centrifugation at 12,000 rcf in a fixed-angle rotor.

1.7.3 I53-50B expression and purification

Pentameric I53-50B was produced recombinantly in *E. coli*. A pET29b expression plasmid encoding I53-50B.4PT1 [27] was synthesized by GenScript using the NdeI and XhoI restriction sites with a double stop codon just before the C-terminal polyhistidine tag. Tagless protein was expressed in Lemo21(DE3) cells (NEB) in LB (10 g Tryptone, 5 g Yeast Extract, 10 g NaCl) grown in a 10 L BioFlo 320 Fermenter (Eppendorf). At inoculation, impeller speed was set to 225 rpm, SPLM set to 5 with O_2 supplementation as part of the dissolved-oxygen aeration cascade, and the temperature set to 37 °C. At the onset of a DO spike (OD ~12), the culture was fed with a bolus addition of 100 mL of 100% glycerol and induced with 1 mM IPTG. During this time, the culture temperature was reduced to 18 °C and O_2 supplementation was ceased, with expression continuing until OD reached ~20. The culture was harvested by centrifugation and the protein was purified from inclusion bodies. First, pellets were resuspended in PBS, homogenized, and then lysed by microfluidization using a Microfluidics M110P at 18,000 psi. Following sample clarification by centrifugation (24,000 g for 30 min), the supernatant was discarded and protein was extracted from the pellet using a series of three washes. The first wash consisted of PBS, 0.1% Triton X-100, pH 8.0. The second wash consisted of PBS, 1 M NaCl, pH 8.0, and the final wash (extraction) consisted of PBS, 2 M urea, 0.75% CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate), pH 8.0. Following extraction, the sample was applied to a DEAE Sepharose FF column (Cytiva) on an AKTA Avant150 FPLC system (Cytiva). After sample binding, the column was washed with 5 column volumes of PBS at pH 8.0 with 0.1% Triton X-100, followed by a wash with 5 column volumes of PBS at pH 8.0 with 0.75% CHAPS. The protein was eluted with 3 column volumes of PBS at pH 8.0 with 500 mM NaCl. After purification, fractions were pooled and concentrated in 10K MWCO centrifugal filters (Millipore), sterile filtered (0.22 µm), aliquoted and flash frozen in liquid nitrogen, and stored at -80 °C until use.

1.7.4 Assembly competency analysis

Single mutations were introduced into the I53-50A trimer [27] by QuikChange™ site-directed mutagenesis. Sequence-verified mutants were expressed at small scale (see above). Clarified lysates were separated from pellets and a 5 µL aliquot set aside for characterization by SDS-PAGE. The pellet was resuspended in lysis buffer and a 5 µL aliquot set aside for characterization by SDS-PAGE. Clarified lysate was immediately mixed with purified I53-50B.4PT1 pentamer. Because trimer expression levels varied from mutant to mutant, pentamer was added at three different concentrations. To 10 µL lysate, 7.5, 2.5, or 0 µL lysis buffer was added, followed by 2.5, 7.5, or 10 µL I53-50B.4PT1 pentamer at 1.8 mg/mL. The assembly reaction was allowed to proceed for 30 minutes at room temperature. A control assembly of previously purified I53-50A trimer and I53-50B.4PT1 pentamer at a 1:1 molar ratio was performed in parallel and included on all native PAGE gels. A 10 µL aliquot of each assembly reaction was mixed 1:1 with Native Sample Buffer (Bio-Rad Laboratories), loaded into precast 4-15% polyacrylamide gels (Bio-Rad Laboratories), and run with 1× Tris-Glycine Native PAGE buffer for 3 hours at 200 V. The gel was stained with GelCode Blue™ (Thermo Fisher Scientific) and destained in water. The lack of an I53-50 nanoparticle band on the native gel indicated single mutations that disrupted either trimer formation or trimer geometry such that the mutant trimer was no longer assembly competent.

1.7.5 Screening of mutant combinations

Single mutants that disrupted I53-50A trimer—and therefore I53-50 nanoparticle—formation were then combined with “rescue” mutations intended to generate pseudosymmetric I53-50A trimers. Synthetic DNA encoding potential combinations were ordered as heterotrimeric operons cloned into pCDB179 from IDT. To facilitate detection of the distinct components of the heterotrimer, a 6×His-SUMO domain was added to one subunit and sfGFP and an avi-tag added to a second subunit via genetic fusion. The third subunit bore a Strep-tag via genetic fusion. Variants were tested for I53-50 nanoparticle formation using trimer-containing *E. coli* lysates and purified I53-50B pentamer as described above. Combinations that formed I53-50 nanoparticles were expressed at large scale (see above) and purified by Ni^{2+} affinity chromatography on a HisTrap FF column (Cytiva). Briefly, clarified lysate was passed through a pre-equilibrated 5 mL HisTrap FF column, washed with 3-5 column volumes of wash buffer (50 mM Tris pH 8.0, 250 mM NaCl, 20 mM imidazole, 1 mM DTT), and heterotrimer was eluted with either a step elution or a gradient over 40 minutes at 3 mL/min flow rate into 100% elution buffer (50 mM Tris pH 8.0, 250 mM NaCl, 500 mM imidazole, 1 mM DTT). Major fractions corresponding to the two observed peaks in the elution profile were pooled seperately, concentrated in a 30 kDa cutoff Amicon concentrator (Millipore), and injected onto a pre-equilibrated Superdex 200 Increase 10/300 column (Cytiva). The SEC buffer was 25 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT. Fractions corresponding to the trimer peak from each chromatogram were collected for analysis by native mass spectrometry (MS). Alternatively the IMAC eluate was pooled and loaded onto a StrepTrap HP column (Cytiva) pre-equilibrated in binding buffer (100 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT). The column was then washed with 10 column volumes of binding buffer, or until the A280 absorbance leveled off at baseline and eluted with a step elution in binding buffer plus 2.5 mM desthiobiotin. Major fractions were analyzed by reducing SDS-PAGE. Westerns were performed on select fractions with either an anti-Avi or anti-Strep primary antibody.

1.7.6 Native mass spectrometry

Trimer purity, identity, and oligomeric state were analyzed by on-line buffer exchange MS [115] in 200 mM ammonium acetate using a Vanish ultra-performance liquid chromatography coupled to a Q Exactive ultra-high mass range Orbitrap mass spectrometer (Thermo Fisher Scientific). The recorded mass spectra were deconvolved with UniDec version 4.2+ [116].

1.7.8 Assembly of I53-50 nanoparticles using pseudosymmetric I53-50A heterotrimer

The native MS-verified pseudosymmetric I53-50A heterotrimer was expressed and purified at medium scale as described above and mixed at a 1:1 molar ratio with purified I53-50B.4PT1 pentamer and allowed to assemble at room temperature for 30 minutes. Assembled nanoparticles were characterized by DLS and negative-stain EM as described below.

1.7.9 Computational design of T=4 nanoparticles

We created a model of the pentasymmetron by extracting five trimers surrounding the icosahedral five-fold from I3-01 [23]. We reverted the interface residues on the unpaired subunit back to the original 1WA3 sequence, mutated 12 residues to negatively charged amino acids to enhance expression and facilitate purification, then combined each trimer into a single chain so that the pentasymmetron could be treated computationally as a simple homopentamer. We used previously described protocols [24,27] to dock and design T=4 nanoparticles. Docked configurations were manually screened to ensure interfaces were between the unpaired pentasymmetron subunit and the homotrimer. Designs were visually inspected and any overly exposed hydrophobic residues introduced during design were reverted to their WT identities.

1.7.10 Screening of T=4 nanoparticles by co-purification

Three tricistronic genes were ordered from IDT. An N-terminal GFP was included on the A subunit of the pentasymmetron heterotrimer as a mass tag. A C-terminal 6x-histidine tag was added to the C subunit. Genes were expressed at medium scale (see above). Clarified lysate was loaded onto 1 mL of Ni-NTA resin (Thermo Fisher Scientific) pre-equilibrated in wash buffer. After washing with 3 column volumes of wash buffer, the protein was eluted with 2 column volumes of elution buffer. Eluate was screened for the presence of all three gene products by SDS-PAGE.

1.7.11 Purification of co-expressed GI_4-F7

GI_4-F7 nanoparticles expressed tricistronically at large scale were purified by loading on a 5 mL HisTrap FF column (Cytiva) equilibrated in wash buffer (50 mM Tris pH 8.0, 250 mM NaCl, 20 mM imidazole, 1 mM DTT). After loading, the column was washed with 3-5 column volumes of wash buffer and protein was eluted with a gradient into 100% elution buffer (50 mM Tris pH 8.0, 250 mM NaCl, 500 mM imidazole, 1 mM DTT) over 40 minutes at 3 mL/min. The major fractions from elution were pooled, concentrated to ~1 mL, and loaded onto an equilibrated Sephacryl S-500 HR 10/300 GL. SEC buffer was 25 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT.

1.7.12 Purification of GI_4-F7 heterotrimeric and homotrimeric components

For *in vitro* assembly, the heterotrimeric component of GI_4-F7, comprising only the A and B chains, was expressed bicistronically. The A chain was modified with an N-terminal 6x-histidine tag. When expressed this way, some AAA nanoparticles and BBB homotrimers likely assemble in addition to AAB and ABB heterotrimers. To purify AAB from ABB heterotrimers, the bicistronic gene was expressed at large scale with the modification that 0.75% CHAPS was added to the lysis buffer and DTT was omitted. Clarified lysate was purified with a 5 mL HisTrap FF column as described above. Elution chromatograms contained three peaks. The first peak was predominantly the ABB heterotrimer, the second peak was predominantly AAB heterotrimer, and the third peak was predominantly AAA homotrimers assembled into an I3-01 like particle, A_{60}. Any BBB homotrimer would be in the flow-through. The first and second peaks were pooled separately and concentrated to ~1 mL. To remove any residual A_{60} we further purified the concentrated fractions on a Superose 6 Increase 10/300 column. The SEC buffer was 25 mM Tris pH 8.0, 150 mM NaCl, 0.75% w/v CHAPS. Glycerol was added to purified heterotrimer to a final concentration of 5%, the concentration was determined by A280, and 1 mL aliquots were flash-frozen in liquid nitrogen. Aliquots were stored at -80 °C until use. The homotrimer components were expressed at large scale and purified by IMAC in the same way as the co-expressed GI_4-F7 nanoparticles except that 1% CHAPS was added to all buffers. It was further purified by SEC on a HiLoad 26/600 Superdex 200 PG column in 25 mM Tris pH 8.0, 150 mM NaCl, 5% Glycerol, 1.0% w/v CHAPS, 1 mM DTT. The total trimer protein concentration was measured by A280, flash-frozen in liquid nitrogen in 1 mL aliquots, and stored at -80 °C until use.

1.7.13 *In vitro* assembly of GI_T-F7 nanoparticles

To assemble GI_T-F7 nanoparticles, components were mixed at various stoichiometries depending on the target assembly state, and dialyzed into 0% CHAPS overnight at room temperature in a 30 kDa cutoff dialysis cassette. Because the AAB and ABB heterotrimers do not directly interact with the BBB homotrimer, those components were mixed first, followed by addition of the CCC homotrimer. Nanoparticles were prepared fresh for each experiment, or stored at 4 °C for up to three days. To assemble BBB-CCC 2D arrays, the components were first individually dialyzed to remove CHAPS, and then mixed at a 1:1 stoichiometric ratio and allowed to assemble overnight at room temperature.

1.7.14 Characterization of assemblies

Assemblies were characterized in solution by dynamic light scattering (DLS). Samples were measured using an UNcle (UNchained Labs) according to the manufacturer's directions. Briefly, 8.8 μL of sample was loaded in triplicate into the capillary cassette. For each replicate, 10 acquisitions 10 seconds in length were collected. Assemblies were further characterized by negative-stain electron microscopy. Samples were diluted to between 0.1 and 0.5 mg/mL total protein depending on the assembly stoichiometry, applied to a glow-discharged thick carbon film 400 mesh copper grid (Electron Microscopy Sciences), and stained with 2% uranyl formate. Care was taken to ensure the stain thickness was sufficient to support the larger assemblies. Micrographs were collected on a Talos L120C (FEI) at up to 48,000× magnification. Individual micrographs were processed with ImageJ.

1.7.15 CryoEM sample preparation, data collection and data processing

Three microliters of 3 mg/mL GI_4-F7, GI_9-F7, and GI_{16}-F16 were loaded onto freshly glow discharged R 2/2 UltrAuFoil grids, prior to plunge freezing using a Vitrobot Mark IV (ThermoFisher Scientific) with a blot force of 0 and 6 sec blot time at 100% humidity and 22°C. Data were acquired using an FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K3 direct detector and Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV. For GI_4-F7 and GI_9-F7, automated data collection was carried out using Leginon [117] at a nominal magnification of 105,000× with a pixel size of 0.843 Å. 7,249 and 2,558 micrographs were collected with a defocus range comprised between -0.5 and -2.5 μm, respectively. The dose rate was adjusted to 15 counts/pixel/s, and each movie was acquired in super-resolution mode fractionated in 75 frames of 40 ms. For the GI_{16}-F7 data set, automated data collection was carried out using Leginon [117] at a nominal magnification of 64,000× with a pixel size of 1.42 Å. 2,268 micrographs were collected with a defocus range between -0.5 and -3.5 μm. The dose rate was adjusted to 15 counts/pixel/s, and each movie was acquired in super-resolution mode fractionated in 50 frames of 100 ms. Movie frame alignment, estimation of the microscope contrast-transfer function parameters, particle picking, and extraction were carried out using Warp [118].

Two rounds of reference-free 2D classification were performed using CryoSPARC [119] to select well-defined particle images. These selected particles were subjected to two rounds of 3D classification with 50 iterations each (angular sampling 7.5° for 25 iterations and 1.8° with local search for 25 iterations) using Relion [120]with an initial model generated with ab-initio reconstruction in cryoSPARC. 3D refinements were carried out using non-uniform refinement along with per-particle defocus refinement in CryoSPARC. Selected particle images were subjected to the Bayesian polishing procedure [121] implemented in Relion 3.1 before performing another round of non-uniform refinement in cryoSPARC followed by per-particle defocus refinement and again non-uniform refinement. To further improve the density of the ASU, the particles were symmetry-expanded and subjected to focus 3D classification without refining angles and shifts. Particles belonging to classes with the best resolved ASU density were selected and then subjected to local refinement using CryoSPARC. Local resolution estimation,and sharpening were carried out using CryoSPARC. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) of 0.143 criterion and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution [122,123].

1.7.16 Model building and refinement

UCSF Chimera [124] and Coot [125] were used to fit atomic models into the cryoEM maps. GI_4-F7 and GI_9-F7 asu models were refined and relaxed using Rosetta using sharpened and unsharpened maps [126,127]. For GI_4-F7 or GI_9-F7 icosahedral model, all of the side chains of GI_4-F7 or GI_9-F7 asu model are truncated except Gly, Cys and Pro residues and the symmetry related copies were generated in ChimeraX with Cryo EM maps.

1.7.17 Alignments and images

To align the cryoEM models to the design model, both models were centered at the origin and their icosahedral symmetry axes aligned in PyMOL [128]. The Cα RMSD was calculated using the “rms_cur” function in PyMOL. To measure deviations in the rigid body degrees of freedom, copies of the pentasymmetron, disymmetron, and trimer (or trimers for GI_9-F7) from the cryoEM model were aligned to the design model using the “super” function in PyMOL. We then calculated the rotations and translations from the transformation matrix between the corresponding component of the original cryoEM model and the aligned cryoEM model. We applied the same approach to the heterotrimer (and homotrimer for GI_9-F7) components to obtain rotations and translations within the pentasymmetron, disymmetron, and homotrimer components, respectively. We found that the “super” function in PyMOL was very sensitive to chain and residue numbering, as well as some of the minor differences between the design model and cryoEM model. Therefore, for all alignments using PyMOL, we made sure to harmonize residue numbering, chain IDs, and remove any residues present in only one model or the other. For that reason, aligned images were generated using the “mm” command in ChimeraX [129] and verified to ensure that the alignments closely matched those generated on the trimmed models created with the “super” function in PyMOL.

1.7.18 Scripts and plots

All data was processed and plotted using python 3.8.8, Matplotlib 3.3.4, and seaborn 0.11.1.

1.8 Supplemental Figures for pseudosymmetric protein nanoparticle design

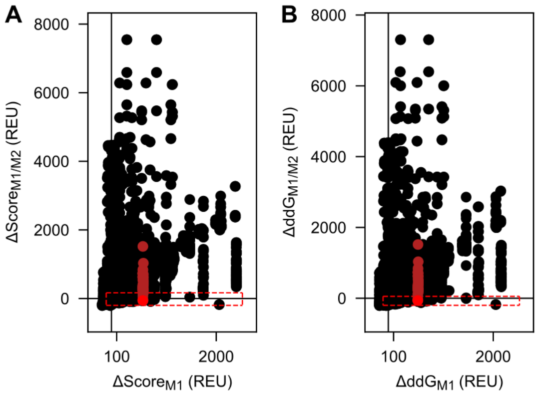


Fig. S1: Rosetta metrics for all single and pairwise 1WA3 trimer interface mutants. (**A**) ΔddG filter metric. (**B**) ΔScore metric. Dark red points correspond to the single mutation P114F. The bright red point corresponds to the double mutant P114F/F131V. The red dotted boxes represent cutoffs used to select mutants for testing.

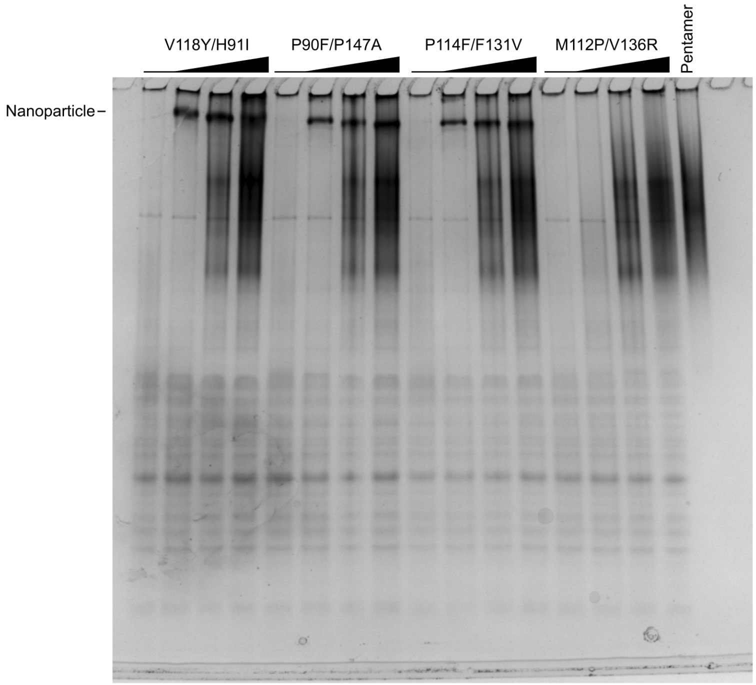


Fig. S2: Native-PAGE screen of double mutants. Recovery of trimer geometry was assayed by assembling double mutant I53-50A trimers in clarified *E. coli* lysates with purified I53-50B pentamer and evaluating the presence or absence of I53-50 nanoparticles by native PAGE. Black wedges indicate increasing pentamer concentration in each series of assembly reactions.

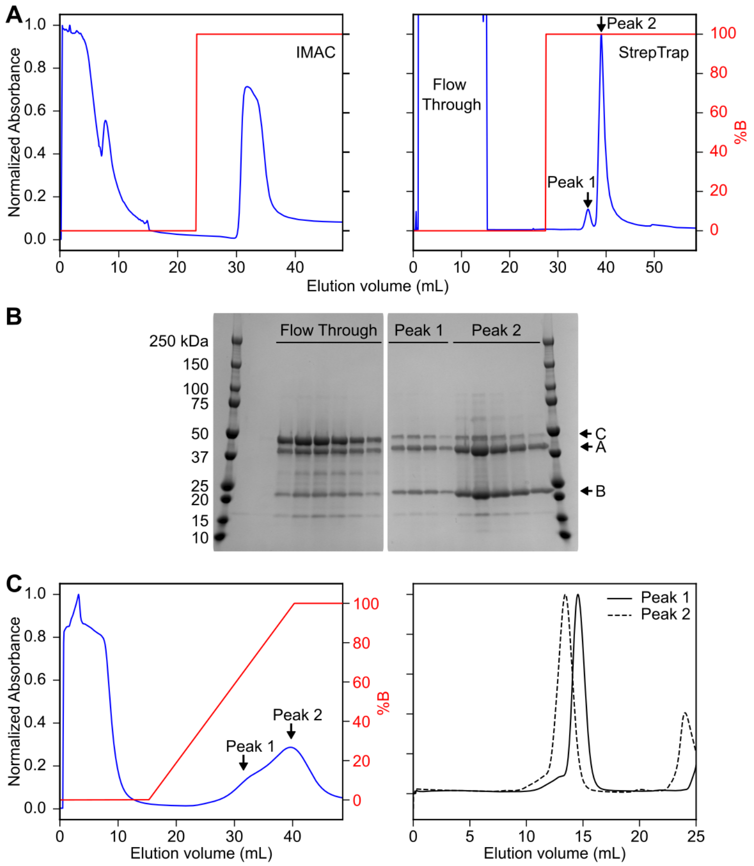


Fig. S3: Purification and characterization of "ABC" tricistronic and "AB" bicistronic constructs. (**A**) The ABC heterotrimer was purified by (*left*) IMAC with a step elution followed by (*right*) StrepTrap purification. The A chain contained a hexa-histidine and SUMO tag, the B chain contained a Strep tag, and the C chain contained sfGFP and avi tags. The eluate of this two-step purification method should therefore only contain trimers that include both the A and B chains. An optimal result would be equimolar amounts of the A, B, and C chains. (**B**) SDS-PAGE of the StrepTrap purification revealed that the eluate contained an excess of the A and B chains and less of the C chain. To test the ability of the A and B chains only to assemble into heterotrimers, we expressed an AB bicistronic gene and (**C**) purified the resulting proteins by (*left*) IMAC with a gradient elution. Two broad and overlapping peaks were observed. The leading half of the first peak and trailing half of the second peak were collected and (*right*) further purified by SEC. Peak 2 has a lower retention volume than peak 1, suggesting a difference in molecular weight. These results are consistent with assembly of an ABB heterotrimer (earlier IMAC elution, later SEC elution) and an AAB heterotrimer (later IMAC elution, earlier SEC elution). We confirmed this interpretation by native mass spectrometry (**Fig. 1G**)

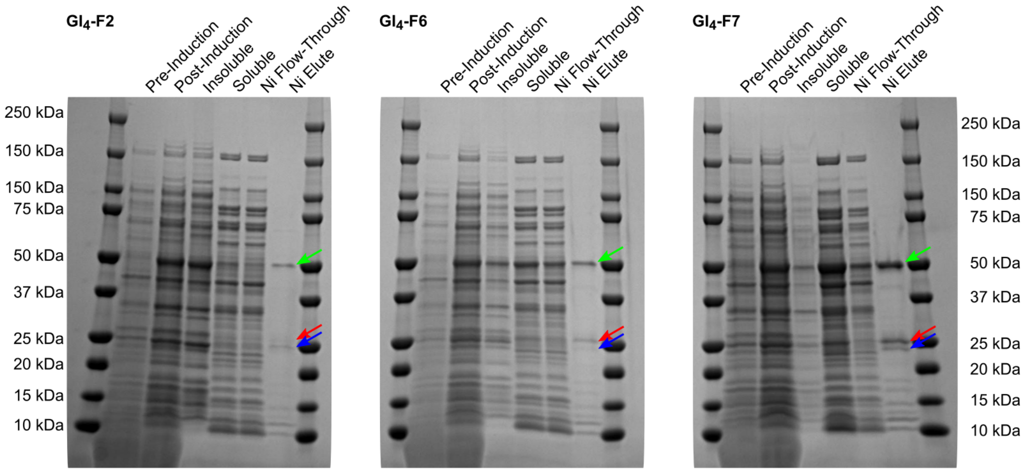


Fig. S4: SDS-PAGE of GI$_4$ designs by Ni^{2+} pull-down assay. Expression and screening by SDS-PAGE for GI$_4$ designs. Bands for chains A (green arrow), B (blue arrow), and C (red) arrow are indicated. The presence of all three bands in the Ni Elute lanes of GI$_4$-F6 and GI$_4$-F7 indicates interactions between the A, B, and C chains.

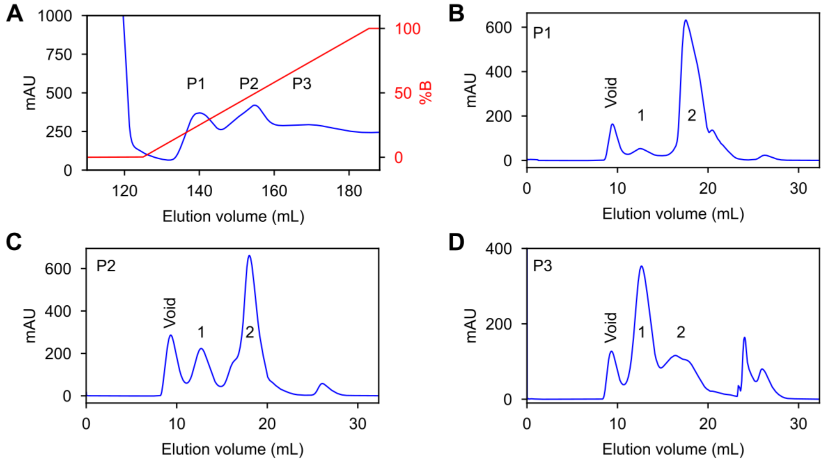


Fig. S5: Purification of heterotrimer components from $\mathbf{A_{60}}$. (**A**) HisTrap elution chromatogram. Blue, absorbance at 280 nm; red, gradient elution. Peak 1 (P1) is predominantly ABB, P2 is predominantly AAB, and P3 is predominantly the A chain, which assembles into 60-subunit I3-01-like nanoparticles. (**B**) Superdex 200 Increase 10/300 chromatogram of P1 from the HisTrap elution. The first peak following the void volume (1) is predominantly I3-01-like nanoparticles and (2) is predominantly ABB heterotrimer. (**C**) Superdex 200 Increase 10/300 chromatogram of P2 from the HisTrap elution. (1) is predominantly I3-01-like nanoparticles and (2) is predominantly AAB heterotrimer. (**D**) Superdex 200 Increase 10/300 chromatogram of P3 from the HisTrap elution. (1) is predominantly I3-01-like nanoparticles and (2) is predominantly AAB heterotrimer.

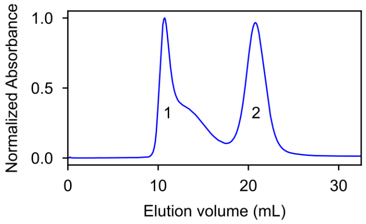


Fig. S6: Purification of GI$_4$-F7 by SEC. SEC purification of GI$_4$-F7 on a Sephacryl S-500 HR 10/300 GL column. Peak 1 contains the assembly while peak 2 is residual homotrimer component.

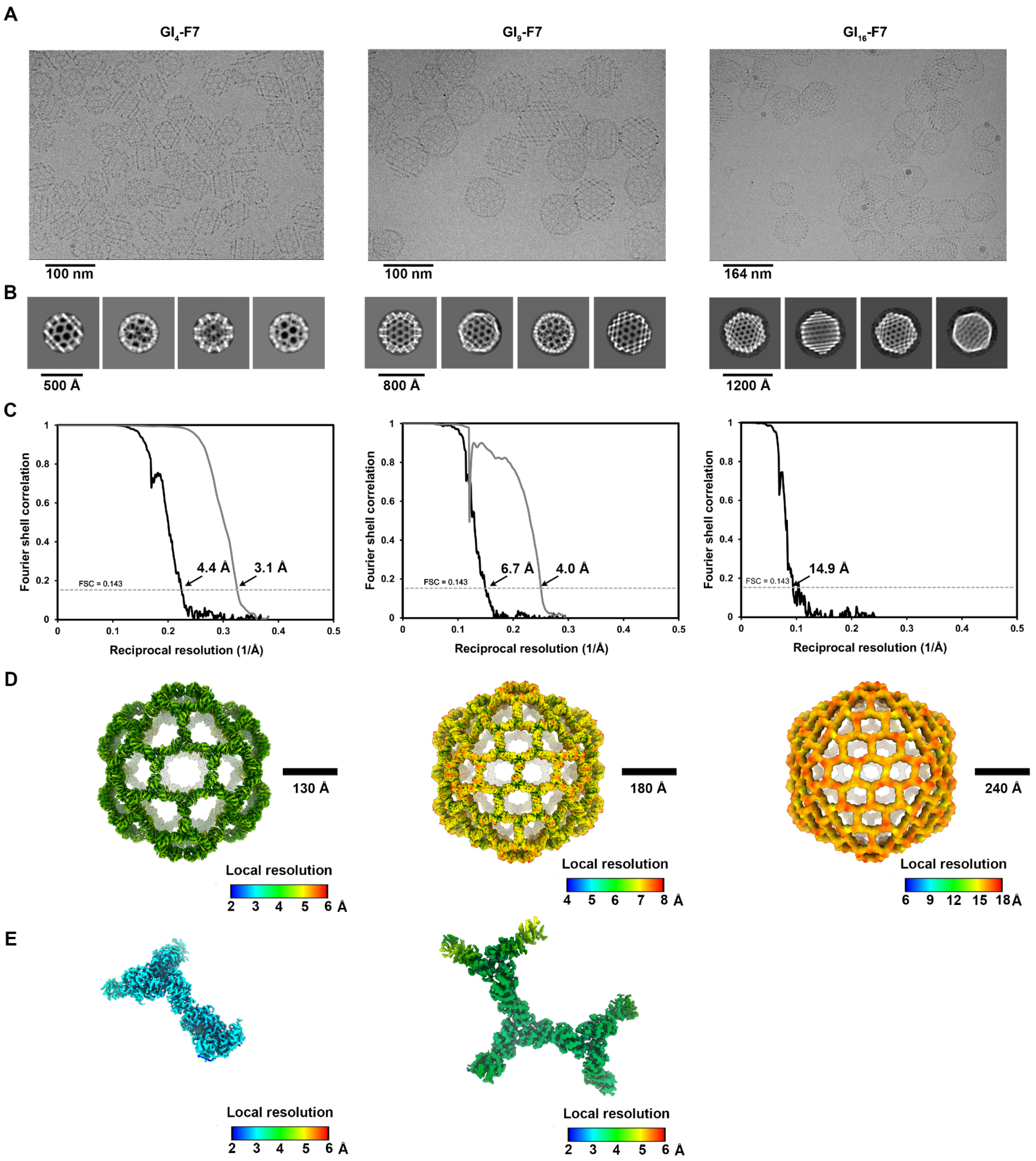


Fig. S7: CryoEM data processing. (**A-B**) Representative electron micrographs (A) and 2D class averages (B) of GI_4-F7 (left), GI_9-F7, (middle) and GI_{16}-F7 (right). (**C**) Gold-standard Fourier shell correlation curves for the 3D reconstructions of GI_4-F7 (left), GI_9-F7 (middle) and GI_{16}-F7 (right) (black line) and locally refined asus (gray lines). (**D-E**) Local resolution maps calculated using cryoSPARC for (D) the 3D reconstructions of GI_4-F7 (left), GI_9-F7 (middle), and

GI_{16}-F7 (right) as well as (E) the locally refined ASUs of GI_4-F7 (left) and GI_9-F7 (middle). The 0.143 cutoff is indicated by a horizontal dashed line.

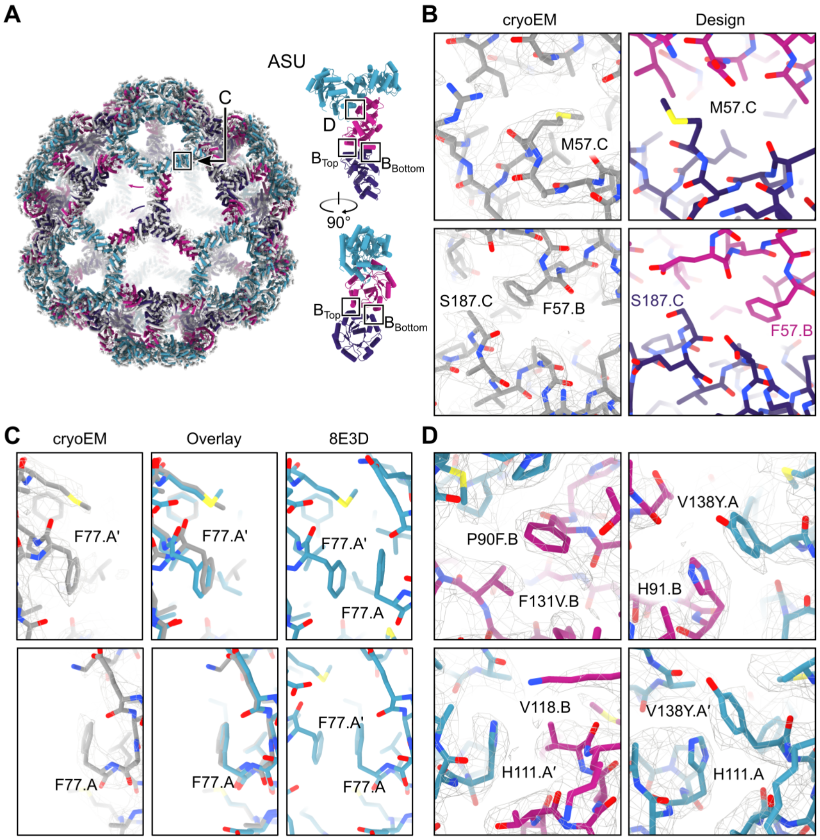


Fig S8: Structural details of GI$_4$-F7. (**A**) Alignment of the complete cryoEM model to the design model. Major rigid-body DoF deviations are indicated with arrows. Two views of the ASU are shown. Approximate locations of each inset (B, C, and D) are indicated. (**B**) Comparison between the cryoEM model (left) and design model (right) of the newly designed nanoparticle (B-C) interface. *Top row*, M57 on the CCC-homotrimer changes rotamer to occupy a void in the interface in the design model. *Bottom row*, F57 on the B chain of the AAB heterotrimer packs against S187 of the CCC homotrimer in the cryoEM model, instead of A190 in the CCC homotrimer as in the design model. (**C**) Comparison of the I3-01 (A-A) interface observed in the cryoEM model to a previously published structure (PDB ID 8ED3). *Top row*, slight rigid-body deviations from perfect two-fold symmetry in one copy of the A chain. *Bottom row,* very little deviation from perfect two-fold symmetry (**D**) Details of the density maps in the regions of the pseudosymmetry-generating mutations within the AAB heterotrimer interface.

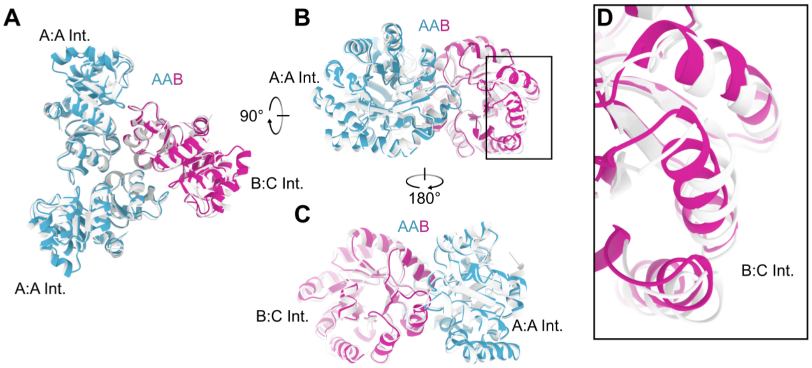


Figure S9: Comparison of the AAB heterotrimer design model to the cryoEM model from GI$_4$-F7. Alignment of the AAB heterotrimer cryoEM model to the design model is viewed (**A**) from the top, towards the center of the nanoparticle along the three-fold symmetry axis; (**B**) from the side, tangential to the nanoparticle surface; and (**C**) from the other side, tangential to the nanoparticle surface. The position of the A:A and newly designed B:C interfaces are indicated. (**D**) Detail of the B side of the B:C interface, highlighting the most significant deviations from the design model.

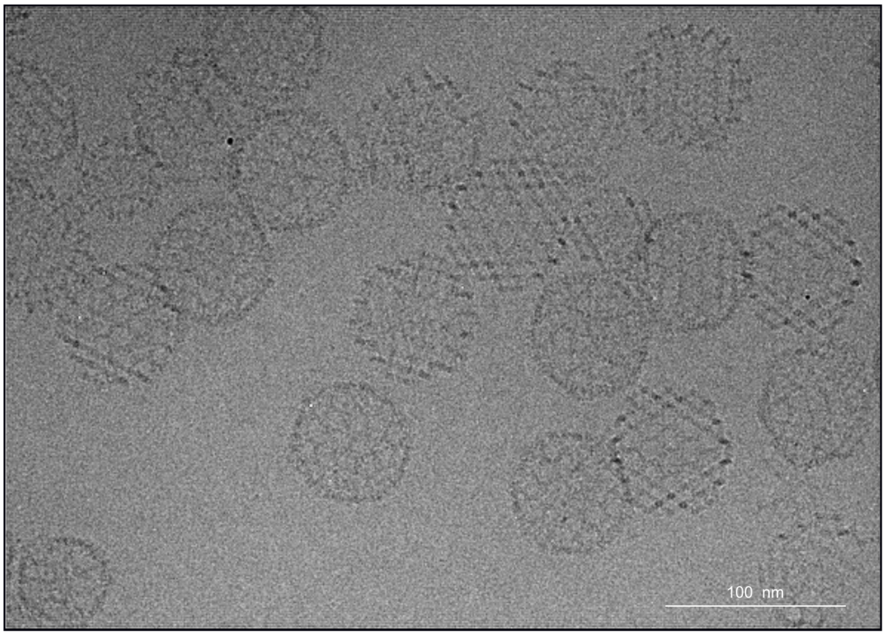


Figure S10: Field view micrograph of GI$_9$-F7. CryoEM field view micrograph of samples enriched for GI$_9$-F7 by SEC purification. Both GI$_9$-F7 (large particles) and GI$_4$-F7 (e.g., bottom-left corner) are clearly visible.

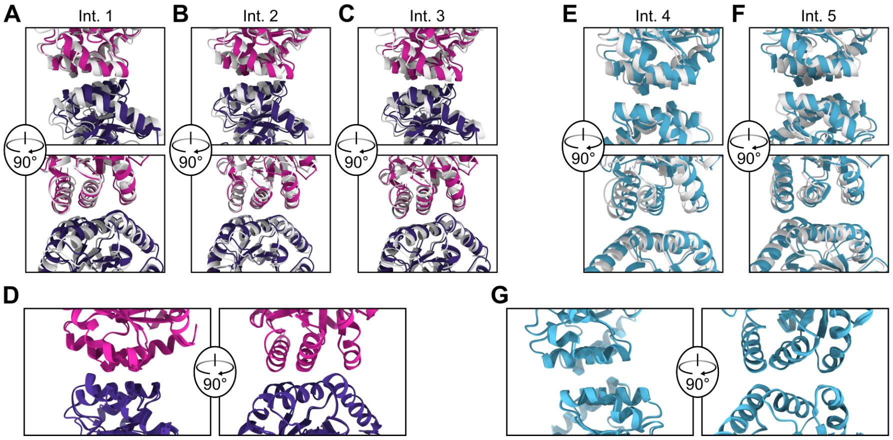


Fig. S11: Structural details of GI$_9$-F7. (**A-C**) Alignment of the GI$_9$-F7 design model chain B (magenta) and chain C (purple) protein-protein interface to the corresponding chains of the cryoEM model (gray). Each of the three interfaces between B and C chains in the ASU are shown. (**D**) The protein-protein interface between chain B and C from the cryoEM model of GI$_4$-F7 (light colors) aligned to the same interface from the cryoEM model of GI$_9$-F7 (dark colors). (**E**) Alignment of design model to the cryoEM model for the I3-01 interface in the pentasymmetron and (**F**) disymmetron. (**G**) Alignment of the I3-01 interface from the cryoEM models of GI$_4$-F7 (light blue) and GI$_9$-F7 (dark blue).

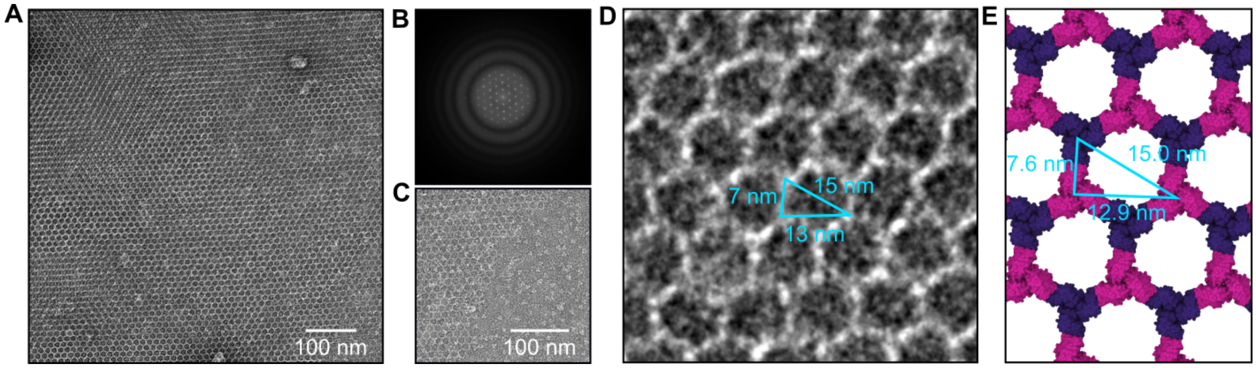


Fig S12: Hexagonal 2D array characterization by negative stain EM. (**A**) An example of the regular hexagonal array formed by mixing BBB and CCC homotrimers by negative stain EM. (**B**) Power spectrum of the micrograph shown in panel A, confirming the periodic arrangement of the array. (**C**) The edge of the array is jagged, with free trimeric components visible. (**D**) Measurement of the array dimensions are consistent with (**E**) the design model.

Supplemental Tables

Supplementary Table S1: Table of all single and double mutants and outcome.

Single mutant	Assembles in lysate?
R17Y	Yes
T69E	Yes
T69G	Yes
T71A	Yes
T71D	Yes
T71E	Yes
T71K	Yes
T71L	Yes
T71N	Yes
T71Y	Yes
P90E	Yes
P90F	**No**
P90K	Yes
P90Q	Yes
P90R	Yes
H91C	Yes
H91D	Yes
H91F	Yes
H91G	Yes
H91I	Yes
H91S	Yes
D93R	Yes
M112P	**No**
M112Q	Yes
M112V	Yes
T113D	Yes
T113F	Yes
T113I	Yes
T113L	Yes
T113M	Yes
T113P	Yes
T113Q	Yes
T113V	Yes
P114Q	Yes
P114S	Yes
P114F	**No**
T115K	Yes
T116C	Yes
T116D	Yes
T116G	Yes
T116L	Yes
T116M	Yes
T116N	Yes
T116Q	Yes
T116S	Yes
T116T	Yes
T116W*	**No**
V118H	Yes
V118K	Yes
V118M	Yes
V118N	Yes
V118S	Yes
V118W	Yes
V118Y	**No**
K122D	Yes
K122E	Yes
F131E	Yes**
E134G	Yes
E134K	Yes
V135P	Yes
V136R	Yes
V136Y	Yes
V136W	Yes
Q139E	Yes
F140K	**No**
F140H	Yes
F140W	Yes
A143D	Yes
A143E	Yes
A143Y	Yes
P147A	Yes
P147D	Yes
P147H	Yes
P147M	Yes
P147R	Yes**
P147S	Yes
F148C	Yes
F148E	Yes
F148H	Yes
F148Q	Yes
F148Y	Yes

Double mutant	Assembles in lysate?
M112P/V136R	No
P90F/P147A	Yes
P114F/F131V	Yes
V118Y/H91I	Yes

* Mutant at three-fold position required screening in a tricistronic gene construct and was not pursued further.

** Did not assemble in one replicate or had a weak assembly band.

Supplementary Table S2. Amino acid sequences for novel proteins used in this study.

>Pseudosymmetric I53-50A A Chain
MGHHHHHHHHHHGSLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAK
RQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGGSEKAAKAEEAARKMEELFKKHKIVA
VLRANSVEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEF
IVSPHLDEEISQFCKEKGVFYMPGVMTPTELYKAMKLGHDILKLFPGEVVGPQFVKAMKGPFPNVKFVP
TGGVNLDNVCKWFKAGVLAVGVGKALVKGKPDEVREKAKKFVKKIRGCTE

>Pseudosymmetric I53-50A B Chain
MKMEELFKEHKIVAVLRANSVEEAISKALAVFAGGVHLIEITFTVPDADQVIKELEFLKEAGAIIGAGTVTS
VEQCREAVESGAEFIVSFHLDEEISQFCKEEGVFYMPGVMTPTELVKAMKLGHTILKL**V**PGEVVGPQFV
EAMKGPFPNVKFVPTGGVNLDNVCEWFEAGVLAVGVGSALVEGEPAEVAELAIRFVEKIRGCTEGSG
WSHPQFEK

>Pseudosymmetric I53-50A C Chain
MKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC
FARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH
KLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLS
KDPNEKRDHMVLLEFVTAAGITHGMDELYKGGSGGSGGKMEELFKKHKIVAVLRANSVEEAIEKAVAVF
AGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPILDEEISQFCKEKG
VFYMPGVMTFTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGAFPNVKFVPTGGVNLDNVCEWFKAG
VLAVGVGSALVKGTPDEVREKAKAFVEKIRGCTEGSGLNDIFEAQKIEWHE

>Pseudosymmetric GI$_T$ A Chain
MGSHHHHHHGSEKAAKAEEAARKMEELFKEHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDA
DTVIKELSFLKEMGAIIGAGTVTSVEQCREAVESGAEFIVSPHLDEEISQFCKEEGVFYMPGVMTPTELY
KAMKLGHTILKLFPGEVVGPQFVEAMKGPFPNVKFVPTGGVNLDNVCEWFEAGVLAVGVGSALVEGT
PVEVAEKAKAFVEKIEGCTE

>Pseudosymmetric GI$_T$ B Chain
MKMEELFKEHKIVAVLRANSVEEAISKALAVFAGGVHLIEITFTVPDADQVIKELEFLKEAGAIIGAGTVTS
VEQCREAVESGAEFIVSFHLDEEISQFCKEEGVFYMPGVMTPTELVKAMKLGHTILKL**V**PGEVVGPQFV
EAMKGPFPNVKFVPTGGVNLDNVCEWFEAGVLAVGVGSALVEGEPAEVAELAIRFVEKIRGCTE

>Homotrimeric GI$_T$ BBB
MKMEELFKEHKIVAVLRANSVEEAISKALAVFAGGVHLIEITFTVPDADQVIKELEFLKEAGAIIGAGTVTS
VEQCREAVESGAEFIVSPHLDEEISQFCKEEGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFV
EAMKGPFPNVKFVPTGGVNLDNVCEWFEAGVLAVGVGSALVEGEPAEVAELAIRFVEKIRGCTELEHH
HHHH

>Homotrimeric GI$_T$ CCC
MKMEELFKEHKIVAVLRANSREEAIEIALAVFAGGVHLIEITFTVPDADEVIKRLEMLKRAGAIIGAGTVTS
VEQCREAVESGAEFIVSPHLDEEISQFCKEEGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFV
EAMKGPFPNVKFVPTGGVNLDNVCEWFEAGVLAVGVGSALVEGKPSEVAEKARRFVKKIRGCTEGSL
EHHHHHH

Appended sequences including SUMO, GFP, and deca- or hexa-histidine, avi- and strep- tags are underlined.

Pseudosymmetrizing mutations are highlighted.

Supplementary Table S3. CryoEM data collection and refinement statistics.

	GI_4-F7	GI_4-F7 (local refinement)	GI_9-F7	GI_9-F7 (local refinement)	GI_{16}-F7
Data collection and processing					
Magnification	105,000	105,000	105,000	105,000	64,000
Voltage (kV)	300	300	300	300	300
Electron exposure ($e^{-}/Å^2$)	60	60	60	60	37
Defocus range (μm)	-0.5 - -2.5	-0.5 - -2.5	-0.5 - -2.5	-0.5 - -2.5	-0.5 - -3.5
Pixel size (Å)	0.843	0.843	0.843	0.843	1.42
Symmetry imposed	I	C1	I	C1	I
Final particle images (no.)	120,979	984,020	1,956	795,360	1,083
Map resolution (Å)	4.4	3.1	6.7	4.0	14.9
FSC threshold	0.143	0.143	0.143	0.143	0.143
Map sharpening *B*factor ($Å^2$)	-269	-143	-530	-138	
Validation					
MolProbity score		1.9		1.06	
Clashscore		5.86		0.71	
Poor rotamers (%)		2.94		0.13	
Ramachandran plot					
Favored (%)		96.52		98.95	
Allowed (%)		3.11		0.5	
Disallowed (%)		0.37		0.55	

Supplementary Table S4: Deviations observed in the cryoEM reconstruction of GI$_4$-F7 compared to the design model.

Component	Axis	Translation (Å)	Rotation (Degrees)
Pentasymmetron	5-fold	5.8	5.9
AAB-Heterotrimer	// Local 3-fold	1.1	1.3
	⊥ Local 3-fold	1.6	–
	AAB:AAB Interface	–	1.6, -2.8
	AAB:CCC Interface	–	-0.2
Trimer	3-fold	4.0	12.4

Supplementary Table S5: Deviations observed in the cryoEM reconstruction of GI$_9$-F7 compared to the design model.

Component	Axis	Translation (Å)	Rotation (Degrees)
Pentasymmetron AAB heterotrimer	Icosahedral 5-fold	7.7	7.2
	∥* Local 3-fold	1.3	1.8
	⊥ Local 3-fold	1.9	--
	AAB:AAB Interface	--	1.9, -3.6
	AAB:CCC Interface	--	0.3
Disymmetron ABB heterotrimer	Icosahedral two-fold	10.0	8.2
	∥ Local 3-fold	0.4	2.3
	⊥ Local 3-fold	1.3	--
	ABB:ABB Interface	--	-1.6
	ABB:CCC Interface	--	-0.8, 3.7
3× Trimer CCC homotrimer	Icosahedral 3-fold	6.5	5.4
	∥ Local 3-fold	0.0	-6.5
	⊥ Local 3-fold	1.23	--
	CCC:ABB Interface	--	0.3, -1.8
	CCC:AAB Interface	--	-0.6*

*The symbol ∥ indicates deviations parallel to the indicated symmetry axis, ⊥ indicates deviations perpendicular to the indicated axis.

Supplementary Table S6: DLS results obtained from assembly reactions corresponding to T numbers 4 to 100.

T	Mean Z-Ave. dia. (nm)	StdDev Z-Ave. dia. (nm)	PDI
4	47.5	0.4	0.048
9	69.9	0.5	0.154
16	95	1	0.121
25	113.2	0.5	0.127
36	129	2	0.134
49	148	2	0.136
64	162	1	0.136
81	174	1	0.126
100	189	1	0.133

Chapter 2. Computational design of constitutively active cGAS

The exceptionally large structures generated with pseudosymmetric proteins represent one end of the length scale of multistate proteins. In that case, independent states exist between protein molecules. Conversely, the multiple states in many enzymes are entirely within a single molecule. cGAS is an enzyme regulated by binding to dsDNA. In the unbound state, cGAS adopts a stable, inactive conformation. When bound to dsDNA adopts an open, active conformation. I set out to design a constitutively active cGAS variant by selectively destabilizing the inactive conformation and stabilizing the active conformation. This work was first published in Nature Structure and Molecular Biology [130], the author manuscript is reproduced here.

2.1 Introduction to cGAS

The ability to induce a controlled and robust innate immune response is highly desirable for prophylactic and therapeutic immunopotentiation. The cGAS-STING pathway [63,67–69,131,132] has rapidly become a promising target, with numerous small molecule STING agonists in development, several of which have entered clinical trials [133,134]. However, several first-generation molecules suffered from low efficacy or adverse reactions [133,135]. With the recent clinical validation of multiple approaches to delivering biologics genetically [136,137], biologic activators of the cGAS-STING pathway are now feasible and may provide an attractive alternative by offering more avenues to engineer activity, localization, and regulation. Highly active cGAS mutants that lack inhibition deriving from chromatin tethering have recently been reported [64]. However, these variants still require the dsDNA ligand for activation. We set out to develop a simple, multi-state computational design protocol to engineer constitutively active cGAS (CA-cGAS) variants that are ligand-independent. CA-cGAS could be a useful alternative therapeutic agent or research reagent because its activity would not depend on factors such as the cell cycle, subcellular localization, or disease state. The tumor regression caused by intratumoral injection of nuclease-resistant, modified STING-agonists [70–72] suggests a potential therapeutic application for CA-cGAS delivered as a biologic.

cGAS is the eponymous member of the cGAS-like receptors (cGLRs), a conserved group of metazoan proteins responsible for double stranded nucleic acid surveillance in the cytosol [138,139]. cGLRs are characterized by an extended N-terminal α-helix, designated the "spine," a mixed α/β nucleotidyltransferase fold domain (NTase core), and an all α-helical C-terminal domain [140]. cGAS recognizes and is allosterically activated by dsDNA in a sequence-independent manner (Fig. 1a), primarily by forming extensive polar contacts with the phosphate backbone [65]. Although somewhat independent of dsDNA length depending on species [65,66,68,141,142], activation occurs most efficiently when cGAS binds at least two strands of dsDNA in roughly parallel orientation, forming extended DNA-protein complexes [66,141,143,144]. One dsDNA strand makes extensive contacts along the spine helix, NTase core, and C-domain; while the other dsDNA strand only contacts the NTase core. Multiple structural studies have shown that recognition of dsDNA initiates three changes within the cGAS structure: 1) a break forms in the spine helix (residues 158–161 in mouse cGAS [mcGAS]), 2) the dynamic active site loop (residues 196–210) becomes significantly more ordered such that it no longer occludes the active site, and 3) the NTase-core domain moves approximately 3.8 Å closer to the C-domain [65,66]. Figure 1b shows a view of the active and inactive conformation with cylindrical helices to visually accentuate these changes. The cGAS-DNA complex also contains cGAS-cGAS contacts, and activation by short dsDNA fragments forms cGAS dimers [66,143]. The sequence of these events and the necessity of each for activation are not fully understood. In contrast to these rearrangements between DNA-free and DNA-bound cGAS, elegant crystallographic studies suggest no significant structural changes during catalysis [145]. These observations indicate that the DNA-bound conformation can be considered the enzymatically active conformation.

We hypothesized that mutations which i) stabilize the dsDNA-bound (active) conformation or ii) destabilize the unbound (inactive) conformation would shift the conformational equilibrium of the enzyme towards the active state, an outcome usually achieved by dsDNA binding. Unlike the rearrangements caused by dsDNA binding, the energetic effects of such mutations would be intrinsic to the designed protein, leading to constitutive, dsDNA-independent enzymatic activity (Fig. 1c). We developed a two-state design approach to identify such mutations and generate CA-cGAS variants. The typical protein design approach seeks to optimize the protein sequence for a single target structure, while disregarding or imperfectly approximating the vast space of all other possible states [77,146]. In contrast, multi-state design seeks to optimize the protein sequence for or against multiple conformations simultaneously. A generalized solution to multi-state protein design that contrasts multiple conformations has been developed, but is complex to use and computationally intensive [11]. Here we developed a simple, knowledge-based two-state design protocol that can be generally applied to stabilize specific conformations of dynamic proteins where target and off-target structures are known. We augment this approach with bioinformatics-inspired mutations derived from proteins with highly divergent sequences and functions that nevertheless retain strongly conserved structural features that closely match the target conformation.

2.2 Computational design of constitutively active cGAS

To identify key residues involved in the structural rearrangements between DNA-free and DNA bound conformations, we calculated a distance difference matrix between the DNA-bound (active; PDB ID 4K97) and -unbound (inactive; PDB ID 4K8V) forms of mcGAS. The matrix measures the change in distance for each residue-residue pair in the two structures. Our analysis highlighted two regions that move significantly and are likely critical for cGAS activation: the spine helix above the active site that bridges the NTase core and C-domain (region 1, residues 155–170), and the active site loop and flanking beta strands (region 2, residues 196–214) (Fig. 1b,d). The dimer interface does not undergo any significant structural rearrangements (Fig. 1d).

To identify potential activating mutations in these regions, we evaluated the energy of all possible single amino acid substitutions in both the active and inactive conformations using Rosetta (660 mutations total) [147]. Mutations were ranked by taking the difference in the energetic impact of each mutation between the active and inactive states, where the energetic impact is defined as the difference between the WT energy and single-mutant energy for a given state (Fig. 1e). This score metric is the central feature of our two-state design approach, as it explicitly favors mutations that stabilize the active conformation, destabilize the inactive conformation, or both. Accordingly, the top-scoring mutations generally fell into two categories: hydrophobic residues exposed to solvent in the inactive conformation and buried in the active conformation (e.g., T197V and Y200I), or proline residues compatible with the backbone geometry of the active conformation but unfavorable in the inactive conformation (e.g., R161P) (Fig. 1f). To generate candidate CA-cGAS designs, combinations of positions with favorable mutations were designed with Rosetta. Independent design trajectories either restricted the possible mutations at each position to those that scored well in the initial screen, or allowed them to mutate freely. We scored these variants in the same way as the individual mutations, and selected 45 designs containing 1–6 mutations for experimental characterization (Supplemental Table 1).

We screened the enzymatic activity of each cGAS variant in HEK293T cells, which do not naturally express cGAS, by transfecting each cGAS construct together with a STING expression vector and a luciferase-based IFN-stimulated response element (ISRE) reporter plasmid (Extended Data Fig. 1, a–d). To ensure that any measured activity was due to constitutive activation of the enzyme and not activation by transfected plasmid DNA, we first knocked out cGAS activity by disrupting DNA binding through introduction of the previously published K395A/K399A mutations [65] and a similar pair of mutations, K395M/K399M. We selected the K395M/K399M mutations for screening variants because they retain packing interactions provided by the aliphatic portion of the native side chains that may be important for stabilizing the nearby active site loop. We confirmed that the K395M/K399M double mutant was as inactive as K395A/K399A (Fig. 1g). The 45 best-scoring computationally designed variants were introduced into this K395M/K399M background. In the initial screen, the double mutant K160D/R161P (CA-cGAS-04) exhibited activity higher than the K395M/K399M mutant, but lower than wild-type (WT) cGAS (Fig. 1g). These mutations were intended to destabilize the inactive conformation by breaking the spine helix (R161P) and stabilize the active conformation by capping the helix spanning residues 161–182 (K160D).

2.3 Bioinformatics-directed refinement of CA-cGAS

To close the activity gap between CA-cGAS-04 and WT cGAS, we used a combination of bioinformatics and computational approaches. We reasoned that proteins with structural homology specific to the active conformation of cGAS may inspire additional active conformation-stabilizing mutations. We used Protein Data Bank in Europe [148] to search for distantly related proteins with structures closely matching the active conformation of cGAS, and then reviewed available structural and functional data to identify those that are not known to undergo large conformational changes. One protein that matched these criteria was MiD51, an ADP-binding mitochondrial receptor that facilitates mitochondrial fission and adopts the same structure in ADP-bound and unliganded crystal forms [149]. MiD51 contains an NTase domain that aligns well with active cGAS (1.96 Å Cα RMSD) (Fig. 2a), though, like many proteins containing NTase domains, it is not an enzyme. The ADP binding site loop and surrounding areas align poorly with the inactive conformation of cGAS, but very well with the active conformation (3.54 Å and 0.34 Å Cα RMSD, respectively; Fig. 2b). Despite this high structural similarity, MiD51 and cGAS have only 18% sequence identity across the aligned domains. Since cGAS undergoes large structural rearrangements upon DNA binding, whereas MiD51 appears to adopt a single static conformation, strongly conserved amino acid differences between the two protein families in cGAS regions 1 and 2 (Fig. 1d) may be related to the functional requirement for movement in cGAS. We hypothesized that mutating these positions in CA-cGAS-04 to the amino acids observed in MiD51-like proteins would further stabilize the active conformation or destabilize the inactive conformation.

To identify such mutations, we generated multiple sequence alignments (MSAs) of varying permissiveness for cGAS and MiD51 using the hhblits algorithm [150]. At low stringency many sequences appeared in both alignments, in keeping with the distant homology between the two proteins, but cGAS-like and MiD51-like clusters were apparent (Fig. 2c). Using a stringent E-value cutoff resulted in a cGAS-like MSA with 468 members and a MiD51-like MSA with 219 members, with no overlap (Fig. 2c and Extended Data Fig. 2). From these two MSAs we calculated the difference in amino acid frequency at each position, as well as a scaled entropy difference that highlights positions with strongly conserved differences between the two families (Fig. 2d). These data identified target positions and amino acid identities likely to stabilize the active conformation of cGAS. Within the regions specified for computational design, 19 positions have strongly conserved differences. The differences at several positions can be explained by the distinct binding or catalytic functions of MiD51 and cGAS or would not be expected to substantially affect cGAS conformational dynamics; these positions were excluded from further analysis. Ten positions remained: 158, 160, 161, 196, 197, 200, 205, 206, 207, and 209. Of these, residue 160 in mcGAS is a conserved lysine or arginine (combined $p=0.70$) that binds dsDNA. This position is also conserved in MiD51, but as a proline ($p=0.70$), which in cGAS would be expected to strongly disfavor formation of the spine helix observed in the inactive conformation, like the R161P mutation in CA-cGAS-04. Residues 196 and 200 in cGAS feature polar amino acids that are partially buried in the active conformation and solvent-exposed in the inactive conformation. Both positions are hydrophobic and well-packed in MiD51, suggesting that they may be useful for stabilizing the active conformation of cGAS. Based on this analysis, a second set of mutants was selected and cloned into the CA-cGAS-04 background.

2.4 In vitro activity of second-generation CA-cGAS

Four of these variants, containing primarily polar to non-polar mutations (Fig. 3a), were constitutively active in the ISRE-luciferase assay (Fig. 3b). Eight hours post-transfection, one variant (CA-cGAS-50) had near-WT levels of activity, while two others (CA-cGAS-22 and CA-cGAS-41) were approximately two-fold lower and CA-cGAS-42 was just above baseline (Extended Data Fig. 3a). 18 hours post-transfection, the activity of all four variants was near WT (Extended Data Fig. 3a).

To further characterize these active variants, we expressed and purified the core catalytic domain of each (d147 CA-cGAS), ensuring that our preparations were free of nucleic acid (Extended Data Fig. 3, b–d), and measured their enzymatic activity under a number of different conditions *in vitro*. Activity was determined by monitoring the change in fATP fluorescence during the cyclization reaction (Fig. 3b) [143]. The most active variant (CA-cGAS-41) was 21-fold less active than WT, and the relative levels of *in vitro* activity between the variants closely matched that observed by ISRE-luciferase assay in 293T cells (Fig. 3c). Critically, while WT cGAS activity was entirely dependent on the presence of dsDNA, CA-cGAS activity was not (Fig. 3, c and d). Furthermore, WT cGAS activity is highly salt-dependent, likely due to disruption of dsDNA binding, while CA-cGAS activity is independent of salt concentration (Fig. 3e). Consistent with previous studies that identified 2′,3′-cGAMP as the major reaction product [67,151,152], WT cGAS, CA-cGAS-41, and CA-cGAS-50 produced substantial 2′,3′-cGAMP *in vitro*, but did not produce detectable levels of 3′,3′-cGAMP (Extended Data Fig. 3, e and f). We analyzed the ability of the CA-cGAS variants to bind dsDNA using two assays: electrophoretic mobility shift assays and size exclusion chromatography (SEC). For the electrophoretic mobility shift assays, we mixed truncated versions of WT cGAS and the CA-cGAS variants lacking the unstructured N-terminal domain with a 17 bp dsDNA fragment at varying molar ratios. For all proteins, no shift was observed until the molar ratio of protein to DNA reached 1:1, and complete shifting was not observed until a ratio of 10:1 (Extended Data Fig. 4a). We observed a consistent shift in retention time by SEC for both WT and CA-cGAS when incubated with excess dsDNA (Extended Data Fig. 4b), consistent with prior studies that identify oligomeric cGAS-dsDNA complexes as the enzymatically active form [141–143]. Together, these data suggest that although our CA-cGAS variants retain dsDNA binding in the presence of molar excesses of dsDNA, their enzymatic activity is independent of oligomerization state. Moreover, we confirmed there was no measurable difference in activity between the truncated (d147) and full-length forms of CA-cGAS in cells by the ISRE assay (Extended Data Fig. 4c). Finally, reverting the CA-cGAS-04 mutations at positions 160 and 161 to WT either individually or together completely eliminated constitutive activity of variant CA-cGAS-41, as measured by ISRE assay (Fig. 3g).

2.5 Structure of CA-cGAS in the absence of bound dsDN

We solved structures of CA-cGAS-41 crystallized in the presence of GTP and CA-cGAS-50 without ligand to 2.18 and 2.60 Å resolution, respectively (Extended Data Fig. 5, a and b and Table 1). Compared to the previously published inactive structure, the NTase cores of CA-cGAS-41 and CA-cGAS-50 were 3.6 and 3.5 Å closer to their C-domains, adopting structures consistent with the active conformation (Fig. 4, a and b) [68]. The designed break in the spine helix in both structures (residues 158–161) also closely matched the active conformation as intended (Fig. 4, a and b, arrows). There was no clear density for the GTP ligand in the CA-cGAS-41 active site, and in both structures the active site loop was poorly resolved. In chain A of each asymmetric unit, where the density was slightly better, the active site loop did not clearly adopt either the inactive or active conformation (Extended Data Fig. 5, c and d). Overall, the conformation of both CA-cGAS variants strongly resembled the active conformation. To our knowledge, the only other structure in which cGAS adopted an active-like conformation in the absence of DNA was obtained when human cGAS was co-crystallized with 2′,3′-cGAMP [145].

In both structures the positions of the CA-cGAS-04 mutations (K160D and R161P) were consistent with the active conformation design model (Fig. 4, c and d). However, the region containing mutations unique to CA-cGAS-41 or -50 (residues 196–203) were in unexpected conformations involved in crystal contacts (Extended Data Fig. 5, e and g) and had weaker electron density and higher B-factors than the rest of the protein. In the CA-cGAS-41 structure the β-strand at residues 193–196 is instead a loop that peels away from the molecule starting at N196L and packs the backbone of G198 and S199, along with the side chain of Y201, against the spine helix of a second molecule in the asymmetric unit (Extended Data Fig. 5f). In the CA-cGAS-50 structure Y200I and Y201 also make crystal packing contacts that change the conformation of this region (Extended Data Fig. 5h). This result is not surprising considering residues 196–203 form part of a dynamic loop and comprise up to four mutations. Nevertheless, these mutations strongly enhance CA-cGAS activity, suggesting they contribute to stabilization of the active conformation in a way that is not captured by the crystal structures.

That only K160D/R161P are well-resolved in the expected conformation suggests that these mutations are primarily responsible for stabilization of the active conformation, which is consistent with the complete loss of ISRE activity when K160D and R161P are together reverted to WT. To explore alternative activating mutations in the spine helix, we also characterized a variant of CA-cGAS-41 containing the bioinformatics consensus mutations L159I/K160P/R161A. This variant is as active as CA-cGAS-41 (Extended Data Fig. 6a).

However, K160P/R161A in the same context is less active, and K160P alone is inactive. Additional mutations we predicted may alter CA-cGAS activity had little beneficial effect (Extended Data Fig. 6, b and c).

The local environment around regions 1 and 2 in cGAS are reasonably well conserved, which suggests activating mutations should generalize to cGAS from different species. Mutations from CA-cGAS-22, -41, -42, and -50 were introduced with K395M/K399M into human cGAS (hcGAS) and activity was measured by ISRE-luciferase assay. All CA-hcGAS variants had near-WT activity (Extended Data Fig. 6d). Given that mcGAS and hcGAS share only 55% sequence identity, we expect that the sequence diversity observed in cGAS proteins from other species should have little impact on the effect of the activating mutations described here.

2.6 In vivo activity of CA-cGAS in a tumor model

To test the activity and therapeutic potential of CA-cGAS *in vivo*, we used the B16-F10 mouse melanoma model. Prior studies have shown that intratumoral injection of nuclease-resistant, modified 2′,3′ cyclic di-AMP causes tumor regression, and that this effect requires STING expression in the host, but not the tumor [70–72]. This approach is currently being evaluated in humans (NCT02675439 and NCT03172936). Instead of injecting immunostimulatory molecules, we sought to determine whether controlled expression of CA-cGAS could generate therapeutically relevant levels of cGAMP in the tumor itself. To do this, we created a clonal line of B16-F10 cells transduced with a lentivirus encoding doxycycline (dox)-inducible CA-cGAS-50. We found that dox-mediated induction of CA-cGAS-50 did not affect cell growth *in vitro* (Fig. 5a). We transplanted these cells into WT or Sting$^{-/-}$ mice and allowed the tumors to grow for ten days before placing a cohort of the mice on dox-containing chow (Fig. 5b). We detected robust levels of cGAMP within tumors harvested from mice after dox induction (Fig. 5c). Moreover, we observed significant regression of CA-cGAS-50–expressing tumors in dox-treated WT mice, but not in dox-treated Sting$^{-/-}$ mice (Fig. 5, d–g), demonstrating that CA-cGAS-50 expression in tumor cells was sufficient to activate STING-dependent immunity in surrounding host cells. Together, these data show that CA-cGAS-50 is functional *in vivo* and could be used as a genetically encoded adjuvant to stimulate anti-tumor immunity.

2.7 Discussion and future directions

Our results establish a general multi-state design framework for stabilizing target conformations in structurally dynamic proteins. In brief, this framework has three steps: 1) Identify target and non-target conformations for the system in question, 2) enumerate all possible mutations in dynamic regions in both conformations to identify those with the largest energetic difference between the two states, and 3) combine top-ranked mutations with additional design and bioinformatics analyses based on conformation-specific homologs to identify supporting mutations and prioritize designs for experimental testing (Extended Data Fig. 7). Most current computational design methods typically explicitly consider only a single, static state, which is fundamentally inconsistent with the physical reality of proteins. This is especially true for enzymes and other dynamic systems, making these difficult targets for design. The computational design framework presented herein overcomes many hurdles in multistate design and should be applicable to other multi-state protein design challenges. The observation that proteins with widely divergent sequences and functions can have strongly conserved structural features (as demonstrated here for cGAS and MiD51), coupled with recent breakthroughs in accurate structure prediction at the genomic scale [74,75], should make our method widely applicable. Our method is also computationally inexpensive and effective even without consideration of possible unknown structural states, such as the unanticipated crystal contacts we observed in our structures. Our designed CA-cGAS variants establish that structural rearrangements alone, without dsDNA binding [65], oligomerization [141,143], or liquid-liquid phase separation [153], are sufficient for enzyme activity. Their independence of dsDNA binding and therefore the biological status of the cell (e.g., infected vs. uninfected) could make CA-cGAS molecules useful tools for better understanding the biological role of the cGAS-STING pathway in various tissues. Finally, our demonstration that induction of CA-cGAS-50 resulted in STING-dependent tumor regression *in vivo* establishes CA-cGAS molecules as biologic alternatives to small molecule STING activators with potential prophylactic and therapeutic applications in infectious disease and cancer.

Acknowledgments:

We thank Marc Lajoie, Marion Pepper, and David Baker for discussions; Donald Mannikko and Stefan Stoll for related investigations; Joshua Woodward for information and related investigations; Ratika Krishnamurty for project management; and members of the King laboratory for comments on the manuscript. Figure 5b was created using BioRender. This work was funded by a grant from the Bill & Melinda Gates Foundation (OPP1156262) to N.P.K. and D.B.S. D.B.S. is a Howard Hughes Medical Institute Faculty Scholar. X-ray data were collected at the Advanced Light Source (ALS) at LBNL, supported by the Howard Hughes Medical Institute and grants from NIH (P30 GM124169-01, ALS-ENABLE P30 GM124169, and S10OD018483), NCI SBDR (CA92584), and DOE-BER IDAT (DE-AC02-05CH11231). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions Statement:

Q.M.D., E.E.G., D.B.S., and N.P.K., conceived the study; Q.M.D. designed CA-cGAS; Q.M.D. and S.O. performed bioinformatics analyses; E.E.G., H.E.V., and S.C. performed ISRE, cellular cGAMP concentration, and in vivo assays; M.R.J performed in vivo assays. Q.M.D. performed biochemical analyses; Q.M.D. and A.K. performed crystal screens and optimization;

B.S. collected the crystallographic data. A.B. and M.B. analyzed and processed the crystallography data; all authors analyzed data; Q.M.D., D.B.S., and N.P.K. wrote and revised the manuscript.

Competing Interests Statement:

Q.M.D., E.E.G., D.B.S., and N.P.K. are inventors on a patent application related to constitutively active cGAS proteins. The King lab has received unrelated sponsored research agreements from Pfizer and GSK. The remaining authors declare no competing interests.

Tables

Table 1: Crystallographic Data Collection and Refinement Statistics

	CA-cGAS-50 (PDB: 7KXS)	**CA-cGAS-41 (PDB: 7LZ3)**
Data collection*		
Space group	$P\ 2_1$	$P\ 2_1$
Cell dimensions		
a, *b*, *c* (Å)	48.02, 109.34, 77.00	47.22, 106.30, 75.67
α, *β*, *γ* (°)	90.00, 91.36, 90.00	90.00, 91.15, 90.00
Resolution (Å)	48.01–2.60 (2.69–2.60)**	47.23–2.18 (2.24–2.18)
R_{merge}	0.108 (1.283)	0.103 (0.909)
$I/\sigma(I)$	7.17 (0.90)	7.2 (1.1)
Completeness (%)	99.60 (97.12)	100.00 (100.00)
Redundancy	7.0 (5.7)	4.6 (4.5)
Refinement		
Resolution (Å)	48.01–2.60 (2.69–2.60)	47.23–2.18 (2.23–2.18)
No. of reflections	24438 (2391)	38957 (2723)
R_{work} / R_{free} (%)	20.89 / 26.15 (33.83 / 37.32)	21.08 / 26.46 (26.98 / 32.37)
No. atoms	5766	6011
Protein	5727	5784
Ligand/ion	2	29
Water	37	207
B-factors (Å^2)		
Protein	85.96	21.40
Ligand/ion	64.85	38.06
Water	61.90	26.04
R.m.s. deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.580	0.456

* Data were collected from one crystal per condition.

** Values in parentheses are for the highest-resolution shell.

Figures

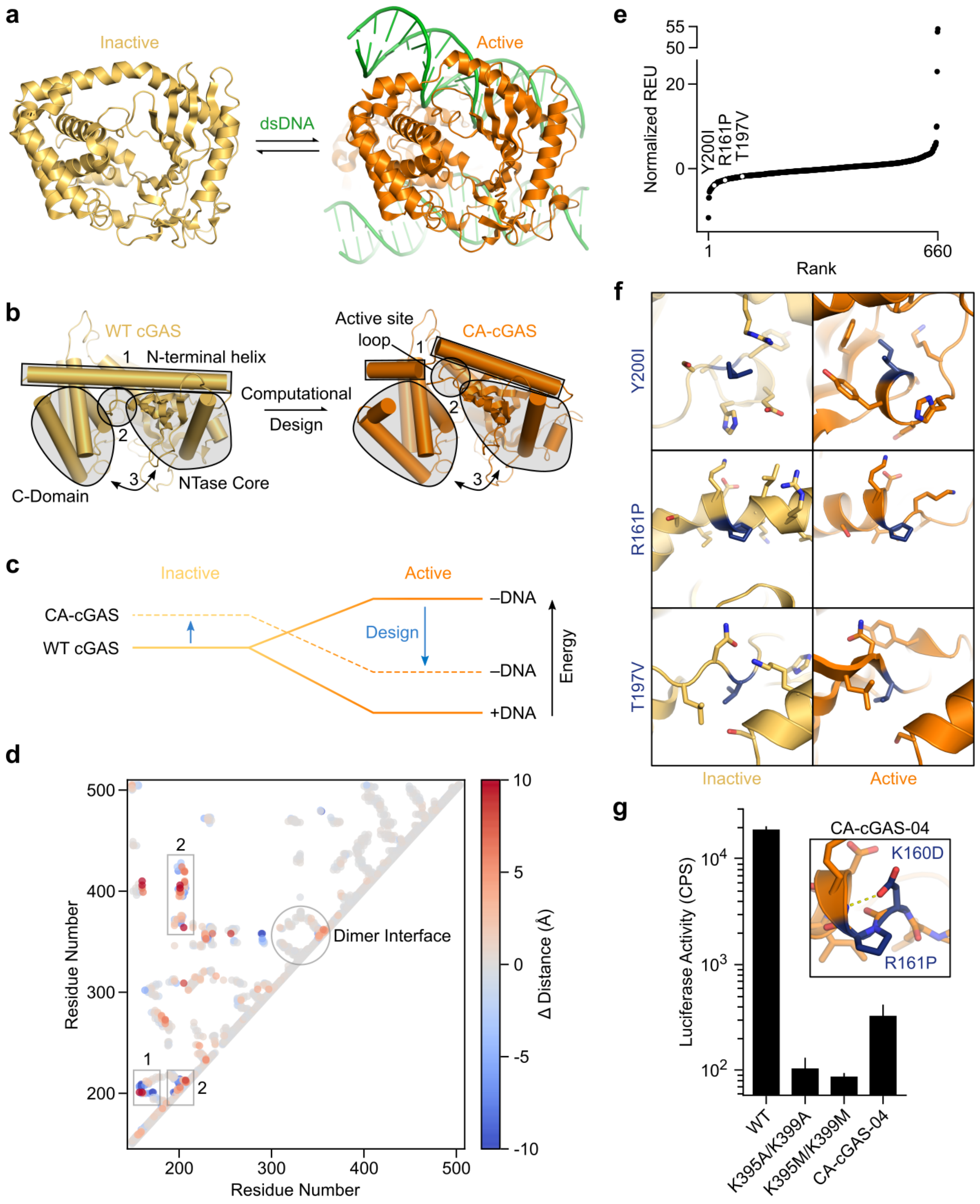

a
Inactive
dsDNA
Active
b
WT cGAS
1 N-terminal helix
2
3
C-Domain
NTase Core
Computational
Design
Active site
loop
CA-cGAS
1
2
3
c
Inactive
Active
CA-cGAS
WT cGAS
Design
–DNA
–DNA
+DNA
Energy
d
2
1
2
Dimer Interface
Residue Number
Residue Number
500
400
300
200
200
300
400
500
10
5
0
-5
-10
Δ Distance (Å)
e
Normalized REU
55
50
20
0
Y200I
R161P
T197V
1
660
Rank
f
Y200I
R161P
T197V
Inactive
Active
g
Luciferase Activity (CPS)
10^4
10^3
10^2
CA-cGAS-04
K160D
R161P
WT
K395A/K399A
K395M/K399M
CA-cGAS-04

Fig. 1: Computational design of a DNA-independent constitutively active cGAS. a, cGAS is activated by binding to dsDNA (inactive PDB ID 4K8V [68], active PDB ID 5N6I [143]). **b,** Conformational changes that occur upon dsDNA binding provide an opportunity to use computational design to stabilize the active conformation in the absence of dsDNA. Local changes are observed in the N-terminal helix (region 1) and active site loop (region 2), and a more global change brings the NTase core and C-domain closer together (region 3). **c,** A two-state design approach for generating CA-cGAS variants destabilizes the inactive conformation and stabilizes the active conformation in the absence of dsDNA. **d,** The residue-residue distance difference matrix of the active (PDB ID 4K97) and inactive (PDB ID 4K8V) conformations of mcGAS highlights two regions (1 and 2) to target for design. The matrix shows pairs of contacting residues (off-diagonal points) that are closer together (blue) or further apart (red) in the active conformation compared to the inactive. **e,** Enumeration of all possible point mutants in regions 1 and 2. The mutations are rank ordered by normalized Rosetta energy, calculated as the difference in the energetic impact of the mutation in the active and inactive states: $(E_{\text{mutant,active}} - E_{\text{WT,active}}) - (E_{\text{mutant,inactive}} - E_{\text{WT,inactive}})$. **f,** Design models for representative high-ranking mutations in the inactive and active conformations. The rank order for these mutations (blue) is highlighted in E. **g,** Activity of CA-cGAS-04 compared to WT, K395A/399A, and K395M/K399M cGAS in cells, measured by ISRE assay. Bars represent the mean of 3 technical replicates. The assay was performed at least twice with similar results. The inset shows the design model for CA-cGAS-04, highlighting the two mutations added to the K395M/K399M background.

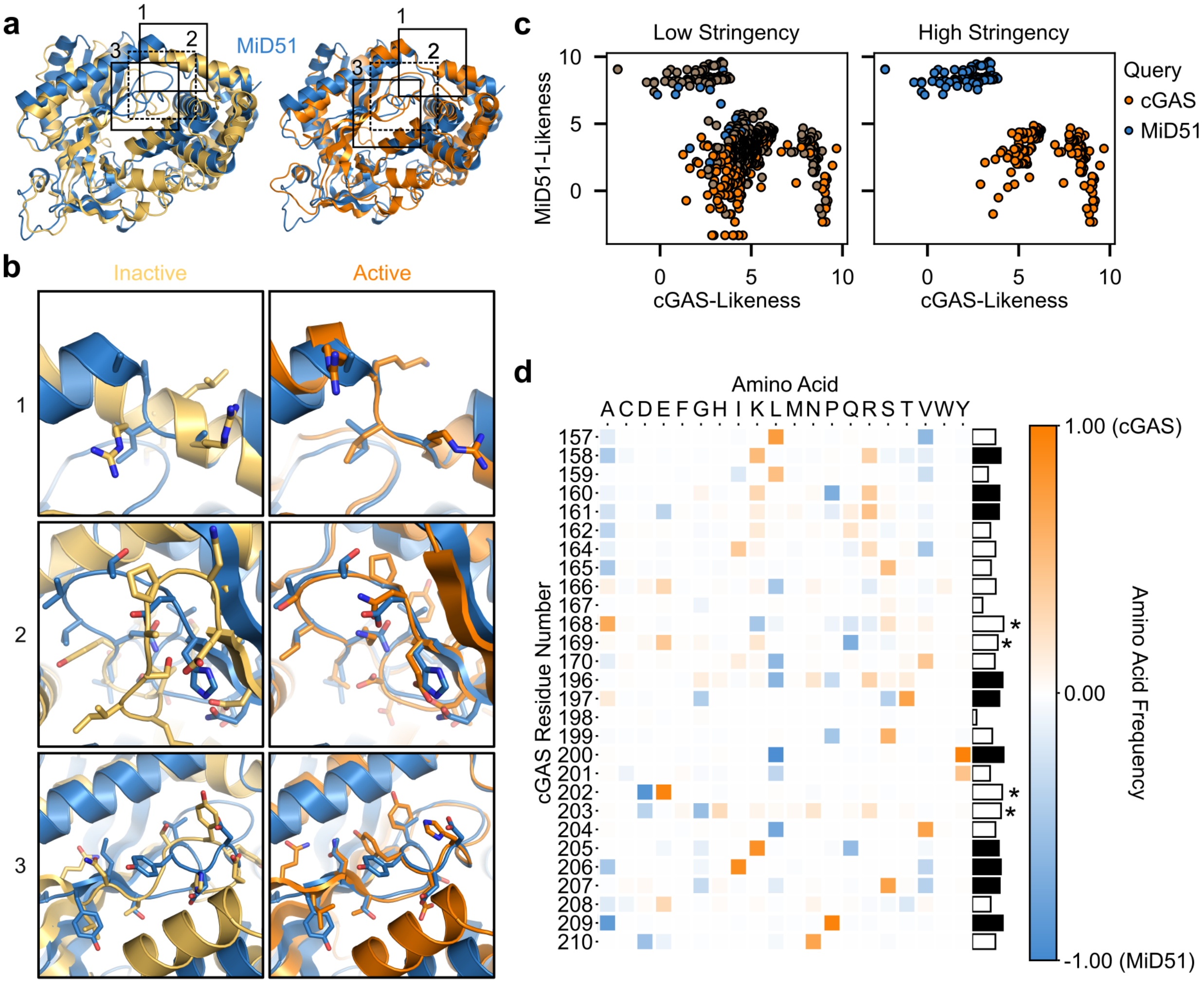


Fig. 2: Structural and bioinformatics comparison between cGAS and the static structural homolog MiD51. a, Full structure alignment between cGAS (yellow, inactive; orange, active) and MiD51 (blue). Boxed areas are shown in panel B. **b,** Local alignments of the boxed areas in panel A. Note that box 1 corresponds to cGAS region 1 and boxes 2 and 3 correspond to cGAS region 2. **c,** Bit scores for each member of a low stringency combined cGAS/MiD51 MSA compared to either the cGAS sequence (cGAS-Likeness) or MiD51 sequence (MiD51-Likeness). Sequences originally from MSAs to cGAS are colored blue, sequences from MSAs to MiD51 colored orange. Sequences appearing in both MSAs colored brown **d,** The difference in position-specific amino acid frequency (heatmap) in the boxed regions highlighted in Fig. 2b, residues 157–179 in box 1 and 196–210 in boxes 2 and 3. The bars show the scaled

entropy difference at each position. Positions excluded because of substrate interactions marked with an asterisk; positions of interest in black.

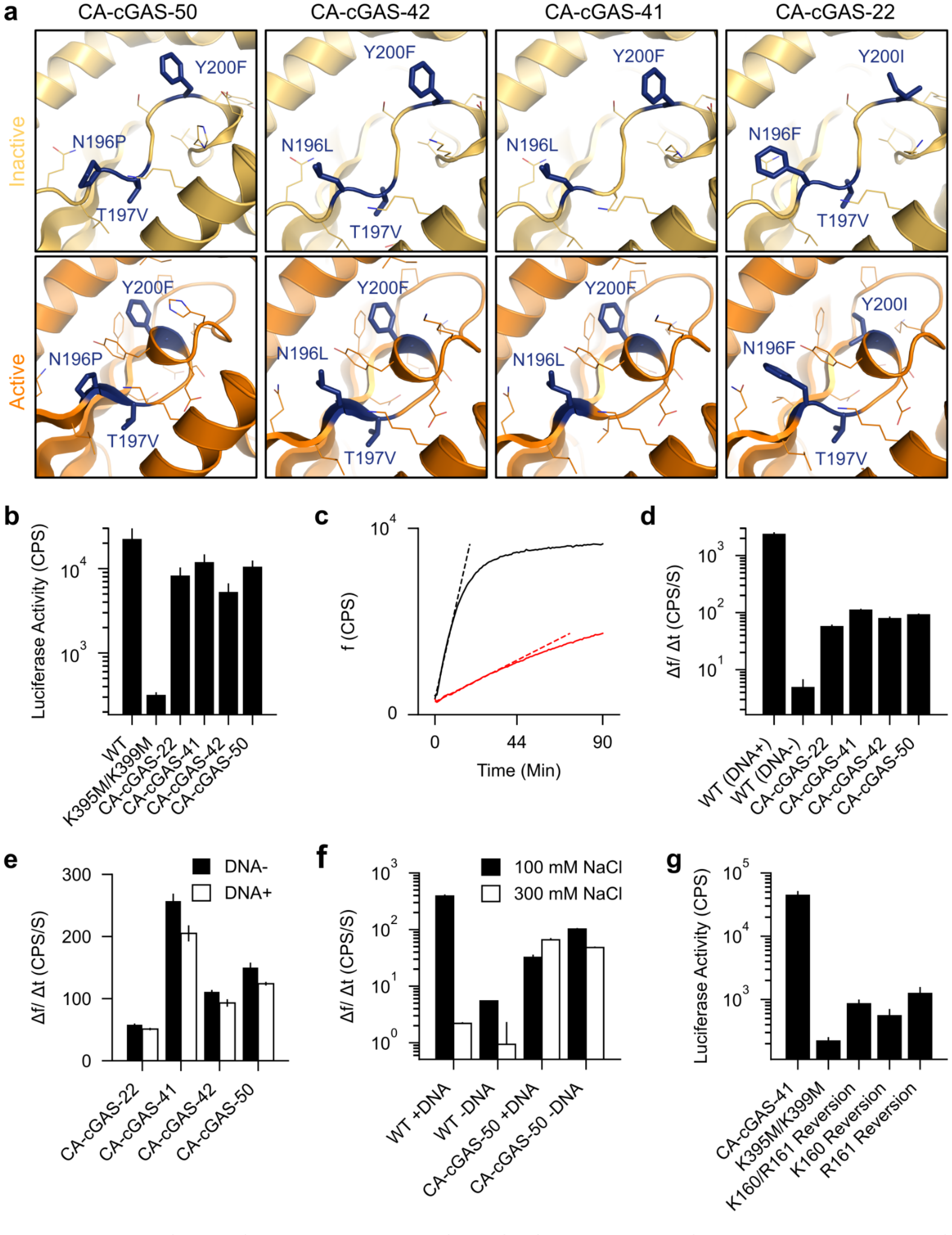


Fig. 3: Activity of bioinformatics-guided CA-cGAS variants. a, Computational models of the activating mutations for each CA-cGAS variant in the active and inactive conformation. Note that CA-cGAS-22, -41, -42, and -50 also contain the CA-cGAS-04 mutations, which are not visible in

this view. **b,** The activity of WT and K395M/K399M cGAS compared to CA-cGAS mutants as measured by ISRE assay 18 hours post transfection. **c,** The normalized fluorescence intensity of fATP changes over time as it reacts with GTP, catalyzed by WT cGAS (black) or CA-cGAS-50 (red). The initial reaction rate is fit with a linear regression. **d,** Initial reaction rates for WT cGAS and CA-cGAS ± dsDNA compared to CA-cGAS mutants without dsDNA. **e,** The effect of the presence of dsDNA on CA-cGAS activity. **f,** Salt dependence of WT cGAS and CA-cGAS-50 activity in the presence or absence of dsDNA. **g,** Activity of CA-cGAS-41 by ISRE assay after reverting one or both of the CA-cGAS-04 mutations to WT. In all plots bars represent the mean. Each assay was performed at least twice with similar results.

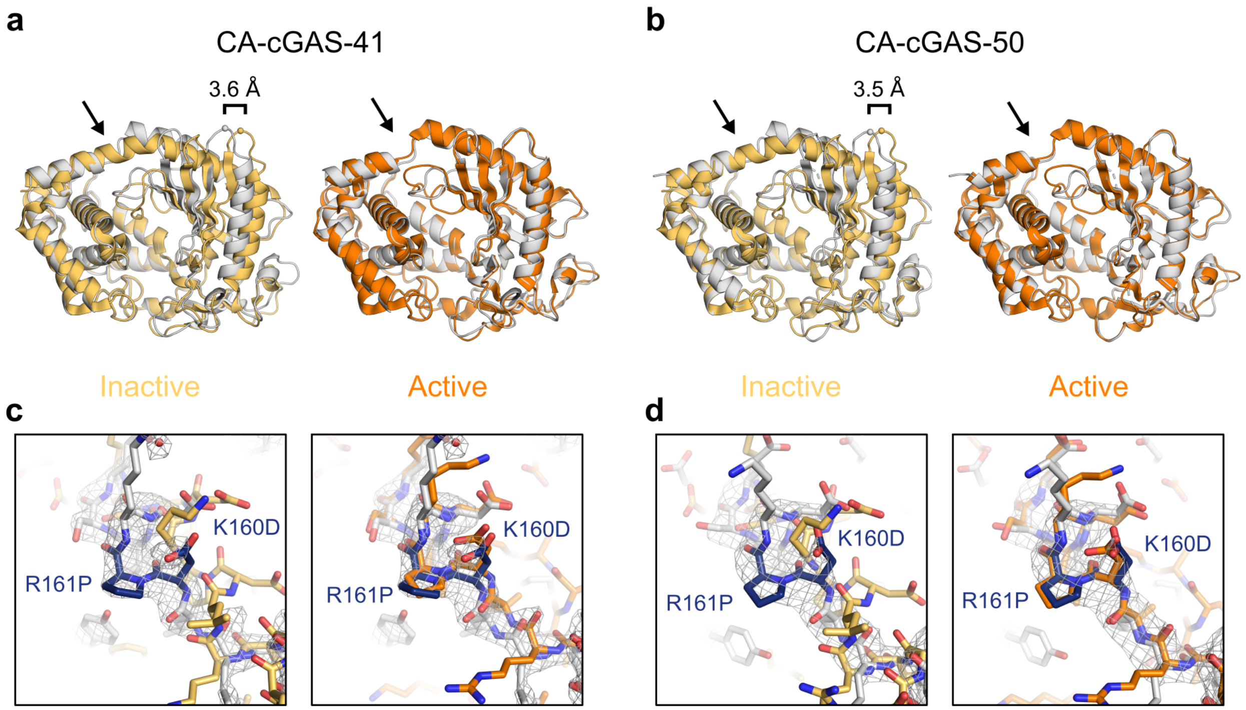


Fig. 4: X-ray crystallography of CA-cGAS variants. a, CA-cGAS-41 co-crystallized with GTP (white) aligned to WT cGAS in the inactive and active conformations. **b,** CA-cGAS-50 apo crystal structure (white) aligned to WT cGAS in the inactive and active conformations. The arrows highlight the conformation of the spine helix. The NTase domain offset between the crystal structure and inactive conformation, defined as the change in position of the Cα of residue 279 (spheres), is also indicated. **c,** Detail of the CA-cGAS-41 and **d,** CA-cGAS-50 $2F_o$-F_c map (contoured at 1.0 sigma) and refined model (white) around residues 158–164 in the spine helix aligned to the spine helix of the inactive and active conformation design models. Mutations are labeled and colored dark blue.

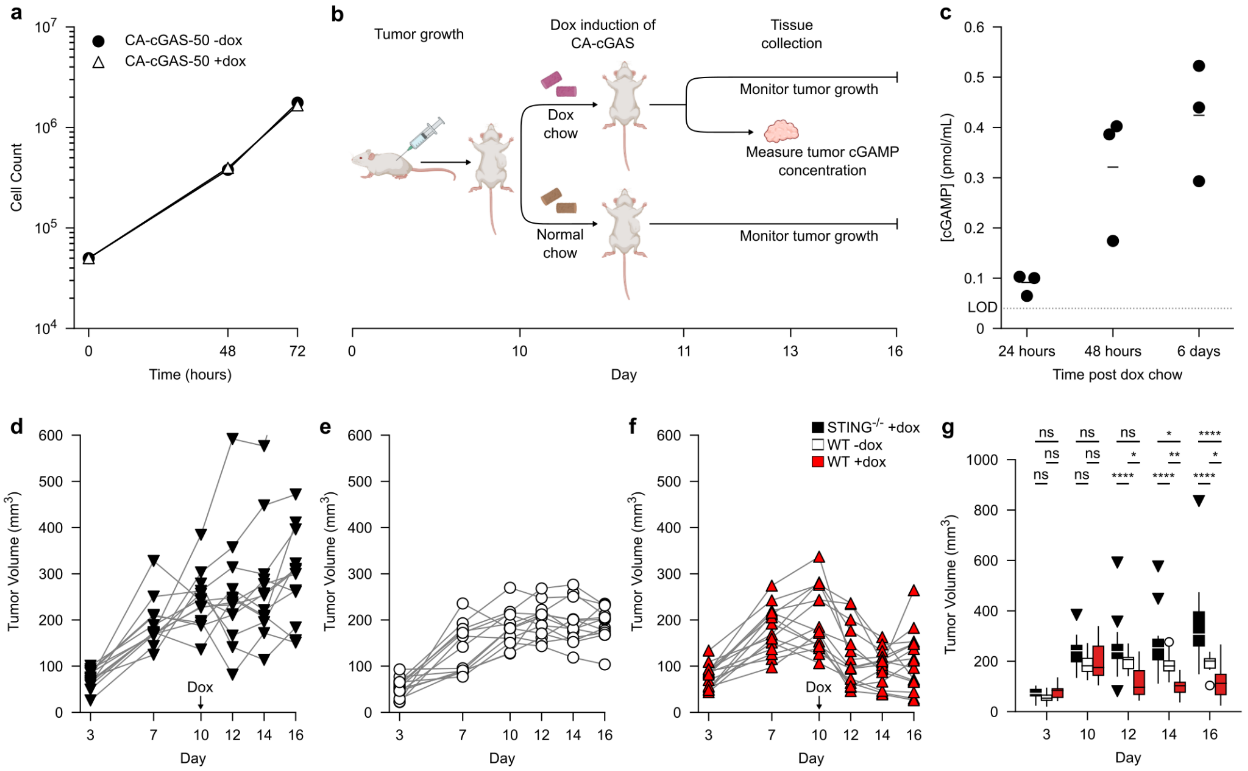
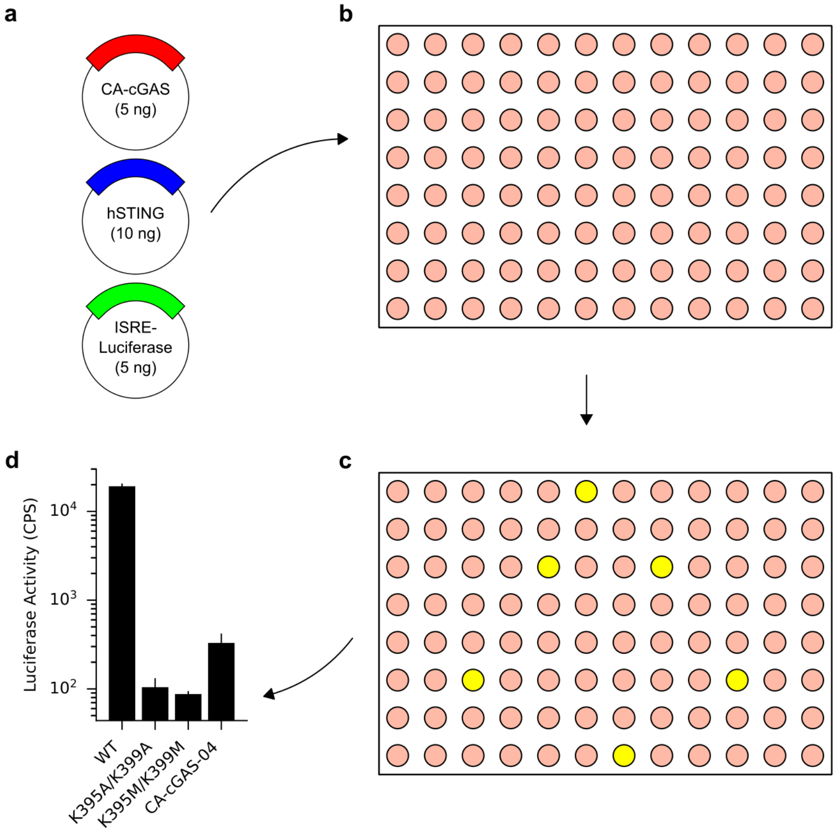
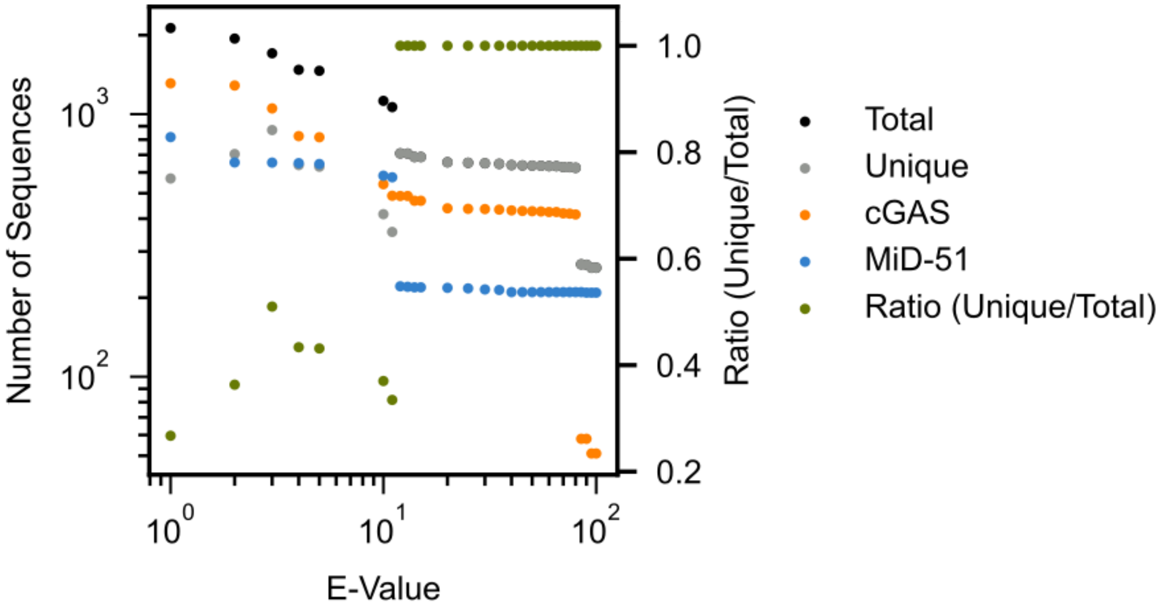


Fig. 5: *In vivo* activity of CA-cGAS-50. a, *In vitro* growth of CA-cGAS-50 B16-F10 cells with and without dox at 1 µg/ml. **b,** Schematic of study design. After implantation, tumors were allowed 10–11 days to establish before switching mice to chow containing dox. **c,** cGAMP concentration in tumor tissue 24 hours, 48 hours, and six days post CA-cGAS-50 induction (12, 13, and 17 days post tumor implantation respectively). For each timepoint, n=3 mice. **d,** Growth of CA-cGAS-50 B16-F10 tumors in STING$^{-/-}$ mice. CA-cGAS-50 expression was induced by switching mice to dox chow on day 10 post tumor implantation. Note one datapoint is out of the displayed range at 16 days (shown in g). **e,** CA-cGAS-50 B16-F10 tumor volume in WT mice without dox chow treatment. **f,** CA-cGAS-50 B16-F10 tumor volume in WT mice with dox induction at day 10 post tumor implantation. **g,** Comparison of tumor volume of CA-cGAS-50–expressing tumors with and without dox induction in WT mice, or with dox induction in STING$^{-/-}$ mice (n=15, 13, and 13 mice, respectively).. The box represents the median, upper, and lower quartile. The whiskers extend to the minimum and maximum data points within 1.5 times the interquartile range. Datapoints beyond that range are shown

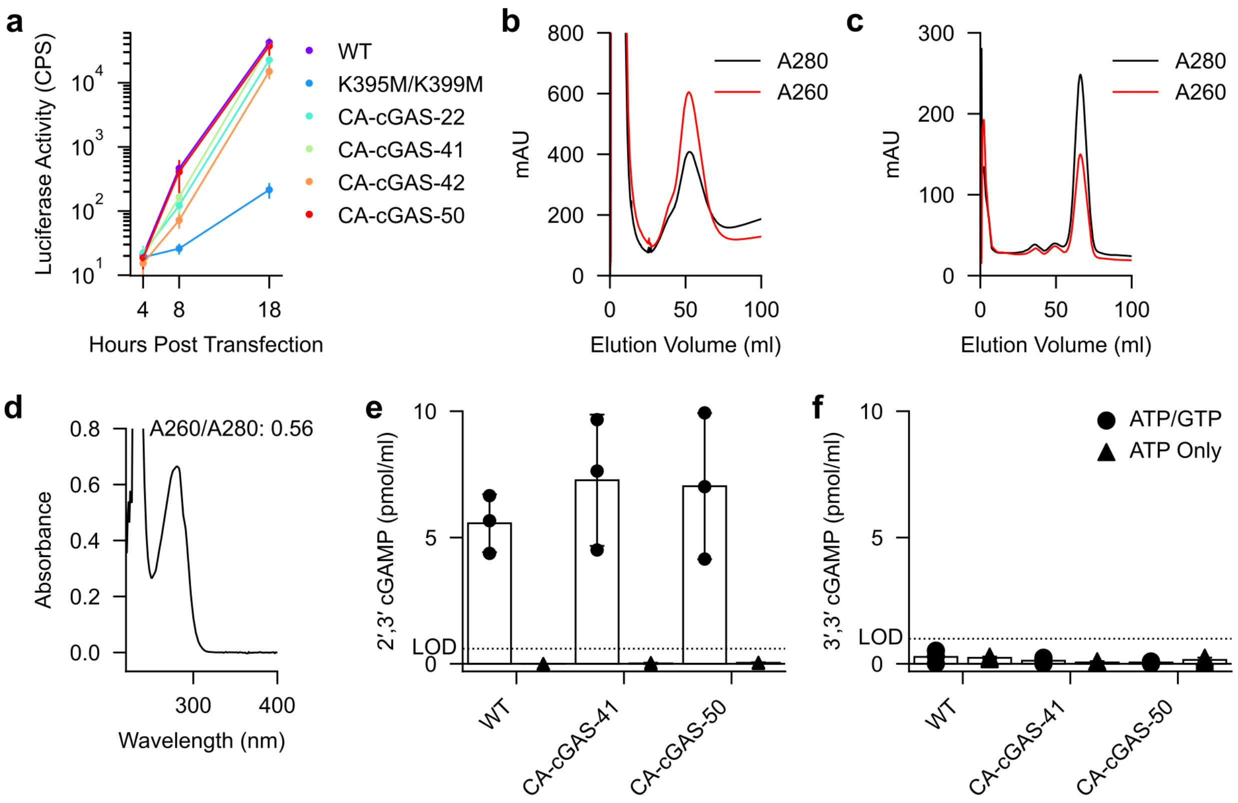
explicitly. Statistical analysis was performed using a two-way ANOVA with Tukey's correction for multiple comparisons. ns, not significant; *, $p<0.5$; **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$. At day 12: WT -dox vs WT +dox $p=0.0203$, WT +dox vs. STING$^{-/-}$ +dox $p<0.0001$; at day 14: WT -dox vs. WT +dox $p=0.0031$, WT -dox vs. STING$^{-/-}$ +dox $p=0.0264$, WT +dox vs. STING$^{-/-}$ +dox $p<0.0001$; at day 16: WT - dox vs. WT + dox $p=0.0148$, WT - dox vs. STING$^{-/-}$ +dox $p<0.0001$, WT + dox vs. STING$^{-/-}$ +dox $p<0.0001$.



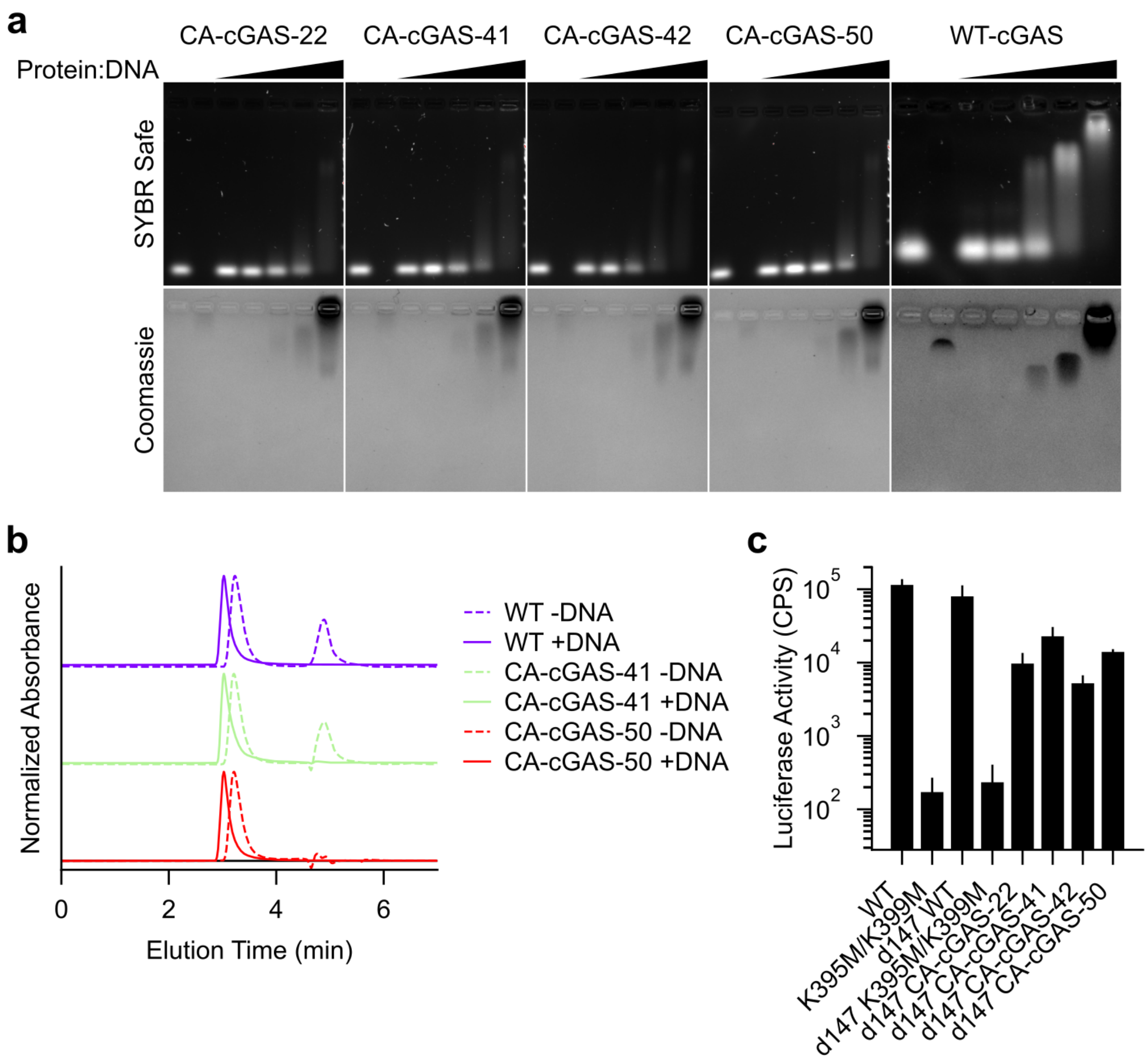
Extended Data Fig. S1: Schematic of the ISRE assay. a, CA-cGAS plasmids are generated, combined with hSTING and ISRE-Luciferase plasmids, and **b,** Transfected into 293T cells. **c,** Cells were incubated for 8 or 16 hours, lysed, and luciferase activity measured. **d,** Luciferase activity of variants was compared to WT cGAS and K395M/K399M cGAS.



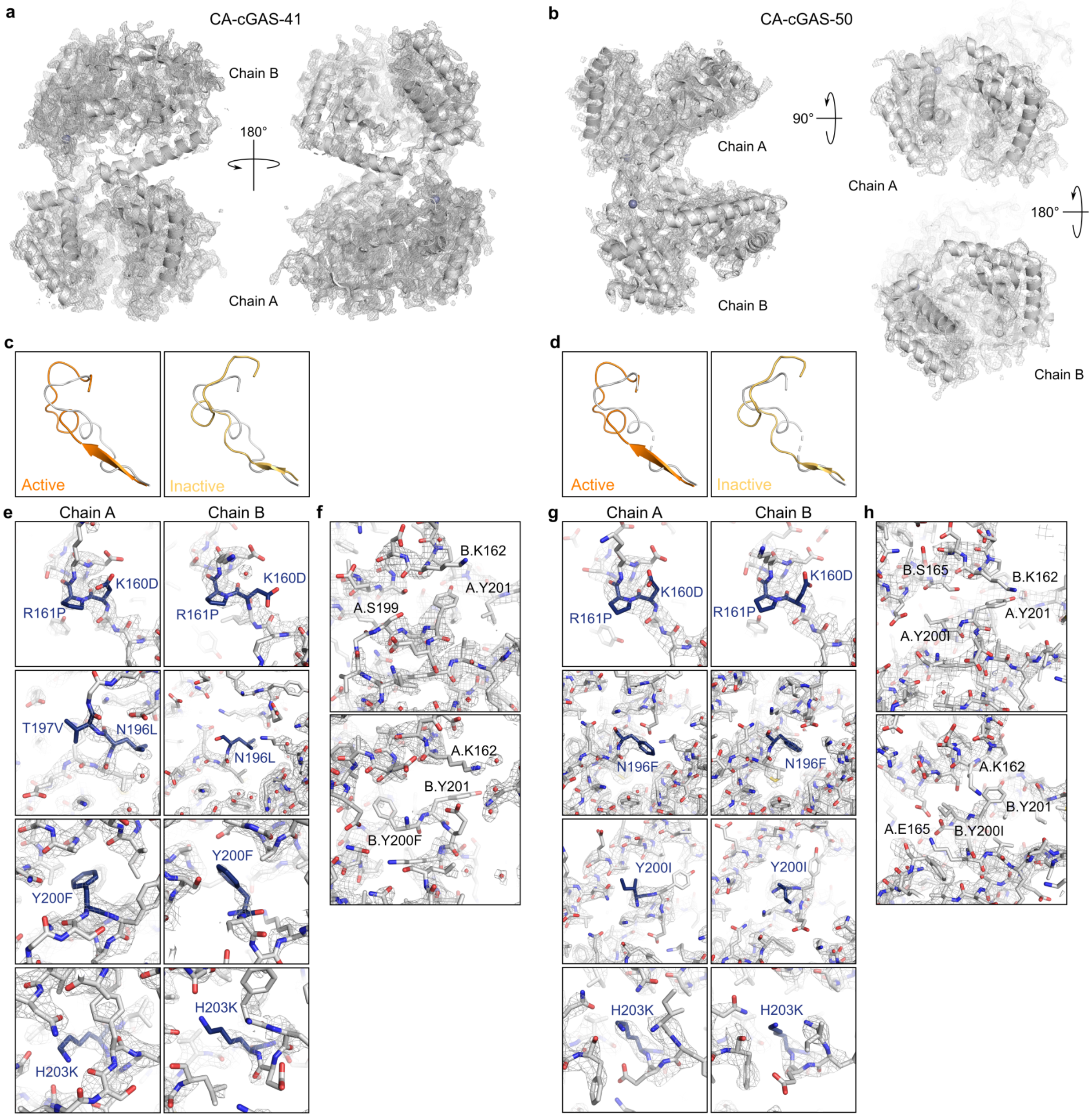
Extended Data Fig. S2: Bifurcation in unique sequences used to find E-value cutoff. The number of sequences in multiple sequence alignments to cGAS (orange) or MiD-51 (blue), as well as the number of unique sequences found only in the cGAS or MiD51 alignment but not both (gray), and the total number of sequences (black) plotted as a function of the E-value cutoff used to generate the alignment. The ratio of unique sequences to total sequences (green) is plotted on the secondary y axis.



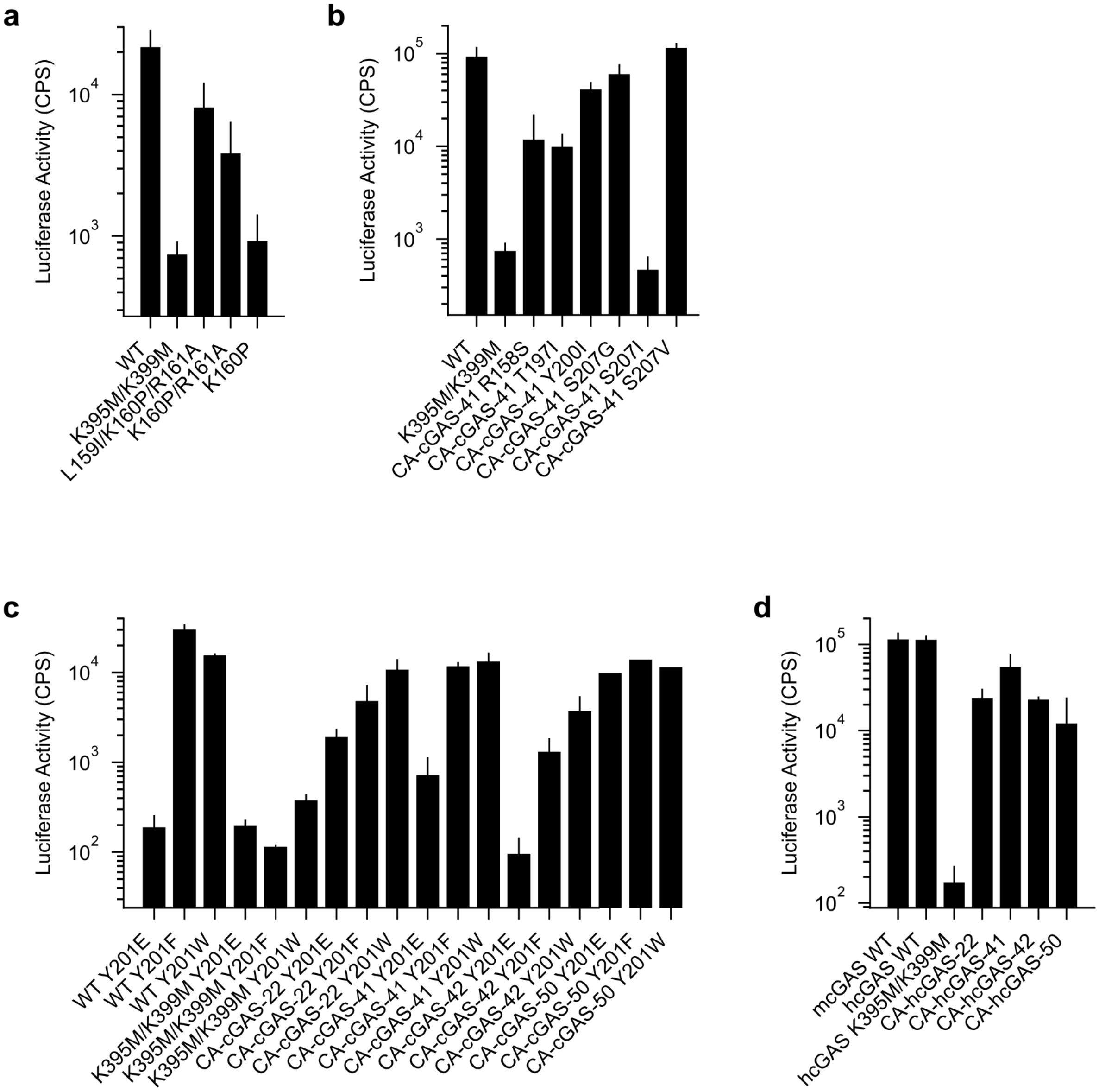
Extended Data Fig. 3: CA-cGAS activity and reaction products in vitro. a, CA-cGAS activity by ISRE assay increases over time post transfection. **b**, Chromatogram of CA-cGAS purification by nickel affinity chromatography. The high absorbance at 260 nm relative to 280 nm indicates co-purification of significant amounts of nucleic acid. **c**, Chromatogram of CA-cGAS purification by heparin affinity chromatography monitored by absorbance at 280 nm and 260 nm. Note the much lower relative absorbance at 260 nm compared to (b). **d**, UV-vis spectrum of nickel-, heparin-, SEC-purified CA-cGAS and the ratio of absorbance to 260 nm to 280 nm, indicating pure protein. **e**, 2′,3′-cGAMP was measured by competition ELISA after in vitro reactions of wild-type cGAS with dsDNA, or CA-cGAS-41 and CA-cGAS-50 without dsDNA and either 100 mM ATP and 100 mM GTP, or 200 mM ATP only. **f**, In vitro reactions failed to produce detectable 3′,3′-cGAMP.



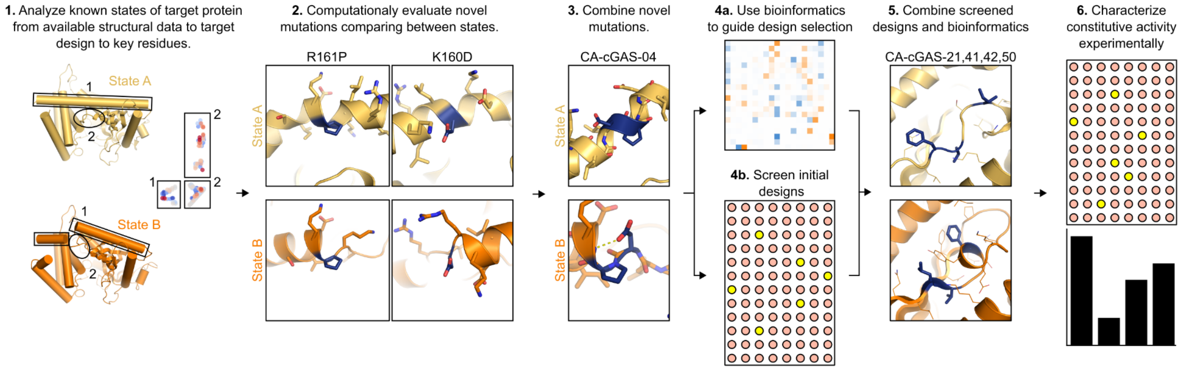
Extended Data Fig. 4: DNA binding and oligomerization of WT and CA-cGAS variants. a, Agarose gel electrophoresis of dsDNA mixed with increasing concentrations of CA-cGAS or WT cGAS. Top row stained with SYBR Safe. Bottom row is the same gel stained with coomassie. Molar ratios are, left to right, 1:0, 0:1, 1:0.1, 1:0.316, 1:1, 1:3.16, and 1:10 dsDNA to CA-cGAS. The increasing molar ratio is indicated by the triangle above the gel images. Electrophoretic mobility shift assays were performed at least three times for each sample with similar results. Representative images are shown. The assay was performed three times with similar results **b**, SEC chromatogram of WT or CA-cGAS with or without incubation with dsDNA oligomers. WT cGAS, CA-cGAS-41, and CA-cGAS-50 all elute as monomers in the absence of dsDNA, shifting to earlier elution volumes in the presence of excess DNA. **c**, Activity in cells is not affected by truncating the unstructured N-terminal domain (compare to Fig. 3b).



Extended Data Fig. 5: Similarities and differences between CA-cGAS crystal structures and the active conformation of mcGAS. The crystallographic asymmetric units of **a**, CA-cGAS-41 and **b**, CA-cGAS-50 with 2Fo-Fc maps after model building and refinement. The activation loop in chain A of the asymmetric unit does not adopt the inactive (PDB ID 4K8V) or active (PDB ID 4K97) conformation in **c**, CA-cGAS-41 or **d**, CA-cGAS-50. **e**, Details of the CA-cGAS-41 mutations (labeled and colored blue). **f**, CA-cGAS-41 crystallographic contacts between chains A and B within the asymmetric unit. **g**, Details of the CA-cGAS-50 mutations (labeled and colored blue). **h**, CA-cGAS-50 crystallographic contacts between copies of the A chain in one asymmetric unit and the B chain



Extended Data Fig. 6: ISRE activity assays of additional mcGAS and hCGAS variants. a, The activity of bioinformatics consensus mutations applied to the break in the spine helix in a CA-cGAS-41 background. **b**, Based on the available activity data, computational, and bioinformatics analysis, we can infer a series of mutations that may enhance the activity of CA-cGAS. Specifically, T197V variants are more active than variants without that mutation, variants with T197L are inactive, here T197I decreased the activity of CA-cGAS-41, suggesting that valine is the largest hydrophobic residue allowed at this position. R158S adds a capping residue on the carbonyl end of the N-terminal segment of the spine helix but does not increase activity. Y200I instead of Y200F makes CA-cGAS-41 more like CA- cGAS-50, testing the interchangeability of the activating mutations. Mutations at position 207 are designed to destabilize the inactive conformation of the active site loop. S207G and S207V are as active as CA-cGAS-41, but S207I is inactive. **c**, Recent work suggests phosphorylation at Y201 retains cGAS in the cytosol [154]. However, in the active conformation this residue is well packed; phosphorylation would likely inhibit activation, but it is unclear how phosphorylation might affect CA-cGAS activity. To mimic phosphorylation at residue 201, we introduced the mutation Y201E. We also made Y201F or Y201W mutations to prevent phosphorylation at that site. Y201E completely knocks out activity in WT cGAS and significantly lowers it in some, but not all, CA-cGAS variants. Mutating Y201 to phenylalanine or tryptophan had little effect on cGAS activity. **d**, The activity of CA-cGAS-22, -41, -42, and -50 mutations in an hcGAS background.



Extended Data Fig. S7: Schematic depiction of multi-state design framework and characterization pipeline. The key steps of the design and characterization pipeline are depicted schematically. The steps at which each CA-cGAS variant was discovered are indicated.

Supplementary Table 1: List of all variants tested and *in vitro* activity

Variant	Active?*
K160D/R161P (CA-cGAS-04)	Yes
L159I/K160D/R161P	No
L159V/K160D/R161P	No
R158P/K160D/R161P	No
R158P/L159V/K160D/R161P	No
R158P/K160D/R161P/I164Q/E166D	No
N196P/Y200F/H203Q/V204I	No
N196P/Y200V/H203K	No
N196P/T197I/Y200F/V204N	No
N196P/Y200F/V204F	No
T197V/Y200I/H203Q/V204F	No
T197I/Y200I/H203Q/V204I	No
Y200F/H203Q/V204I	No
Y200I/H203K/V204I	No
Y200F/H203K/V204I	No
T197L/Y200F/V203F	No
T197I/Y200F/V204I	No
N196L/Y200F/H203Q/V204I	No
N196L/Y200I/H203K/V204I	No
N196L/T197V/Y200F/H203K	No
N196L/Y200F/H203K	No
N196R/T197I/Y200I/H203K/V204I	No
N196F/T197L/Y200F/H203G/V204N	No

N196F/T197V/Y200I/H203K	No
N196F/T197L/Y200F/H203R	No
N196F/Y200F/V204I	No
K205L	No
R158P	No
K158P	No
K160P	No
R161P	No
N196W	No
N196F	No
Y200I	No
Y200F	No
K160D/R161P/N196P/T197V/Y200F/H203K (CA-cGAS-22)	Yes
K160D/R161P/T197I/Y200I/H203Q/V204I	No
K160D/R161P/Y200F/H203Q/V204I	No
K160D/R161P/Y200I/H203K/V204I	No
K160D/R161P/Y200F/H203K/V204I	No
K160D/R161P/Y200I/H203R/V204I	No
K160D/R161P/T197L/Y200F/V204F	No
K160D/R161P/Y200F/H203G/V204F	No
K160D/R161P/N196L/Y200F/H203Q/V204I	No
K160D/R161P/N196L/Y200I/H203K/V204I	No
K160D/R161P/N196L/T197V/Y200F/H203K (CA-cGAS-41)	Yes
K160D/R161P/N196L/Y200F/H203K (CA-cGAS-42)	Yes
K160D/R161P/N196R/Y200F/H203Q/V204I	No

K160D/R161P/N196R/Y200F/V204F	No
K160D/R161P/N196F/T197V/Y200I/H203K (CA-cGAS-50)	Yes
K160D/R161P/N196F/T197L/Y200F/H203R	No
K160D/R161P/K205L	No
K160D/R161P/N196W/I206V	No

***Defined as above K395M/K399M baseline**

2.8 Methods for CA-cGAS design

2.8.1 Computational Design

Structures of mcGAS in the active (PDB ID 4K97) and inactive (PDB ID 4K8V) conformations were downloaded from RCSB. The residue-residue distance difference matrix was generated by calculating the alpha-carbon or side-chain heavy atom distances for the active and inactive conformations. The difference is the inactive conformation residue-residue distance subtracted from the active conformation residue-residue distance. To prepare the PDB files for computational design, all heteroatoms and DNA atoms were removed. For PDB 4K8V all chains except chain A were removed. The Rosetta computational design methodology (version 3.1) was used to introduce point mutants at every residue in regions of the enzyme that undergo putatively relevant conformational changes during activation. Each residue was mutated to every other residue and any residues in a 10 Å sphere were allowed to repack (sample rotamer conformations) to accommodate the mutation. The entire pose was then minimized, but backbone atoms were constrained outside of a 15 Å sphere around the target residue. This protocol was applied to both the active and inactive conformation. The difference in total energy in Rosetta Energy Units (REU) was calculated for the mutant and WT active conformations (dActive), and for the mutant and WT inactive conformations (dInactive). The difference between dActive and dInactive (normalized REU) was used to rank order mutants. A mutant with negative normalized REU should bias the structure towards the active conformation. The best scoring single mutants were combined or improved by either allowing any other site where a beneficial mutation was observed to mutate to one of those beneficial mutations or allowing design to any residue at those sites. Designs were scored as described above.

2.8.2 Bioinformatics

We also took a bioinformatics approach to identify likely variants. First we searched the PDB for cGAS structural homologues that had low sequence conservation using the protein structure comparison service PDBeFold at the European Bioinformatics Institute, paying close attention to the alignment of the active site and dynamic regions identified in cGAS. We then performed a literature search for information about the biochemistry of hits, looking specifically for a protein that did not undergo structural rearrangements, and identified a suitable protein, MiD51 (PDB ID 4OAF). Using MiD51 and cGAS sequences excluding the unstructured N-terminal domain as inputs, we generated two overlapping multiple sequence alignments (MSAs) using the HHblits algorithm (hhsuite v3.0-beta3) with a high e-value cutoff. We then combined the two alignments and calculated bit-scores for cGAS and MiD51. We clustered the bit-score from alignments to cGAS and MiD51 using a Gaussian mixture model with Dirichlet process. We then generated MSAs with increasing e-value cutoff until there was no overlap between the cGAS MSA and MiD51 MSA, resulting in MSAs containing cGAS-like and MiD51-like proteins. We calculated the site-specific amino acid frequencies for each group, and the frequency difference between groups. To evaluate the magnitude of the differences in site specific amino acid frequencies we calculated the divergence. The site specific amino acid frequency divergence between cGAS and MiD51 multiple sequence alignments was calculated from the site specific entropy $S = -\sum pa \times log_2(pa)$. The divergence was then calculated according to $D = \sqrt{H_{mixed} - \frac{H_{cGAS} + H_{MiD51}}{2}}$ where H_{mixed} is calculated from the average frequency $pa_{mixed} = \frac{pa_{cGAS} + pa_{MiD51}}{2}$ where *pa* is the site specific amino acid frequency. By calculating the difference in site specific amino acid frequency between cGAS-like and MiD51-like proteins, we could identify key regions and mutations that were likely to explain the difference in function between the two proteins, specifically the presence or absence of a conformational change. Using those data we combined sets of single mutations to make constitutively active cGAS variants for experimental testing. Some heuristics were applied when selecting design variants for experimental testing, such as minimizing distance in primary structure between mutations to facilitate primer design for Kunkel mutagenesis. Structural alignments between MiD51 and cGAS were generated in PyMOL with the “super” command applied to the full structure, or for details, by local alignments. The spine helix alignment was made between cGAS (4k97) residues 158-161 and MiD51 residues 148-151. The active site loop alignment was made between cGAS residues 194-203 and MiD51 residues 184-193. The inactive conformation of cGAS (4K8V) was then aligned to the active conformation to compare MiD51 to inactive cGAS.

2.8.3 Plasmid Preparation

The sequences for mouse cGAS (uniprot: Q8C6L5) and human cGAS (uniprot: Q8N884) were optimized for expression in mammalian cells and purchased as synthetic gBlocks from IDT. The gene fragment was amplified using Phusion Polymerase per the manufacturers instructions with primers incorporating homology regions for the 5′ or 3′ ends of the target cut vector (pCDNA3, pET28b, or pCDB179) and for the gBlock. For pCDNA3.1 and pET28b vectors the primers were homologous to the 5′ and 3′ ends of the gBlock. In some applications the sequence for a myc tag (GAACAGAAACTGATTAGCGAAGAAGAT) was included in the antisense primer. For pCDB179 the sense primer was homologous for the 3′ end of the cut vector, and for the 436th-454th base pairs in the gBlock, corresponding to a truncation at the 148th amino acid. Because cloning into pCDB179 introduces a SUMO domain on the N terminus of the protein and the 148th amino acid is a proline, a non-homologous serine codon (TCG) was added between the primer and gBlock homology regions. When the SUMO domain is cleaved from the peptide, the serine residue remains. Constructs with this truncation are referred to as d147 cGAS. The antisense primer was homologous for the 5′ cut end of the vector. The resulting amplicons were assembled with restriction-digested (NdeI and XhoI) pET28b, (XhoI and BamHI) pCDB179, or (XhoI and NotI) pCDNA3.1 using Gibson Assembly and transformed into chemically competent *E. coli* DH5α cells (NEB). Colonies were verified by Sanger sequencing. To identify CA-cGAS, the ability to bind and be activated by DNA was first removed by introducing the mutations K395M and K399M into the wild-type sequence. The mutations were made by QuickChange site-directed mutagenesis. The resulting amplified plasmids were transformed into chemically competent *E. coli* DH5α cells (NEB) and colonies verified by Sanger sequencing.

Activating mutations identified by computational design were generated by site directed mutagenesis using the Kunkel method. Single stranded, uracilated plasmid was generated by transforming genes in pCDNA3 plasmid into chemically competent CJ236 cells and inoculating six colonies into 3 mL of LB with carbenicillin. Cultures were allowed to expand for three to four hours at 37 °C, shaking at 200 rpm before adding 3×10^9 plaque forming units of M13K07 helper phage. After an additional hour the culture was expanded 1:50 and grown overnight at 37 °C, shaking at 200 rpm. Phage was isolated from the culture medium by first clarifying the medium and then pelleting the phage in 3.3% PEG 8000, 420 mM NaCl. Single-stranded DNA was harvested using a Qiagen Qiaprep M13 kit (#27704). Oligonucleotide primers containing the desired mutation or mutations were obtained from IDT. Oligonucleotides were phosphorylated with T4 polynucleotide kinase (NEB) and diluted to an appropriate working concentration. Oligonucleotides were annealed to the single stranded template by slowly lowering the temperature from 95 °C to 25 °C at 1 °C/min. The phosphorylated and annealed oligonucleotides were used to prime in vitro DNA synthesis, with T4 DNA ligase and T7 DNA polymerase (NEB). The product, double stranded heteroduplex plasmid, was transformed into chemically competent *E. coli* DH5α cells (NEB). Colonies were verified by Sanger sequencing.

CA-cGAS genes were cloned from pCDNA3 to pCDB179 or pET28b+ by *in vitro* amplification with an appropriate set of primers containing homology for the target plasmid, and then Gibson assembled into cut vector. pET28b was cut with XhoI and NdeI. pCDB179 was cut with XhoI and BamH1. For crystallography and *in vitro* activity assays, a truncated d147 variant was cloned into pCDB179 with a single serine residue inserted between the C-terminal SUMO glycine and N-terminal P147 residue in cGAS. The serine residue facilitates cleavage during purification.

2.8.4 Cell-based ISRE assay

To screen likely constitutively active mutants, pCDNA3 plasmid containing the mutant cGAS gene was purified from chemically competent *E. coli* DH5α cells (NEB) using a Qiagen Miniprep kit. 293T cells were seeded into 96-well flat bottom plates at 2.5×10^4 per well. Cells were transiently transfected with three plasmids containing mutant cGAS (5 ng/well), STING (20 ng/well), and luciferase under regulation by an Interferon Stimulated Response Element (5 ng/well). The cells were incubated for 8 to 18 hours before being lysed and luciferase activity assessed using the Luciferase Assay System (Promega E4550) according to the manufacturer's instructions.

2.8.5 Production and purification

cGAS mutants showing activity by ISRE assay were purified recombinantly from chemically competent *E. coli* T7 cells (NEB). A small culture of TB with kanamycin was inoculated and grown overnight at 37 °C. Cultures were then expanded 1:50 into 500 mL of TB with kanamycin in a 2 L baffled shake flask and grown up to approximately OD 0.6 at 37 °C, shaking at 220 rpm. Once the indicated cell density was achieved, expression was induced by the addition of β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. Expression proceeded for 18 h at 18 °C with shaking at 220 r.p.m.. Cultures were harvested by centrifugation at 5,000 r.c.f. for 10 minutes, and the supernatant discarded. The pellet was resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM PMSF, 1 mM DTT, 0.1 mg/ml DNAse, and 0.1 μM RNase) and lysed at 4 °C by sonication with a probe sonicator at 70% power for two minutes or microfluidization with a single pass at 20,000 psi.

Lysate was clarified by centrifugation at 12,000 r.c.f for 30 minutes. Clarified lysate was further filtered through a 0.22 µm PVDF membrane before loading onto a 5 mL HisTrap™ HP (Cytiva Life Sciences) column. The column was washed with approximately 25 mL of running buffer (50 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM DTT) at 3 mL per minute and the protein eluted with a linear gradient from 0% to 100% elution buffer (50 mM Tris pH 8.0, 300 mM NaCl, 500 mM imidazole, 1 mM DTT) over 40 minutes. Protein elution was monitored by absorbance at 280 nm and purity estimated by monitoring absorbance at 260 nm.

To remove any bound dsDNA, the major elution fractions were pooled further purified by Heparin affinity chromatography. Pooled fractions were diluted in 20 mM Tris pH 8.0 to bring the final NaCl concentration down to approximately 200 mM. For full-length cGAS constructs the NaCl concentration must not drop below 200 mM, otherwise precipitation occurs. Truncated d147 cGAS constructs are more tolerant of low ionic strength. The diluted protein was then loaded onto a 5 mL HiTrap™ Heparin HP column (Cytiva Life Sciences). Nucleic acids were eluted by washing the column with 25 mL of wash buffer (20 mM Tris pH 8.0, 250 mM NaCl, 1 mM DTT), then the protein was eluted with a linear gradient from 0% to 100% elution buffer (20 mM Tris pH 8.0, 1000 mM NaCl, 1 mM DTT) over 40 minutes. Elution was monitored by absorbance at 260 nm (nucleic acids) and 280 nm (protein).

For d147 cGAS constructs, the SUMO domain was cleaved with a custom Ulp1 variant produced as previously described [155]. The protein concentration was first estimated by absorbance at 280 nm. After correcting for scattering, the concentration was calculated based on the theoretical molar extinction coefficient, assuming all cysteine residues are reduced. One milligram of protease per every 100 mg cGAS was added directly to the pooled fractions from the heparin column, without modification of the buffer. The cleavage reaction was allowed to proceed overnight at 4 °C. The extent of cleavage was followed by SDS-PAGE. To remove SUMO, Ulp1 protease, and uncleaved cGAS the entire reaction volume was loaded onto a 5 mL HisTrap™ HP column and the flowthrough collected. The column was then washed with 25 mL of running buffer at pH 7.5 and protein elution monitored at 280 nm. The major wash fractions were pooled with the flow through.

Full length or cleaved d147 cGAS constructs were further purified by SEC. Nickel or Heparin affinity eluates were pooled and concentrated in a spin concentrator to approximately one ml volume or a final concentration of ~20 mg/mL, whichever was larger. One ml aliquots were injected onto a Superdex™ 75 10/300 GL column. For d147 cGAS constructs the column was equilibrated and run with low salt SEC buffers (20 mM Tris pH 7.5, 100 mM NaCl, 1 mM DTT). If the protein was to be frozen for storage, the buffer included 250 mM glycine and 5% v/v glycerol, except for d147 WT and d147 K395M/K399M cGAS, which was eluted in low salt SEC buffer, without glycine. Full-length constructs were eluted into low-salt SEC buffers. Again, if the protein was to be frozen, it was eluted in buffer with 250 mM glycine and 5% v/v glycerol, except WT and K395M/K399M cGAS, which was always eluted in high-salt (300 mM NaCl) SEC buffer. Full-length constructs eluted at 10.8 mL, and d147 constructs eluted at 12.6 mL. Major fractions were pooled and quantified by absorbance at 280 nm. Presence of DNA was evaluated by 260/280 nm ratio. Protein identity and purity was confirmed by SDS-PAGE, Liquid Chromatography-Mass Spectrometry, and UV-Vis spectroscopy. Constructs were frozen by diluting in SEC buffer to an appropriate concentration and aliquoting into cryo-safe tubes. Tubes were flash frozen in liquid nitrogen and stored at -80 °C.

2.8.6 *In vitro* cGAS activity assay

In vitro cGAS activity was determined using the method previously reported by Andreeva et al. [143]. Activity was measured by combining 1 μM cGAS, 500 μM GTP, and 50 μM 2-aminopurine riboside-5′-O-triphosphate (fATP) (Biolog) with or without 2.6 ng/μL plasmid dsDNA. The reaction was initiated by spiking in 5 mM $MgCl_2$. The fluorescence intensity of fATP was monitored for three hours at 32 °C on a Synergy Neo2 plate reader (Biotek), with λ_{ex} 305 nm and λ_{em} 363 nm. As cGAS converts GTP and fATP into cyclic GMP-fAMP, the fluorescence intensity decreases. The fluorescence intensity was normalized to a buffer-only sample with 50 μM fATP. The initial rate was fit with a linear model.

2′,3′ cGAMP and 3′,3′ cGAMP production *in vitro* were measured by competition ELISA (Arbor Assays 2′,3′-Cyclic GAMP ELISA Kit #K067-H5W and 3′,3′-Cyclic GAMP ELISA Kit #K073-H5W) per the manufacturer’s instructions. 2′,3′ cGAMP was produced by incubating 1 μM cGAS with 5 μM $MgCl_2$, 100 μM GTP and ATP, or only 200 μM ATP. For WT cGAS, 2.6 ng/μL plasmid dsDNA was included in the reaction mixture, but not for CA-cGAS reactions. The reaction was allowed to proceed for 90 minutes at 37 °C before adding 1 μl Benzonase® (>250 U/μl, Benzonase® Nuclease, Sigma Aldrich) and incubating for an additional 30 minutes at 37 °C. All reactions were heat-inactivated at 95 °C for 10 minutes and clarified by centrifugation at 21,000 rcf for 10 minutes. Supernatant was split into 20 μl aliquots, flash frozen in liquid nitrogen, and stored at -80 °C prior to ELISA analysis.

2.8.7 Electrophoretic mobility shift assay

The interaction between cGAS and dsDNA was evaluated by electrophoretic mobility shift. A 1% agarose gel was prepared in 1× Tris/Glycine buffer (Bio-Rad). CA-cGAS was mixed with a 40 bp dsDNA fragment purchased from IDT in 20 mM Tris pH 7.5, 100 mM NaCl buffer. Mixtures were set up such that the final concentration of dsDNA was constant for all ratios. After incubating CA-cGAS and dsDNA mixtures for 30 minutes, 2 μL of glycerol was added to 10 μL of the reaction and loaded into the pre-prepared agarose gel. Gels were run in 1× Tris/Glycine buffer at 50 volts for one hour. The gel was stained for 15 minutes in 1× SYBR™ Safe (Invitrogen) in the dark and imaged with a UV transilluminator. The gel was then stained for protein with GelCode™ Blue Stain Reagent (Thermo Scientific) for 15 minutes, de-stained overnight in water, and imaged.

2.8.8 Size-exclusion chromatography of cGAS-dsDNA complexes

cGAS-dsDNA complex oligomerization state was evaluated by SEC. 20 µM cGAS was incubated in 20 mM Tris pH 7.5, 100 mM NaCl buffer, with or without a 10-fold molar excess of 40 bp dsDNA fragment purchased from IDT. After incubation for 30 minutes at room temperature, cGAS or cGAS-dsDNA complexes were analyzed by SEC-HPLC using an Agilent 1260 HPLC equipped with a 4.6×150 mm Agilent 100A Bio3 column (P/N 5190-2504), monitoring absorbance at 280 nm.

2.8.9 X-ray crystallography and structure determination

CA-cGAS-50 was crystallized by sitting drop vapor diffusion in 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 8.5, and 20% PEG 3350 after three days at 20 °C. The crystals were soaked in 5% PEG 300 diluted in reservoir solution at 1.2× concentration, flash frozen, and stored in liquid nitrogen. Alternatively, 10 mM $MgCl_2$ and 10 mM GTP were added to 142 µM CA-cGAS-41 and crystallized by sitting drop vapor diffusion in 0.45 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 7.0, and 20% PEG 3350. Crystals formed rapidly and were looped after three days at 20 °C. Crystals were soaked in a 1.2× reservoir solution with 10% v/v diethylene glycol before being flash frozen and stored in liquid nitrogen. X-ray diffraction data were collected at the Advanced Light Source. Data sets were processed with XDS and merged with XSCALE (Version Feb. 14, 2019). The crystal structure was solved by molecular replacement with PHASER using two copies of the active conformation design model. Iterative model building and refinement was done with the macromolecular graphics and model building tool *Coot*. Comparisons to the design model were made in PyMOL. The spine helix was aligned using the “align” command, aligning residues 147-157 and 162-180 of the design model were aligned to the same residues of the A chain in the crystal structure. Comparisons between CA-cGAS crystal structures and inactive (4k8v) or active (4k97) cGAS conformations were generated by aligning the A chains using the “super” command.

2.8.10 Lentiviral Vector

Mouse cGAS (mcGAS) mut50 was cloned into the pSLIK vector as described (12). Specifically, the pSLIK vector was cut with the enzyme AsiSI (NEB catalog #R0630s) and mcGAS mut50 was inserted using In-Fusion Cloning according to manufacturer's instructions (Takara catalog #638911). PCR primers used for amplification of mcGAS were BG537 (TGATCACTAGGCGATCGCGCCACCATGGAAGATCCGCG) and BG538 (AGTCTTCCAATTGCGATCGCTCACAGATCTTCTTCGCTAATCA).

2.8.11 Cell culture

B16-F10 cells were maintained in Dulbecco's modification of Eagle medium (DMEM) supplemented with 10% FBS, 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate, 0.05 mM beta-mercaptoethanol, 100 IU penicillin, and 100 µg/ml streptomycin (complete DMEM). A clonal line of parental B16-F10 cells was created by lentiviral transduction of a construct that constitutively expressed the mWasabi fluorescent protein fused to a nuclear localization signal and a peptide derived from ovalbumin (SIINFEKL). These cells were transduced with pSLIK CA-cGAS-50. Single clones were grown up and tested for expression levels of CA-cGAS-50. A clone with low basal cGAMP production and high inducible expression of CA-cGAS-50 was selected for further experiments. For quantification of *in vitro* cell growth and CA-cGAS-50 induction, doxycycline was used at 1 µg/ml in complete DMEM.

2.8.12 Mice

Female C57BL6/J mice 6–8 weeks of age were purchased from Jackson Labs and allowed at least one week to acclimate prior to experiment initiation. STING KO mice (Tmem173-/-) were generated as described (39) and backcrossed to C57BL/6J (40). All mice were maintained in a specific pathogen-free (SPF) barrier facility at the University of Washington, and all experiments were done in accordance with protocols approved by the

Institutional Animal Care and Use Committee of the University of Washington (IACUC protocol #4190-01). Mice are housed with a light cycle of 14 hours of light and 10 hours of darkness with acceptable temperature and humidity ranges of 68 °F–79 °F and 30%–70% respectively.

2.8.13 Mouse Tumor Implantation

9-13 week old female C57BL6/J WT mice (n=28) or STING KO (Tmem173-/-) mice (n=13) were injected subcutaneously into the flank with 1×10^5 B16-F10 tumor cells mixed 1:1 with Matrigel Basement Matrix (Corning #354248) for a final injection volume of 100 μL. Tumor volume was calculated using the following formula: Volume = Short axis 2 × Long axis × 0.523 (41). Mice receiving doxycycline were switched to doxycycline chow (Envigo #TD.01306 625 mg/kg irradiated) starting at day 10–11 post tumor implantation. Maximum total tumor burden allowed per mouse was 2000 mm^3 and was not exceeded in this study.

2.8.14 Tumor cGAMP measurements

Tumors were dissected out of the mice and cut into small pieces in 3 ml of dissociation buffer: 2.7 mg/mL Collagenase A (Sigma #110088793001), 23 U/mL DNase I (Sigma #D4263-5VL), 2 mM $CaCl_2$ in 1× PBS. Tumors were then incubated at 37 °C with shaking for 30 minutes. 3 mL of termination buffer (2% FCS, 5 mM EDTA in 1× PBS) was then added. Samples were then filtered through a 70 mm mesh strainer, washed twice in 1× PBS, and lysed in 200 μL RIPA buffer: 150 mM NaCl, 1.0% NP-40 substitute, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0). 50 μL of sample was used per well to measure cGAMP content (Arbor Assays 2′,3′-Cyclic GAMP ELISA Kit #K067-H5W) per the manufacturer’s instructions.

2.8.15 Statistics and data analysis

Tumor growth and cGAMP quantification data were visualized and analyzed using GraphPad Prism 9 software. Statistical tests used to analyze data are noted in the figure legends and were performed using GraphPad Prism 9. Data were analyzed and plotted with Matplotlib 3.3.3 and Seaborn 0.11.2. Crystallography data were processed with XDS/XSCALE (Version Feb 14, 2019), PHASER 2.8.3, and Coot 0.9.

2.8.16 Data availability:

X-ray coordinates and structure factors are deposited in the RCSB Protein Data Bank under IDs 7LZ3 (CA-cGAS-41) and 7KXS (CA-cGAS-50). All raw data are provided as source data files.

2.8.17 Code availability:

The Rosetta software and source code is freely available for non-profit use and can be licensed for commercial use. See https://www.rosettacommons.org.

References

1. Chen, M. Y., Butler, S. S., Chen, W. & Suh, J. Physical, chemical, and synthetic virology: Reprogramming viruses as controllable nanodevices. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **11**, e1545 (2019).
2. Richard, J. P. Protein Flexibility and Stiffness Enable Efficient Enzymatic Catalysis. *J. Am. Chem. Soc.* **141**, 3320–3331 (2019).
3. Caspar, D. L. D. & Klug, A. Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor symposia on quantitative bioloqy : [proceedings]* **27**, 1–24 (1962).
4. Tetter, S. *et al.* Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* **372**, 1220–1224 (2021).
5. Terasaka, N., Azuma, Y. & Hilvert, D. Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5432–5437 (2018).
6. Harrison, S. C. The familiar and the unexpected in structures of icosahedral viruses. *Curr. Opin. Struct. Biol.* **11**, 195–199 (2001).
7. Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
8. Lajoie, M. J. *et al.* Designed protein logic to target cells with precise combinations of surface antigens. *Science* **369**, 1637–1643 (2020).
9. Quijano-Rubio, A. *et al.* De novo design of modular and tunable protein biosensors. *Nature* **591**, 482–487 (2021).
10. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).
11. Leaver-Fay, A., Jacak, R., Stranges, P. B. & Kuhlman, B. A generic program for multistate protein design. *PLoS One* **6**, e20937 (2011).
12. Lisanza, S. L. *et al.* Joint Generation of Protein Sequence and Structure with RoseTTAFold Sequence Space Diffusion. *bioRxiv* 2023.05.08.539766 (2023) doi:10.1101/2023.05.08.539766.
13. Lewis, S. M. *et al.* Generation of bispecific IgG antibodies by structure-based design of an orthogonal Fab interface. *Nat. Biotechnol.* **32**, 191–198 (2014).

14. Leaver-Fay, A. *et al.* Computationally Designed Bispecific Antibodies using Negative State Repertoires. *Structure* **24**, 641–651 (2016).
15. Chiesa, G., Kiriakov, S. & Khalil, A. S. Protein assembly systems in natural and synthetic biology. *BMC Biol.* **18**, 35 (2020).
16. Pieters, B. J. G. E., van Eldijk, M. B., Nolte, R. J. M. & Mecinović, J. Natural supramolecular protein assemblies. *Chem. Soc. Rev.* **45**, 24–39 (2016).
17. Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **92**, 291–294 (1998).
18. Douglas, T. & Young, M. Viruses: Making Friends with Old Foes. *Science* **312**, 873–875 (2006).
19. Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* **22**, 485–491 (2011).
20. Lee, E. J., Lee, N. K. & Kim, I.-S. Bioengineered protein-based nanocage for drug delivery. *Adv. Drug Deliv. Rev.* **106**, 157–171 (2016).
21. López-Sagaseta, J., Malito, E., Rappuoli, R. & Bottomley, M. J. Self-assembling protein nanoparticles in the design of vaccines. *Comput. Struct. Biotechnol. J.* **14**, 58–68 (2016).
22. Zhu, J. *et al.* Protein Assembly by Design. *Chem. Rev.* (2021) doi:10.1021/acs.chemrev.1c00308.
23. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
24. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
25. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
26. Divine, R. *et al.* Designed proteins assemble antibodies into modular nanocages. *Science* **372**, (2021).
27. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
28. Ueda, G. *et al.* Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *Elife* **9**, (2020).
29. Lutz, I. D. *et al.* Top-down design of protein nanomaterials with reinforcement learning. *bioRxiv*

(2022) doi:10.1101/2022.09.25.509419.

30. Wang, J. Y. J. *et al.* Improving the secretion of designed protein assemblies through negative design of cryptic transmembrane domains. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2214556120 (2023).
31. Watson, J. L. *et al.* Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *bioRxiv* 2022.12.09.519842 (2022) doi:10.1101/2022.12.09.519842.
32. Kobayashi, N. *et al.* Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric de Novo Protein. *J. Am. Chem. Soc.* **137**, 11285–11293 (2015).
33. Sciore, A. *et al.* Flexible, symmetry-directed approach to assembling protein cages. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8681–8686 (2016).
34. Cristie-David, A. S. *et al.* Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability. *J. Am. Chem. Soc.* **141**, 9207–9216 (2019).
35. Lai, Y.-T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336**, 1129 (2012).
36. Lai, Y.-T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6**, 1065–1071 (2014).
37. Patterson, D. P., Desai, A. M., Holl, M. M. B. & Marsh, E. N. G. Evaluation of a symmetry-based strategy for assembling protein complexes. *RSC Adv.* **1**, 1004–1012 (2011).
38. Patterson, D. P. *et al.* Characterization of a highly flexible self-assembling protein system designed to form nanocages. *Protein Sci.* **23**, 190–199 (2014).
39. Cristie-David, A. S., Koldewey, P., Meinen, B. A., Bardwell, J. C. A. & Marsh, E. N. G. Elaborating a coiled-coil-assembled octahedral protein cage with additional protein domains. *Protein Sci.* **27**, 1893–1900 (2018).
40. Sinclair, J. C., Davies, K. M., Vénien-Bryan, C. & Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6**, 558–562 (2011).
41. Laniado, J. *et al.* Geometric Lessons and Design Strategies for Nanoscale Protein Cages. *ACS Nano* **15**, 4277–4286 (2021).

42. Vulovic, I. *et al.* Generation of ordered protein assemblies using rigid three-body fusion. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

43. Hsia, Y. *et al.* Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021).

44. Malay, A. D. *et al.* An ultra-stable gold-coordinated protein cage displaying reversible assembly. *Nature* **569**, 438–442 (2019).

45. Cristie-David, A. S. & Marsh, E. N. G. Metal-dependent assembly of a protein nano-cage. *Protein Sci.* **28**, 1620–1629 (2019).

46. Miyamoto, T. *et al.* Domain-swapped cytochrome cb562 dimer and its nanocage encapsulating a Zn-SO4 cluster in the internal cavity. *Chem. Sci.* **6**, 7336–7342 (2015).

47. Ni, T. W. & Tezcan, F. A. Structural characterization of a microperoxidase inside a metal-directed protein cage. *Angew. Chem. Int. Ed Engl.* **49**, 7014–7018 (2010).

48. Brouwer, P. J. M. *et al.* Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle. *Nat. Commun.* **10**, 4272 (2019).

49. Bruun, T. U. J., Andersson, A.-M. C., Draper, S. J. & Howarth, M. Engineering a Rugged Nanoscaffold To Enhance Plug-and-Display Vaccination. *ACS Nano* **12**, 8855–8866 (2018).

50. Rahikainen, R. *et al.* Overcoming Symmetry Mismatch in Vaccine Nanoassembly through Spontaneous Amidation. *Angew. Chem. Int. Ed Engl.* **60**, 321–330 (2021).

51. Boyoglu-Barnum, S. *et al.* Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature* **592**, 623–628 (2021).

52. Cohen, A. A. *et al.* Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* **371**, 735–741 (2021).

53. Walls, A. C. *et al.* Elicitation of broadly protective sarbecovirus immunity by receptor-binding domain nanoparticle vaccines. *Cell* **184**, 5432–5447.e16 (2021).

54. Marcandalli, J. *et al.* Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell* **176**, 1420–1431.e17 (2019).

55. Walls, A. C. *et al.* Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382.e17 (2020).

56. Song, J. Y. *et al.* Safety and immunogenicity of a SARS-CoV-2 recombinant protein nanoparticle vaccine (GBP510) adjuvanted with AS03: A randomised, placebo-controlled, observer-blinded phase 1/2 trial. *EClinicalMedicine* **51**, 101569 (2022).
57. Ben-Sasson, A. J. *et al.* Design of biologically active binary protein 2D materials. *Nature* **589**, 468–473 (2021).
58. De Colibus, L. *et al.* Assembly of complex viruses exemplified by a halophilic euryarchaeal virus. *Nat. Commun.* **10**, 1456 (2019).
59. Liu, H. *et al.* Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science* **329**, 1038–1043 (2010).
60. Veesler, D. *et al.* Atomic structure of the 75 MDa extremophile Sulfolobus turreted icosahedral virus determined by CryoEM and X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5504–5509 (2013).
61. Yu, X. *et al.* Cryo-EM structure of human adenovirus D26 reveals the conservation of structural organization among human adenoviruses. *Sci Adv* **3**, e1602670 (2017).
62. Kibler, R. D. *et al.* Stepwise design of pseudosymmetric protein hetero-oligomers. *bioRxiv* (2023) doi:10.1101/2023.04.07.535760.
63. Motwani, M., Pesiridis, S. & Fitzgerald, K. A. DNA sensing by the cGAS-STING pathway in health and disease. *Nat. Rev. Genet.* **20**, 657–674 (2019).
64. Volkman, H. E., Cambier, S., Gray, E. E. & Stetson, D. B. Tight nuclear tethering of cGAS is essential for preventing autoreactivity. *Elife* **8**, (2019).
65. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332–337 (2013).
66. Zhang, X. *et al.* The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep.* **6**, 421–430 (2014).
67. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339**, 826–830 (2013).
68. Gao, P. *et al.* Cyclic [G(2’,5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153**, 1094–1107 (2013).

69. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).

70. Corrales, L. *et al.* Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity. *Cell Rep.* **11**, 1018–1030 (2015).

71. Sivick, K. E. *et al.* Magnitude of Therapeutic STING Activation Determines CD8+ T Cell-Mediated Anti-tumor Immunity. *Cell Rep.* **25**, 3074–3085.e5 (2018).

72. Francica, B. J. *et al.* TNFα and Radioresistant Stromal Cells Are Essential for Therapeutic Efficacy of Cyclic Dinucleotide STING Agonists in Nonimmunogenic Tumors. *Cancer Immunol Res* **6**, 422–433 (2018).

73. Pancotti, C. *et al.* Predicting protein stability changes upon single-point mutation: a thorough comparison of the available tools on a new dataset. *Brief. Bioinform.* **23**, (2022).

74. Humphreys, I. R. *et al.* Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021).

75. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

76. Dowling, Q. M. *et al.* Hierarchical design of pseudosymmetric protein nanoparticles. *Manuscript in preparation* (2023).

77. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

78. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).

79. Butterfield, G. L. *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420 (2017).

80. Edwardson, T. G. W., Mori, T. & Hilvert, D. Rational Engineering of a Designed Protein Cage for siRNA Delivery. *J. Am. Chem. Soc.* **140**, 10439–10442 (2018).

81. Edwardson, T. G. W., Tetter, S. & Hilvert, D. Two-tier supramolecular encapsulation of small molecules in a protein cage. *Nat. Commun.* **11**, 5410 (2020).

82. Mohan, K. *et al.* Topological control of cytokine receptor signaling induces differential effects in

hematopoiesis. *Science* **364**, (2019).

83. Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.* **10**, 1864 (2019).
84. Liu, Y., Gonen, S., Gonen, T. & Yeates, T. O. Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3362–3367 (2018).
85. Castells-Graells, R. *et al.* Rigidified Scaffolds for 3 Angstrom Resolution Cryo-EM of Small Therapeutic Protein Targets. *bioRxiv* (2022) doi:10.1101/2022.09.18.508009.
86. McConnell, S. A. *et al.* Designed Protein Cages as Scaffolds for Building Multienzyme Materials. *ACS Synth. Biol.* **9**, 381–391 (2020).
87. Wukovitz, S. W. & Yeates, T. O. Why protein crystals favour some space-groups over others. *Nat. Struct. Biol.* **2**, 1062–1067 (1995).
88. King, N. P. & Lai, Y.-T. Practical approaches to designing novel protein assemblies. *Curr. Opin. Struct. Biol.* **23**, 632–638 (2013).
89. Laniado, J. & Yeates, T. O. A complete rule set for designing symmetry combination materials from protein molecules. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 31817–31823 (2020).
90. Lindstedt, S. & Nishikawa, K. Huxleys' Missing Filament: Form and Function of Titin in Vertebrate Striated Muscle. *Annu. Rev. Physiol.* **79**, 145–166 (2017).
91. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
92. Cramer, P. *et al.* Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **37**, 337–352 (2008).
93. Wirth, C., Brandt, U., Hunte, C. & Zickermann, V. Structure and function of mitochondrial complex I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1857**, 902–914 (2016).
94. Fullerton, S. W. B. *et al.* Mechanism of the Class I KDPG aldolase. *Bioorg. Med. Chem.* **14**, 3002–3010 (2006).
95. Wang, J. Y. (john) *et al.* Improving the secretion of designed protein assemblies through negative design of cryptic transmembrane domains. *bioRxiv* (2022) doi:10.1101/2022.08.04.502842.
96. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **2014**, 1–21 (2014).

97. Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences* **110**, 15674–15679 (2013).
98. Sinkovits, R. S. & Baker, T. S. A tale of two symmetrons: rules for construction of icosahedral capsids from trisymmetrons and pentasymmetrons. *J. Struct. Biol.* **170**, 109–116 (2010).
99. Goldberg, M. A Class of Multi-Symmetric Polyhedra. *Tohoku Mathematical Journal, First Series* **43**, 104–108 (1937).
100. Wargacki, A. J. *et al.* Complete and cooperative in vitro assembly of computationally designed self-assembling protein nanomaterials. *Nat. Commun.* **12**, 883 (2021).
101. Mannige, R. V. & Brooks, C. L., 3rd. Periodic table of virus capsids: implications for natural selection and design. *PLoS One* **5**, e9423 (2010).
102. Adachi, K. *et al.* 37. Capacity of Viral Genome Packaging and Internal Volumes of AAV Viral Particles. *Mol. Ther.* **23**, S17 (2015).
103. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2217–2221 (2001).
104. Furukawa, H., Cordova, K. E., O'Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal-organic frameworks. *Science* **341**, 1230444 (2013).
105. Jiang, H., Alezi, D. & Eddaoudi, M. A reticular chemistry guide for the design of periodic solids. *Nature Reviews Materials* **6**, 466–487 (2021).
106. Wagenbauer, K. F., Sigl, C. & Dietz, H. Gigadalton-scale shape-programmable DNA assemblies. *Nature* **552**, 78–83 (2017).
107. Tikhomirov, G., Petersen, P. & Qian, L. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature* **552**, 67–71 (2017).
108. Sigl, C. *et al.* Programmable icosahedral shell system for virus trapping. *Nat. Mater.* **20**, 1281–1289 (2021).
109. Wintersinger, C. M. *et al.* Multi-micron crisscross structures grown from DNA-origami slats. *Nat. Nanotechnol.* 1–9 (2022).
110. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* eadd1964 (2022).

111. Courbet, A. *et al.* Computational design of mechanically coupled axle-rotor protein assemblies. *Science* **376**, 383–390 (2022).
112. Dawson, W. M. *et al.* Coiled coils 9-to-5: rational de novo design of α-helical barrels with tunable oligomeric states. *Chem. Sci.* **12**, 6923–6928 (2021).
113. Fallas, J. A. *et al.* Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).
114. Bermeo, S. *et al.* De novo design of obligate ABC-type heterotrimeric proteins. *Nat. Struct. Mol. Biol.* **29**, 1266–1276 (2022).
115. VanAernum, Z. L. *et al.* Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. *Nat. Protoc.* **15**, 1132–1157 (2020).
116. Marty, M. T. *et al.* Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).
117. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
118. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
119. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
120. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
121. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
122. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
123. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
124. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J.*

Comput. Chem. **25**, 1605–1612 (2004).

125. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
126. Frenz, B. *et al.* Automatically Fixing Errors in Glycoprotein Structures with Rosetta. *Structure* **27**, 134–139.e3 (2019).
127. Wang, R. Y.-R. *et al.* Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5**, (2016).
128. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. Preprint at (2015).
129. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
130. Dowling, Q. M. *et al.* Computational design of constitutively active cGAS. *Nat. Struct. Mol. Biol.* **30**, 72–80 (2023).
131. Ablasser, A. *et al.* cGAS produces a 29-59-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, (2013).
132. Decout, A., Katz, J. D., Venkatraman, S. & Ablasser, A. The cGAS-STING pathway as a therapeutic target in inflammatory diseases. *Nat. Rev. Immunol.* **21**, 548–569 (2021).
133. Le Naour, J., Zitvogel, L., Galluzzi, L., Vacchelli, E. & Kroemer, G. Trial watch: STING agonists in cancer therapy. *Oncoimmunology* **9**, 1777624 (2020).
134. Wu, J.-J., Zhao, L., Hu, H.-G., Li, W.-H. & Li, Y.-M. Agonists and inhibitors of the STING pathway: Potential agents for immunotherapy. *Med. Res. Rev.* **40**, 1117–1141 (2020).
135. Larkin, B. *et al.* Cutting Edge: Activation of STING in T Cells Induces Type I IFN Responses and Cell Death. *J. Immunol.* **199**, 397–402 (2017).
136. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. L. & Gao, G. Viral vector platforms within the gene therapy landscape. *Signal Transduct Target Ther* **6**, 53 (2021).
137. Buschmann, M. D. *et al.* Nanomaterial Delivery Systems for mRNA Vaccines. *Vaccines (Basel)* **9**, (2021).
138. Holleufer, A. *et al.* Two cGAS-like receptors induce antiviral immunity in Drosophila. *Nature* **597**, 114–118 (2021).

139. Slavik, K. M. *et al.* cGAS-like receptors sense RNA and control 3’2'-cGAMP signalling in Drosophila. *Nature* **597**, 109–113 (2021).

140. Hornung, V., Hartmann, R., Ablasser, A. & Hopfner, K.-P. OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat. Rev. Immunol.* **14**, 521–528 (2014).

141. Li, X. *et al.* Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization. *Immunity* **39**, 1019–1031 (2013).

142. Zhou, W. *et al.* Structure of the Human cGAS-DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell* **174**, 300–311.e11 (2018).

143. Andreeva, L. *et al.* cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein–DNA ladders. *Nature* (2017) doi:10.1038/nature23890.

144. Xie, W. *et al.* Human cGAS catalytic domain has an additional DNA-binding interface that enhances enzymatic activity and liquid-phase condensation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11946–11955 (2019).

145. Hall, J. *et al.* The catalytic mechanism of cyclic GMP-AMP synthase (cGAS) and implications for innate immunity and inhibition. *Protein Sci.* **26**, 2367–2380 (2017).

146. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).

147. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).

148. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268 (2004).

149. Richter, V. *et al.* Structural and functional analysis of mid51, a dynamin receptor required for mitochondrial fission. *J. Cell Biol.* **204**, 477–486 (2014).

150. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2011).

151. Whiteley, A. T. *et al.* Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).

152. Kranzusch, P. J., Lee, A. S.-Y., Berger, J. M. & Doudna, J. A. Structure of human cGAS reveals a

conserved family of second-messenger enzymes in innate immunity. *Cell Rep.* **3**, 1362–1368 (2013).

153. Du, M. & Chen, Z. J. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* **1022**, 1–10 (2018).

154. Liu, H. *et al.* Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature* (2018) doi:10.1038/s41586-018-0629-6.

155. Lau, Y.-T. K. *et al.* Discovery and engineering of enhanced SUMO protease enzymes Downloaded from SUMO protease engineering. **5**, (2018).

VITA

Quinton Dowling grew up in Seattle Washington and completed his undergraduate studies at the College of Idaho in Caldwell Idaho in 2009, where he received a Bachelors of Science with honors, majoring in Biology and Chemistry. At the College of Idaho Quinton performed undergraduate research using molecular dynamics to investigate heparin sulfate proteoglycans. Quinton also studied heavy metal contamination in aquatic organisms using a variety of analytical tools. Upon completion of his degree, and summer working as a fly fishing guide in Alaska, Quinton worked at the Infectious Disease Research Institute (IDRI) in Seattle Washington as a research associate. There Quinton developed analytical methods for characterizing protein-adjuvant interactions and became an expert in nanoparticle and protein characterization in complex matrices. Quinton also developed a passion for vaccines and global health while at IDRI. After joining the Bioengineering graduate program at the University of Washington, Seattle, Quinton pursued his long-standing interest in proteins and the application of protein design to vaccines in the labs of Neil Kind and David Baker. Quinton is now an expert in computational protein engineering and design, protein assembly theory, and numerous analytical and experimental methods for protein characterization. Outside of the lab Quinton is a father who enjoys spending time with his family, especially in the forests, mountains, and waterways surrounding Seattle.