Journal of Medicinal Chemistry

pubs.acs.org/jmc

Article

Toward the Design of Allosteric Effectors: Gaining Comprehensive Control of Drug Properties and Actions

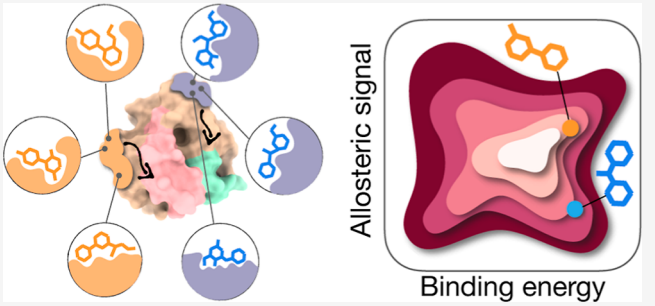
Wei-Ven Tee,*[,§] Sylvester J. M. Lim,[§] and Igor N. Berezovsky*

Cite This: https://doi.org/10.1021/acs.jmedchem.4c01043

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: While the therapeutic potential of allosteric drugs is increasingly realized, the discovery of effectors is largely incidental. The rational design of allosteric effectors requires new state-of-the-art approaches to account for the distinct characteristics of allosteric ligands and their modes of action. We present a broadly applicable computational framework for obtaining allosteric site−effector pairs, providing targeted, highly specific, and tunable regulation to any functional site. We validated the framework using the main protease from SARS-CoV-2 and the K-Ras^{G12D} oncoprotein. High-throughput per-residue quantification of the energetics of allosteric signaling and effector binding revealed known drugs capable of inducing the required modulation upon binding. Starting from fragments of known well-characterized drugs, allosteric effectors and binding sites were designed and optimized simultaneously to achieve targeted and specific signaling to distinct functional sites, such as, for example, the switch regions of K-Ras^{G12D}. The generic framework proposed in this work will be instrumental in developing allosteric therapies aligned with a precision medicine approach.



INTRODUCTION

Started from early observations in a number of enzymes, a decades-long journey and progress have unveiled to us diverse forms and mechanisms of allosteric regulation[1−3]—an omnipresent phenomenon where the protein activity is modulated by distant perturbations. There have also been increasing efforts toward exploring and realizing the potential of allostery for therapeutic interventions,[4−6] as allosteric effectors/drugs provide several advantages compared to traditional orthosteric drugs.[7,8] Traditional drug discovery typically focuses on identifying molecules that bind in the active sites of proteins and block interactions with endogenous substrates. A key distinction between allosteric and orthosteric modes of action is that in the former, conformational changes at the location of the functional site are induced by long-range signaling through the protein dynamics upon a perturbation by the distant binding of the effector molecule.[1,9] Remote modulation of the protein function makes it possible to target proteins with active/functional sites proved to be challenging for conventional drug discovery.[4,5,10]

The nonconservative nature of allosteric sites and drugs/effectors provides a foundation for obtaining specific and selective medicines with lower off-target toxicity and higher target specificity in the paradigm of precision medicine.[5] This is especially important for major drug targets that are highly conserved such as G protein-coupled receptors (GPCRs) and kinases.[11,12] In general, varied allosteric responses can be specifically elicited by different perturbations, e.g., effector binding at distinct or overlapping sites. It was proposed that all allosterically triggered processes, such as side-chain rearrangements and alterations of electrostatic interactions, to name a few, can be universally described through the energy redistribution caused by external perturbations.[13] A computational protocol was demonstrated to identify rescue mutations in the marginally stable von Hippel−Lindau tumor suppressor by analyzing correlated fluctuations from molecular dynamics simulations.[14] It was suggested that drugs mimicking the stabilization effects of these distant rescue mutations may be designed to recover the function of the oncoprotein. A site-directed chemical screening approach was used to identify various selective small molecules that target a functionally conserved allosteric site on the protein kinase PDK1, known as the PIF pocket.[15] Based on the bidirectional allosteric communication between the ATP-binding site and regulatory PIF pocket in PDK1 kinase, small molecules binding to the ATP site were shown to produce allosteric effects at the PIF pocket, suggesting their therapeutic potential.[16] The mechanisms, energetics, and the synergy of allosteric signaling were characterized[17] using a combination of amide hydrogen−deuterium exchange mass spectrometry (HDX-MS) and computational modeling based on the

Received: May 3, 2024
Revised: September 3, 2024
Accepted: September 9, 2024

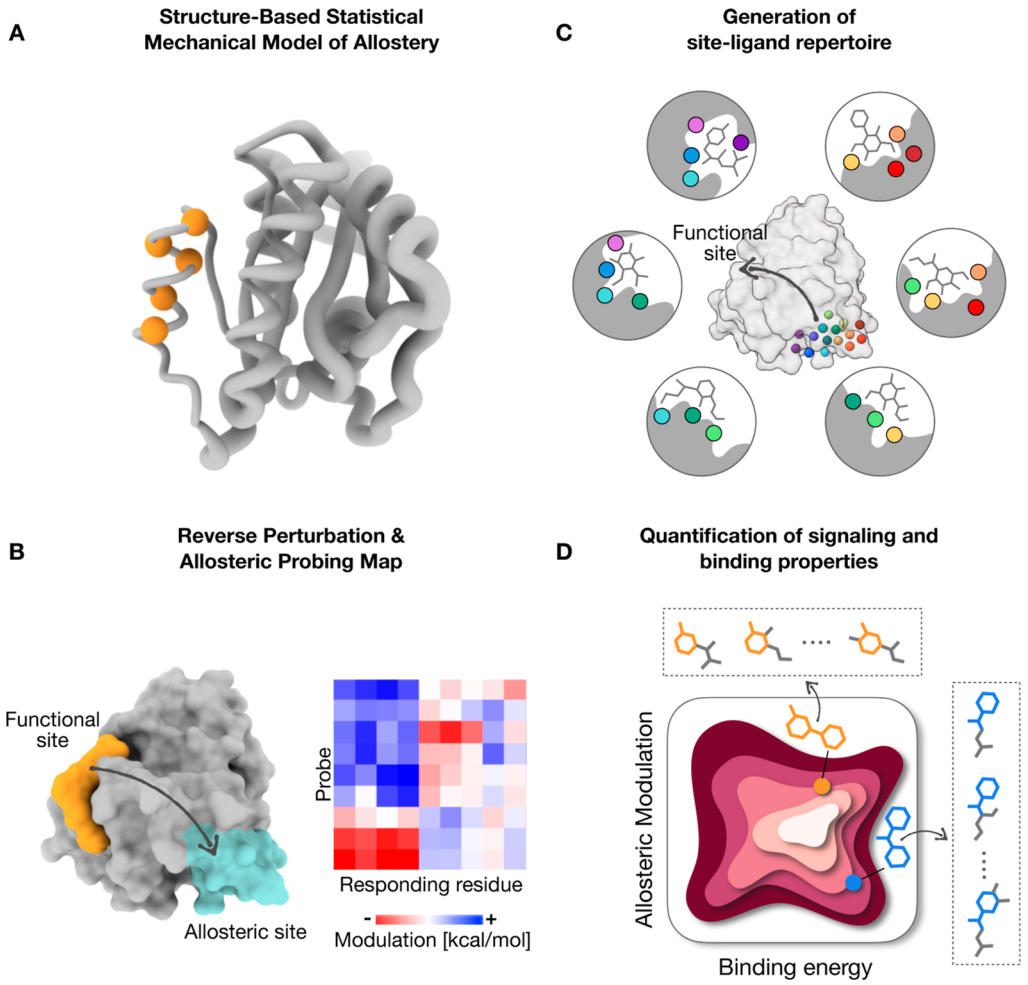


Figure 1. Key components of the generic design framework. (A) Illustration of the per-residue energetics caused by the ligand binding at residues shown as orange spheres, quantified using the SBSMMA. The radius of the tube-like representation depicts the corresponding per-residue allosteric modulation values (kcal/mol). In this example, binding of a ligand (yellow spheres) stabilizes the surrounding residues (thinner tube) and results in increased dynamics (thicker tube) at distant residues including those involved in the protein function. (B) Left: based on the reversibility of allosteric communication, a perturbation (e.g., ligand binding, yellow) at a functional site originates a signal to distant allosterically coupled locations (light blue) which may contain known or latent allosteric sites. Right: the APM is a high-throughput method that simulates binding of a small probe ligand to every three-residue segment and quantifies resulted allosteric modulation (blue: positive modulation pointing to potential conformational changes; red: negative values indicating stabilization) at every residue. It provides a comprehensive description of allosteric signaling in a protein structure allowing identification of residues that elicit a desired modulation to any residue(s). Both methods can be used independently or in conjunction to yield a list of residues that likely constitute latent allosteric sites. (C) Residues shortlisted in part (B) shown as spheres allow obtaining a repertoire of corresponding site−ligand pairs via methods including molecular docking. Each site−ligand pair describes a bound ligand and a composition of interacting residues. (D) The site−ligand pairs are subsequently evaluated and optimized for allosteric modulation, binding free energy, and other properties. For instance, further derivation and design of potential site−effector pairs can be achieved with fragment-based design using substructures of selected allosteric effectors and obtaining corresponding site residues.

structure-based statistical mechanical model of allostery (SBSMMA[18,19]). Using a disulfide-fragment-based screening approach, the laboratories of Wells and Shokat identified K-Ras^{G12C} covalent inhibitors that bind irreversibly to the mutant cysteine at a pocket just outside the switch II to disrupt both switch I and II.[20] One of the small molecules has led to the covalent inhibitor sotorasib (described below), and more recently, a follow-up noncovalent inhibitor binding preferentially to the inactive state of K-Ras has been developed[21] based on the mode of action of sotorasib. Above approaches, as well as most of the other methods, utilize unique aspects of target proteins to identify potential effectors/drug candidates. There are, however, no generic screening/development frameworks for rational design of allosteric drugs so far, besides the recently proposed *the directed design protocol*.[5] In this protocol, we argued for simultaneous design and adjustment of corresponding allosteric sites and prospective effectors based on the exhaustive quantification of allosteric signaling and optimization of other properties. To this end, the chief goal of this work is to implement this protocol in a broadly applicable computational framework for exploring and designing allosteric communication in proteins and for early phase identification/design of new effector drugs, which could be instrumental in facilitating experimental screening and validation in the overall drug design efforts.

We have previously developed the SBSMMA[18,19] which quantifies the energetics of allosteric signaling upon ligand binding,[22−24] mutations,[22,25] and other perturbations.[26] SBSMMA has been extensively benchmarked[25,27] and widely employed to investigate a range of allosteric phenomena. It is

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

also implemented as a suite of freely accessible online resources known as AlloSigMA[28] and AlloMAPS.[29] The exhaustive description of allosteric communication at a single-residue resolution by SBSMMA, a major component in the framework presented here, provides important advantages in the fine-tuning of effector candidates via fragment-based design and designated binding sites. A library of well-characterized drugs/drug candidates was used as a starting points for the identification of effector drugs through virtual screening on the allosteric sites identified by the SBSMMA, followed by the high-throughput quantification of the allosteric modulation and binding free energy.

We focus here on two case studies: the main protease (M^{Pro}) from SARS-CoV-2 and K-Ras^{G12D} oncoprotein are exemplified as targets for allosterically acting drug candidates. These proteins are therapeutic targets under intense clinical focus for COVID-19[30] and cancers,[31] respectively. The former is vital for the proteolytic processing of viral polyproteins, and the latter is a key GTPase in signal transduction pathways. They were previously considered as challenging drug targets largely due to the stereochemical properties of the active sites, which hampered developments of orthosteric drugs.[32,33] The inhibitors for M^{Pro} and K-Ras are being actively and widely pursued;[30,34] currently available medicines include the nirmatrelvir/ritonavir combination,[35] which orthosterically targets M^{Pro} (the former is a covalent inhibitor targeting the proteolytic site) and sotorasib (an approved allosteric drug) for patients harboring the K-Ras^{G12C} mutation.[36] Nevertheless, ritonavir can lead to side effects via the inhibition of CYP3A enzymatic activity[37] within the antiviral treatment, and sotorasib relies on forming a covalent bond with the active site cysteine presents only in the K-Ras^{G12C} mutant.

In this work, we show how new allosteric effectors can be built from molecular fragments of known drugs and/or drug candidates to achieve fine-tuning of allosteric modulation. We also show that signaling originating from an allosteric site can be adjusted to target different remote sites, an important characteristic and advantage of allosteric effectors. For instance, additional interactions with three residues in an allosteric site in K-Ras lead to a specificity in the action of effectors, exemplified by the targeted modulation of switches I and II. Overall, this work marks a change of strategy in the quest of allosteric effectors in multiple aspects: a transition from the screening of existing ligand libraries to fragment-based rational design, and adjustments of novel compounds to bind designated binding sites for obtaining highly specific and tunable allosteric signaling.

RESULTS

Generic Computational Framework for the Identification and Design of Allosteric Sites and Effectors. We developed and implemented here a computational framework, which is based on *the directed design protocol* that we recently proposed for allosteric drug development.[5] The framework identifies allosteric sites on a target protein and yields a set of allosteric ligands that bind to a designated site. Figure 1 illustrates key components of the framework.

Structure-Based Statistical Mechanical Model of Allostery. The computational framework developed here is based on the SBSMMA[18,19]—a physical model that describes causality of allosteric communication and quantifies it via free energy of the allosteric signaling (in real energy units, kcal/mol). The allosteric signal can originate from different types of perturbations, including but not limited to ligand binding and mutations (Figure 1a). The SBSMMA consists of three major steps. First, it takes a protein structure as an input and constructs a C_α elastic network representing it, the dynamics of which is simulated on the basis of the C_α harmonic model (see Methods for the energy function in eq 1). Ligand binding is modeled as a perturbation of the energy function through an inclusion of harmonic restraining terms among binding residues (shown as yellow spheres), which simulates a stabilization effect due to interactions with a ligand. Here, the protein dynamics in native and perturbed states are characterized by respective sets of low-frequency normal modes. In other implementations, protein dynamics may also be explored using principal components of the covariance matrices of trajectories obtained in molecular dynamics simulations. Second, the normal modes are used as an input to calculate the per-residue allosteric potential in free and perturbed states (see eq 2 in the Methods), allowing us to evaluate elastic work exerted on a residue as a result of the dynamics of neighboring residues in both states. Lastly, the allosteric potential is used for calculation of the partition function (eq 3 and Methods) to estimate the per-residue free energy of the allosteric signal Δg on targeted residues originated by the ligand binding (eq 4 and Methods). To estimate the work caused by purely allosteric signaling, the per-residue allosteric modulation Δh (kcal/mol) is calculated as the difference of Δg from the average value over all residues of the protein chain (eq 5 and Methods). A positive modulation value ($\Delta h > 0$) indicates work exerted on involved residues, which might lead to conformational changes, whereas a negative value ($\Delta h < 0$) points to decrease in dynamics of the affected residues, suggesting overstabilization of corresponding parts of the protein. To compare the effects at different sites, allosteric modulation can be evaluated by averaging the per-residue Δh over all residues forming the site of interest.

Reverse Perturbation and APM. The framework utilizes the reverse perturbation[27,38] and allosteric probing map (APM[28]) approaches previously proposed and developed by us. A generic reverse perturbation concept[38] was implemented for a large set of proteins[27] to detect latent/known allosteric sites using information on functional sites and for characterizing their regulatory functions and drug design potential.[8,22,24] Noteworthy, an independent experimental work[16] demonstrated a presence of bidirectional allosteric communication between the ATP-binding and PIF pocket in kinases. Alterations of protein dynamics associated with the reversible signaling across both sites were also characterized using HDX-MS and SBSMMA.[17] Recently, the reverse perturbation concept has been adapted and utilized by other groups for various tasks.[39,40] Here, we use the reverse perturbation approach for identifying distant protein residues that are allosterically coupled to a site of interest, including the functional or active site of the protein (Figure 1b, left). First, a binding perturbation to a known functional site is performed using a protein structure and a set of known functional site residues as inputs. Then, distant residues (with C_α–C_α distance at least 11 Å to all functional site residues, the cutoff distance used in the elastic network model of protein) exhibiting the strongest allosteric response (i.e., high-magnitude modulation values) are shortlisted. The exhaustive APM is computed by simulating binding of a small probe ligand to every three-residue segment/triplet in a protein chain and evaluating the resulting modulation at each residue (Figure 1b, right). It allows identification of positions in a protein structure that originate a desired signaling to any single residue or residue

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

group in the targeted site. The computation of APM requires only a protein structure as an input and information on a functional site is not necessary.

Depending on the tasks, one may shortlist segments/residues that elicit required modulation to a distant functional site known a priori or specific residue(s) that may play a functional role. Importantly, due to the extensivity of the free energy, the exhaustive APM may be utilized to explore and estimate the allosteric signal as a sum of contributions from various combinations of probed three-residue segments. The APM and reverse perturbation are typically used as complementary approaches. The former can be used to evaluate allosteric signaling that originated from binding to individual segments, including residues strongly linked to the functional site determined by the latter approach. Analyzing allosteric communication linking regulated and regulatory sites in opposite directions, both APM and reverse perturbation methods can be used independently or in conjunction. A combination of top-ranking distant residues/segments shortlisted by the reverse perturbation and/or APM method yields a list of residues strongly coupled to the remote functional site, delineating protein regions/locations consisting of latent allosteric sites with varied residual composition.

Generation of Site−Ligand Repertoire. In this component, residues in allosterically linked locations identified by the reverse perturbation and APM approaches form a basis for the screening or design of the corresponding ligands. Structure-based molecular docking of compound libraries and de novo molecular design can be used to sample interactions of various ligands to (non)overlapping sets of binding residues with sufficient estimated binding energy (Figure 1c). Virtual screening is demonstrated here with AutoDock Vina[41] (see also Methods for details), and docking results with very low estimated binding free energy were discarded. Since the reverse perturbation and APM approaches are based on the analysis of protein dynamics on a per-residue resolution, they reveal and quantify signaling to and from single residues in sites/pockets apparent in static crystal structures as well as residues that may form relatively transient binding sites. Noteworthy, commonly used definitions and prediction for the so-called cryptic sites,[42−45] emerging in protein dynamics as transient binding pockets, generally utilize properties of known binding pockets of functional sites without explicit characterization on the allosteric signaling and not considering many other potential allosteric sites with properties distinct to those of functional pockets.[5,8] Recognizing the signaling as a prerequisite for defining the site as allosteric and in a full agreement with the original work[46] by Bowman and Geissler, which shows a multitude of distant regulatory sites capable of originating allosteric signaling to functional sites, including the "allosteric cryptic sites", we evaluate the allosteric modulation and binding strength characterizing each ligand and corresponding site (see below).

Quantification of Signaling and Binding Properties. The simultaneous evaluation of corresponding pairs of binding sites and ligands is an important part of the framework (Figure 1d). It estimates the energetics of allosteric signaling induced by an alteration of the protein dynamics upon binding, which should be optimized in parallel with binding affinity and other relevant properties. Shortlisted pairs of binding sites and ligands with required specifics of allosteric modulation and affinity, allosteric site and effector pairs, serve as starting points for further derivation and design of the chemical structures of effectors and residual composition of binding sites. It is followed by evaluation and fine-tuning of the signaling, binding, and other properties. Importantly, statistics and descriptors on characteristics of known allosteric drugs/drug candidates are utilized to complement the screening/design protocols and for shortlisting of site−effector pairs within the framework.

In the current implementation of the generic framework outlined above, we used the advantages of well-characterized drugs/drug candidates to efficiently utilize the information and benefits of characterized "drug-likeness", pharmacokinetics, and effects of available therapeutics for new modes/mechanisms of actions or targets. Here, molecular docking of the Drug Repurposing Hub (DRH)[47] revealed approved and clinical trial drugs binding to latent allosteric sites on protein(s) that are not original targets for these drugs. Noteworthily, while the pharmacological and therapeutic profiles of repositioned orthosteric drugs are typically well-characterized, allosteric and orthosteric compounds/sites can exhibit distinct physicochemical and molecular properties,[8,48] which should be taken into account. On the basis of known drugs providing a source of drug-like molecular fragments for fragment-based design, new "allosteric drug-like" effectors were derived for various (non)-overlapping allosteric sites with diverse features and requirements.

We have previously argued for the simultaneous design and adjustment of site−effector pairs by evaluation and optimization of allosteric signaling and other properties, termed as the directed design protocol.[5] Here, the framework based on the directed design protocol yields allosteric site−effector pairs after optimizing them toward desired characteristics, including but not limited to (i) sufficient "drug-likeness" with respect to orally administered medicines and (ii) "allosteric drug-likeness" based on stereotypical properties of known allosteric drugs/effectors;[8] (iii) strong binding affinity to regulatory sites; (iv) required allosteric modulation of the protein function; and (v) specific long-range signaling to precisely target protein residues/sites. The framework is demonstrated using M^{Pro} and K-Ras^{G12D} as case studies, each showing important aspects in the discovery and design of allosteric medicines. For example, the M^{Pro} case study illustrates how one can derive new compounds with molecular properties typical for allosteric effectors and the required level of drug-likeness, based on fragments of previously known orthosteric drugs instrumental for originating targeted allosteric responses. In the latter case study, a unique approach on the design of new effectors highly specific in the distinct regulation of switch regions I and II was presented.

CASE STUDY 1: SARS-COV-2 M^{PRO}

Revealing Potential Allosteric Sites Including the Domain I/II Hinge Region. The M^{Pro} harbors a catalytic dyad formed by His41 and Cys145 in the catalytic cleft situated between 2 six-stranded β barrels known as domains I (residues 10−99) and II (residues 100−182). Domain III (residues 198−306) with five α helices is involved in the dimerization of the enzyme, which is functional only as homodimers. In an X-ray crystallographic screen, clinical trial drugs pelitinib and AT7519 were discovered to bind outside of the catalytic cleft and showed moderate antiviral activity in viral reduction assays.[49] However, the mechanism underlying the antiviral activity, specifically the involvement of allosteric communication, was not elucidated from the crystallographic structures. Information on the existing allosteric inhibitors, including their binding sites, allowed us to first evaluate to what extent the SBSMMA, a key component of the framework, accounts for the allosteric modulation. Binding

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

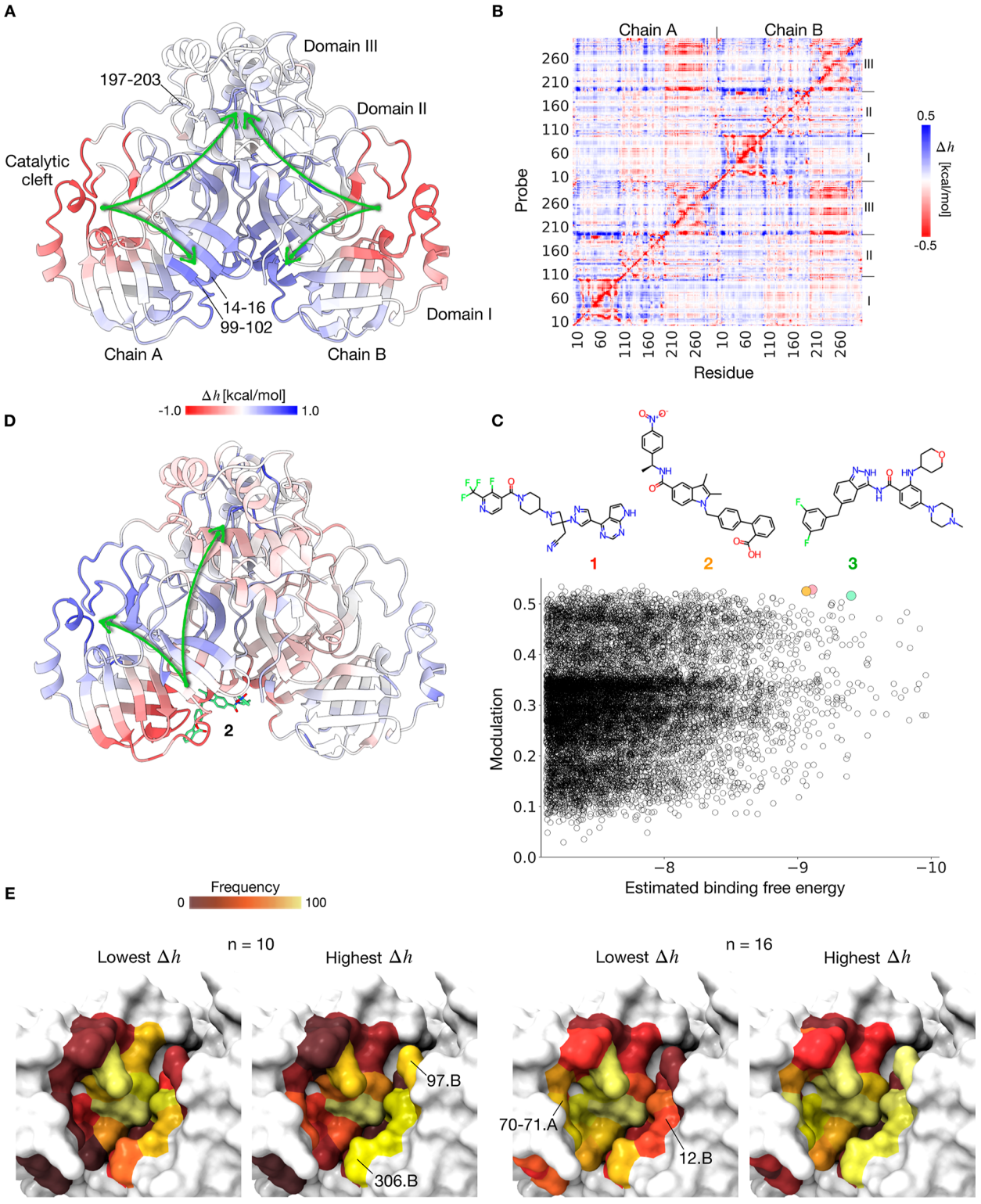


Figure 2. Case study of the SARS-CoV-2 M^{Pro}. (A) Modulation values upon reverse perturbation at the catalytic clefts are illustrated on the crystal structure of M^{Pro} (PDB: 7jp1). Positive modulation shown in blue points to increased dynamics, whereas negative values in red indicate stabilization. Green arrows indicate signaling emanated to the domain I/II hinge and domain III interface. (B) APM containing exhaustive quantification of allosteric modulation (kcal/mol) originated by the binding of a small probe to every consecutive residue triplet (*y* axis) in each protein chain to all responding residues (*x* axis) of the dimer. The residue range of monomers A/B and domains I−III of the M^{Pro} dimer are indicated. The exhaustive APM can be computed using the publicly available AlloSigMA Web server[28] within 1 h, and results can be downloaded in a machine-readable format. (C) Analysis of the modulation value and binding free energy (both in kcal/mol) of drugs/drug candidates from DRH docked to $L1_{MPro}$ of chain A. Compounds **1**−**3** with large modulations and binding energies are illustrated. (D) Allosteric modulation induced by binding of **2**. (E) Heatmaps of contact frequencies of binding site residues in $L1_{MPro}$ obtained from 100 site−effector pairs (n = 10 and 16) with the highest and lowest modulation values. A spectrum from dark red to bright yellow is used to color residues from low to high contact frequencies.

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

residues within 4 Å from any heavy atom of pelitinib and AT7519 were obtained from the respective ligand-bound crystal structures (PDBs: 7axm and 7aga). Then, SBSMMA was used to measure the allosteric modulation caused by perturbations of the binding sites (see Methods for detail) in a substrate and inhibitor-free M^{Pro} dimer structure (PDB: 7jp1). Both pelitinib and AT7519 cause negative allosteric modulation in the helical portion of domain III, suggesting stabilization near their binding sites (Figure S1). We also observed positive modulation at the distant catalytic cleft and domain III dimer interface (0.42 and 1.2 kcal/mol), suggesting conformational changes upon AT7519 binding. For pelitinib, the interface is also positively modulated but not the catalytic cleft. Overall, we found that pelitinib and AT7519, effectors identified from previous experimental screening, allosterically destabilize the distant functional sites via alterations in the structure and dynamics, affecting the enzyme activity and/or dimer formation.

The computational framework was applied to demonstrate a rational approach (Figure 1) for the identification and design of latent allosteric sites and prospective effector drugs. Reverse perturbation by simulating substrate binding at catalytic clefts showed positive modulation exerted predominantly to residues in the domain I/II hinge and, to a lesser extent, the domain III interface (Figure 2a). The allosterically coupled regions identified by the method also overlap with the AT7519 effector binding site, corroborating the allosteric mode of action. The APM for the M^{Pro} dimer (chains A and B) was computed to obtain an exhaustive per-residue description of allosteric signaling upon simulated probe binding to each residue triplet (Figure 2b). It revealed major patterns and specifics of allosteric communication across regions, domains, and monomers, for example, (i) probe binding positions in domain I often result in positive modulation in domain II of the same monomer; (ii) probe binding to domain III positively modulates domain I/II of both monomers; and (iii) stabilizing the flexible connecting loop between domain II/III (residues 190−200) of each monomer (by interactions with probes) may result in decreased dynamics of residues in domain III (negative modulation); at the same time, it sends allosteric signals to domains I/II of chains A and B that may lead to conformational changes. Using the APM, we identified residue triplets that can induce the strongest allosteric modulation to distant catalytic cleft residues upon perturbation by a small probe. Two clusters of triplet positions, which might form allosteric sites, were found to exert positive modulation in the catalytic cleft, indicating conformational changes (Figure S1). The first location $L1_{MPro}$ is centered at residues 14−16 and 99−102 near the domain I/II hinge (opposite face of the monomer relative to the catalytic cleft, indicated in Figures 2a and S1). The second area $L2_{MPro}$ located in the vicinity of residues 197−203 in domain III causes weaker modulation in the catalytic cleft (Figure S1). Together, the results from APM and reverse perturbation simultaneously point to the $L1_{MPro}$ allosteric location at the domain I/II hinge, consistent with previous experimental and computational findings.[50,51]

Toward the Repurposing of Drugs for Allosteric Regulation of the M^{Pro} Catalytic Cleft. Molecular docking was performed on $L1_{MPro}$ to obtain binding poses of existing drugs from the DRH[47] (with about 6800 clinical and launched drugs). For each ligand, several docked poses were considered to elucidate how compositional differences of binding sites (all residues within 4 Å from any heavy atom of the ligand) may fine-tune the signaling, providing the information for designing new allosteric effectors and sites. Prioritizing those with sufficient interaction strength (within the top 20% of all evaluated binding free energies), we obtained a repertoire of site−effector pairs with varied combinations of residues bound by different effectors. Subsequently, respective allosteric modulations at the distant catalytic cleft (averaged per-residue Δh) originated by corresponding site−effector pairs were calculated using SBSMMA.[18,19] The results show that site−effector pairs at $L1_{MPro}$ induce a range of positive modulation values (Figure 2c). The majority of site−effector pairs lead to modulation values of about 0.3 kcal/mol at each catalytic cleft residue, and some pairs cause even stronger modulation than the known effector AT7519 described above (Figure S1). This result illustrates that while ligands may bind strongly to a location, the perturbation caused by the ligand binding may not necessarily induce a large modulation at targeted distant sites. The two-step action of allosteric effectors—binding and sending a modulatory allosteric signal—is an important distinction in comparison to traditional orthosteric drugs, which works via competitive binding to a functional site. The strengths of binding affinity and of allosteric modulation depend individually on characteristics of the targeted protein—the former is largely determined by the intermolecular interactions and the latter resulted from an alteration of protein dynamics. Both binding and signaling aspects are chiefly independent of each other, being determined by a number of characteristics, such as structure, size, and oligomerization state of the protein, structures and compositions of its allosteric and functional sites, and others. The binding and signaling values may, thus, be individually controlled and adjusted in designing of prospective allosteric drugs. The site−effector repertoire can be further analyzed to identify those with optimal modulation and binding capability, as well as other properties. For example, itacitinib, SR-1664, and entrectinib (compounds **1**−**3**) showed large modulations and binding energies. Compound **2** is an antagonist for the PPARγ nuclear receptor, whereas compounds **1** and **3** are kinase inhibitors. Details on interacting residues in respective binding sites of compounds **1**−**3** and elicited per-residue modulations can be found in Figure S1 and Table S1. Figure 2d illustrates that long-range allosteric signaling induced by binding of compound **2** (SR-1664) at $L1_{MPro}$ of chain A (which also interacts with chain B residues) causes positive modulations at both the catalytic cleft and dimer interface, indicating potential conformational changes that disrupt the protein function and dimerization. The parts A and D of Figure 2 also illustrate that allosteric communication is reversible: perturbation of a functional site originates a signaling to an allosteric site and vice versa.

Elucidating Key Residues in Allosteric Communication for the Design of New Effectors. Prospective effectors should be able to initiate long-range signaling to a targeted site upon binding to an allosteric site with a sufficient affinity. Figure 2c shows that while some ligands may bind tightly to a protein site, the perturbation caused by their binding induces rather weak modulation. Therefore, methods for obtaining and tuning of site−effector pairs at the molecular level should consider two important aspects—affinity and allosteric signaling,[5,8] which are chiefly determined by specifics of perturbed site residues and intermolecular interactions. We focus here on elucidating key determinants of the strength of allosteric signaling exerted to the catalytic cleft. The catalytic cleft modulation is largely uncorrelated with the number of perturbed residues n at the distant $L1_{MPro}$ (Figure S2). On the other hand, the heatmaps in Figure 2e show that some site−effector pairs with specific

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

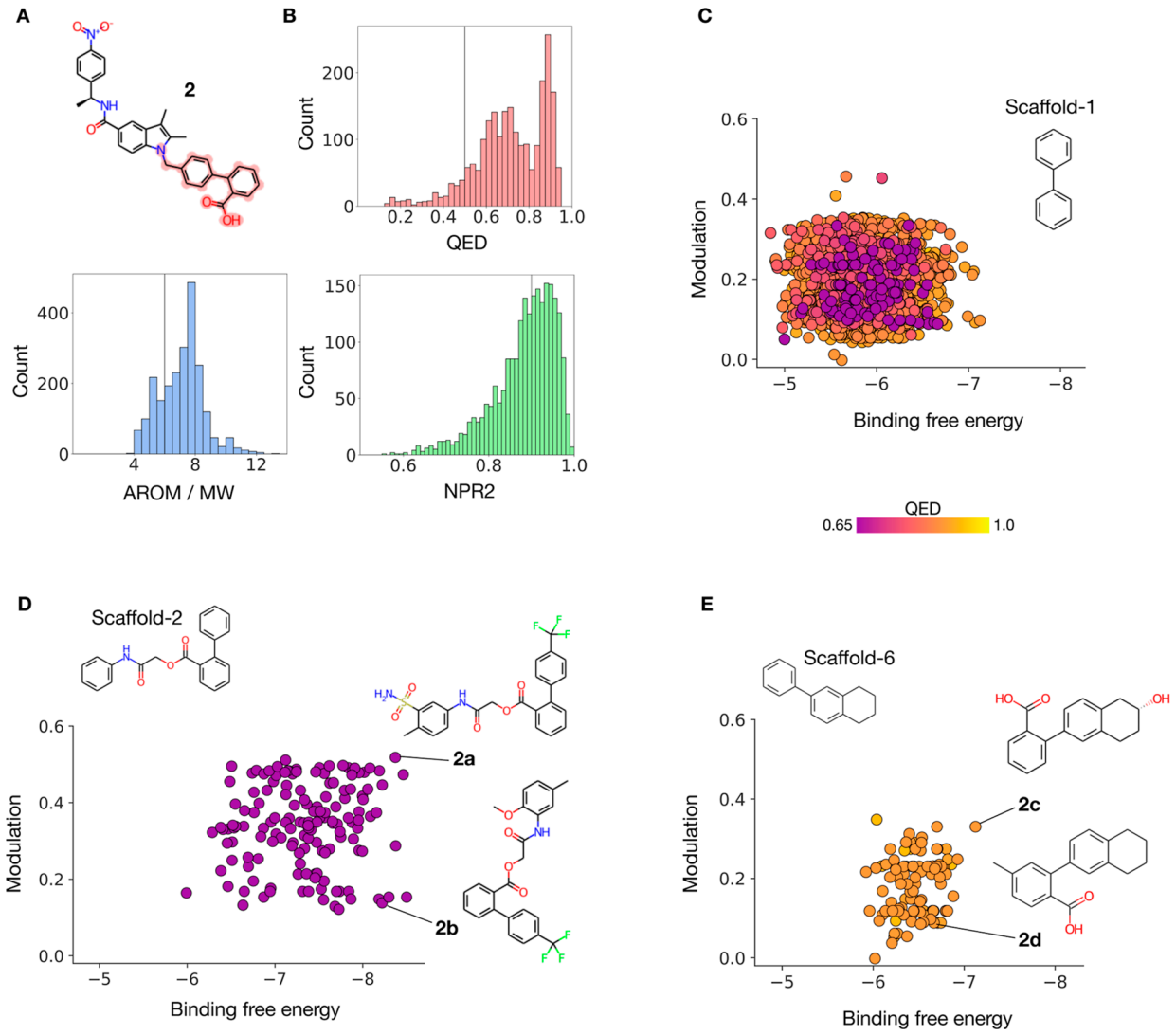


Figure 3. "Drug-likeness" in a transition from orthosteric to allosteric drug development. (A) Structure of **2** and the fragment used for analogue search is colored in red. (B) QED, AROM/MW, and NPR2 distributions of analogues derived from the fragment of **2** are colored in red, blue, and green, respectively. Vertical lines indicate criteria used for selecting analogues for subsequent SBSMMA and docking analysis. (C–E) Catalytic cleft modulation and binding free energy (kcal/mol) of all analogues with selected Bemis–Murcko scaffolds. A spectrum from purple to bright yellow is used to indicate increasing QED values of analogues. Analogues **2a**–**2d** are shown as examples.

interactions/contacts can induce a large allosteric response in the catalytic cleft. Comparing the frequencies of contact residues in 100 site–effector pairs ranked top in modulation versus the lowest ones (using n = 10 and 16, for instance) shows that perturbing residues Lys12.B, Lys97.B, and Gln306.B (in chain B) is crucial for inducing stronger allosteric signals. Full contact analysis with n from 10 to 16 also revealed other key residues including chain A residues Ala70.A and Gly71.A (Figure S2). Noteworthy, the above residues are also involved in the interactions with compounds **1**–**3** obtained from the DRH.[47] Specific residues obtained in the analysis of the binding sites further emphasize the demand for mutual design of site–effector pairs, in which adjustment and tuning of both, the site and the effector components, are of similar importance.

To obtain effectors for diverse allosteric sites, which, in principle, can exist at any location on a protein structure, one may explore beyond the chemical space populated by existing drugs and drug candidates that work orthosterically. We previously showed in a statistical analysis that allosteric and orthosteric drug candidates exhibit distinct properties. The statistics and knowledge of the salient properties of known allosteric compounds can be used to prioritize promising candidates. For example, allosteric effectors generally contain more aromatic rings and are typically flatter.[8] Using compound **2** (SR-1664), we show how new effectors with "allosteric drug-like" properties can be derived from existing drugs. We obtained molecular fragments of compound **2** using the retrosynthetic combinatorial analysis procedure (RECAP) method[52] (Figure S3) and selected the 2-phenylbenzoic acid part, which binds near key residues Ala70 and Gly71, for subsequent substructure-based analogue search (Figure 3a). The fragment is queried against the large-scale ZINC20 database[53] and yielded 2531 analogues for further investigation. Pairwise Tanimoto similarity shows the majority of analogue pairs with a coefficient at about 0.2, suggesting diversity in chemical structures (Figure S3). The quantitative estimate of the drug-likeness (QED), providing a continuous measurement of compound quality and potential as oral drugs,[54] allows elimination of those with very low likelihood of success in development. The criteria on flatness and aromatic ring contents were employed here for a preferential selection of

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX–XXX

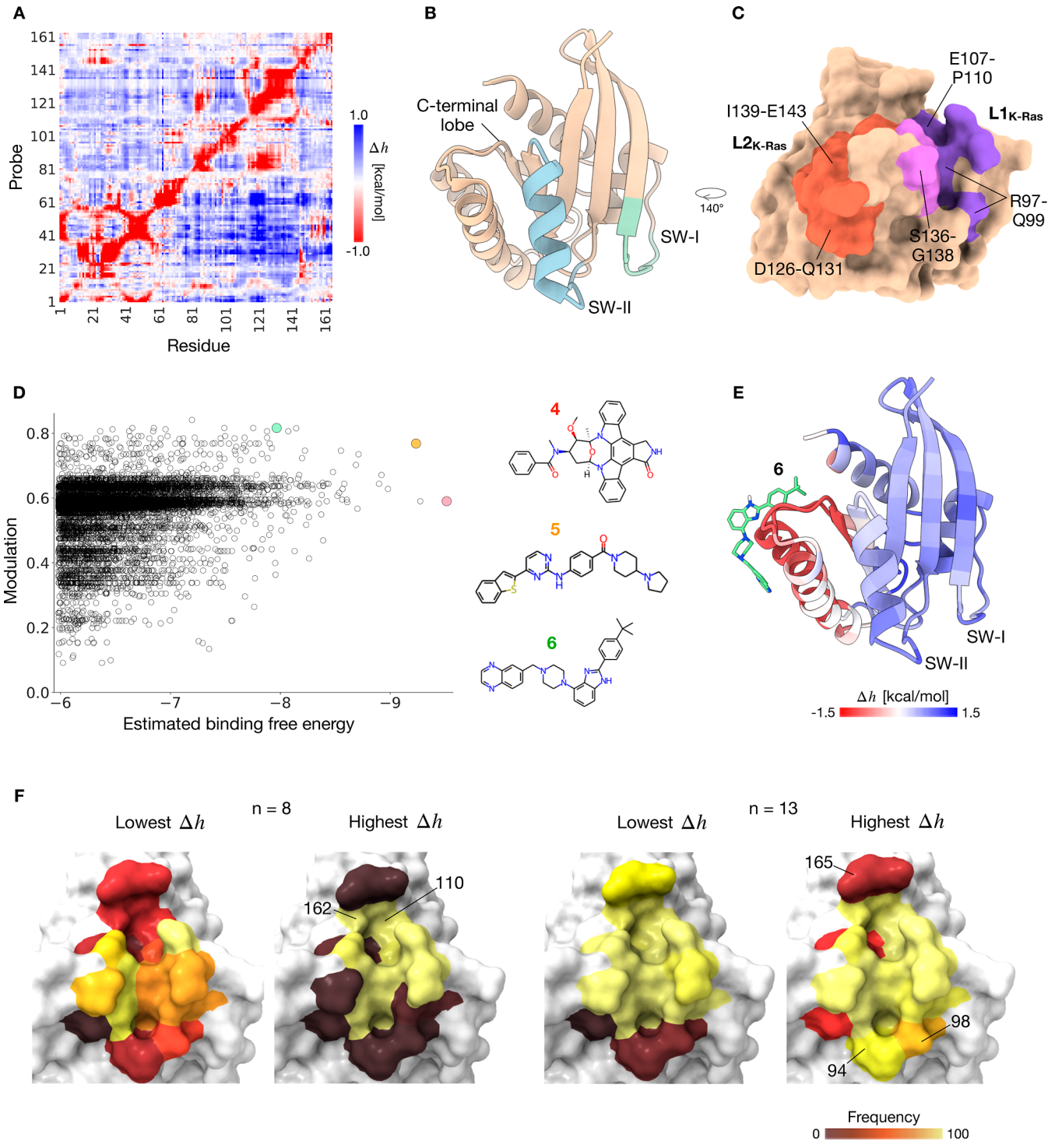


Figure 4. Case study of the K-Ras oncoprotein. (A) APM containing exhaustive quantification of allosteric modulation (kcal/mol) from all probed triplets to every residue in K-Ras^{G12D}. The exhaustive APM can be computed using the publicly available AlloSigMA Web server[28] in approximately 10 min and downloaded. (B) Crystal structure of K-Ras^{G12D} (PDB: 6gof) with SW-I/II and C-terminal lobe highlighted. (C) Residues constituting allosteric locations $L1_{K\text{-}Ras}$ (violet and magenta) and $L2_{K\text{-}Ras}$ (red and magenta). Both locations are overlapping at residues 136−138 (magenta). (D) Analysis of SW-I modulation and binding free energy of known drugs/drug candidates docked to $L1_{K\text{-}Ras}$ of K-Ras^{G12D}. For instance, compounds **4**−**6** with large modulations and binding energies are shown. (E) Allosteric modulation induced by binding of **6**. (F) Heatmaps of contact frequencies of binding site residues in $L1_{K\text{-}Ras}$ obtained from 100 site−effector pairs (n = 8 and 13) with the highest and lowest modulation values.

"allosteric drug-like" effectors (Figure 3b). To assess molecular shapes, we first calculated the principal moments of inertia (PMI) of all obtained analogues (Figure S3) and used the second normalized PMI ratio (NPR2) as a measure for flatness.[55] The number of aromatic rings normalized by the molecular weight was also evaluated. Additional data on the distributions from the normalized PMI analysis and on the molecular properties are available in Figure S3. After discarding potentially promiscuous analogues using the pan assay interference compounds (PAINS) filter[56] and those containing

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

chemical moieties with undesirable toxicity and pharmacokinetics using the Brenk filter,[57] followed by a selection of those satisfying the above criteria (QED ≥ 0.5, AROM/MW ≥ 6 kDa^{-1}, NPR2 ≤ 0.9; Figure 3b), 708 analogues of compound **2** were collected.

We next performed virtual screening (i.e., molecular docking) of the analogues on $L1_{MPro}$ followed by a quantification of modulations and binding energies. Stratifying site−effector pairs by structural scaffolds of analogues allows one to explore how modifications within same scaffold types can lead to different binding free energies and modulations. The Bemis−Murcko method was used to characterize scaffolds/frameworks defined as the union of ring systems and linker atoms.[58] The modulation and estimated binding free energy of every site−effector pair belonging to scaffolds 1, 2, and 6 are shown in Figure 3c−e. Results on 6 selected scaffolds out of a total of 170 are shown in Figure S4. Analogues featuring a biphenyl scaffold (scaffold 1), which is the simplest and most populated one, display wide spectra of modulations, binding energies, and QED values, reflecting a potential diversity of structural analogues that can be derived. Analogues belonging to scaffold 2 show uniformly lower QED values, but many of the site−effector pairs result in modulation and affinity stronger than those from scaffold 1. Comparison of analogues **2a** and **2b** indicates that minor modifications in effector structures can lead to different site−effector pairs with distinct binding modes (see Figure S4) and respective large changes in allosteric response (Figure 3d). Conversely, scaffold-6 analogues generally feature higher QED (more oral drug-like) and show a different distribution relative to scaffold-2—typically weaker modulation and binding strength. While analogue **2d** binds solely to residues in chain A, **2c** binds to both monomers and forms a salt bridge with Lys12 in chain B (Figure S4), which may induce strong positive modulation (Figure 2e).

The major result of the design approach, starting from targeting the identified allosteric sites with orthosteric drugs to deriving new effectors, is that strong binders and those showing "drug-like" characteristics established based on orthosteric ligands are not necessarily optimal as allosteric ligands. The latter should also satisfy requirements on specifics of allosteric signaling and perturbations (site−effector pairs) required for inducing the desired signaling. Depending on the tasks, therefore, the molecular characteristics of allosteric effectors should guide explorations of the chemical space, aimed at obtaining new structural species with both necessary binding affinity and potential for exerting an allosteric signal.

CASE STUDY 2: K-RAS ONCOPROTEIN

Identification of Allosteric Sites Specific for the Switch-I and Switch-II Regions. K-Ras is a monomeric GTPase with a catalytic domain that is highly conserved in the Ras family. Conformational changes in the flexible switch I (SW-I, residues Asp30-Asp38) and switch II (SW-II, residues Gly60-Glu76) regions of the catalytic domain play instrumental roles in various aspects of K-Ras regulation and function.[32] SW-I/II and the P loop constitute the catalytic site, and the former (mainly SW-I residues) interacts and activates different downstream effector proteins. Moreover, K-Ras activity is tightly regulated by guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) which interact with SW-I/II to facilitate GDP dissociation and GTP hydrolysis, respectively.[59] We focus here on the identification of allosteric sites capable of emanating specific and strong signaling to the SW-I and/or SW-II regions of the oncogenic $K\text{-}Ras^{G12D}$ (4B splice isoform), relative to the wildtype. To this end, comprehensive APMs were computed for the catalytic domains of $K\text{-}Ras^{G12D}$ (Figure 4a) and $K\text{-}Ras^{WT}$ (Figure S5). The results show probing residue triplets in the C-terminal lobe (residues 87−166, Figure 4b) generally causes large positive allosteric modulation in SW-I/II of both structures, suggesting conformational changes at these functional regions. From the APMs, we also obtained distant probed positions that originate much stronger modulation in the SW-I/II region of the $K\text{-}Ras^{G12D}$ variant, compared to the modulation in $K\text{-}Ras^{WT}$ (tabulated in Figure S5). Locations $L1_{K\text{-}Ras}$ and $L2_{K\text{-}Ras}$ were delineated based on these positions (Figure 4c), potentially forming allosteric sites to remotely modulate the protein dynamics of distant SW-I and SW-II of $K\text{-}Ras^{G12D}$, respectively. $L1_{K\text{-}Ras}$ (colored in violet and magenta) and $L2_{K\text{-}Ras}$ (red and magenta) are overlapping; perturbing residues 136−138 (magenta) is instrumental for sending allosteric signals to both switch regions. To our knowledge, both locations have received little attention for therapeutic developments. Emerging experimental data have shown that a small molecule termed KAL-21404358 (KAL)[60] and a monobody known as NS1[61] bind locations overlapping with $L1_{K\text{-}Ras}$ and $L2_{K\text{-}Ras}$, respectively, and allosterically inhibited K-Ras function. Here, SBSMMA shows that the binding of KAL and NS1 can remotely modulate the dynamics of SW-I/II and other distal regions (Figure S5).

Toward Designing New Effectors for Allosteric Regulation of the Switch-I Region. We next performed molecular docking of known drugs and drug candidates on $L1_{K\text{-}Ras}$, which is allosterically coupled with the SW-I region. Using the same protocol as that for the M^{Pro}, site−effector pairs were evaluated for modulation values at the remote SW-I region and predicted binding free energies at $L1_{K\text{-}Ras}$ (Figure 4d). Compounds **4**−**6** (midostaurin, IKK-16, and WAY-207024, respectively) exhibiting high modulations and binding free energies are shown here. The binding of **5** and **6** to $L1_{K\text{-}Ras}$ induces stronger allosteric modulation to SW-I/II than to KAL (Figure S5). For example, compound **6**, an antagonist for the gonadotropin releasing hormone receptor, binds to and stabilizes $L1_{K\text{-}Ras}$. At the same time, it causes extensive positive modulation in the N-terminal lobe containing functional sites, especially the SW-I/II regions (Figure 4e). Residues important for signaling were elucidated by comparing their contact frequencies in site−effector pairs ranked as the top and the lowest in terms of modulation (Figure 4f). Binding sites with n = 8 and 13, which result in modulation values ranging from 0.2 to 0.6 kcal/mol and from 0.45 to 0.7 kcal/mol, respectively, are shown here as examples (see Figure S2 for complete data). The heatmaps show that in cases of lower numbers of contact residues (indicative of smaller ligands, such as n = 8), intermolecular interactions in top-ranking site−effector pairs are rather confined to specific preferential residues, for example, Pro110 and Arg162. The critical role of Pro110 and Arg162 in allosteric signaling is corroborated by several experiments. Both residues form the binding site, dubbed as the P110 site by authors, for KAL.[60] Using deep mutational scanning, it was also shown that mutations at Pro110 and Arg162 disrupted the binding of Raf-1 effector protein from a distance.[62] In another example (n = 13, Figure 4f), perturbing residues His94 and Glu98 increases the modulation at SW-I at the expense of interactions with Lys165, showing the fine-tuning of signaling by mutual adjustments of the sites and effectors.

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

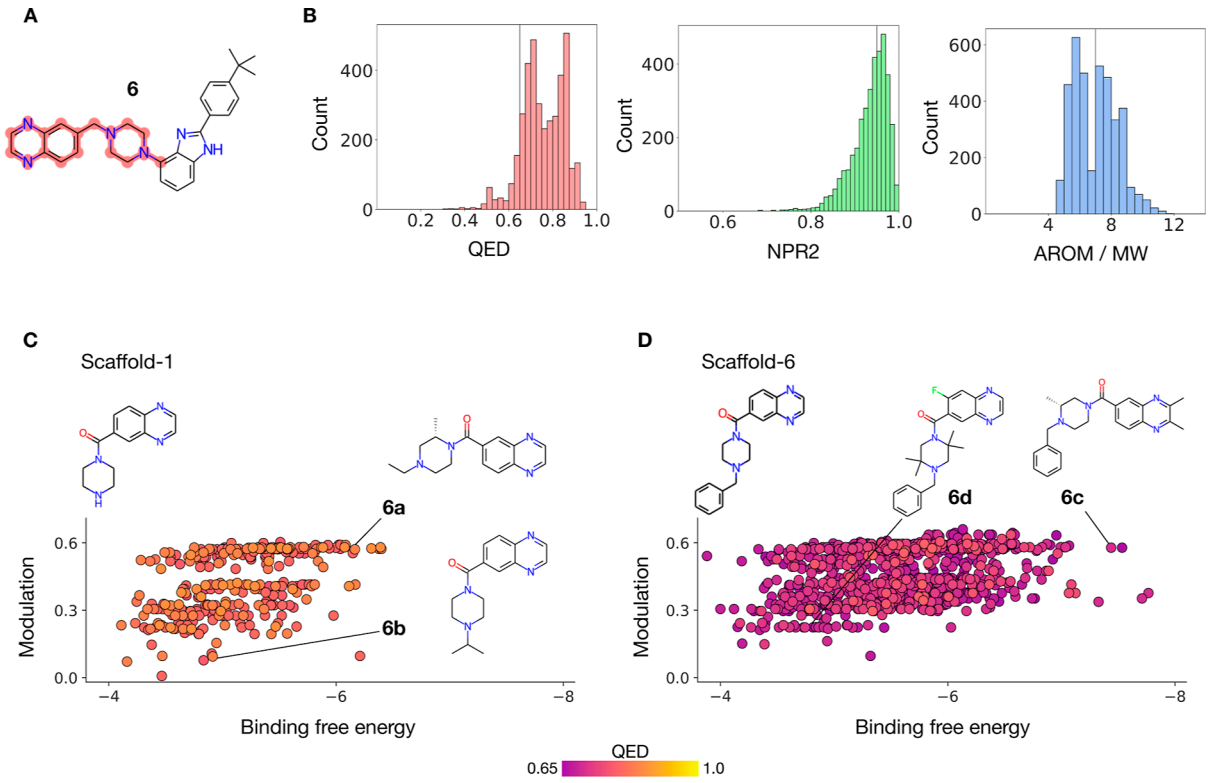


Figure 5. Designing new specific effectors for allosteric regulation of the K-Ras switch I. (A) Structure of **6** and the fragment used for analogue search is colored in red. (B) QED, NPR2, and AROM/MW distributions of analogues derived from the fragment of **6** are colored in red, green, and blue, respectively. (C,D) SW-I modulation and binding free energy of all analogues of Bemis−Murcko scaffold 1 and 6. Analogues **6a**−**6d** are shown as examples.

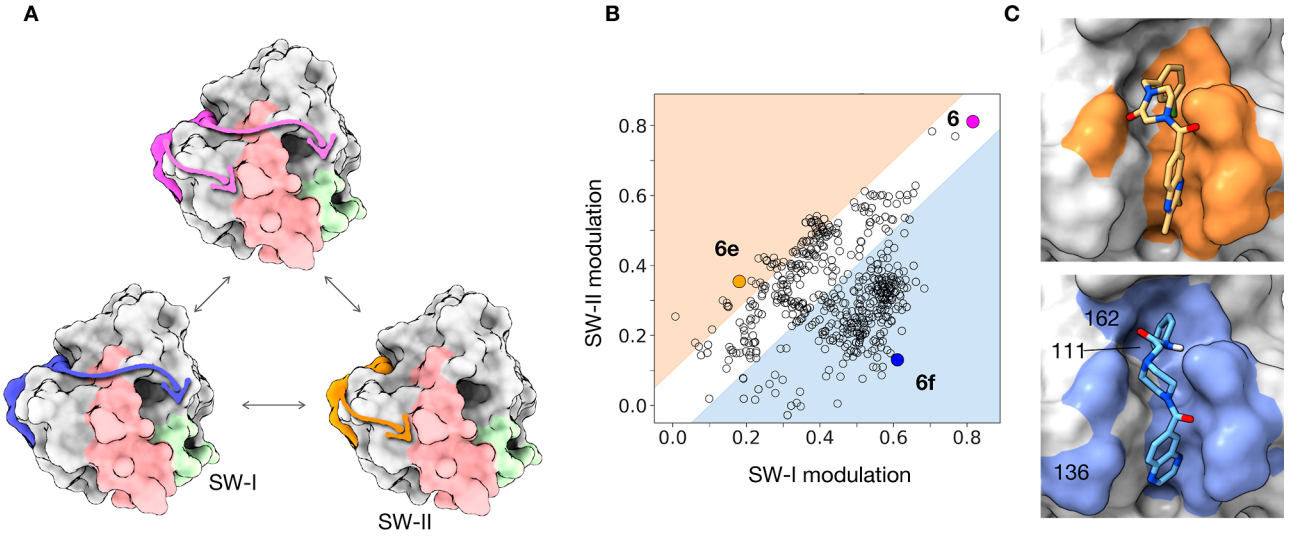


Figure 6. Gaining and tuning the signal specificity of allosteric effectors for the K-Ras switch I and switch II. (A) Illustration depicting a tuning of the specificity of allosteric signaling to distant K-Ras switches by modifications of site−effector pairs within a general location. The effector in magenta modulates both SW-I (green) and SW-II (red) regions, whereas effectors in blue and orange specifically target the SW-I and SW-II, respectively. (B) Characterization of signal specificity of analogues based on **6** via a comparison of the strength of respective modulations at SW-I and SW-II. The blue and orange zones are delineated by $y = x \pm 0.05$ to indicate the extent of effector specificity for switch I and switch II, respectively. Site−effector pairs that are highly specific for targeting SW-I are located near the bottom-right corner of the plot, whereas those specifically modulating SW-II are identified closer to the top-left corner. Compound **6**, a broadly modulating effector, is colored in magenta. (C) **6e** and **6f** binding at $L1_{K\text{-}Ras}$ are shown with binding residues, colored in orange and blue, respectively. The additional residues 111, 136, and 162 in the binding site of **6f** are also indicated.

Subsequently, we explored the design of new site−effector pairs at $L1_{K\text{-}Ras}$ on the basis of compound **6**. Analogue search yielded about 5000 compounds containing a 6-(1-piperazinylmethyl)quinoxaline group (red, Figure 5a). Similar to the case of M^{Pro}, we used several criteria to prioritize flat analogues with a higher aromatic ring content and discarding those with low likelihood as orally administered drugs (AROM/MW ≥ 7 kDa^{-1}, NPR2 ≤ 0.95, QED ≥ 0.65; Figures 5b and S6).

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

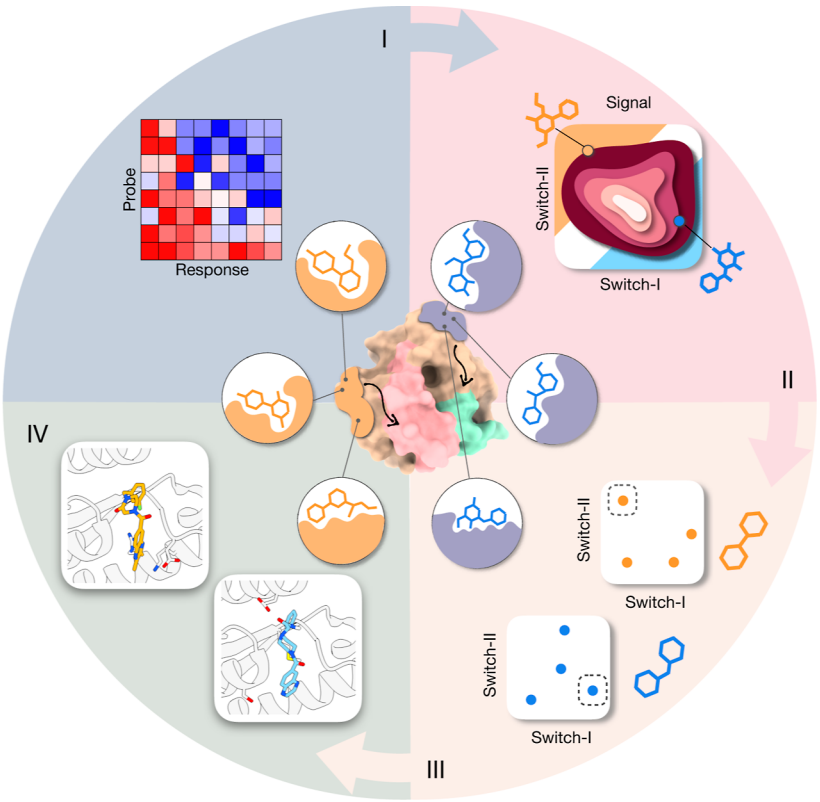


Figure 7. Design framework for development of allosteric drug candidates. An implementation of *the directed design protocol* is illustrated here; Part I (blue): a quantification of allosteric signaling using the APM and reverse perturbation based on the SBSMMA to identify or design new allosteric sites, followed by a screening or design of effectors. Part II (red and center): a repertoire of site−effector pairs is yielded, which leads to an evaluation of signaling, binding, and other molecular properties, such as a signal specificity for a site of interest (shown here using the SW-I/II regions of K-Ras). Part III (beige): potential site−effector pairs with desired characteristics are evaluated and selected and further design or derivation may be performed based on molecular fragments, structural scaffold, and/or other salient features of the previously selected ones in an iterative procedure. Part IV (green): simultaneous fine-tuning, adjustments, and optimization of site−effector pairs produce multiple allosteric drug candidates with distinct binding sites and properties, as well as specific allosteric modes of action (center).

Molecular docking followed by SBSMMA-based evaluation of modulation was performed for all site−effector pairs obtained from 841 shortlisted analogues. The results from site−effector pairs belonging to scaffolds 1 and 6 revealed repertoires of analogues causing a range of positive modulations at SW-I upon interacting with different combinations of residues at $L1_{K\text{-}Ras}$ (Figure 5c). Comparison among analogues **6a−d** shows that interactions of **6a** and **6c** with both allosterically coupled residues Pro110 and Arg162 described above (Figure S7 and Table S2) likely contribute to larger modulation values and binding free energies compared to **6b** (Figure 5c) and **6d** (Figure 5d). While all analogues show slightly weaker modulations (with the highest being 0.6 kcal/mol) compared to that of precursor **6** (0.8 kcal/mol), other aspects, such as the estimate of drug-likeness (QED of **6** = 0.36), were improved considerably in these analogues. Similar to the case of M^{Pro}, we showed that modifications of chemical structures and interacting residues of the site−effector pairs allow optimization of the molecular properties and functional outcomes, such as affinity to the binding site and induced allosteric modulation. One can conclude that the directed design approach implemented here is universal, and it can be used in future protocols for designing new allosteric drugs/effectors for proteins with different functions and intrinsic mechanisms of their regulation.

Tuning the Signal Specificity of K-Ras Effectors for SW-I/II Regions. As one of the critical hubs in cellular signal transduction, activated Ras proteins regulate and, at the same time, are regulated by other proteins, such as GEFs and GAPs, via different interaction interfaces formed by various residues in SW-I/II. In principle, precise modulation of conformational dynamics at interface sites/residues would allow one to specifically target dysregulated interaction(s) mediated by a signaling protein with minimal interferences in the other pathways. Therefore, the specificity of intra- and intermolecular signaling, a quintessential advantage of prospective allosteric effectors, should be considered as one of the major goals in their design. Depending on a required modulation and targeted residue(s), effectors inducing broad modulation or those modulating only specific sites/residues may be favored one over another (Figure 6a). To characterize the signal specificity of compound **6** and its derived analogues, we evaluated the modulation at both switches (Figure 6b). Binding of compound **6** (magenta) results in a modulation of 0.8 kcal/mol at switch I and switch II, which may affect interactions with different proteins. Noteworthy, compound **6** interacts with Ser136-Gly138, which can originate signaling to both switches as revealed by the APM (Figure 4c). For the analogues, while the majority of site−effector pairs at $L1_{K-Ra}$ causes larger modulation in SW-I than SW-II, varied modes and extents of switch specificity can be observed. Analogues **6e** and **6f** (orange and blue) are exemplified here. They differentially modulate SW-I and SW-II upon binding to subsets of residues bound by compound **6**. Binding of **6f** induces large modulation at switch I but not switch II, whereas the opposite was observed for **6e**. The

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

former binds three additional residues Met111, Ser136, and Glu162 compared to the latter (Figure 6c). As described above, Glu162 and, to a lesser extent, Met111 are both instrumental for signaling to SW-I (Figures 4f and S2), indicating that perturbing these residues generally increases the dynamics of SW-I residues and likely exerts the opposite effects in SW-II. Information about the binding sites is tabulated in Table S2. Altogether, the results here demonstrate that fine-tuning of the effectors and the composition of perturbed residues allow one to change the signal specificity of site−effector pairs toward distinct functional sites.

DISCUSSION AND CONCLUSIONS

Six decades after the seminal work by Monod, Changeux, and Jacob,[63] in which they astutely deduced from a handful of enzymes that allosteric effects are entirely due to conformational alterations induced in the protein upon binding with an effector,[63] the quest for comprehensive and specific control of allosteric communication still remains open. The potential and implication of the inducing of an allosteric response to any distant location provide strong impetuses toward achieving allosteric control over protein activity, particularly in the developments of effector drugs. Importantly, the paradigm of allosteric drugs and effects necessitates a shift from contemporary approaches that are used in the identification of orthosteric drugs. The latter is chiefly focused on finding molecules, in rather restrictive chemical space, that are able to interact with specific molecular structures in a functional site and with a strong affinity that allows them to effectively compete with the binding of endogenous substrates. On the other hand, it is increasingly recognized that the tasks of identifying allosteric sites and effectors can be generically formulated as the inducing of required modulation at a functional site via perturbations of the protein dynamics from various locations.[24,27,46] The list of proteins featuring multiple (non)-overlapping allosteric sites for differential modulation of their activity is ever growing, including the major drug targets GPCRs and kinases, associated with expanding arrays of diverse endogenous and synthetic modulators.[64,65] Therefore, the diversity and multiplicity of regulatory sites, and their combined allosteric effects, while adding to the sheer scale and complexity in the tasks for inducing optimal allosteric signaling, offer broad possibilities in designing effector drugs targeting designated residues.[5] Moreover, signaling originating from ligand binding to regulatory sites can also be complemented and modulated by the allosteric effects of mutations.

In this work, we proposed and demonstrated two-sided conceptual and methodological developments. First, we showed that allosteric modulation can be rationally induced by simultaneous design and adjustments of the corresponding allosteric sites and effectors. Second, we developed a generic computational framework for the directed design of optimal site−effector interactions to achieve a desired allosteric modulation. The framework begins with identifying allosteric sites and obtaining known drugs/drug candidates targeting these sites, and it is used as a source of molecular fragments for the fragment-based design of new allosteric effectors. We showed that the catalytic cleft of the M^{Pro} can be remotely modulated by the binding of itacitinib, SR-1664, and entrectinib at the domain I/II hinge. For K-Ras^{G12D}, midostaurin, IKK-16, and WAY-207024 bind a distant surface located on the C-terminal lobe that is allosterically coupled with the SW-I region. On the basis of known drugs, we demonstrated that new effectors can be designed by accounting for molecular properties, binding modes, and various aspects of allosteric modulation including the signal specificity.

It is established that the nonconservatism of allosteric sites provides a foundation for high selectivity of prospective drugs, which may alleviate the issue of off-target toxicity. However, an equally important requirement and advantage, the specificity of allosteric signaling and effects, is much less investigated. The case study on K-Ras showed that some effectors may remotely affect multiple regions, while the others target only certain locations (Figure 6c). Therefore, it is essential to evaluate to what extent a resulting allosteric modulation is localized within a targeted functional site, in order to optimize the specificity of signaling and to minimize the possibility of affecting other critical regions. The schematic diagram in Figure 7 illustrates key components in a design framework directed toward highly specific and tunable regulation. To start with, SBSMMA used here provides a great advantage of comprehensive control allowing us to obtain a map of allosteric communication (APM) in a target protein on a per-residue level (Figure 7, part I). The APM reveals locations allosterically coupled to different sites, quantifying the allosteric modulation in energetic terms (kcal/mol) via calculation of the allosteric free energy obtained on each residue of a protein.[18,19] For example, in this work, we mapped two distinct locations for the binding of effectors specific to SW-I and SW-II of K-Ras^{G12D}. The reverse perturbation method may also be utilized independently or in conjunction with the APM. Based on these locations allosterically coupled to SW-I (green; Figure 7, center) and SW-II (red), one can obtain a repertoire of site−effector pairs in each of the locations allowing analysis of intermolecular interactions and allosteric responses caused by them (Figure 7, part II). Then, in the directed design protocol, lead compounds and their shortlisted analogues are iteratively sampled, evaluated, selected, and further modified for continuous optimization of selected traits, including stereochemical properties of site−effector pairs as well as the specificity and strength of allosteric modulation (Figure 7, part III). Finally, the fine-tuning and optimization of these characteristics and outputs should comprise the simultaneous design and adjustments of effector structures and perturbed residue compositions (Figure 7, part IV), eventually yielding arrays of site−effector pairs with unique characteristics and specific allosteric modes of action (center).

To conclude, being complementary rather than contradictory to the design of traditional orthosteric drugs, the development of allosteric medicines requires a certain paradigm change in the design approaches, which is driven by several important factors. First, there is the existence of multiple potential allosteric sites in every considered drug target, each with its own repertoire of effectors with distinct characteristics and capabilities. The set of allosteric site−effector pairs is an indispensable resource for personalized and tunable treatments, allowing one to reach specificity and to avoid undesired side effects in the framework of precision medicine. Another distinction of allosteric effectors is their mode of action, which includes two equally important steps: binding to the site and originating a signal that causes a modulation on a targeted functional site. The uniqueness of the allosteric drugs and corresponding target sites prompts us to switch from screening existing ligand libraries for detection of binding sites/pockets to rational design approaches for allosteric sites and effectors. The latter, based on the in-depth knowledge of physical−chemical fundamentals of protein−ligand interactions and the understanding of the role of protein dynamics in the protein function and its regulation, could be facilitated by the design power of generative artificial intelligence models and high-throughput experimental techniques in the quest for new chemotypes[5] and characteristics. Undoubtedly, experimental testing and validation play critical roles in the overall design efforts. This work, in which we implemented *the directed design protocol* recently proposed by us,[5] is only a small first step in the beginning of a long and exciting endeavor in the yet uncharted field of allosteric drug design. Starting here from solving immediate tasks formulated above, we will gain a better understanding of allosteric drug actions and will face new challenges, the answer of which will eventually bring us a spectrum of new highly specific and tunable medicines capable to act on difficult or so far undruggable targets.

METHODS

Structure-Based Statistical Mechanical Model of Allostery. The SBSMMA[18,19] was used here to quantify the energetics of allosteric communication at the per-residue resolution. The physical model was thoroughly benchmarked in multiple studies[25,27] accounting for the causality of allosteric signalling, and it describes the allosteric mechanisms caused by general perturbations *P*, including ligand/probe binding,[17,19,22−24,28] mutation(s),[22,25] order−disorder transitions,[26] post-translational modifications,[22] and their combinations. The

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

model consists of three major steps. First, it takes a protein structure as an input to represent it via the Elastic Network Model (ENM) and to construct the C_α atom-based harmonic models for both native 0 and perturbed P states of a protein. The energy function for the perturbed P state is

$$E^{(P)}(\boldsymbol{r}) = \frac{1}{2}\sum_{i,j} k_{ij}(d_{ij} - d_{ij}^0)^2 + \alpha_P V^{(P)}(\boldsymbol{r}) \quad (1)$$

The first term corresponds to the native state with d_{ij} and d_{ij}^0 as interatomic distances between C_α atoms in respective generic r and reference r^0 structures. The k_{ij} is the distance-dependent force constant that decays as $(1/d_{ij}^0)^6$ with a global distance cutoff at 25 Å, with a summation over pairs of neighboring C_α atoms i and j. The second term $V^{(P)}(\boldsymbol{r}) = \sum_{\langle i,j\rangle \in S} k_{ij}(d_{ij} - d_{ij}^0)^2$ is an additional perturbation harmonic term for the ligand binding, where S contains a list of C_α atoms representing a binding site. Ligand binding is indirectly modeled with a stiffening α_P factor at 100, affecting residue pairs forming S and resulting in a local stabilization effect. Orthonormal modes $\boldsymbol{e}_\mu$ characterizing the configurational ensemble of a native/perturbed state are obtained from the Hessian matrix $K = \partial^2 E/\partial \boldsymbol{r}_i \partial \boldsymbol{r}_j$. For each state, allosteric potential which evaluates the effect of a perturbation on a residue i as elastic work exerted due to changes in the displacements of proximal residues (with $C_\alpha - C_\alpha$ distance not more than 11 Å) described by the first ten normal modes $\boldsymbol{e}_\mu^{(P)}$ and $\boldsymbol{e}_\mu^{(0)}$ reads

$$U_i(\sigma) = \frac{1}{2}\sum_\mu \varepsilon_{\mu,i}\sigma_\mu^2 \quad (2)$$

where $\varepsilon_{\mu,i} = \sum_j |\boldsymbol{e}_{\mu,i} - \boldsymbol{e}_{\mu,j}|^2$ and $\sigma = (\sigma_1, ..., \sigma_\mu, ...)$ is a set of Gaussian distributed amplitudes with variance $1/\varepsilon_{\mu,i}$. As the generic displacement of a residue i is $\boldsymbol{r}_i(\sigma) - \boldsymbol{r}_i^0 = \sum_\mu \sigma_\mu \boldsymbol{e}_{\mu,i}$, vector σ can be considered as a configurational state of residue i. In the last part, integrating over the ensemble of all possible configurations σ of a residue, the partition function on the level of a single residue is obtained as

$$\begin{aligned} z_i &= \int d\sigma e^{-U_i(\sigma)/k_B T} \\ &= \prod_\mu \int d\sigma_\mu e^{-\sigma_\mu^2 e_{\mu,i}/2k_B T} \\ &= \prod_\mu \left(\frac{\pi 2k_B T}{\varepsilon_{\mu,i}}\right)^{1/2} \end{aligned} \quad (3)$$

The per-residue partition function z_i allows one to estimate the free energy associated with an allosteric signal $g_i = -k_B T \ln z_i$. The free energy difference (kcal/mol) reads

$$\Delta g_i^{(P)} = \frac{1}{2}k_B T \sum_\mu \ln \varepsilon_{\mu,i}^{(P)}/\varepsilon_{\mu,i}^{(0)} \quad (4)$$

which quantifies the work exerted on residue i as a result of perturbation P. A positive $\Delta g_i^{(P)}$ value indicates an increase of configurational work exerted on residue i which may result in alteration of the protein structures and dynamics; a negative value suggests overstabilization of the residue. Lastly, allosteric modulation $\Delta h_i^{(P)}$ which evaluates exerted work due to purely allosteric signaling is obtained as a difference $\Delta g_i^{(P)}$ from its mean over all residues of the protein chain containing i

$$\Delta h_i^{(P)} = \Delta g_i^{(P)} - \langle \Delta g_i^{(P)} \rangle_{chain} \quad (5)$$

A positive modulation value points to potential conformational changes, whereas a negative one indicates that conformational changes are precluded. Importantly, per-residue $\Delta h_i^{(P)}$ values with magnitude above thermal fluctuations $k_B T$ (0.6 kcal/mol) reflect significant allosteric communication. Generally, the $k_B T$ value may be used to discriminate between an efficient signal and signaling noise due to thermal fluctuations. Nevertheless, in some cases, $\Delta h_i^{(P)}$ values close to $k_B T$ may still be considered in a homogeneously affected protein region due to the extensivity of the free energy that may result in strengthening of the signal by small additional contributions. While large magnitudes of modulation values may suggest strong allosteric effects on a protein function, the functional outcomes caused by binding of effectors should be validated in experiments or compared with known experimental data from other works. To compare allosteric modulation at different sites, a modulation on a site of interest S is given by $\Delta h_S^{(P)} = \langle \Delta h_i^{(P)} \rangle_S$ where residue $i \in S$. We have previously shown that perturbations simulated at ligand-free input structures initially obtained from separate crystallographic experiments in apo and holo states typically lead to comparable results between them due to the overlapping conformational ensembles sampled in both states. Moreover, the SBSMMA may be applied to protein structures with or without apparent binding sites/pockets, as the model quantifies the per-residue energetics of signaling to and from single residues regardless of specific structural properties or dynamics of protein surfaces. The SBSMMA has been implemented in the AlloSigMA Web server[28] and AlloMAPS database,[29] widely used by experimental and theoretical researchers as explorative and analytical tools. Users should expect a computing time ranging from minutes to several hours to obtain downloadable data from AlloSigMA, which depends on the number of residues in the target protein and the specification of the tasks. More details on the computing time can be found in the Benchmarking section in the Tutorial page.

Identification of Allosterically Linked Locations. Crystal structures of M^{Pro}, K-Ras^{G12D}, and K-Ras^{WT} (PDB: 7JP1, 6GOF, and 6GOD) were used by the framework to identify potential allosteric locations. Residues constituting the SW-I/II regions in K-Ras and the catalytic cleft and dimer interface of the M^{Pro} were collected from the literature.[66,67] In reverse perturbation, substrate binding was simulated, and $\Delta h_i^{(Bind)}$ was measured for all residues. The APM is obtained by simulating binding of a small probe ligand to every consecutive three-residue segment along a protein chain and evaluating the modulation $\Delta h_i^{(Probe)}$ at each residue i. A combination of top-ranking residues/segments, with the largest $|\Delta h|$ and with $C_\alpha - C_\alpha$ distance at least 11 Å (the operational cutoff distance for the elastic network model of the protein) to all targeted site residues, is compiled by the reverse perturbation and/or APM methods. It presents a list of residues strongly coupled to the targeted site, allowing delineation of protein regions/locations containing allosteric sites with different compositions of residues.

Molecular Docking and Fragment-Based Design of Effectors. DRH[47] containing approximately 6800 approved and clinical trial drugs with molecular weight not more than 600 Da was used for molecular docking. Structures of all compounds in PDBQT format were converted from SMILES strings using OpenBabel[68] and protonated at pH 7. The protein structures were prepared using AutoDockTools[69] prior to docking. Molecular docking was performed at the geometric center of shortlisted strongly modulating residues/triplets on the crystal structures with a cubic box of lengths of 18 and 20 Å for K-Ras and M^{Pro}, respectively. AutoDock Vina[41] was used with default scoring functions and parameters, providing an estimation of the free energy of binding. For each ligand, up to 10 docked poses were examined. Allosteric site residues were delineated based on a distance cutoff of 4 Å from any heavy atom of a bound ligand, which were then used for computing the allosteric modulation $\Delta h_S^{(Bind)}$ at a distant regulated site using the SBSMMA formalism. Substructures/fragments of shortlisted molecules from DRH were obtained with the RECAP method, which performs fragmentation based on chemical knowledge. Subsequently, the SMILES strings of selected fragments were used to carry out an analogue search on the ZINC20 database.[53] The RDKit package[70] was used to flag and remove analogues using implementations of the PAINS and Brenk filters, to obtain Bemis–Murcko scaffolds[58] of derived analogues and to evaluate properties of compounds. In particular, the pairwise Tanimoto similarity was calculated using the Morgan fingerprints implemented in RDKit. The default average descriptor weights were employed in the QED calculation. The calculation of NPMI values was performed by the Descriptors3D module in RDKit after the embedding of a 3D conformation, which was set to prioritize known experimental torsional angles. The SBSMMA was used to

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

evaluate the modulation caused by derived analogues. Details on compounds, binding sites, and modulations are tabulated in Tables S1 and S2.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.4c01043.

SBSMMA-based identification of latent allosteric sites; analysis of signaling by known and new allosteric effectors/drug candidates for M^{Pro} and K-Ras^{G12D}; design protocols and characterization of derived analogues; and tables of SMILES, binding site residues, and modulation values associated with all described compounds (PDF)

PDB files of discussed effector candidates docked to predicted allosteric sites of M^{Pro} and K-Ras^{G12D} (ZIP)

List of molecular formula strings of design effector candidates discussed (CSV)

■ AUTHOR INFORMATION

Corresponding Authors

Wei-Ven Tee − *Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01, Matrix, Singapore 138671, Singapore*; Phone: (65) 6478 8305; Email: teewv@bii.a-star.edu.sg; Fax: (65) 6478 9047

Igor N. Berezovsky − *Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01, Matrix, Singapore 138671, Singapore; Department of Biological Sciences (DBS), National University of Singapore (NUS), Singapore 117579, Singapore*; orcid.org/0000-0002-3315-8483; Phone: (65) 6478 8269; Email: igorb@bii.a-star.edu.sg; Fax: (65) 6478 9047

Author

Sylvester J. M. Lim − *Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01, Matrix, Singapore 138671, Singapore*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jmedchem.4c01043

Author Contributions

[§]W.-V.T. and S.J.M.L. contributed equally and W.-V.T. and S.J.M.L. are first co-authors. W.-V.T. performed the research, analyzed the data, and wrote and edited the manuscript. S.J.M.L. performed the research, analyzed the data, and wrote and edited the manuscript. I.N.B. analyzed the data and wrote and edited the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research is supported by A*STAR C210812031 to W.-V.T. and by MOH-001402-00 to I.N.B.

■ ABBREVIATIONS USED

Ala	alanine
APM	allosteric probing map
Arg	arginine
AROM	aromatic ring
Asp	aspartic acid
COVID-19	coronavirus disease 2019
CYP3A	cytochrome 3A
Cys	cysteine
DRH	drug repurposing hub
GAP	GTPase-activating protein
GEF	guanine nucleotide exchange factor
Gln	glutamine
Glu	glutamic acid
Gly	glycine
GPCR	G protein-coupled receptor
His	histidine
Lys	lysine
Met	methionine
M^{Pro}	main protease
NPR	normalized principal moment of inertia ratio
MW	molecular weight
K-Ras	Kirsten rat sarcoma virus
PAINS	pan assay interference compounds
PMI	principal moments of inertia
PPAR γ	peroxisome proliferator-activated receptor gamma
Pro	proline
QED	quantitative estimate of drug-likeness
Ras	rat sarcoma virus
RECAP	retrosynthetic combinatorial analysis procedure
SARS-CoV	severe acute respiratory syndrome coronavirus
SBSMMA	structure-based statistical mechanical model of allostery
Ser	serine
SW	switch

■ REFERENCES

(1) Guarnera, E.; Berezovsky, I. N. On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr. Opin. Struct. Biol.* **2019**, *56*, 18−27.

(2) Gunasekaran, K.; Ma, B.; Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* **2004**, *57* (3), 433−443.

(3) Wodak, S. J.; Paci, E.; Dokholyan, N. V.; Berezovsky, I. N.; Horovitz, A.; Li, J.; Hilser, V. J.; Bahar, I.; Karanicolas, J.; Stock, G.; et al. Allostery in Its Many Disguises: From Theory to Applications. *Struct.* **2019**, *27* (4), 566−578.

(4) Nussinov, R.; Tsai, C. J. Allostery in disease and in drug discovery. *Cell* **2013**, *153* (2), 293−305.

(5) Tee, W. V.; Berezovsky, I. N. Allosteric drugs: New principles and design approaches. *Curr. Opin. Struct. Biol.* **2024**, *84*, 102758.

(6) Hardy, J. A.; Wells, J. A. Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol.* **2004**, *14* (6), 706−715.

(7) Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E. Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **2016**, *116* (11), 6370−6390.

(8) Tan, Z. W.; Tee, W. V.; Berezovsky, I. N. Learning About Allosteric Drugs and Ways to Design Them. *J. Mol. Biol.* **2022**, *434* (17), 167692.

(9) Nussinov, R.; Tsai, C. J. The different ways through which specificity works in orthosteric and allosteric drugs. *Curr. Pharm. Des.* **2012**, *18* (9), 1311−1316.

(10) Berezovsky, I. N. Thermodynamics of allostery paves a way to allosteric drugs. *Biochim. Biophys. Acta* **2013**, *1834* (5), 830−835.

(11) Pan, Y.; Mader, M. M. Principles of Kinase Allosteric Inhibition and Pocket Validation. *J. Med. Chem.* **2022**, *65* (7), 5288−5299.

(12) Conn, P. J.; Christopoulos, A.; Lindsley, C. W. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nat. Rev. Drug Discov.* **2009**, *8* (1), 41−54.

(13) Liu, J.; Nussinov, R. Energetic redistribution in allostery to execute protein function. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (29), 7480−7482.

(14) Liu, J.; Nussinov, R. Allosteric effects in the marginally stable von Hippel-Lindau tumor suppressor protein and allostery-based rescue mutant design. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (3), 901−906.
(15) Rettenmaier, T. J.; Sadowsky, J. D.; Thomsen, N. D.; Chen, S. C.; Doak, A. K.; Arkin, M. R.; Wells, J. A. A small-molecule mimic of a peptide docking motif inhibits the protein kinase PDK1. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (52), 18590−18595.
(16) Schulze, J. O.; Saladino, G.; Busschots, K.; Neimanis, S.; Süß, E.; Odadzic, D.; Zeuzem, S.; Hindie, V.; Herbrand, A. K.; Lisa, M. N.; et al. Bidirectional Allosteric Communication between the ATP-Binding Site and the Regulatory PIF Pocket in PDK1 Protein Kinase. *Cell Chem. Biol.* **2016**, *23* (10), 1193−1205.
(17) Ghode, A.; Gross, L. Z. F.; Tee, W. V.; Guarnera, E.; Berezovsky, I. N.; Biondi, R. M.; Anand, G. S. Synergistic Allostery in Multiligand-Protein Interactions. *Biophys. J.* **2020**, *119* (9), 1833−1848.
(18) Guarnera, E.; Berezovsky, I. N. Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLoS Comput. Biol.* **2016**, *12* (3), No. e1004678.
(19) Guarnera, E.; Berezovsky, I. N. Toward Comprehensive Allosteric Control over Protein Activity. *Struct.* **2019**, *27* (5), 866−878.e1.
(20) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **2013**, *503* (7477), 548−551.
(21) Kim, D.; Herdeis, L.; Rudolph, D.; Zhao, Y.; Böttcher, J.; Vides, A.; Ayala-Santos, C. I.; Pourfarjam, Y.; Cuevas-Navarro, A.; Xue, J. Y.; et al. Pan-KRAS inhibitor disables oncogenic signalling and tumour growth. *Nature* **2023**, *619* (7968), 160−166.
(22) Tan, Z. W.; Tee, W. V.; Samsudin, F.; Guarnera, E.; Bond, P. J.; Berezovsky, I. N. Allosteric perspective on the mutability and druggability of the SARS-CoV-2 Spike protein. *Struct.* **2022**, *30* (4), 590−607.e4.
(23) Tee, W. V.; Tan, Z. W.; Guarnera, E.; Berezovsky, I. N. Conservation and Diversity in Allosteric Fingerprints of Proteins for Evolutionary-inspired Engineering and Design. *J. Mol. Biol.* **2022**, *434* (17), 167577.
(24) Tee, W. V.; Tan, Z. W.; Lee, K.; Guarnera, E.; Berezovsky, I. N. Exploring the Allosteric Territory of Protein Function. *J. Phys. Chem. B* **2021**, *125* (15), 3763−3780.
(25) Tee, W. V.; Guarnera, E.; Berezovsky, I. N. On the Allosteric Effect of nsSNPs and the Emerging Importance of Allosteric Polymorphism. *J. Mol. Biol.* **2019**, *431* (19), 3933−3942.
(26) Tee, W. V.; Guarnera, E.; Berezovsky, I. N. Disorder driven allosteric control of protein activity. *Curr. Res. Struct. Biol.* **2020**, *2*, 191−203.
(27) Tee, W. V.; Guarnera, E.; Berezovsky, I. N. Reversing allosteric communication: From detecting allosteric sites to inducing and tuning targeted allosteric response. *PLoS Comput. Biol.* **2018**, *14* (6), No. e1006228.
(28) Tan, Z. W.; Guarnera, E.; Tee, W. V.; Berezovsky, I. N. AlloSigMA 2: paving the way to designing allosteric effectors and to exploring allosteric effects of mutations. *Nucleic Acids Res.* **2020**, *48* (W1), W116−w124.
(29) Tan, Z. W.; Tee, W. V.; Guarnera, E.; Berezovsky, I. N. AlloMAPS 2: allosteric fingerprints of the AlphaFold and Pfam-trRosetta predicted structures for engineering and design. *Nucleic Acids Res.* **2023**, *51* (D1), D345−d351.
(30) Hu, Q.; Xiong, Y.; Zhu, G. H.; Zhang, Y. N.; Zhang, Y. W.; Huang, P.; Ge, G. B. The SARS-CoV-2 main protease (M(pro)): Structure, function, and emerging therapies for COVID-19. *MedComm* **2022**, *3* (3), No. e151.
(31) Nussinov, R.; Jang, H. Direct K-Ras Inhibitors to Treat Cancers: Progress, New Insights, and Approaches to Treat Resistance. *Annu. Rev. Pharmacol. Toxicol.* **2024**, *64*, 231−253.
(32) Huang, L.; Guo, Z.; Wang, F.; Fu, L. KRAS mutation: from undruggable to druggable in cancer. *Signal Transduction Targeted Ther.* **2021**, *6* (1), 386.
(33) Steuten, K.; Kim, H.; Widen, J. C.; Babin, B. M.; Onguka, O.; Lovell, S.; Bolgi, O.; Cerikan, B.; Neufeldt, C. J.; Cortese, M.; et al. Challenges for Targeting SARS-CoV-2 Proteases as a Therapeutic Strategy for COVID-19. *ACS Infect. Dis.* **2021**, *7* (6), 1457−1468.
(34) Lu, S.; Jang, H.; Muratcioglu, S.; Gursoy, A.; Keskin, O.; Nussinov, R.; Zhang, J. Ras Conformational Ensembles, Allostery, and Signaling. *Chem. Rev.* **2016**, *116* (11), 6607−6665.
(35) Hammond, J.; Leister-Tebbe, H.; Gardner, A.; Abreu, P.; Bao, W.; Wisemandle, W.; Baniecki, M.; Hendrick, V. M.; Damle, B.; Simon-Campos, A.; et al. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *N. Engl. J. Med.* **2022**, *386* (15), 1397−1408.
(36) Skoulidis, F.; Li, B. T.; Dy, G. K.; Price, T. J.; Falchook, G. S.; Wolf, J.; Italiano, A.; Schuler, M.; Borghaei, H.; Barlesi, F.; et al. Sotorasib for Lung Cancers with KRAS p.G12C Mutation. *N. Engl. J. Med.* **2021**, *384* (25), 2371−2381.
(37) Akinosoglou, K.; Schinas, G.; Gogos, C. Oral Antiviral Treatment for COVID-19: A Comprehensive Review on Nirmatrelvir/Ritonavir. *Viruses* **2022**, *14* (11), 2540.
(38) Guarnera, E.; Berezovsky, I. N. Allosteric sites: remote control in regulation of protein activity. *Curr. Opin. Struct. Biol.* **2016**, *37*, 1−8.
(39) Zhang, Q.; Chen, Y.; Ni, D.; Huang, Z.; Wei, J.; Feng, L.; Su, J. C.; Wei, Y.; Ning, S.; Yang, X.; et al. Targeting a cryptic allosteric site of SIRT6 with small-molecule inhibitors that inhibit the migration of pancreatic cancer cells. *Acta Pharm. Sin. B* **2022**, *12* (2), 876−889.
(40) Xiao, S.; Alshahrani, M.; Gupta, G.; Tao, P.; Verkhivker, G. Markov State Models and Perturbation-Based Approaches Reveal Distinct Dynamic Signatures and Hidden Allosteric Pockets in the Emerging SARS-Cov-2 Spike Omicron Variant Complexes with the Host Receptor: The Interplay of Dynamics and Convergent Evolution Modulates Allostery and Functional Mechanisms. *J. Chem. Inf. Model.* **2023**, *63* (16), 5272−5296.
(41) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61* (8), 3891−3898.
(42) Cimermancic, P.; Weinkam, P.; Rettenmaier, T. J.; Bichmann, L.; Keedy, D. A.; Woldeyes, R. A.; Schneidman-Duhovny, D.; Demerdash, O. N.; Mitchell, J. C.; Wells, J. A.; et al. CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J. Mol. Biol.* **2016**, *428* (4), 709−719.
(43) Kimura, S. R.; Hu, H. P.; Ruvinsky, A. M.; Sherman, W.; Favia, A. D. Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics. *J. Chem. Inf. Model.* **2017**, *57* (6), 1388−1401.
(44) Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138* (43), 14257−14263.
(45) He, J.; Liu, X.; Zhu, C.; Zha, J.; Li, Q.; Zhao, M.; Wei, J.; Li, M.; Wu, C.; Wang, J.; et al. ASD2023: towards the integrating landscapes of allosteric knowledgebase. *Nucleic Acids Res.* **2024**, *52* (D1), D376−d383.
(46) Bowman, G. R.; Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (29), 11681−11686.
(47) Corsello, S. M.; Bittker, J. A.; Liu, Z.; Gould, J.; McCarren, P.; Hirschman, J. E.; Johnston, S. E.; Vrcic, A.; Wong, B.; Khan, M.; et al. The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat. Med.* **2017**, *23* (4), 405−408.
(48) Wang, Q.; Zheng, M.; Huang, Z.; Liu, X.; Zhou, H.; Chen, Y.; Shi, T.; Zhang, J. Toward understanding the molecular basis for chemical allosteric modulator design. *J. Mol. Graph. Model.* **2012**, *38*, 324−333.
(49) Gunther, S.; Reinke, P. Y. A.; Fernandez-Garcia, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; et al. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372* (6542), 642−646.
(50) Stromich, L.; Wu, N.; Barahona, M.; Yaliraki, S. N. Allosteric Hotspots in the Main Protease of SARS-CoV-2. *J. Mol. Biol.* **2022**, *434* (17), 167748.

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX

(51) Yuce, M.; Cicek, E.; Inan, T.; Dag, A. B.; Kurkcuoglu, O.; Sungur, F. A. Repurposing of FDA-approved drugs against active site and potential allosteric drug-binding sites of COVID-19 main protease. *Proteins* **2021**, *89* (11), 1425−1441.
(52) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP–retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (3), 511−522.
(53) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60* (12), 6065−6073.
(54) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, *4* (2), 90−98.
(55) Sauer, W. H.; Schwarz, M. K. Molecular shape diversity of combinatorial libraries: a prerequisite for broad bioactivity. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (3), 987−1003.
(56) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53* (7), 2719−2740.
(57) Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I. H.; Frearson, J.; Wyatt, P. G. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **2008**, *3* (3), 435−444.
(58) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887−2893.
(59) Bos, J. L.; Rehmann, H.; Wittinghofer, A. GEFs and GAPs: critical elements in the control of small G proteins. *Cell* **2007**, *129* (5), 865−877.
(60) Feng, H.; Zhang, Y.; Bos, P. H.; Chambers, J. M.; Dupont, M. M.; Stockwell, B. R. K-Ras(G12D) Has a Potential Allosteric Small Molecule Binding Site. *Biochemistry* **2019**, *58* (21), 2542−2554.
(61) Spencer-Smith, R.; Koide, A.; Zhou, Y.; Eguchi, R. R.; Sha, F.; Gajwani, P.; Santana, D.; Gupta, A.; Jacobs, M.; Herrero-Garcia, E.; et al. Inhibition of RAS function through targeting an allosteric regulatory site. *Nat. Chem. Biol.* **2017**, *13* (1), 62−68.
(62) Weng, C.; Faure, A. J.; Escobedo, A.; Lehner, B. The energetic and allosteric landscape for KRAS inhibition. *Nature* **2024**, *626* (7999), 643−652.
(63) Monod, J.; Changeux, J. P.; Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **1963**, *6*, 306−329.
(64) Wold, E. A.; Chen, J.; Cunningham, K. A.; Zhou, J. Allosteric Modulation of Class A GPCRs: Targets, Agents, and Emerging Concepts. *J. Med. Chem.* **2019**, *62* (1), 88−127.
(65) Lu, X.; Smaill, J. B.; Ding, K. New Promise and Opportunities for Allosteric Kinase Inhibitors. *Angew. Chem., Int. Ed. Engl.* **2020**, *59* (33), 13764−13776.
(66) Ostrem, J. M.; Shokat, K. M. Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design. *Nat. Rev. Drug Discov.* **2016**, *15* (11), 771−785.
(67) Goyal, B.; Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci.* **2020**, *22* (6), 297−305.
(68) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(69) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **2016**, *11* (5), 905−919.
(70) RDKit: Open-source cheminformatic, Version 2022.03.2. https://www.rdkit.org (accessed April 25, 2022).

https://doi.org/10.1021/acs.jmedchem.4c01043
J. Med. Chem. XXXX, XXX, XXX−XXX