Microbially guided discovery and biosynthesis of biologically active natural products

Ankur Sarkar
University of Colorado Boulder
Edward Kim
National Cancer Institute, NIH
Taehwan Jang
Korea Advanced Institute of Science and Technology
Akarawin Hongdusit
University of Colorado Boulder
Hyungjun Kim
Korea Advanced Institute of Science and Technology https://orcid.org/0000-0001-8261-9381
Jeong-Mo Choi
Korea Advanced Institute of Science and Technology https://orcid.org/0000-0003-2656-4851
Jerome Fox (jerome.fox@colorado.edu)
University of Colorado Boulder https://orcid.org/0000-0002-3739-1899

Article

Keywords: biologically active natural products, terpenoid inhibitors, protein tyrosine phosphatase 1B (PTP1B)

Posted Date: March 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-92116/v2

Abstract

The design of small molecules that inhibit disease-relevant proteins represents a longstanding challenge of medicinal chemistry. Here, we describe an approach for encoding this challenge—the inhibition of a human drug target—into a microbial host and using it to guide the discovery and biosynthesis of targeted, biologically active natural products. This approach identified two previously unknown terpenoid inhibitors of protein tyrosine phosphatase 1B (PTP1B), an elusive therapeutic target for the treatment of diabetes and cancer. Both inhibitors appear to target an allosteric site, which confers selectivity, and can inhibit PTP1B in living cells. A screen of 24 uncharacterized terpene synthases from a pool of 4,464 genes uncovered additional hits, demonstrating a scalable discovery approach, and the incorporation of different PTPs into the microbial host yielded alternative PTP-specific detection systems. Findings illustrate the potential for using microbes to discover and build natural products that exhibit precisely defined biochemical activities yet possess unanticipated structures and/or binding sites.

Introduction

Despite advances in structural biology and computational chemistry, the design of small molecules that bind tightly and selectively to disease-relevant proteins remains exceptionally difficult[1]. The free energetic contributions of rearrangements in the molecules of water that solvate binding partners and structural changes in the binding partners themselves are particularly challenging to predict and, thus, to incorporate into molecular design[2,3]. Drug development, as a result, often begins with screens of large compound libraries[4].

Nature has endowed living systems with the catalytic machinery to build an enormous variety of biologically active molecules—a diverse natural library[5]. These molecules evolved to carry out important metabolic and ecological functions (e.g., the phytochemical recruitment of predators of herbivorous insects[6]) but often also exhibit useful medicinal properties. Over the years, screens of environmental extracts and natural product libraries—augmented, on occasion, with combinatorial (bio)chemistry[7,8]—have uncovered a diverse set of therapeutics, from aspirin to paclitaxel[9]. Unfortunately, these screens tend to be resource intensive[10], limited by low natural titers[11], and largely subject to serendipity[12]. Bioinformatic tools, in turn, have permitted the identification of biosynthetic gene clusters[13,14], where co-localized resistance genes can reveal the biochemical function of their products[15,16]. The therapeutic applications of many natural products, however, differ from their native functions[17], and many biosynthetic pathways can, when appropriately reconfigured, produce entirely new and, perhaps, more effective therapeutic molecules[18,19]. Methods for efficiently identifying and building natural products that inhibit specific disease-relevant proteins remain largely undeveloped.

Protein tyrosine phosphatases (PTPs) are an important class of drug targets that could benefit from new approaches to inhibitor discovery. These enzymes catalyze the hydrolytic dephosphorylation of tyrosine residues and, together with protein tyrosine kinases (PTKs), contribute to an enormous number of

diseases (e.g., cancer, autoimmune disorders, and heart disease, to name a few)[20,21]. The last several decades have witnessed the development of many potent inhibitors of PTKs, which are targets for over 30 approved drugs[22]. Therapeutic inhibitors of PTPs, by contrast, have lagged behind. These enzymes possess well conserved, positively charged active sites that make them difficult to inhibit with selective, membrane-permeable molecules[23]; they lack targeted therapeutics of any kind.

In this study, we describe an approach for using microbial systems to find natural products that inhibit difficult-to-drug proteins. We focused on protein tyrosine phosphatase 1B (PTP1B), a therapeutic target for the treatment of type 2 diabetes, obesity, and HER2-positive breast cancer[24]. PTP1B possesses structural characteristics that are generally representative of the PTP family[25] and regulates a diverse set of physiological processes (e.g., energy expenditure[26], inflammation[27], and neural specification in embryonic stem cells[28]). In brief, we assembled a strain of *Escherichia coli* with two genetic modules—(i) one that links cell survival to the inhibition of PTP1B and (ii) one that enables the biosynthesis of structurally varied terpenoids. In a study of five well-characterized terpene synthases, this strain identified two previously unknown terpenoid inhibitors of PTP1B. Both inhibitors were selective for PTP1B, exhibited distinct binding mechanisms, and increased insulin receptor phosphorylation in mammalian cells. A screen of 24 uncharacterized terpene synthases from eight phylogenetically diverse clades uncovered additional hits, demonstrating a scalable approach for inhibitor discovery, and the exchange of PTP genes in our genetically encoded detection system permitted its facile extension to new targets. Our findings illustrate a versatile approach for using microbial systems to find targeted, readily synthesizable inhibitors of disease-relevant enzymes.

Results And Discussion

Development of a Genetically Encoded Objective

E. coli is a versatile platform for building natural products from unculturable or low-yielding organisms[29,30]. We hypothesized that a strain of *E. coli* programmed to detect the inactivation of PTP1B (i.e., a genetically encoded objective) might enable the discovery of natural products that inhibit it (i.e., molecular solutions to the objective). To program such a strain, we assembled a bacterial two-hybrid (B2H) system in which PTP1B and Src kinase control gene expression (Figure 1A). In this system, Src phosphorylates a substrate domain, enabling a protein-protein interaction that activates transcription of a gene of interest (GOI). PTP1B dephosphorylates the substrate domain, preventing that interaction, and the inactivation of PTP1B re-enables it. *E. coli* is a particularly good host for this detection system because its proteome is sufficiently orthogonal to the proteome of *H. sapiens* to minimize off-target growth defects that can result from the regulatory activities of Src and PTP1B (SI Note 1)[31].

We carried out B2H development in several steps. To begin, we assembled a luminescent “base” system in which Src modulates the binding of a substrate domain to a Src homology 2 (SH2) domain (Figure 1B); this system, which includes a chaperone that helps Src to fold (Cdc37)[32], is similar to other B2H designs

that detect protein-protein binding[33]. Unfortunately, our initial system did not yield a phosphorylation-dependent transcriptional response, so we complemented it with inducible plasmids—each harboring a different system component—to identify proteins with suboptimal expression levels (Figure 1B). Interestingly, secondary induction of Src increased luminescence, an indication that insufficient substrate phosphorylation and/or weak substrate-SH2 binding depressed GOI expression in our base system. We modified this system by swapping in different substrate domains, by adding mutations to the SH2 domain that enhance its affinity for phosphopeptides[34], and by removing the gene for Src, a modification that allowed us to control its expression exclusively from a second plasmid. With this configuration, induction of Src increased luminescence most prominently for the MidT substrate (Figure 1C), and simultaneous induction of both Src and PTP1B prevented that increase—an indication of intracellular PTP1B activity (Figure 1D). We finalized the MidT system by incorporating genes for PTP1B and Src, by adjusting promoters and ribosome binding sites to amplify its transcriptional response further (Figure 1D, S1-2), and by adding a gene for spectinomycin resistance (SpecR) as the GOI. The final plasmid-borne detection system required the inactivation of PTP1B to permit growth at high concentrations of antibiotic (Figure 1E).

Biosynthesis of PTP1B Inhibitors

To search for inhibitors of PTP1B that bind outside of its active site, we coupled the B2H system with metabolic pathways for terpenoids (Figure 2A). We chose this class of natural products for three reasons: (i) Previous studies have identified terpenoids that inhibit PTP1B, though their mechanisms of inhibition are incompletely resolved[35,36]; we hypothesized that these molecules, which are largely nonpolar, would avoid the conserved, positively charged regions of the active site that confer poor selectivity to substrate analogs. (ii) Terpenoids include over 80,000 known compounds and represent nearly one-third of all characterized natural products[37] (the basis of approximately 50% of clinically approved drugs[38]); they thus define a rich molecular landscape for inhibitor discovery. (iii) Terpenoids can be synthesized and functionalized in *E. coli*[39,40]. To begin, we focused on a handful of structurally diverse terpenoids (Figure 2B): amorphadiene, α-humulene, β-bisabolene, abietadiene, and taxadiene. Abietadiene is a precursor to abietic acid, a weak inhibitor of PTP1B[36]. The other molecules lack established inhibitory effects. Each terpenoid pathway comprises two plasmid-borne modules: (i) the mevalonate-dependent isoprenoid pathway from *S. cerevisiae* (optimized for expression in *E. coli*[41]) and (ii) a terpene synthase capable of producing one of the five terpenoids[41–45]. We also supplemented each diterpene synthase with a gene for geranylgeranyl diphosphate synthase (GGPPS). These modules generated terpenoids at titers of 0.3-18 mg/L in *E. coli* (Figure S3A).

We screened each pathway for its ability to produce inhibitors of PTP1B by transforming *E. coli* with plasmids harboring both the pathway of interest and the B2H system (Figure 2C). To our surprise, pathways for amorphadiene and β-bisabolene, but not abietadiene, permitted survival at high concentrations of antibiotic. Critically, GC-MS traces confirmed that all pathways generated terpenoids in the presence of the B2H system (Figure 2D, S3B,C), and maximal resistance of the amorphadiene- and β-

bisabolene-producing strains required both an active terpene synthase and a functional B2H system (Figure S3D).

We confirmed the inhibitory effects of purified terpenoids by examining their influence on PTP1B-catalyzed hydrolysis of *p*-nitrophenyl phosphate (*p*NPP; Figure 2E, Table S13). The IC_{50}s for amorphadiene and β-bisabolene were 53 ± 8 μM and 13 ± 2 μM, respectively, in 10% DMSO (Figure 2F). These IC_{50}s are surprisingly strong for small, unfunctionalized hydrocarbons (i.e., the ligand efficiencies are high; Table S16); they resemble the IC_{50}s of larger molecules that form hydrogen bonds and other stabilizing interactions with PTP1B[20,46]. Importantly, both IC_{50}s are similar to terpenoid concentrations in liquid culture (Figure 2G), a finding consistent with *in vivo* inhibition. Terpenoids also tend to accumulate intracellularly, where their concentrations may be much higher than in the media[47]. Our growth-coupled assays, kinetic assays, and production measurements, taken together, indicate that amorphadiene and β-bisabolene activate the B2H system by inhibiting PTP1B inside the cell.

Biophysical Analysis of PTP1B Inhibitors

Allosteric inhibitors of PTPs are valuable starting points for drug development. These molecules bind outside of the structurally similar, positively charged active sites of PTPs and tend to have improved selectivities and membrane permeabilities over substrate analogs[20]. Motivated by these considerations, an early screen identified a benzbromarone derivative that inhibited PTP1B weakly (IC_{50} = 350 μM) without competing with substrates; subsequent optimization of this compound led to two improved inhibitors (IC_{50}s = 8 and 22 μM) that bind to an allosteric site[48] (Figure 3A). Over the next 15 years, efforts to find new inhibitors that bind to this or other allosteric regions on the catalytic domain have been largely unsuccessful[49]. Benzbromarone derivatives are the only publicly reported allosteric inhibitors with crystallographically verified binding sites. (Although, an allosteric inhibitor that binds to a disordered region of the full-length protein has been characterized with NMR[24]). New approaches for finding allosteric inhibitors are clearly needed.

Our microbial system could grant access to new compounds that bind in unexpected ways. Amorphadiene and β-bisabolene are good examples. They are highly nonpolar and, thus, incapable of engaging in the hydrogen bonds and electrostatic interactions on which most other PTP inhibitors rely[20,46]. We used X-ray crystallography to examine their binding mechanisms in detail. Unfortunately, only amorphadiene allowed us to identify a binding site; β-bisabolene could not be soaked into crystals, and β-bisabolol, a soluble analog, did not yield sufficient electron density for confident ligand placement (Figures S8-9). Intriguingly, amorphadiene binds to the same allosteric site targeted by benzbromarone derivatives; both molecules stabilize the WPD loop in an open conformation (Figure 3A)[48]. Amorphadiene, however, binds in two distinct ways: (i) It causes the α7 helix to reorganize to create a hydrophobic cleft (Figure 4B); this type of reorganization is interesting because it is typically slow (micro- to millisecond)[50] and difficult to incorporate into computational screens[51]. (ii) It appears to adopt multiple bound conformations (i.e., the electron density indicates regions of disorder; Figure S8). This behavior, which is

supported by molecular dynamics simulations (Figure S10), is not unusual for hydrocarbon moieties, which can exhibit enhanced mobility in protein binding pockets[52].

We probed the binding of amorphadiene and β-bisabolene further by carrying out a diverse set of kinetic analyses. First, we examined the inhibition of PTP1B by dihydroartemisinic acid. This structural analogue of amorphadiene has a carboxyl group in a position likely to interfere with binding to the hydrophobic cleft (Figure 3C). The IC_{50} of this molecule was eight-fold higher than that of amorphadiene, a reduction in potency consistent with its crystallographic pose (Figure 3D , Figure S11L). Second, we examined competition between amorphadiene and two inhibitors that bind to the active site: (i) TCS401, which causes the WPD loop to adopt a closed conformation, and (ii) orthovanadate, which does not. This analysis allowed us to determine if amorphadiene functions similarly to benzbromarones, which stabilize the WPD loop in an open conformation that is incompatible with the binding of TCS401, but not orthovanadate. Indeed, amorphadiene behaved in a similar manner (Figure 3E,F), a finding consistent with a shared allosteric mechanism. Third, we examined the sensitivity of amorphadiene and β-bisabolene to the removal of the α7 helix, which is proximal to the C-terminal allosteric site. This truncation reduced the potency of both terpenoids by four- to five-fold; the α7 helix thus appears to contribute to terpenoid binding (Figure S11). Fourth, we assessed the selectivity of amorphadiene and β-bisabolene for PTP1B over TC-PTP, its closest homolog (the two catalytic domains share 68% sequence identity[53]). Both molecules inhibited TC-PTP five- to six-fold less potently than PTP1B (Figure 3G, Figure S11A-K), a finding consistent with binding to the poorly conserved allosteric site. This selectivity may seem modest, but it matches or exceeds the selectivities of most pre-optimized inhibitors (e.g., including benzbromarones) and is rare for unfunctionalized hydrocarbons[54]. Finally, we assessed the contribution of the α7 helix to selectivity by removing the equivalent region from both PTP1B and TC-PTP (Figure 3G). This modification caused a four-fold reduction in the selectivity of amorphadiene, but not β-bisabolene; the contribution of the α7 helix to binding thus appears to differ between the two inhibitors. Our kinetic analyses, taken together, indicate that amorphadiene and β-bisabolene bind to the C-terminal allosteric site of PTP1B. Full resolution of the interaction between β-bisabolene and PTP1B will require additional experiments.

Amorphadiene and β-bisabolene are lipophilic molecules that could be valuable for their ability to pass through the membranes of mammalian cells. To examine the biological activity of these molecules, we incubated them with HEK293T/17 cells and used an enzyme-linked immunosorbent assay to measure shifts in insulin receptor (IR) phosphorylation. IR is a receptor tyrosine kinase that undergoes PTP1B-mediated dephosphorylation from the cytosolic side of the plasma membrane (PTP1B, in turn, localizes to the endoplasmic reticulum of the cell). Both molecules increased IR phosphorylation over a negative control (Figure 3G, Figure S13). We checked for off-target contributions to this signal, in turn, by repeating the ELISA with equivalent concentrations of dihydroartemisinic acid and β-bisabolol, which exhibit potencies that are eight- and twenty-fold lower than those of amorphadiene and β-bisabolene, respectively (Figure S11). To our satisfaction, both molecules led to a reduction in signal consistent with their reduced potencies for PTP1B.

Other PTPs can promote IR dephosphorylation; SHP1 and SHP2 are important examples[55–57]. To assess the potential contribution of these enzymes to the increase in IR phosphorylation observed in our ELISA experiments, we measured their inhibition by amorphadiene and β-bisabolene (Figure S12). Amorphadiene inhibited SHP2 three-fold less potently than PTP1B; its inhibition of SHP1 was too weak to measure. The low potency of β-bisabolene against SHP1 and SHP2, in turn, also precluded experimental measurement. These weak inhibitory effects, alongside the reduced ELISA signal afforded by structural analogs, suggests that the increase in IR phosphorylation caused by amorphadiene and β-bisabolene results primarily from their inhibition of PTP1B.

A Scalable Approach to Molecular Discovery

Our microbial strain provides a powerful tool for screening genes for their ability to generate novel PTP1B inhibitors. Most terpenoids, as a case study, are not commercially available, and even when their metabolic pathways are known, their biosynthesis, purification, and *in vitro* analysis is a resource-intensive process that is difficult to parallelize with existing methods[58]. Our B2H system offers a potential solution: It can identify inhibitor-synthesizing genes with a simple growth-coupled assay. We explored its application to discovery efforts by using it to screen a diverse set of uncharacterized biosynthetic genes. In brief, we carried out a bioinformatic analysis of the largest terpene synthase family (PF03936) by building and annotating a cladogram of its 4,464 constituent members (Figures S4, S5); from here, we synthesized three uncharacterized genes from each of eight clades: six with no characterized genes and two with some characterized genes (Figure 4A). We reasoned that these 24 phylogenetically diverse genes (8 from fungi, 13 from plants, and 3 from bacteria) might encode enzymes that generate distinct product profiles and, perhaps, novel sesquiterpene scaffolds.

Guided by our initial screen, we searched for sesquiterpene inhibitors by pairing each of the uncharacterized genes with the FPP pathway. To our surprise, six genes conferred a significant survival advantage (Figure 4B), and maximal resistance required an active B2H system (Figure S6). Each hit generated distinct product profiles (Figure S7); we focused our analysis on A0A0C9VSL7, which produced mostly (+)-δ-cadinene as a major product (Figure 4C, D). This terpenoid is a structural analogue of amorphadiene but has a weaker potency (IC_{50} = 165±33 μM, Figure 4E); a titer of 33±18 μM suggests that intracellular accumulation may allow it to inhibit PTP1B inside the cell. Our ability to detect a weak inhibitor suggests that the B2H system can capture a broad set of scaffolds in molecular discovery efforts. The purification and analysis of additional hits, the incorporation of alternative isoprenoid substrates of different sizes (e.g., C10 and C20 isoprenoid diphosphates generated via geranyl pyrophosphate synthase or GGPPS, respectively), or the inclusion of a larger set of uncharacterized genes could expand the scope of such efforts.

Design of Alternative PTP-specific Objectives

We explored the versatility of our B2H system by assessing its ability to detect the inactivation of several other disease-relevant PTPs. In short, we swapped out the gene for PTP1B with genes for PTPN2, PTPN6,

or PTPN12; these enzymes are targets for immunotherapeutic enhancement[59], the treatment of ovarian cancer[60], and acute myocardial infarction[61], respectively.

Their catalytic domains share 32-68% sequence identity with the catalytic domain of PTP1B. Interestingly, the new B2H systems were immediately functional; PTP inactivation permitted growth at high concentrations of spectinomycin (Figure 5A). This finding suggests that our detection system can be easily extended to other members of the PTP family.

PTP-specific B2H systems could facilitate the identification of natural products that selectively inhibit one PTP over another. We explored this application by comparing the antibiotic resistance conferred by PTP1B- and TC-PTP-specific systems in response to metabolic pathways for amorphadiene and β-bisabolene (Figure 5B). As expected, the PTP1B-specific system permitted growth at higher concentrations of antibiotic, a result consistent with the selectivity of both terpenoids for PTP1B. Indistinguishable terpenoid titers between the two strains suggest that this survival advantage does not result from differences in intracellular concentration (Figure 5C). Findings thus indicate that a simple comparison of B2H systems—a potential secondary screen—offers a simple approach for evaluating the selectivity of PTP-inhibiting metabolites. Notably, high concentrations of inhibitors in two strains could swamp out selective effects; in such cases, terpenoid levels could be reduced with lower mevalonate concentrations.

Conclusions

This study addresses an important challenge of medicinal chemistry—the design of molecular structures that inhibit disease-relevant enzymes—by using a desired biochemical activity (i.e., an objective) as a genetically encoded constraint to guide molecular biosynthesis. This approach enabled the identification of two selective, biologically active inhibitors of PTP1B, an elusive drug target[65]. These molecules are not drugs, but they are promising scaffolds for lead development. The mechanism of modulation of amorphadiene—which elicits allosteric conformational changes yet appears to rely on loose, conformationally flexible binding—is unusual (and computationally elusive[66]), and demonstrates the ability of microbial systems to find new solutions to difficult challenges in molecular design. The binding of amorphadiene (and, likely, β-bisabolene) to the allosteric site of PTP1B also distinguishes it from abietic acid, a weak, nonselective inhibitor with a carboxyl group that directs it to the active site[36]; this discrepancy suggests that unfunctionalized terpenoids may facilitate the identification of functional, nonpolar patches on protein surfaces. Our identification of unusual inhibitors in relatively small libraries, in turn, indicates that microbial systems can access a rich molecular landscape that is not efficiently explored by existing approaches to molecular discovery.

The B2H system at the core of our approach is a valuable tool for identifying biologically active natural products, which are structurally complex, difficult to synthesize, and often hidden in cryptic gene clusters[67]. It has several key advantages over contemporary approaches to inhibitor discovery: (i) It incorporates synthesizability as a search criterion—an important attribute of drug leads[68]. (ii) It is

scalable. We used a growth-coupled assay to screen 24 uncharacterized terpene synthases; this type of assay is compatible with the very large libraries (e.g., 10^{10} mutants) often sought for directed evolution[69]. (iii) It can use cellular machinery to stabilize proteins (e.g., CDC37 for Src); this capability could facilitate the integration of unstable and/or disordered targets. Future efforts to exploit these advantages by incorporating large libraries of mutated and/or reconfigured pathways, alternative biosynthetic enzymes (e.g., cytochromes P450, halogenases, and methyltransferases), or new classes of disease-relevant enzymes would be informative.

The B2H system also has important limits. When used alongside metabolic pathways, it links survival not only to the potency of metabolites, but also to their titers, off-target effects, and pathway toxicities. These limitations can be beneficial; they bias the discovery process toward potent, readily synthesizable inhibitors and could facilitate post-discovery efforts to improve the titers of interesting molecules[70]. Nonetheless, they will exclude some types of structurally complex molecules that are difficult to synthesize in *E. coli*. The use of similar activity-based screens in other organisms (e.g., *Streptomyces*) could be interesting.

The compatibility of our discovery approach with different PTPs is valuable in light of their increasingly well validated potential as a rich—and essentially untapped—source of new therapeutic targets[71]. We anticipate that some PTPs will require the use of chaperones and/or transcriptional adjustments to be incorporated into B2H systems. Our systematic optimization of the PTP1B-based system provides an experimental framework for exploring these modifications. Side-by-side comparisons of B2H systems, in turn, offer a promising strategy for evaluating inhibitor selectivity in secondary screens. In future work, new varieties of objectives (e.g., B2H systems or genetic circuits that detect the selective inhibition—or, perhaps, activation—of one PTP over another) could facilitate the discovery of molecules with sophisticated mechanisms of modulation in primary screens. The versatility of genetically encoded objectives highlights the power of using microbial systems to find targeted, biologically active molecules.

Methods

Bacterial Strains. We used *E. coli* DH10B, chemically competent NEB Turbo, or electrocompetent One Shot Top10 (Invitrogen) to carry out molecular cloning and to perform preliminary analyses of terpenoid production; we used *E. coli* BL2-DE31 to express proteins for *in vitro* studies; and we used *E. coli* s1030[72] for our luminescence studies and for all experiments involving terpenoid-mediated growth (i.e., evolution studies).

For all strains, we generated chemically competent cells by carrying out the following steps: (i) We plated each strain on LB agar plates with the required antibiotics. (ii) We used one colony of each strain to inoculate 1 mL of LB media (25 g/L LB with appropriate antibiotics listed in Table 2) in a glass culture tube, and we grew this culture overnight (37°C, 225 RPM). (iii) We used the 1-mL culture to inoculate 100-300 mL of LB media (as above) in a glass shake flask, and we grew this culture for several hours (37°C, 225 RPM). (iv) When the culture reached an OD of 0.3-0.6, we centrifuged the cells (4,000 x g for 10

minutes at 4°C), removed the supernatant, resuspended them in 30 mL of ice cold TFB1 buffer (30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% v/v glycerol, water to 200 mL, pH=5.8, sterile filtered), and incubated the suspension at 4°C for 90 min. (v) We repeated step iv, but resuspended in 4 mL of ice cold TFB2 buffer (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% glycerol, water to 50 mL, pH=6.5, sterile filtered). (iv) We split the final suspension into 100 µL aliquots and froze them at -80°C until further use.

We generated electrocompetent cells by following an approach similar to the one above. In step iv, however, we resuspended the cells in 50 mL of ice cold MilliQ water and repeated this step twice—first with 50 mL of 20% sterile glycerol (ice cold) and, then, with 1 mL of 20% sterile glycerol (ice cold). We froze the pellets as before.

Materials. We purchased methyl abietate from Santa Cruz Biotechnology; trans-caryophyllene, tris(2-carboxyethyl)phosphine (TCEP), bovine serum albumin (BSA), M9 minimal salts, phenylmethylsulfonyl fluoride (PMSF), and DMSO (dimethyl sulfoxide) from Millipore Sigma; glycerol, bacterial protein extraction reagent II (B-PERII), and lysozyme from VWR; cloning reagents from New England Biolabs; amorphadiene from Ambeed, Inc.; and all other reagents (e.g., antibiotics and media components) from Thermo Fisher. Taxadiene was a kind gift from Phil Baran of the The Scripps Research Institute. We prepared mevalonate by mixing 1 volume of 2 M DL-mevalanolactone with 1.05 volumes of 2 M KOH and incubating this mixture at 37°C for 30 minutes.

Cloning and Molecular Biology. We constructed all plasmids by using standard methods (i.e., restriction digest and ligation, Golden Gate and Gibson assembly, Quikchange mutagenesis, and circular polymerase extension cloning). Table S1 describes the source of each gene; table S2 describes the composition of all final plasmids.

We began construction of the B2H system by integrating the gene for HA4-RpoZ from pAB094a into pAB078d and by replacing the ampicillin resistance marker of pAB078d with a kanamycin resistance marker (Gibson Assembly). We modified the resulting “combined” plasmid, in turn, by replacing the HA4 and SH2 domains with kinase substrate and substrate recognition (i.e., SH2) domains, respectively (Gibson assembly), and by integrating genes for Src kinase, CDC37, and PTP1B in various combinations (Gibson assembly). We finalized the functional B2H system by modifying the SH2 domain with several mutations known to enhance its affinity for phosphopeptides (K15L, T8V, and C10A, numbered as in Kaneko et. al.[34]), by exchanging the GOI for luminescence (LuxAB) with one for spectinomycin resistance (SpecR), and by toggling promoters and ribosome binding sites to enhance the transcriptional response (Gibson assembly and Quickchange Mutagenesis, Agilent Inc.). For the last step, we also converted Pro1 to ProD by using the Quikchange protocol. When necessary, we constructed plasmids with arabinose-inducible components by cloning a single component from the B2H system into pBAD (Golden Gate assembly). Tables S3-S6 list the primers and DNA fragments used to construct each plasmid.

We assembled pathways for terpenoid biosynthesis by purchasing plasmids encoding the first module (pMBIS) and various sesquiterpene synthases (ADS or GHS in pTRC99a) from Addgene, and by building the remaining plasmids. We replaced the tetracycline resistance in pMBIS with a gene for chloramphenicol resistance to create $pMBIS_{CmR}$. We integrated genes for ABS, TXS, ABA, and GGPPS into pTRC99t (i.e., pTRC99a without BsaI sites). Tables S3-S6 list the primers and DNA fragments used to construct each plasmid.

Luminescence Assays. We characterized preliminary B2H systems (which contained LuxAB as the GOI) with luminescence assays. In brief, we transformed necessary plasmids into *E. coli* s1030 (Table S2), plated the transformed cells onto LB agar plates (20 g/L agar, 10 g/L tryptone, 10 g/L sodium chloride, and 5 g/L yeast extract with antibiotics described in Table S2), and incubated all plates overnight at 37°C. We used individual colonies to inoculate 1 ml of terrific both (TB at 2%, or 12 g/L tryptone, 24 g/L yeast extract, 12 mL/L 100% glycerol, 2.28 g/L KH_2PO_4, 12.53 g/L K_2HPO_4, pH = 7.3, and antibiotics described in Table S2), and we incubated these cultures overnight (37°C and 225 RPM). The following morning, we diluted each culture by 100-fold into 1 ml of TB media (above), and we incubated these cultures in individual wells of a deep 96-well plate for 5.5 hours (37°C, 225 RPM). (When pBAD was present, we supplemented the TB media with 0-0.02 w/v % arabinose). We transferred 100µL of each culture into a single well of a standard 96-well clear plate and measured both OD_{600} and luminescence on a Biotek Synergy plate reader (gain: 135, integration time: 1 second, read height: 1 mm). Analogous measurements of cell-free media allowed us to measure background signals, which we subtracted from each measurement prior to calculating OD-normalized luminescence (i.e., Lum / OD_{600}).

Analysis of Antibiotic Resistance. We evaluated the spectinomycin resistance conferred by various B2H systems in the absence of terpenoid pathways by carrying out the following steps: (i) We transformed *E. coli* with the necessary plasmids (Table S2) and plated the transformed cells onto LB agar plates (20 g/L agar, 10 g/L tryptone, 10 g/L sodium chloride, 5 g/L yeast extract, 50 µg/ml kanamycin, 10 µg/ml tetracycline). (ii) We used individual colonies to inoculate 1-2 ml of TB media (12 g/L tryptone, 24 g/L yeast extract, 12 mL/L 100% glycerol, 2.28 g/L KH_2PO_4, 12.53 g/L K_2HPO_4, 50 µg/ml kanamycin, 10 µg/ml tetracycline, pH = 7.3), and we incubated these cultures overnight (37°C, 225 RPM). In the morning, we diluted each culture by 100-fold into 4 ml of TB media (as above) with 0-500 µg/ml spectinomycin (we used spectinomycin in the liquid culture only for Figure S2), and we incubated these cultures in deep 24-well plates until wells containing 0 µg/ml spectinomycin reached an OD_{600} of 0.9-1.1. (iv) We diluted each 4-ml culture by 10-fold into TB media with no antibiotics and plated 10-µL drops of the diluent onto agar plates with various concentrations of spectinomycin. (v) We incubated plates overnight (37°C) and photographed them the following day.

To examine terpenoid-mediated resistance, we began with steps i and ii as described above with the addition of 34 µg/ml chloramphenicol and 50 µg/ml carbenicillin in all liquid/solid media. We then proceeded with the following steps: (iii) We diluted samples from 1-ml cultures to an OD_{600} of 0.05 in 4.5 ml of TB media (supplemented with 12 g/L tryptone, 24 g/L yeast extract, 12 mL/L 100% glycerol, 2.28

g/L KH_2PO_4, 12.53 g/L K_2HPO_4, 50 µg/ml kanamycin, 10 µg/ml tetracycline, 34 µg/ml chloramphenicol, and 50 µg/ml carbenicillin), which we incubated in deep 24-well plates (37°C, 225 RPM). (iv) At an OD_{600} of 0.3-0.6, we transferred 4 ml of each culture to a new well of a deep 24-well plate, added 500 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 20 mM of mevalonate, and incubated for 20 hours (22°C, 225 RPM). (v) We diluted each 4-ml culture to an OD_{600} of 0.1 with TB media and plated 10 µL of the diluent onto either LB or TB plates supplemented with 500 µM IPTG, 20 mM mevalonate, 50 µg/ml kanamycin, 10 µg/ml tetracycline, 34 µg/ml chloramphenicol, 50 µg/ml carbenicillin, and 0-1200 µg/ml spectinomycin (for both plates, we used 20 g/L agar with media and buffer components described above).

Terpenoid Biosynthesis. We prepared *E. coli* for terpenoid production by transforming cells with plasmids harboring requisite pathway components (Table S2) and plating them onto LB agar plates (20 g/L agar, 10 g/L tryptone, 10 g/L sodium chloride, and 5 g/L yeast extract with antibiotics described in Table S2). We used one colony from each strain to inoculate 2 ml TB (12 g/L tryptone, 24 g/L yeast extract, 12 mL/L 100% glycerol, 2.28 g/L KH_2PO_4, 12.53 g/L K_2HPO_4, pH = 7.0, and antibiotics described in Table S2) in a glass culture tube for ~16 hours (37°C and 225 RPM). We diluted these cultures by 75-fold into 10 ml of TB media and incubated the new cultures in 125 mL glass shake flasks (37°C and 225 RPM). At an OD_{600} of 0.3-0.6, we added 500 µM IPTG and 20 mM mevalonate. After 72-88 hours of growth (22°C and 225 RPM), we extracted terpenoids from each culture as outlined below. Table S9 lists exact sample sizes, culture volumes, and fermentation times.

Protein Expression and Purification. We expressed and purified PTPs as described previously[73]. Briefly, we transformed *E. coli* BL21(DE3) cells with pET16b or pET21b vectors (see Table S2 for details), and we induced with 500 µM IPTG at 22°C for 20 hours. We purified PTPs from cell lysate by using desalting, nickel affinity, and anion exchange chromatography (HiPrep 26/10, HisTrap HP, and HiPrep Q HP, respectively; GE Healthcare). We stored the final protein (30–50 µM) in HEPES buffer (50 mM, pH 7.5, 0.5 mM TCEP) in 20% glycerol at −80°C.

Extraction and Purification of Terpenoids. We used hexane to extract terpenoids generated in liquid culture. For 10-mL cultures, we added 14 mL of hexane to 10 ml of culture broth in 125-mL glass shake flasks, shook the mixture (100 RPM) for 30 minutes, centrifuged it (4000 x g), and withdrew 10 mL of the hexane layer for further analysis. For 4-mL cultures, we added 600 µL hexane to 1 mL of culture broth in a microcentrifuge tube, vortexed the tubes for 3 minutes, centrifuged the tubes for 1 minute (17000 x g), and saved 300-400 µL of the hexane layer for further analysis.

To purify amorphadiene, □-bisabolene, and (+)-δ-cadinene, we supplemented 500-1000 mL culture broth with hexane (16.7% v/v), shook the mixture for 30 minutes (100 RPM), isolated the hexane layer with a separatory funnel, centrifuged the isolated organic phase (4000 x g), and withdrew the hexane layer. To concentrate the terpenoid products, we evaporated excess hexane in a rotary evaporator to bring the final volume to 500 µL, and we passed the resulting mixture over a silica gel 1-3 times (Sigma-Aldrich; high purity grade, 60 Å pore size, 230-400 mesh particle size). We analyzed elution fractions (100% hexane) on

the GC/MS and pooled fractions with the compound of interest (amorphadiene). Once purified, we dried pooled fractions under a gentle stream of air, resuspended the concentrated terpenoids in DMSO, and quantified the final samples as outlined below. We repeated the purification process until samples (in DMSO) were >95% pure by GC/MS unless otherwise noted.

GC-MS Analysis of Terpenoids. We measured terpenoids generated in liquid culture with a gas chromatograph / mass spectrometer (GC-MS; a Trace 1310 GC fitted with a TG5-SilMS column and an ISQ 7000 MS; Thermo Fisher Scientific). We prepared all samples in hexane (directly or through a 1:100 dilution of DMSO) with 20 µg/ml of caryophyllene as an internal standard. Highly concentrated samples were diluted 10-20x prior to preparation to bring concentrations within the MS detection limit. When the peak area of an internal standard exceeded ± 40% of the average area of all samples containing that standard, we re-analyzed the corresponding samples. For all runs, we used the following GC method: hold at 80°C (3 min), increase to 250°C (15°C/min), hold at 250°C (6 min), increase to 280°C (30°C/min), and hold at 280°C (3 min). To identify various analytes, we scanned m/z ratios from 50 to 550.

We examined sesquiterpenes generated by variants of ADS by using select ion mode (SIM) to scan for the molecular ion (m/z =204). For quantification, we used Eq. 1: where A_i

$$C_i = C_{std} * \frac{A_i}{A_{std}} * R \qquad \text{(Eq. 1)}$$

$$R = \frac{A_{std,o}/C_{std,o}}{A_{ref,o}/C_{ref,o}} \qquad \text{(Eq. 2)}$$

is the area of the peak produced by analyte i, A_{std} is the area of the peak produced by C_{std} of caryophyllene in the sample, and R is the ratio of response factors for caryophyllene and amorphadiene in a reference sample. Table S12 provides the concentrations of all standards and reference compounds used in this analysis.

We quantified diterpenoids by, once again, accompanying our general procedure with several modifications: We scanned for a different molecular ion (m/z = 272) and an ion common to both diterpenoids and caryophyllene (m/z=93); we used a ratio of response factors for pure taxadiene (a kind gift from Phil Baran) and caryophyllene at m/z = 93; and we calculated peak areas m/z = 93. For all analyses, we examined only peaks with areas that exceeded 1% of the total area of all peaks at m/z=272.

We identified molecules by using the NIST MS library and, when necessary, confirmed this identification with analytical standards or mass spectra reported in the literature. The assumption of a constant response factor for different terpenoids (that is, the assumption that all sesquiterpenes and diterpenes ionize like amorphadiene and taxadiene, respectively) can certainly yield error in estimates of their concentrations; our analyses, which are consistent with those of other studies of terpenoid production in microbial systems[74,75], supply rough estimates of concentrations for all compounds except amorphadiene and taxadiene (which had analytical standards).

Bioinformatics. We used a bioinformatic analysis to identify a phylogenetically diverse set of terpene synthases. Briefly, we downloaded (i) all constituent genes of PF03936 (the largest terpene synthase family grouped by a C-terminal domain) from the PFAM Database and (ii) all enzymes with Enzyme Commission (EC) number of 4.2.3.# from the UniProt Database; this string, which defines carbon oxygen lyases that act on phosphates, includes terpene synthases. We cleaned both datasets in Excel (i.e., we ensured that every identifier had only one row), and we used a custom R script to designate each PF03936 member as characterized (i.e., in possession of a UniProt-based EC number) or uncharacterized. We accessed UniProt on May 17, 2020; this database is updated every eight weeks[76], so our estimate of characterized genes likely reflects an underestimate of the true number. Finally, we used FastTree[77] with default settings to create a phylogenetic tree of the PF03936 family and the R-package ggtree[78] to visualize the resulting tree and function data as a cladogram and heatmap.

After annotating the cladogram by hand, we selected three genes from each of six clades: six with no characterized genes and two with some characterized genes. We avoided clades proximal to known monoterpene synthases or diterpene synthases known to act on GGPP isomers absent in our system (e.g., ent-copalyl diphosphate); these enzymes are unlikely to act on FPP, the primary product of pMBIS$_{CmR}$. When selecting enzymes within clades, we biased our choice towards bacterial/fungal species and selected genes with a minimal number of common ancestors within the clade. The selected genes were synthesized and cloned into the pTrc99a vector by Twist Biosciences and assayed for antibiotic resistance as described above.

Enzyme Kinetics. To examine terpenoid-mediated inhibition, we measured phosphatase-catalyzed hydrolysis of *p*-nitrophenyl phosphate (*p*NPP) or 4-methylumbelliferyl phosphate (4-MUP, used when K_M for *p*NPP was large) in the presence of various concentrations of terpenoids. Each reaction included PTP (0.05 μM PTP1B/TCPTP or 0.1 μM SHP1/SHP2 in 50 mM HEPES, 0.5 mM TCEP, 50 μg/ml BSA), *p*NPP (0.33, 0.67, 2, 5, 10, and 15 mM) or 4-MUP (0.13, 0.27, 0.8, 2.27, 2.93, 4.53, 7.07, and 8 mM), inhibitor (with concentrations listed in the figures), buffer (50 mM HEPES pH=7.3, 50 μg/ml BSA), and DMSO at 10% v/v. We monitored the formation of *p*-nitrophenol by measuring absorbance at 405 nm every 10 seconds for 5 minutes on a SpectraMax M2 plate reader and the formation of 4-methylumbelliferyl by measuring fluorescence at 450 nm (370 nm ex, 435 nm cutoff, medium gain). We report exact sample sizes (i.e., the number of independently prepared reactions) in Table S10.

We used a custom MATLAB script to process all raw kinetic data. This script removed all concentration values that fell outside of either (i) the range of our standard curve (absorbance/fluorescence vs. μM; Figure S21) or (ii) the initial rate regime (>10% of the *p*NPP or 4-MUP concentration used in the assay). When this step reduced kinetic dataset to fewer than ten points, we re-measured those datasets to collect at least ten. We fit final datasets, in turn, with a linear regression model (using Matlab's backslash operator).

We evaluated kinetic models in three steps: (i) We fit initial-rate measurements collected in the absence and presence of inhibitors to Michaelis-Menten and inhibition models, respectively (here, we used the

nlinfit and *fminsearch* functions from MATLAB; Table S13). (ii) We used an F-test to compare the mixed model to the single-parameter model with the least sum squared error (here, we used the *fcdf* function from MATLAB to assign p-values), and we accepted the mixed model when $p < 0.05$. (iii) We used the Akaike's Information Criterion (AIC) to compare the best-fit single parameter model to each alternative single parameter model, and we accepted the "best-fit" model when the difference in AIC (Δ_i) exceed 5 for all comparisons.[79] For amorphadiene, β-bisabolene, and (+)-δ-cadinene, this criterion was not met; both noncompetitive and uncompetitive models, however, yielded indistinguishable IC50s.

We estimated the half maximal inhibitory concentration (IC_{50}) of inhibitors by using the best-fit kinetic models to determine the concentration of inhibitor required to reduce initial rates of PTP-catalyzed hydrolysis of 15 mM of *p*NPP by 50%. We used the MATLAB function "nlparci" to determine the confidence intervals of kinetic parameters, and we propagated those intervals to estimate corresponding confidence intervals for each IC_{50}.

X-ray Crystallography. We prepared crystals of PTP1B by using hanging drop vapor diffusion. In brief, we added 2 μL of PTP1B (~600 μM PTP1B, 50 mM HEPES, pH 7.3) to 6 μL of crystallization solution (100 mM HEPES, 200 mM magnesium acetate, and 14% polyethylene glycol 8000, pH 7.5) and incubated the resulting droplets over crystallization solution for one week at 4°C (EasyXtal CrystalSupport, Qiagen). We soaked crystals with ligand by transferring them to droplets formed with 6 μL of crystallization solution and 1 μL of ligand solution (10 mM in DMSO); for our apo control, in turn, we used 1 μL of ligand-free DMSO. We soaked these crystals for 2-5 days at 4°C. We prepared all crystals for freezing by soaking them in cryoprotectant formed from a 70/30 (v/v) mixture of buffer (100 mM HEPES, 200 mM magnesium acetate, and 25% polyethylene glycol 8000, pH 7.5) and glycerol.

We collected X-ray diffraction data through the Collaborative Crystallography Program at Lawrence Berkeley National Lab (ALS ENABLE, beamline 8.2.1, 100 K, 1.00003 Å). We performed integration, scaling, and merging of X-ray diffraction data using the xia2 software package[80], and we carried out molecular replacement and structure refinement with the PHENIX graphical interface,[81] supplemented with manual model adjustment in COOT[82] and one round of PDB-REDO[83] (the latter, only for the PTP1B-amorphadiene complex). Table S14 reports the statistics for X-ray crystallography.

Molecular Dynamics (MD) Simulations. Full-length PTP1B contains a disordered region that extends beyond the α7 helix (i.e., 299-435). In this study, we used a well-studied truncation variant (i.e., $PTP1B_{1-321}$) that includes residues from the disordered region. Starting from a crystal structure of the PTP1B-AD complex (which does not include the disordered tail), we used CAMPARI v.2[84] to generate structures of the disordered region (i.e., residues 288-321 for PTP1B-AD). To quickly thermalize the tail structures, we ran short Monte Carlo (MC) simulations using the ABSINTH implicit-solvent force field[85,86], constraining the coordinates of the atoms in the ligand and the protein core.

We performed MD simulations on the tail-reconstructed PTP1B-AD complex using GROMACS 2020[87]. We used the CHARMM36m protein force field[88], a CHARMM-modified TIP3P water model[89], and ligand parameters generated by CGenFF[90,91]. We solvated the protein-ligand complex in a dodecahedral box with edges positioned ≥ 10 Å from the surface of the complex, and we added six sodium ions to neutralize the system. We used the LINCS algorithm[92] to constrain all bonds involving hydrogen atoms, the Verlet leapfrog algorithm to numerically integrate equations of motion with a 2-fs time step, and the particle-mesh Ewald summation[93] (cubic interpolation with a grid spacing of 0.16 nm) to calculate long-range electrostatic interactions; we used a cutoff of 1.2 nm, in turn, for short-range electrostatic and Lennard-Jones interactions. We independently coupled the protein-ligand complex and solvent molecules to a temperature bath (300K) using a modified Berendsen thermostat[94] with a relaxation time of 0.1 ps, and we fixed pressure coupling to 1 bar using the Parrinello–Rahman algorithm[95] with a relaxation time of 2 ps and isothermal compressibility of 4.5×10^{-5} bar^{-1}.

We carried out 30 independent MD simulations to reduce sampling bias. For each MD trajectory, we minimized energy using the steepest decent method followed by 100-ps solvent relaxation in the NVT ensemble and 100-ps solvent relaxation in the NPT ensemble. After an additional 5-ns NPT equilibration, we carried out production runs for 5 ns in the NPT ensemble and registered coordinate data every 10 ps.

Analysis of PTP1B Inhibition in HEK293T Cells . We prepared HEK293T/17 cells for an enzyme-linked immunosorbent assay (ELISA) by growing them in 75 cm^2 culture flasks (Corning) with DMEM media supplemented with 10% FBS, 100 units/ml penicillin, and 100 units/ml streptomycin. We replaced the media every day for 3-5 days until the cells reached 80-100% confluency.

We measured the influence of inhibitors on insulin receptor (IR) phosphorylation by using an IR-specific ELISA (Figure S13A). Briefly, we starved cells for 48 hours in FBS-free media and incubated the with inhibitors (all at 3% DMSO) for 10 minutes. After incubation, we lysed cells with lysis buffer (9803, Cell Signaling Technology) supplemented with 1X halt phosphatase inhibitor cocktail and 1X halt protease inhibitor cocktail (Thermo Fisher Scientific) for 10 min, pelleted the cell debris, and used the lysis buffer to dilute each sample to 60 mg/ml total protein. We measured IR phosphorylation in subsequent dilutions of the 60 mg/ml samples with the PathScan® Phospho-Insulin Receptor β (panTyr) Sandwich ELISA Kit (Cell Signaling Technology; #7082). To identify biologically active concentrations of □-bisabolene and amorphadiene, we screened several concentrations and chose those that gave the highest signal (405 μM for □-bisabolene and 930 μM for amorphadiene); similar concentrations of weak inhibitors did not yield a detectable signal (Figure S13B,C).

Sequence Alignment. We used EMBOSS Needle to carry out sequence alignments[96]. Briefly, we aligned the catalytic domain of PTP1B (residues 1-301) with the corresponding regions of TC-PTP, SHP1, and PTPN12. The calculated sequence identities were 32% (PTP1B and PTPN12), 33% (PTP1B and SHP1), and 68% (PTP1B, and TCPTP).

Statistical Analysis and Reproducibility. We determined statistical significance (Fig 3H) with a two-tailed Student's t-test (details in Tables S11 and S15), and we used an F-test to compare one- and two-parameter models of inhibition (Table S13).

Data Availability. Source data for our figures is available as follows: Table S7 (Figures 1B-D, S1), Table S8 (Figures 1E, 2C, 4B, 5A,B, S2, S3D ,S6, S14), Table S9 (Figures 2G, 4C, 5C, S3A), Table S10 (Figures 2E, 3D-F, 4E, S11D-M, S12), Table S11 (Figure 3H, S13B-C). Table S14 provides refinement statistics for both protein crystal structures. All code generated for data analysis is available upon request.

Declarations

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications Website.

Accession Codes

The plasmids generated in this study are available on Addgene (https://www.addgene.org/) or from the authors. See Tables S1-S2 for accession numbers. The crystal structures determined in this study are available from the RCSB Protein Data Bank (PDB entries 6w30 and 7LFO).

AUTHOR INFORMATION

Corresponding Author

Jerome M. Fox – *Department of Chemical and Biological Engineering, University of Colorado – Boulder, Boulder, Colorado 80303, United States*;orcid.org/0000-0002-3739-1899; Email: jerome.fox@colorado.edu

Author Contributions

A.S. and J.M.F. conceived of research. A.S., E.Y.K, J.M.F, and A.H. designed experiments. A.S. and E.Y.K constructed plasmids, mutant libraries, and *E. coli* strains. A.S. carried out metabolic engineering, evolution studies, kinetic measurements, and GC/MS analyses. A.H. performed ELISA experiments and assisted with protein purification. A.S. grew crystals, and J.M.F. assisted with structural refinement. T.J., H.K., and J.M.C. carried out molecular dynamics simulations. All authors analyzed data. A.S. and J.M.F. wrote the paper.

ACKNOWLEDGEMENTS

This work was supported by funds provided by the National Science Foundation (A.S., E.Y.K, and J.M.F., award 1750244) and (A.H., award 1804897), and by the National Research Foundation of Korea (J.M.C. – Basic Science Research Program, award 2020R1I1A1A01070805; T.J. and H.K. – Creative Materials

Discovery Program, award 2017M3D1A1039378). We thank Banumathi Sankaran for assistance with data collection (Advanced Light Source; beamline 8.2.1). The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The Pilatus detector was funded under NIH grant S10OD021832. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169).

NOTES

A.S., E.Y.K, A.H., and J.M.F. are inventors on patent applications that include data from this manuscript. J.M.F. is a founder of Think Bioscience. The remaining authors declare no competing interests.

References

(1) Olsson, T. S. G.; Williams, M. a.; Pitt, W. R.; Ladbury, J. E. The Thermodynamics of Protein-Ligand Interaction and Solvation: Insights for Ligand Design. *J. Mol. Biol.***2008**, *384* (4), 1002–1017.

(2) Fox, J. M.; Zhao, M.; Fink, M. J.; Kang, K.; Whitesides, G. M. The Molecular Origin of Enthalpy/Entropy Compensation in Biomolecular Recognition. *Annu. Rev. Biophys.***2018**, *47*.

(3) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.***2017**, *46*, 531–558.

(4) Hert, J.; Irwin, J. J.; Laggner, C.; Keiser, M. J.; Shoichet, B. K. Quantifying Biogenic Bias in Screening Libraries. *Nat. Chem. Biol.***2009**, *5*, pages479–483.

(5) Smanski, M. J.; Zhou, H.; Claesen, J.; Shen, B.; Fischbach, M. A.; Voigt, C. A. Synthetic Biology to Access and Expand Nature's Chemical Diversity. *Nature Reviews Microbiology*. 2016, pp 135–149.

(6) Fürstenberg-Hägg, J.; Zagrobelny, M.; Bak, S. Plant Defense against Insect Herbivores. *Int. J. Mol. Sci.***2013**, *14* (5), 10242–10297.

(7) Maier, M. E. Design and Synthesis of Analogues of Natural Products. *Organic and Biomolecular Chemistry*. Royal Society of Chemistry May 2015, pp 5302–5343.

(8) Chen, M. S.; White, M. C. A Predictably Selective Aliphatic C-H Oxidation Reaction for Complex Molecule Synthesis. *Science (80-.).***2007**, *318* (5851), 783–787.

(9) Atanasov, A. G.; Waltenberger, B.; Pferschy-Wenzig, E. M.; Linder, T.; Wawrosch, C.; Uhrin, P.; Temml, V.; Wang, L.; Schwaiger, S.; Heiss, E. H.; et al. A Historical Overview of Natural Products in Drug Discovery. *Metabolites*. 2012, pp 1582–1614.

(10) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to Improve RD Productivity: The Pharmaceutical Industry's Grand Challenge. *Nature Reviews Drug Discovery*. Nature Publishing Group March 2010, pp 203–214.

(11) Li, J. W. H.; Vederas, J. C. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science*. July 2009, pp 161–165.

(12) Jensen, P. R.; Chavarria, K. L.; Fenical, W.; Moore, B. S.; Ziemert, N. Challenges and Triumphs to Genomics-Based Natural Product Discovery. *J. Ind. Microbiol. Biotechnol.***2014**, *41* (2), 203–209.

(13) Medema, M. H.; Blin, K.; Cimermancic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. AntiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. *Nucleic Acids Res.***2011**, *39* (suppl_2), W339–W346.

(14) Jensen, P. R. Natural Products and the Gene Cluster Revolution. *Trends Microbiol.***2016**, *24*, 968–977.

(15) Yan, Y.; Liu, Q.; Zang, X.; Yuan, S.; Bat-Erdene, U.; Nguyen, C.; Gan, J.; Zhou, J.; Jacobsen, S. E.; Tang, Y. Resistance-Gene-Directed Discovery of a Natural-Product Herbicide with a New Mode of Action. *Nature***2018**, *559*, 415–418.

(16) Culp, E. J.; Waglechner, N.; Wang, W.; Fiebig-Comyn, A. A.; Hsu, Y. P.; Koteva, K.; Sychantha, D.; Coombes, B. K.; Van Nieuwenhze, M. S.; Brun, Y. V.; et al. Evolution-Guided Discovery of Antibiotics That Inhibit Peptidoglycan Remodelling. *Nature***2020**, *578* (7796), 582–587.

(17) Zhabinskii, V. N.; Khripach, N. B.; Khripach, V. A. Steroid Plant Hormones: Effects Outside Plant Kingdom. *Steroids*. Elsevier Inc. 2015, pp 87–97.

(18) Li, Y.; Li, S.; Thodey, K.; Trenchard, I.; Cravens, A.; Smolke, C. D. Complete Biosynthesis of Noscapine and Halogenated Alkaloids in Yeast. *Proc. Natl. Acad. Sci. U. S. A.***2018**, *115* (17), E3922–E3931.

(19) Luo, X.; Reiter, M. A.; D'Espaux, L.; Wong, J.; Denby, C. M.; Lechner, A.; Zhang, Y.; Grzybowski, A. T.; Harth, S.; Lin, W.; et al. Complete Biosynthesis of Cannabinoids and Their Unnatural Analogues in Yeast. *Nature***2019**, *567*, 123–126.

(20) He, R.; Yu, Z.; Zhang, R.; Zhang, Z. Protein Tyrosine Phosphatases as Potential Therapeutic Targets. *Acta Pharmacol. Sin.***2014**, *35* (10), 1227–1246.

(21) Paul, M. K.; Mukhopadhyay, A. K. Tyrosine Kinase – Role and Significance in Cancer. *Int. J. Med. Sci.***2012**, *1* (2), 101–115.

(22) Ferguson, F. M.; Gray, N. S. Kinase Inhibitors: The Road Ahead. *Nature Reviews Drug Discovery*. 2018, pp 353–376.

(23) Stanford, S. M.; Bottini, N. Targeting Tyrosine Phosphatases: Time to End the Stigma. *Trends in Pharmacological Sciences*. 2017, pp 524–540.

(24) Krishnan, N.; Koveal, D.; Miller, D. H.; Xue, B.; Akshinthala, S. D.; Kragelj, J.; Jensen, M. R.; Gauss, C.-M.; Page, R.; Blackledge, M.; et al. Targeting the Disordered C Terminus of PTP1B with an Allosteric Inhibitor. *Nat. Chem. Biol.***2014**, *10* (7), 558–566.

(25) Barr, A. J.; Ugochukwu, E.; Lee, W. H.; King, O. N. F.; Filippakopoulos, P.; Alfano, I.; Savitsky, P.; Burgess-Brown, N. A.; Müller, S.; Knapp, S. Large-Scale Structural Analysis of the Classical Human Protein Tyrosine Phosphatome. *Cell***2009**, *136* (2), 352–363.

(26) Banno, R.; Zimmer, D.; De Jonghe, B. C.; Atienza, M.; Rak, K.; Yang, W.; Bence, K. K. PTP1B and SHP2 in POMC Neurons Reciprocally Regulate Energy Balance in Mice. *J. Clin. Invest.***2010**, *120* (3), 720–734.

(27) Zabolotny, J. M.; Kim, Y. B.; Welsh, L. A.; Kershaw, E. E.; Neel, B. G.; Kahn, B. B. Protein-Tyrosine Phosphatase 1B Expression Is Induced by Inflammation in Vivo. *J. Biol. Chem.***2008**, *283* (21), 14230–14241.

(28) Matulka, K.; Lin, H. H.; Hříbková, H.; Uwanogho, D.; Dvořák, P.; Sun, Y. M. PTP1B Is an Effector of Activin Signaling and Regulates Neural Specification of Embryonic Stem Cells. *Cell Stem Cell***2013**, *13* (6), 706–719.

(29) Zhang, H.; Wang, Y.; Wu, J.; Skalina, K.; Pfeifer, B. A. Complete Biosynthesis of Erythromycin A and Designed Analogs Using E. Coli as a Heterologous Host. *Chem. Biol.***2010**, *17* (11), 1232–1240.

(30) Antosch, J.; Schaefers, F.; Gulder, T. A. M. Heterologous Reconstitution of Ikarugamycin Biosynthesis in *E. Coli*. *Angew. Chemie Int. Ed.***2014**, *53* (11), 3011–3014.

(31) Montalibet, J.; Kennedy, B. P. Using Yeast to Screen for Inhibitors of Protein Tyrosine Phosphatase 1B. *Biochem. Pharmacol.***2004**, *68* (9), 1807–1814.

(32) Piserchio, A.; Cowburn, D.; Ghose, R. Expression and Purification of Src-Family Kinases for Solution NMR Studies. *Methods Mol. Biol.***2012**, *831*, 111–131.

(33) Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P.; Kain, W.; Nance, A. M.; Evdokimov, A.; Moshiri, F.; Turner, K. H.; et al. Continuous Evolution of Bacillus Thuringiensis Toxins Overcomes Insect Resistance. *Nature***2016**, *533* (7601), 58–63.

(34) Kaneko, T.; Huang, H.; Cao, X.; Li, X.; Li, C.; Voss, C.; Sidhu, S. S.; Li, S. S. C. C. Superbinder SH2 Domains Act as Antagonists of Cell Signaling. *Sci. Signal.***2012**, *5* (243), ra68–ra68.

(35) Jiang, C. S.; Liang, L. F.; Guo, Y. W. Natural Products Possessing Protein Tyrosine Phosphatase 1B (PTP1B) Inhibitory Activity Found in the Last Decades. *Acta Pharmacol. Sin.***2012**, *33* (10), 1217–1245.

(36) Hjortness, M. K.; Riccardi, L.; Hongdusit, A.; Ruppe, A.; Zhao, M.; Kim, E. Y.; Zwart, P. H.; Sankaran, B.; Arthanari, H.; Sousa, M. C.; et al. Abietane-Type Diterpenoids Inhibit Protein Tyrosine Phosphatases by Stabilizing an Inactive Enzyme Conformation. *Biochemistry***2018**, *57* (40), 5886–5896.

(37) Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.***2006**, *106* (8), 3412–3442.

(38) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*. 2016, pp 629–661.

(39) Dietrich, J. A.; Yoshikuni, Y.; Fisher, K. J.; Woolard, F. X.; Ockey, D.; McPhee, D. J.; Renninger, N. S.; Chang, M. C. Y.; Baker, D.; Keasling, J. D. A Novel Semi-Biosynthetic Route for Artemisinin Production Using Engineered Substrate-Promiscuous P450BM3. *ACS Chem. Biol.***2009**, *4* (4), 261–267.

(40) Mafu, S.; Jia, M.; Zi, J.; Morrone, D.; Wu, Y.; Xu, M.; Hillwig, M. L.; Peters, R. J. Probing the Promiscuity of *Ent* -Kaurene Oxidases via Combinatorial Biosynthesis. *Proc. Natl. Acad. Sci.***2016**, *113* (9), 2526–2531.

(41) Martin, V. J. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D. Engineering a Mevalonate Pathway in Escherichia Coli for Production of Terpenoids. *Nat. Biotechnol.***2003**, *21* (7), 796–802.

(42) Morrone, D.; Lowry, L.; Determan, M. K.; Hershey, D. M.; Xu, M.; Peters, R. J. Increasing Diterpene Yield with a Modular Metabolic Engineering System in E. Coli: Comparison of MEV and MEP Isoprenoid Precursor Pathway Engineering. *Appl. Microbiol. Biotechnol.***2010**, *85* (6), 1893–1906.

(43) Williams, D. C.; Wildung, M. R.; Jin, A. Q.; Dalal, D.; Oliver, J. S.; Coates, R. M.; Croteau, R. Heterologous Expression and Characterization of a “pseudomature” Form of Taxadiene Synthase Involved in Paclitaxel (Taxol) Biosynthesis and Evaluation of a Potential Intermediate and Inhibitors of the Multistep Diterpene Cyclization Reaction. *Arch. Biochem. Biophys.***2000**, *379* (1), 137–146.

(44) Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D. Designed Divergent Evolution of Enzyme Function. *Nature***2006**, *440*, 1078–1082.

(45) Peralta-Yahya, P. P.; Ouellet, M.; Chan, R.; Mukhopadhyay, A.; Keasling, J. D.; Lee, T. S. Identification and Microbial Production of a Terpene-Based Advanced Biofuel. *Nat. Commun.***2011**, *2* (1).

(46) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; et al. Allosteric Inhibition of Protein Tyrosine Phosphatase 1B. *Nat. Struct. Mol. Biol.***2004**, *11* (8), 730–737.

(47) Zhang, C.; Chen, X.; Stephanopoulos, G.; Too, H. P. Efflux Transporter Engineering Markedly Improves Amorphadiene Production in Escherichia Coli. *Biotechnol. Bioeng.***2016**, *113* (8), 1755–1763.

(48) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; et al. Allosteric Inhibition of Protein Tyrosine Phosphatase 1B. *Nat. Struct. Mol. Biol.***2004**, *11*,

730–737.

(49) Keedy, D. A.; Hill, Z. B.; Biel, J. T.; Kang, E.; Rettenmaier, T. J.; Brandao-Neto, J.; Pearce, N. M.; Delft, F. von; Wells, J. A.; Fraser, J. S. An Expanded Allosteric Network in PTP1B by Multitemperature Crystallography, Fragment Screening, and Covalent Tethering. *Elife***2018**, *7* (e36307), doi: 10.7554/eLife.36307.

(50) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.***2012**, *134* (19), 8148–8161.

(51) Amamuddy, O. S.; Veldman, W.; Manyumwa, C.; Khairallah, A.; Agajanian, S.; Oluyemi, O.; Verkhivker, G. M.; Bishop, Ö. T. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. *Int. J. Mol. Sci.***2020**, *21* (3), 847.

(52) Brandt, T.; Holzmann, N.; Muley, L.; Khayat, M.; Wegscheid-Gerlach, C.; Baum, B.; Heine, A.; Hangauer, D.; Klebe, G. Congeneric but Still Distinct: How Closely Related Trypsin Ligands Exhibit Different Thermodynamic and Structural Properties. *J. Mol. Biol.***2011**, *405* (5), 1170–1187.

(53) PTP1B and TCPTP--nonredundant phosphatases in insulin signaling and glucose homeostasis - PubMed https://pubmed.ncbi.nlm.nih.gov/22404968/ (accessed Jan 17, 2021).

(54) Shimada, T.; Yamazaki, H.; Foroozesh, M.; Hopkins, N. E.; Alworth, W. L.; Guengerich, F. P. Selectivity of Polycyclic Inhibitors for Human Cytochrome P450s 1A1, 1A2, and 1B1. *Chem. Res. Toxicol.***1998**, *11* (9), 1048–1056.

(55) Goldstein, B. J.; Bittner-Kowalczyk, A.; White, M. F.; Harbeck, M. Tyrosine Dephosphorylation and Deactivation of Insulin Receptor Substrate-1 by Protein-Tyrosine Phosphatase 1B. Possible Facilitation by the Formation of a Ternary Complex with the GRB2 Adaptor Protein. *J. Biol. Chem.***2000**, *275*, 4283–4289.

(56) Choi, E.; Kikuchi, S.; Gao, H.; Brodzik, K.; Nassour, I.; Yopp, A.; Singal, A. G.; Zhu, H.; Yu, H. Mitotic Regulators and the SHP2-MAPK Pathway Promote IR Endocytosis and Feedback Regulation of Insulin Signaling. *Nat. Commun.***2019**, *10* (1473).

(57) Dubois, M. J.; Bergeron, S.; Kim, H. J.; Dombrowski, L.; Perreault, M.; Fournès, B.; Faure, R.; Olivier, M.; Beauchemin, N.; Shulman, G. I.; et al. The SHP-1 Protein Tyrosine Phosphatase Negatively Modulates Glucose Homeostasis. *Nat. Med.***2006**, *12* (5), 549–556.

(58) Hubert, J.; Nuzillard, J. M.; Renault, J. H. Dereplication Strategies in Natural Product Research: How Many Tools and Methodologies behind the Same Concept? *Phytochemistry Reviews*. 2017, pp 55–95.

(59) Manguso, R. T.; Pope, H. W.; Zimmer, M. D.; Brown, F. D.; Yates, K. B.; Miller, B. C.; Collins, N. B.; Bi, K.; La Fleur, M. W.; Juneja, V. R.; et al. In Vivo CRISPR Screening Identifies Ptpn2 as a Cancer Immunotherapy Target. *Nature***2017**, *547*, 413–418.

(60) Varone, A.; Spano, D.; Corda, D. Shp1 in Solid Cancers and Their Therapy. *Frontiers in Oncology*. 2020, p 935.

(61) Yang, C. F.; Chen, Y. Y.; Singh, J. P.; Hsu, S. F.; Liu, Y. W.; Yang, C. Y.; Chang, C. W.; Chen, S. N.; Shih, R. H.; Hsu, S. T. D.; et al. Targeting Protein Tyrosine Phosphatase PTP-PEST (PTPN12) for Therapeutic Intervention in Acute Myocardial Infarction. *Cardiovasc. Res.***2020**, *116* (5), 1032–1046.

(62) Li, H.; Yang, F.; Liu, C.; Xiao, P.; Xu, Y.; Liang, Z.; Liu, C.; Wang, H.; Wang, W.; Zheng, W.; et al. Crystal Structure and Substrate Specificity of PTPN12. *Cell Rep.***2016**, *15* (6), 1345–1358.

(63) Paling, N. R. D.; Welham, M. J. Role of the Protein Tyrosine Phosphatase SHP-1 (Src Homology Phosphatase-1) in the Regulation of Interleukin-3-Induced Survival, Proliferation and Signalling. *Biochem. J.***2002**, *368* (3), 885–894.

(64) Van Vliet, C.; Bukczynska, P. E.; Puryer, M. A.; Sadek, C. M.; Shields, B. J.; Tremblay, M. L.; Tiganis, T. Selective Regulation of Tumor Necrosis Factor-Induced Erk Signaling by Src Family Kinases and the T Cell Protein Tyrosine Phosphatase. *Nat. Immunol.***2005**, *6* (3), 253–260.

(65) Zhang, S.; Zhang, Z. Y. PTP1B as a Drug Target: Recent Developments in PTP1B Inhibitor Discovery. *Drug Discov. Today***2007**, *12*, 373–381.

(66) 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.

(67) Rutledge, P. J.; Challis, G. L. Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters. *Nature Reviews Microbiology*. 2015, pp 509–523.

(68) Hartenfeller, M.; Schneider, G. De Novo Drug Design. In *Chemoinformatics and Computational Chemical Biology*; Bajorath, J., Ed.; Humana Press: Totowa, NJ, 2011; pp 299–323.

(69) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.***2015**, *16* (7), 379–394.

(70) Johnston, C. W.; Badran, A. H.; Collins, J. J. Continuous Bioactivity-Dependent Evolution of an Antibiotic Biosynthetic Pathway. *Nat. Commun.***2020**, *11*, 4202.

(71) Chen, M. J.; Dixon, J. E.; Manning, G. Genomics and Evolution of Protein Phosphatases. *Sci. Signal.***2017**, *10* (474), 1–17.

(72) Carlson, J. C.; Badran, A. H.; Guggiana-Nilo, D. A.; Liu, D. R. Negative Selection and Stringency Modulation in Phage-Assisted Continuous Evolution. *Nat. Chem. Biol.***2014**, *10* (3), 216–222.

(73) Hjortness, M. K.; Riccardi, L.; Hongdusit, A.; Zwart, P. H.; Sankaran, B.; Vivo, M. De; Fox, J. M.; De Vivo, M.; Fox, J. M.; Vivo, M. De; et al. Evolutionarily Conserved Allosteric Communication in Protein Tyrosine

Phosphatases. *Biochemistry***2018**, *57* (45), 6443–6451.

(74) Chen, X.; Zhang, C.; Zou, R.; Zhou, K.; Stephanopoulos, G.; Too, H. P. Statistical Experimental Design Guided Optimization of a One-Pot Biphasic Multienzyme Total Synthesis of Amorpha-4,11-Diene. *PLoS One***2013**, *8* (11), e79650.

(75) Edgar, S.; Zhou, K.; Qiao, K.; King, J. R.; Simpson, J. H.; Stephanopoulos, G. Mechanistic Insights into Taxadiene Epoxidation by Taxadiene-5α-Hydroxylase. *ACS Chem. Biol.***2016**, *11* (2), 460–469.

(76) UniProt Consortium. UniProt.org.

(77) Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2 - Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One***2010**, *5* (3), e9490.

(78) Yu, G.; Smith, D. K.; Zhu, H.; Guan, Y.; Lam, T. T. Y. Ggtree: An r Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data. *Methods Ecol. Evol.***2017**, *8* (1), 28–36.

(79) Burnham, K. P.; Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edn. Springer-Verlag, New York*; 2002; Vol. 60.

(80) Winter, G. Xia2: An Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Crystallogr.***2010**, *43* (1), 186–190.

(81) Afonine, P. V; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. D. Biol. Crystallogr.***2012**, *68* (Pt 4), 352–367.

(82) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.***2004**, *60* (12 I), 2126–2132.

(83) Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. The *PDB_REDO* Server for Macromolecular Structure Model Optimization. *IUCrJ***2014**, *1* (4), 213–220.

(84) Vitalis, A.; Pappu, R. V. Chapter 3 Methods for Monte Carlo Simulations of Biomacromolecules; Wheeler, R. A. B. T.-A. R. in C. C., Ed.; Elsevier, 2009; Vol. 5, pp 49–76.

(85) Vitalis, A.; Pappu, R. V. ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *J. Comput. Chem.***2009**, *30* (5), 673–699.

(86) Choi, J.-M.; Pappu, R. V. Improvements to the ABSINTH Force Field for Proteins Based on Experimentally Derived Amino Acid Specific Backbone Conformational Statistics. *J. Chem. Theory Comput.***2019**, *15* (2), 1367–1382.

(87) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX***2015**, *1* (2), 19–25.

(88) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods***2016**, *14* (71–73), 71–73.

(89) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B***1998**, *102* (18), 3586–3616.

(90) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.***2010**, *31* (4), 671–690.

(91) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. Extension of the CHARMM General Force Field to Sulfonyl-Containing Compounds and Its Utility in Biomolecular Simulations. *J. Comput. Chem.***2012**, *33* (31), 2451–2468.

(92) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.***1997**, *18* (12), 1463–1472.

(93) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.***1993**, *98* (12), 10089.

(94) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.***2007**, *126*, 014101.

(95) Parrinello, M. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.***1981**, *52* (12), 7182.

(96) Rice, P.; Longden, L.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.***2000**, *16* (6), 276–277.

Figures

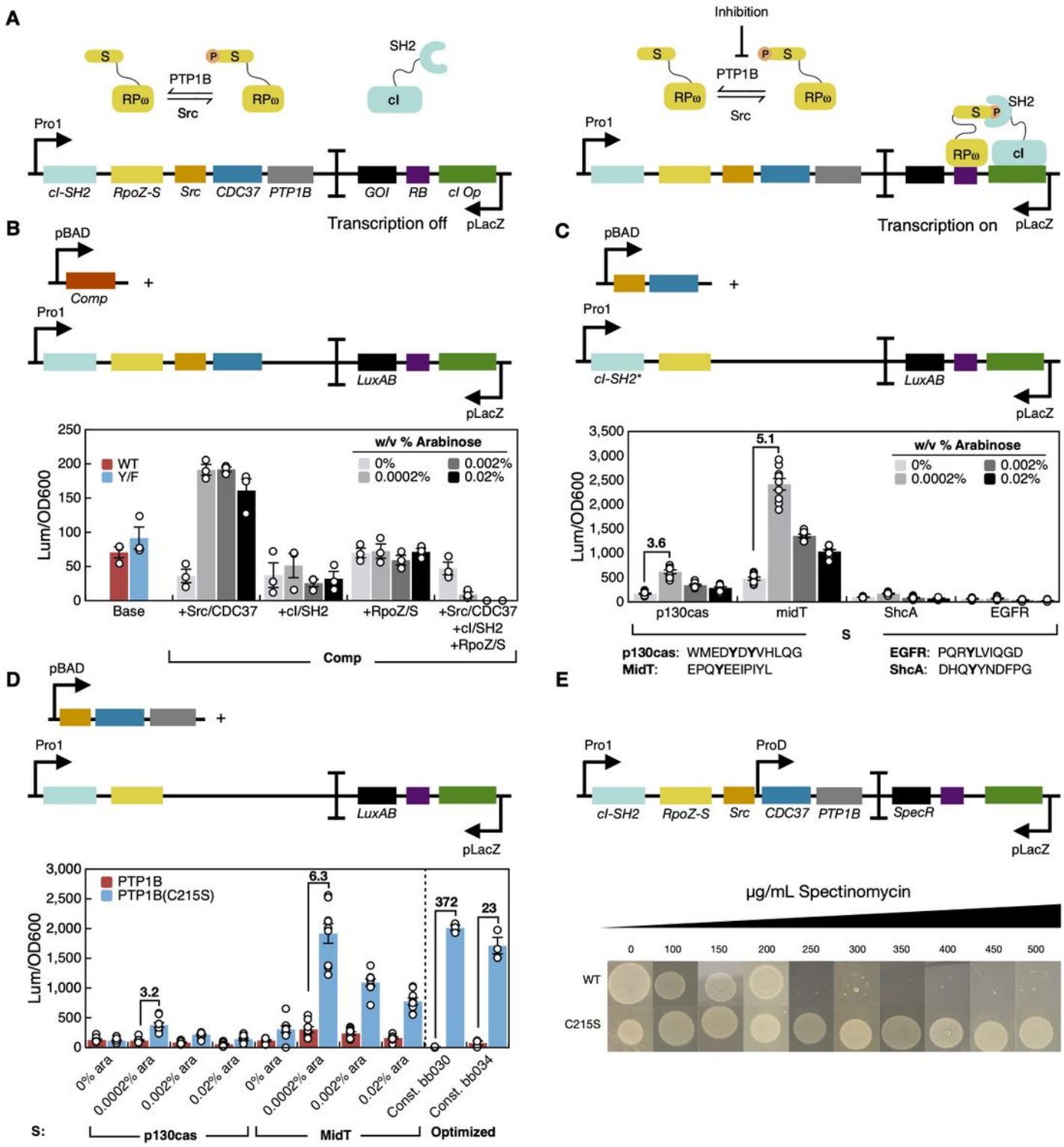

Figure 1

Development of a bacterial-two hybrid system that links the inhibition of PTP1B to antibiotic resistance. (A) A bacterial two-hybrid (B2H) system in which a phosphorylation-dependent protein-protein interaction modulates transcription of a gene of interest (GOI, black). Major components include (i) a substrate domain fused to the omega subunit of RNA polymerase (yellow), (ii) an SH2 domain fused to the 434 phage cI repressor (light blue), (iii) Src kinase and PTP1B, (iv) an operator for 434cI (dark green), (v) a

binding site for RNA polymerase (purple), and (vi) a gene of interest (GOI, black). (B) The luminescence generated by a B2H system with a p130cas substrate, LuxAB as the GOI, and no PTP1B. “Y/F” denotes a tyrosine-to-phenylalanine mutation in the substrate domain that prevents its phosphorylation. Gray bars include a WT substrate. We used an inducible plasmid to increase expression of specific components. (C) The luminescence generated by B2H systems with an SH2 domain that exhibits enhanced affinity for phosphopeptides (SH2*), one of four substrate domains, LuxAB as the GOI, and no Src or PTP1B. We used an inducible plasmid to control the expression of Src. (D) The B2H system from (C) with either p130cas or MidT substrates. We used a second plasmid to control the expression of Src and an active or inactive (C215) variant of PTP1B. Right: Two optimized single-plasmid systems. (E) The final B2H system. Inactivation of PTP1B enabled a strain of E. coli harboring this system to survive at high concentrations of spectinomycin (> 250 μg/ml). Error bars in (B)-(D) denote standard error with n = 3 biological replicates. tables S2, S7, and S8 detail the plasmids used in each B2H version.

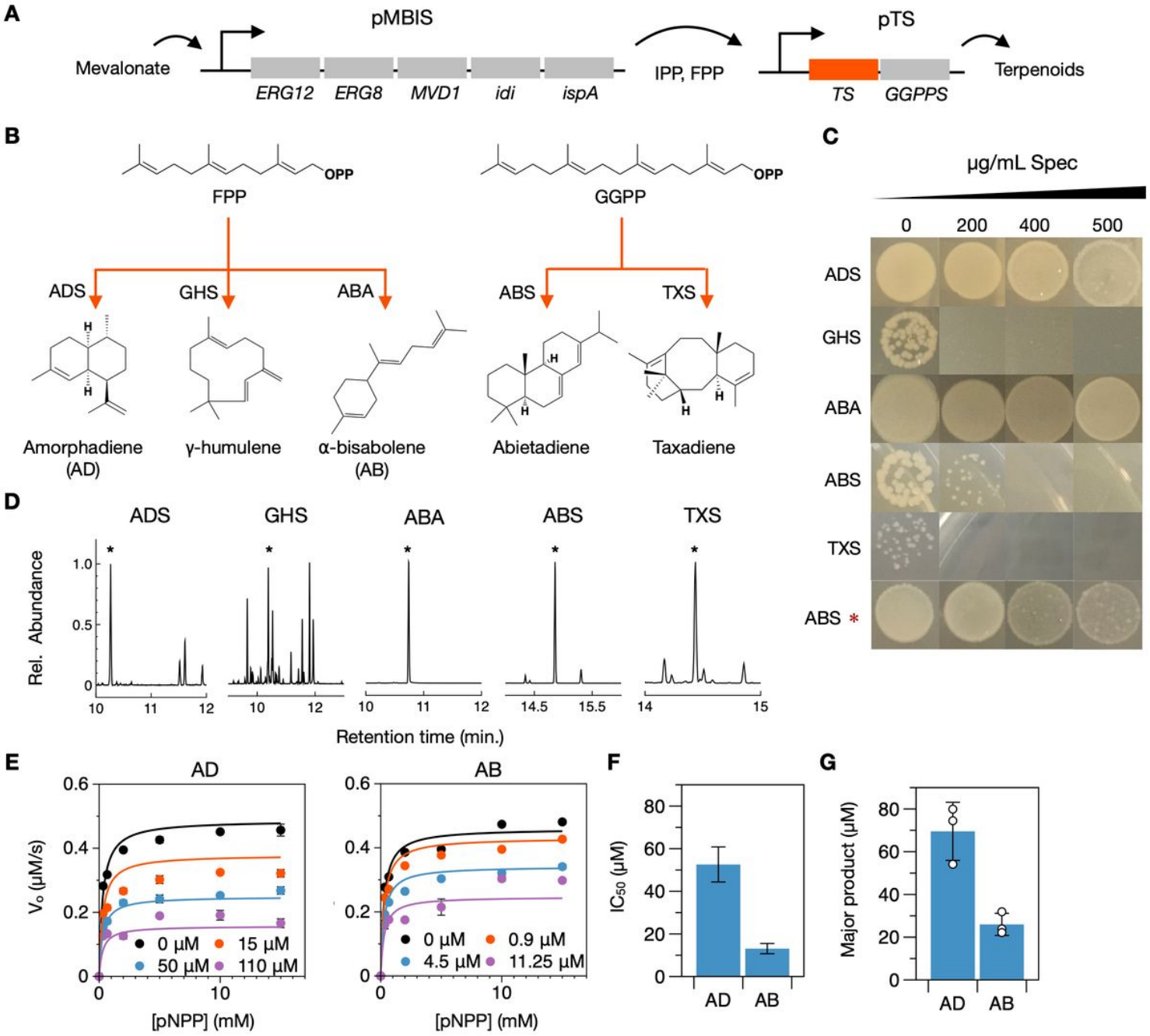

Figure 2

Biosynthesis of PTP1B-inhibiting terpenoids enables cell survival. (A) The plasmid-borne pathway for terpenoid biosynthesis: (i) pMBISCmR¬, which harbors the mevalonate-dependent isoprenoid pathway of S. cerevisiae, converts mevalonate to isopentyl pyrophosphate (IPP) and farnesyl pyrophosphate (FPP). (ii) pTS, which encodes a terpene synthase (TS) and, when necessary, a geranylgeranyl diphosphate synthase (GGPPS), converts IPP and FPP to sesquiterpenes or diterpenes. (B) Five terpene synthases examined in this study: amorphadiene synthase (ADS), γ-humulene synthase (GHS), □-bisabolene synthase (ABS), abietadiene synthase (ABS), and taxadiene synthase (TXS). (C) The spectinomycin resistance of strains of E. coli that harbor both (i) the bacterial two-hybrid (B2H) system (ii) a TS-specific terpenoid pathway. Note: ABS*, a positive control, has a constitutively active B2H (i.e., it includes

PTP1BC215S). (D) Chromatograms show expected major products (i.e., namesake; *) for each TS-specific strain from c in the presence of the B2H system. Values are normalized to the largest peak within a given sample. (E) Initial rates of PTP1B-catalyzed hydrolysis of pNPP in the presence of increasing concentrations of (AD) amorphadiene or (AB) β-bisabolene. Lines show the best-fit kinetic models of inhibition (Table S13). (F) Estimated IC50s. (G) Titers of the major products generated by ADS and ABA. Error bars denote (E) standard error and (F) 95% confidence intervals for $n \geq 3$ independent measurements (exact sample sizes reported in Table S10), and (G) standard deviation for n = 3 biological replicates.

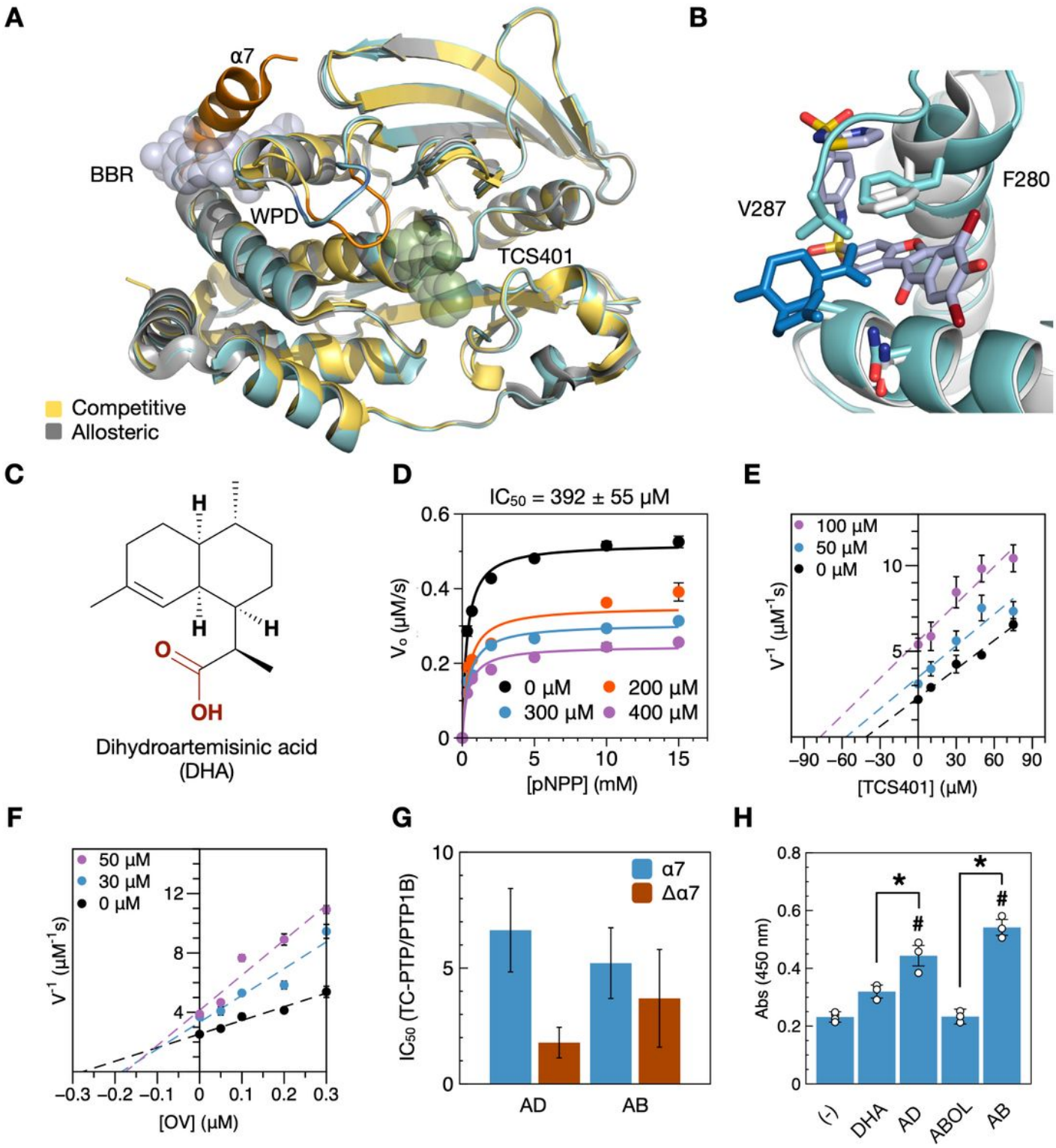

Figure 3

Biophysical analysis of terpenoid-mediated inhibition. (A) Aligned X-ray crystal structures of PTP1B bound to a competitive inhibitor (TCS401; yellow protein, orange highlights, and green spheres; pdb entry 5k9w), an allosteric inhibitor (BBR; gray protein, blue highlights, and light blue spheres; pdb entry 1t4j), and amorphadiene (inhibitor absent for clarity; cyan protein; PDB entry 6w30). (B) Aligned crystal structures of PTP1B bound to BBR (white protein and light blue ligand) and amorphadiene (cyan protein

and dark blue ligand, pdb entry 6W30). (C) Dihydroartemisinic acid (DHA), a structural analogue of amorphadiene with a carboxyl group in a position likely to disrupt binding to the hydrophobic cleft. (D) DHA is eight-fold less potent than amorphadiene. Lines show the best-fit kinetic models of inhibition (Table S13). (E) Dixon plot showing Vo-1 vs. [TCS401] at various concentrations of amorphadiene (black, blue, purple markers). The parallel lines indicate that TCS401 and amorphadiene cannot bind simultaneously. (F) Dixon plot showing Vo-1 vs. [orthovanadate] at various concentrations of amorphadiene (black, blue, purple markers). The intersecting lines indicate that orthovanadate and amorphadiene can bind simultaneously. (G) Both amorphadiene and □-bisabolene inhibit PTP1B much more potently than TC-PTP; the removal of the □7 helix (or equivalent) from both enzymes reduces the selectivity of amorphadiene, but not □-bisabolene. (H) Amorphadiene (930 μM) and □-bisabolene (405 μM) stimulate IR phosphorylation in HEK293T/17 cells; at the same concentrations, DHA and □-bisabolol (ABol) exhibit reduced signals consistent with their reduced potencies (#: $p<0.05$, compared to negative control,*: $p<0.05$). All inhibitors are dissolved in 3% DMSO (v/v; negative control). Error bars in (D)-(F) denote standard error for $n \geq 3$ independent measurements with a 95% confidence interval for the IC50. Error bars in (G) denote propagated 95% confidence intervals for $n > 3$ independent measurements. Error bars in (H) denote standard error propagated from a buffer-only control (n = 3 biological replicates). Tables S10 and S11 show exact sample sizes for all measurements.

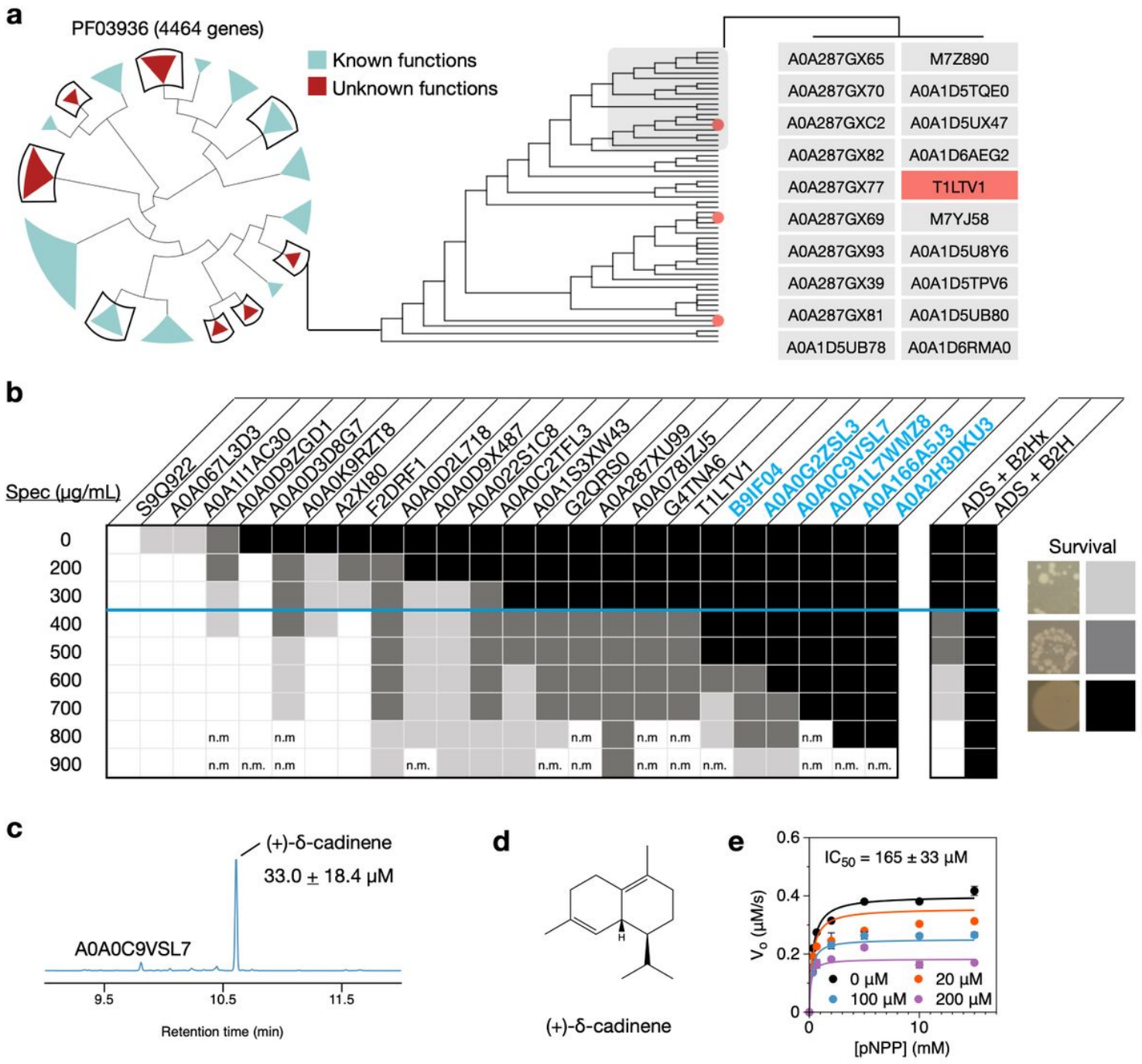

Figure 4

Analysis of uncharacterized terpene synthase genes. (A) A bioinformatic analysis of terpene synthases. We assembled a cladogram of 4,464 members of the largest terpene synthase family (PF03936) and annotated it with functional data. We selected three genes from each of eight clades (curved boxes): six with no characterized genes (i.e., genes with known functions) and two with some characterized genes. (B) The spectinomycin resistance conferred by the selected genes alongside pMBISCmR and the B2H system. Hits with robust growth beyond 400 ug/mL spectinomycin appear in blue ("n.m." indicates a condition that was not measured). (C) A0A0C9VSL7 produces (+)-δ-cadinene as a dominant product

(m/z=204). (D), Structure of (+)-δ-cadinene. (E) The inhibition of PTP1B by (+)-δ-cadinene (85% purity, 10% DMSO). Lines show the best-fit kinetic models of inhibition (Table S13).

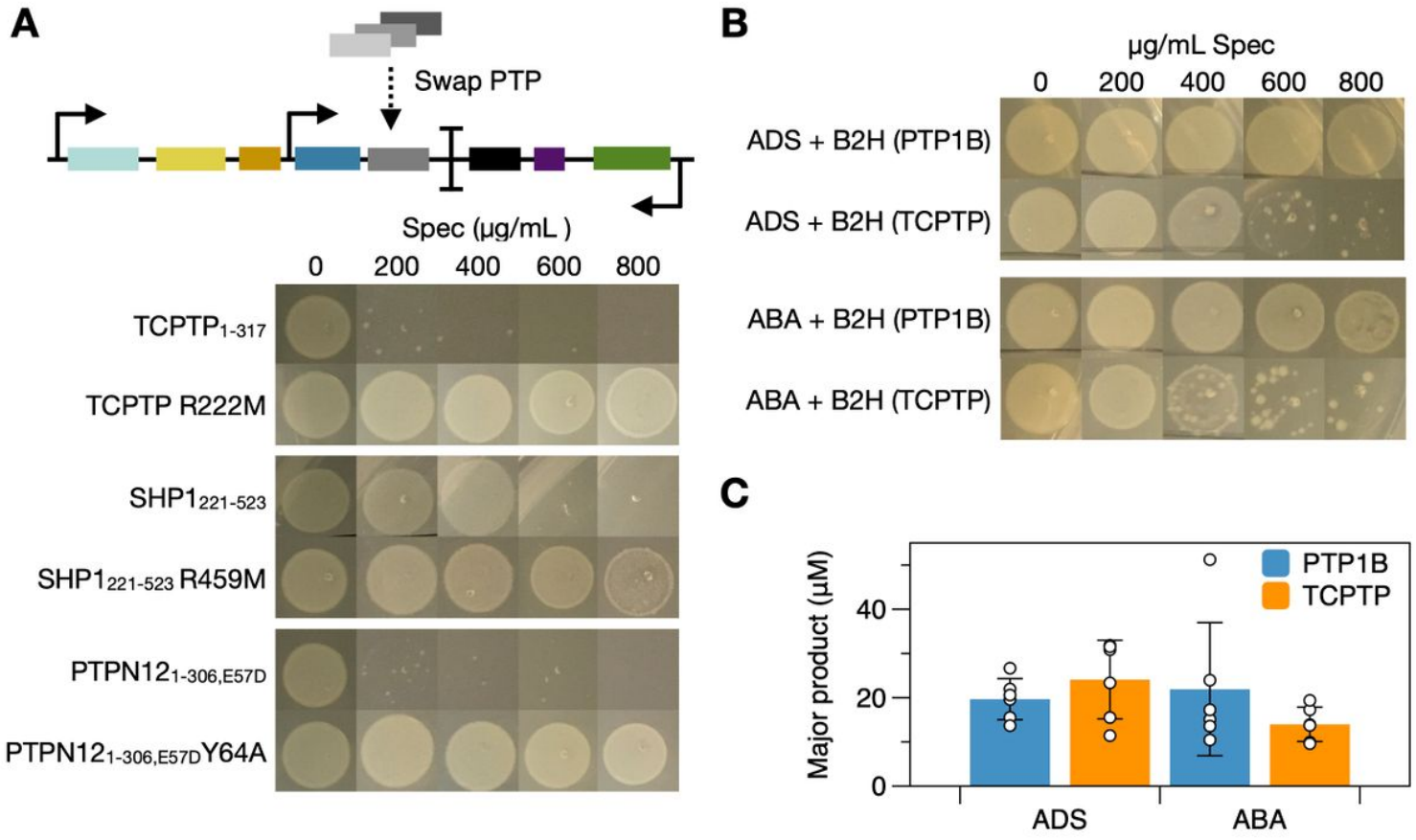

Figure 5

Extension to other disease-related PTPs. (A) The spectinomycin resistance of strains harboring B2H systems modified to detect the inactivation of different disease-relevant PTPs. Inactivating mutations62–64 confer survival at high concentrations of antibiotic. (B) A comparison of the resistance conferred by PTP1B- and TC-PTP-specific B2H systems in the presence of metabolic pathways for amorphadiene and -bisabolene (i.e., pMBISCmR + ADS or ABA). The PTP1B-specific system exhibits a prominent survival advantage, a finding consistent with the selectivity of both terpenoids for this enzyme. (C) The titers of amorphadiene and -bisabolene in strains harboring both the B2H systems and associated metabolic pathways are indistinguishable between strains.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- TableS7.xlsx
- TableS8.xlsx

- TableS9.xlsx
- TableS10.xlsx
- TableS11.xlsx