

# 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](mailto: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)

**DOI:** <https://doi.org/10.21203/rs.3.rs-92116/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# 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 target an allosteric site, which confers unusual 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<sup>1</sup>. 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<sup>2,3</sup>. Drug development, as a result, often begins with screens of large compound libraries<sup>4</sup>.

Nature has endowed living systems with the catalytic machinery to build an enormous variety of biologically active molecules—a diverse natural library<sup>5</sup>. These molecules evolved to carry out important metabolic and ecological functions (e.g., the phytochemical recruitment of predators of herbivorous insects<sup>6</sup>) 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<sup>7–9</sup>—have uncovered a diverse set of therapeutics, from aspirin to paclitaxel<sup>10</sup>. Unfortunately, these screens tend to be resource intensive<sup>11</sup>, limited by low natural titers<sup>12</sup>, and largely subject to serendipity<sup>13</sup>. Bioinformatic tools, in turn, have permitted the identification of biosynthetic gene clusters<sup>14,15</sup>, where co-localized resistance genes can reveal the biochemical function of their products<sup>16,17</sup>. The therapeutic applications of many natural products, however, differ from their native functions<sup>18</sup>, and many biosynthetic pathways can, when appropriately reconfigured, produce entirely new and, perhaps, more effective therapeutic molecules<sup>19,20</sup>. 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)<sup>21,22</sup>. The last several decades have witnessed the construction of many potent inhibitors of PTKs, which are targets for over 30 approved drugs<sup>23</sup>. Therapeutic inhibitors of PTPs, by contrast, have proven difficult to develop. These enzymes possess well conserved, positively charged active sites that make them difficult to inhibit with selective, membrane-permeable molecules<sup>24</sup>; 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<sup>25</sup>. PTP1B possesses structural characteristics that are generally representative of the PTP family<sup>26</sup> and regulates a diverse set of physiological processes (e.g., energy expenditure<sup>27</sup>, inflammation<sup>28</sup>, and neural specification in embryonic stem cells<sup>29</sup>). 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 finding inhibitor-synthesizing genes. A simple exchange of PTP genes, in turn, permitted the facile extension of our genetically encoded detection system to new targets. Our findings illustrate a versatile approach for using microbial systems to find targeted, readily synthesizable inhibitors of disease-relevant enzymes.

## Results

### Development of a genetically encoded objective

*E. coli* is a versatile platform for building natural products from unculturable or low-yielding organisms<sup>30,31</sup>. 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 (Fig. 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)<sup>32</sup>.

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 an Src homology 2 (SH2) domain (Fig. 1b); this system, which includes a chaperone that helps Src to fold (Cdc37)<sup>33</sup>, is similar to other B2H designs

that detect protein-protein binding<sup>34</sup>. 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 (Fig. 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<sup>35</sup>, and by removing the gene for Src—a modification that allowed us to control expression exclusively from a second plasmid. With this configuration, induction of Src increased luminescence most prominently for the MidT substrate (Fig. 1c), and simultaneous induction of both Src and PTP1B prevented that increase—an indication of intracellular PTP1B activity (Fig. 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 (Fig. 1d; Supplementary Figs. 1-2), 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 (Fig. 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, a structurally diverse class of secondary metabolites with largely nonpolar structures (Fig. 2a). Terpenoids include over 80,000 known compounds and represent nearly one-third of all characterized natural products<sup>36</sup> (the basis of approximately 50% of clinically approved drugs<sup>37</sup>). To begin, we focused on a handful of structurally diverse terpenoids without established inhibitory effects (Fig. 2b): amorphadiene,  $\alpha$ -humulene,  $\alpha$ -bisabolene, abietadiene, and taxadiene. Each terpenoid pathway consisted of two plasmid-borne modules: (i) the mevalonate-dependent isoprenoid pathway from *S. cerevisiae* (optimized for expression in *E. coli*<sup>38</sup>) and (ii) a terpene synthase supplemented, when necessary for diterpenoid production, with a geranylgeranyl diphosphate synthase. These modules generated terpenoids at titers of 0.3-18 mg/L in *E. coli* (Supplementary Fig. 3a).

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 (Fig. 2c). To our surprise, pathways for amorphadiene and  $\alpha$ -bisabolene permitted survival at high concentrations of antibiotic. Critically, GC-MS traces confirmed that all pathways generated terpenoids in the presence of the B2H system (Fig. 2d, Supplementary Fig. 3b,c), and maximal resistance of the amorphadiene- and  $\alpha$ -bisabolene-producing strains required both an active terpene synthase and a functional B2H system (Supplementary Fig. 3d).

We confirmed the inhibitory effects of purified terpenoids by examining their influence on PTP1B-catalyzed hydrolysis of p-nitrophenyl phosphate (pNPP; Fig. 2e, Supplementary Table 13). The IC<sub>50</sub>'s for amorphadiene and  $\alpha$ -bisabolene were 53 ± 8 μM and 13 ± 2 μM, respectively, in 10% DMSO (Fig. 2f). These IC<sub>50</sub>'s are surprisingly strong for unfunctionalized hydrocarbons (i.e., they are similar to the IC<sub>50</sub>'s of inhibitors that form hydrogen bonds and other stabilizing interactions with PTP1B<sup>21,39</sup>) and,

importantly, resemble terpenoid concentrations in liquid culture (Fig. 2g), a finding consistent with *in vivo* inhibition. We note: Our estimates of potency and concentration are probably conservative. DMSO, which we used as a cosolvent for kinetic assays, increases the solubility of nonpolar inhibitors but tends to reduce their apparent potency (likely by increasing free energetic cost of desolvation<sup>40</sup>), and terpenoids tend to accumulate intracellularly, where their concentrations can exceed extracellular levels by an order of magnitude<sup>41</sup>. Our growth-coupled assays, kinetic assays, and production measurements, taken together, indicate that amorphadiene and  $\alpha$ -bisabolene activate the B2H system by inhibiting PTP1B inside the cell.

## 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<sup>42</sup>. 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 (Supplementary Figs. 4,5); from here, we synthesized three uncharacterized genes from each of eight clades: six with no characterized genes and two with some characterized genes (Fig. 3a). We reasoned that these 24 phylogenetically diverse genes might encode enzymes with different product profiles.

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 (Fig. 3B), and maximal resistance required an active B2H system (Supplementary Fig. 6). Each hit generated distinct product profiles (Supplementary Fig. 7); we focused our analysis on A0A0C9VSL7, which produced mostly (+)-1(10),4-cadinadiene as a major product (Fig. 3c, d). This terpenoid is a structural analog of amorphadiene but has a weaker potency ( $IC_{50} = 165 \pm 33 \mu M$ ); a titer is  $33 \pm 18 \mu M$  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 isoprenoid substrates of different sizes (e.g., GGPP), and the inclusion of more uncharacterized genes could expand the scope of such efforts.

## Biophysical analysis of PTP1B inhibitors

Allosteric inhibitors of PTPs are valuable starting points for drug development. These molecules bind outside of the well conserved, positively charged active sites of PTPs and tend to have improved selectivities and membrane permeabilities over substrate analogs<sup>21</sup>. Motivated by these considerations, an early screen identified a benzbromarone derivative that inhibited PTP1B weakly ( $IC_{50} = 350 \mu M$ )

without competing with substrates; subsequent optimization of this compound led to two improved inhibitors ( $IC_{50}$ 's = 8 and 22  $\mu$ M) that bind to an allosteric site<sup>39</sup> (Fig. 4a). Over the next 15 years, efforts to find new inhibitors that bind to this site—or other allosteric regions on the catalytic domain—have been largely unsuccessful<sup>43</sup>. Benzbromarone derivatives are the only 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<sup>25</sup>). 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  $\beta$ -bisabolene provide examples. They are highly nonpolar and, thus, incapable of engaging in the hydrogen bonds and electrostatic interactions on which most other PTP inhibitors rely<sup>21,39</sup>. To examine their binding mechanisms in detail, we collected X-ray crystal structures of PTP1B bound to amorphadiene and  $\beta$ -bisabolol, a soluble analogue of  $\beta$ -bisabolene that can be soaked into protein crystals. Intriguingly, both molecules bind to the allosteric site targeted by benzbromarone derivatives (Fig. 4a). Amorphadiene, however, causes the  $\alpha$ 7 helix to reorganize to create a hydrophobic cleft (Fig. 4b); this type of reorganization is interesting because it is typically slow (micro- to millisecond)<sup>44</sup> and difficult to incorporate into computational ligand design<sup>45</sup>. By contrast,  $\beta$ -bisabolol wraps around F280 with an orientation orthogonal to that of the crystallized benzobromarone derivatives (Fig. 4c). We note:  $\beta$ -bisabolol is a  $\sim$  20-fold weaker inhibitor than  $\beta$ -bisabolene and may adopt a different bound pose (Supplementary Fig. 9m); unfortunately, the low solubility of  $\beta$ -bisabolene hindered direct structural studies.

Our crystal structures suggest that both terpenoids adopt multiple bound conformations (i.e., the electron density indicates regions of disorder; Supplementary Fig. 8 a,b). Molecular dynamics simulations provide additional support for this behavior (Supplementary Fig. 8 c-4e) which, we note, is difficult to incorporate into computational screens<sup>46</sup>. To probe the binding of amorphadiene and  $\beta$ -bisabolene further, we carried out several additional analyses. First, we examined the inhibition of PTP1B by dihydroartemisinic acid. This structural analogue of amorphadiene has a carboxyl group that, according to our crystal structures, should interfere with binding to the hydrophobic cleft created by the  $\alpha$ 7 helix (Fig. 4d). The  $IC_{50}$  of this molecule was eight-fold higher than that of amorphadiene, a reduction in potency consistent with its crystallographic pose (Fig. 4e, Supplementary Fig. 9l). Second, we assessed the inhibitory effects of amorphadiene and  $\beta$ -bisabolene against TC-PTP, the most closely related phosphatase to PTP1B. These molecules inhibited TC-PTP five- to six-fold less potently than PTP1B (Fig. 4f, Supplementary Fig. 9a-9k)—a finding consistent with binding to the poorly conserved allosteric site. Importantly, this selectivity may seem modest, but it matches or exceeds the selectivities of most pre-optimized inhibitors (including benzobromarone derivatives) and is exceedingly rare for unfunctionalized hydrocarbons (particularly, in light of their comparatively modest molecular weights)<sup>47</sup>. We assessed the contribution of the  $\alpha$ 7 helix to selectivity by removing the equivalent region from both PTP1B and TC-PTP (Fig. 4f). This modification caused a four-fold reduction in the selectivity of amorphadiene, but not  $\beta$ -bisabolene—an effect consistent with the unique involvement of the  $\alpha$ 7 helix in the binding of amorphadiene.

Amorphadiene and  $\beta$ -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 (Fig. 4g, Supplementary Fig. 10). We checked for off-target contributions to this signal, in turn, by repeating the ELISA with equivalent concentrations of dihydroartemisinin and  $\beta$ -bisabolol. To our satisfaction, both molecules led to a reduction in signal consistent with their reduced potencies.

### 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 diseases-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<sup>48</sup>, the treatment of ovarian cancer<sup>49</sup>, and acute myocardial infarction<sup>50</sup>, respectively. Their catalytic domains share 31-65% 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 (Fig. 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  $\beta$ -bisabolene (Fig. 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 difference in intracellular concentration (Fig. 5c). Findings thus indicate that a simple comparison of B2H systems—a potential secondary screen—offers a simple approach for evaluating the selectivity PTP-inhibiting gene products. 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.

## Discussion

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<sup>54</sup>. These molecules are not drugs, but they are promising scaffolds for lead development. Their mechanisms of modulation—which elicit allosteric conformational changes yet appear to rely on loose, conformationally flexible binding—are unusual (and computationally elusive<sup>55</sup>), and demonstrate the ability of microbial systems to find new

solutions to difficult challenges in molecular design. Our identification of unusual inhibitors in relatively small libraries, in turn, suggests 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<sup>56</sup>. 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<sup>57</sup>. (ii) It is scalable. We used a growth-coupled assay to screen 24 uncharacterized terpene synthases; this type of assay is compatible with very large libraries (e.g.,  $10^{10}$ )<sup>58</sup>. (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, thus, facilitate post-discovery efforts to improve the titers of interesting molecules<sup>59</sup>. 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<sup>60</sup>. 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<sup>61</sup> 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 S2) 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<sub>2</sub>, 50 mM MnCl<sub>2</sub>, 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<sub>2</sub>, 10 mM RbCl, 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; Tables S2 and S3 describe 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.<sup>35</sup>), 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.). We note: 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 S4-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<sub>CmR</sub>. We integrated genes for ABS, TXS, ABA, and GGPPS into pTRC99t (i.e., pTRC99a without BsaI sites). Tables S4-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<sub>2</sub>PO<sub>4</sub>, 12.53 g/L K<sub>2</sub>HPO<sub>4</sub>, 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). (We note: 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<sub>600</sub> 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<sub>600</sub>).

**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<sub>2</sub>PO<sub>4</sub>, 12.53 g/L K<sub>2</sub>HPO<sub>4</sub>, 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<sub>600</sub> 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<sub>600</sub> 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<sub>2</sub>PO<sub>4</sub>, 12.53 g/L K<sub>2</sub>HPO<sub>4</sub>, 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<sub>600</sub> 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<sub>600</sub> 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<sub>2</sub>PO<sub>4</sub>, 12.53 g/L K<sub>2</sub>HPO<sub>4</sub>, 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<sub>600</sub> 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 (Hjortness et al., 2018). 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, β-bisabolene, and (+)-1(10),4-cadinadiene, 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  $\mu$ L, 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 terpenoid solids 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  $\mu$ g/ml of caryophyllene or methyl abietate 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  $\pm$  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  $A_i$

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

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

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. Tables S12-14 provide the concentrations of all standards and reference compounds used in this analysis.

We quantified sesquiterpenes generated by variants of GHS by using the aforementioned procedure with several modifications: We used methyl abietate as an internal standard (several mutants of GHS generate caryophyllene as a product); we scanned for both m/z = 204 and m/z = 121, a common ion between sesquiterpenes and methyl abietate; we used a ratio of response factors for amorphadiene and methyl abietate at m/z = 121 for R; and we calculated peak areas at m/z = 121. We focused our analysis on peaks with areas that exceeded 1% of the total area of all peaks at m/z=204.

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. We note: 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<sup>62,63</sup>, 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. Finally, we used FastTree<sup>64</sup> to create a phylogenetic tree of the PF03936 family and the R-package ggtree<sup>65</sup> 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<sub>CmR</sub>. 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 PTP1B or TCPTP-catalyzed hydrolysis of p-nitrophenyl phosphate (pNPP) in the presence of various concentrations of terpenoids. Each reaction included PTP (0.05  $\mu$ M in 50 mM HEPES, 0.5 mM TCEP, 50  $\mu$ g/ml BSA), pNPP (0.33, 0.67, 2, 5, 10, and 15 mM), inhibitor (with concentrations listed in the figures), buffer (50 mM HEPES pH=7.3, 50  $\mu$ g/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. We report exact sample sizes (i.e., the number of independently prepared reactions) in Supplementary Table 10.

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 vs.  $\mu\text{M}$ ; Supplementary Fig. 18) or (ii) the initial rate regime ( $>10\%$  of the pNPP 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; Supplementary Table 13). (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 ( $\Delta_i$ ) exceed 5 for all comparisons.<sup>66</sup> We note: For amorphadiene,  $\beta$ -bisabolene, and (+)1-(10),4-cadinadiene this criterion was not met; both noncompetitive and uncompetitive models, however, yielded indistinguishable  $\text{IC}_{50}$ 's.

We estimated the half maximal inhibitory concentration ( $\text{IC}_{50}$ ) 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 pNPP 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  $\text{IC}_{50}$ .

**X-ray crystallography.** We prepared crystals of PTP1B by using hanging drop vapor diffusion. In brief, we added 2  $\mu\text{L}$  of PTP1B ( $\sim 600 \mu\text{M}$  PTP1B, 50 mM HEPES, pH 7.3) to 6  $\mu\text{L}$  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  $4^\circ\text{C}$  (EasyXtal CrystalSupport, Qiagen). We soaked crystals with ligand by transferring them to droplets formed with 6  $\mu\text{L}$  of crystallization solution and 1  $\mu\text{L}$  of ligand solution (10 mM in DMSO), which we incubated for 2-5 days at  $4^\circ\text{C}$ . We prepared all ligands 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,  $1.00003 \text{ \AA}$ ). We performed integration, scaling, and merging of X-ray diffraction data using the xia2 software package<sup>67</sup>, and we carried out molecular replacement and structure refinement with the PHENIX graphical interface,<sup>68</sup> supplemented with manual model adjustment in COOT<sup>69</sup> and one round of PDB-REDO<sup>70</sup> (the latter, only for the PTP1B-amorphadiene complex).

**Molecular dynamics (MD) simulations.** We performed MD simulations using GROMACS 2020<sup>71</sup>. Briefly, we used the CHARMM36m protein force field<sup>72</sup>, a CHARMM-modified TIP3P water model<sup>73</sup>, and ligand

parameters generated by CGenFF<sup>74,75</sup>. We solvated each PTP1B-ligand complex (initialized from the corresponding crystal structure) in a dodecahedral box with edges positioned  $\geq 10$  Å from the surface of the complex, and we added sodium ions (three for amorphadiene and one for  $\beta$ -bisabolol) to neutralize each system. We used the LINCS algorithm<sup>76</sup> 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<sup>77</sup> (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<sup>78</sup> with a relaxation time of 0.1 ps, and we fixed pressure coupling to 1 bar using the Parrinello–Rahman algorithm<sup>79</sup> with a relaxation time of 2 ps and isothermal compressibility of  $4.5 \times 10^{-5}$  bar<sup>-1</sup>.

For each system, 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 1-ns NPT equilibration, we carried out production runs for 1 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<sup>2</sup> 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 (Supplementary Fig. 10a). 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  $\beta$  (panTyr) Sandwich ELISA Kit (Cell Signaling Technology; #7082). We note: To identify biologically active concentrations of  $\beta$ -bisabolene and amorphadiene, we screened several concentrations and chose those that gave the highest signal (405  $\mu$ M for  $\beta$ -bisabolene and 930  $\mu$ M for amorphadiene); similar concentrations of weak inhibitors did not yield a detectable signal (Supplementary Fig. 10b,c).

**Statistical analysis and reproducibility.** We determined statistical significance (Figs 3g) with a two-tailed Student's t-test (details in Supplementary Tables 11 and 15), and we used an F-test to compare one- and two-parameter models of inhibition (Supplementary Table 13).

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

**Data availability.** The plasmids generated in this study are available on Addgene (<https://www.addgene.org/>) or from the authors. All code generated for data analysis is available upon request. Source data for our figures is available as follows: Supplementary Table 7 (Fig. 1b-d, Supplementary Fig. 1), Supplementary Table 8 (Figs. 1e, 2c, 4b; Supplementary Figs. 2, 3d, 9,11,12), Supplementary Table 9 (Fig. 2g, 4c, 5c; Supplementary Fig. 4a), Supplementary Table 10 (Fig. 2e, 3g, 4e; Supplementary Fig. 5d-o), Supplementary Table 11 (Fig. 4g; Supplementary Fig. 6b-c). The crystal structures determined in this study are available from the RCSB Protein Data Bank (PDB entry 6w30, 6w31). Table 14 provides refinement statistics for both structures.

## Declarations

### 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 2019R1A6A1A10073887; 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 S100D021832. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169).

### 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.

### COMPETING INTERESTS

A.S., 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.

### ADDITIONAL INFORMATION

**Supplementary information** available in the online version of this paper



Correspondence and requests for materials should be addressed to J.M.F

## 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.* **384**, 1002–1017 (2008).
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.* **47**, (2018).
3. Mobley, D. L. & Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **46**, 531–558 (2017).
4. Hert, J., Irwin, J. J., Laggner, C., Keiser, M. J. & Shoichet, B. K. Quantifying biogenic bias in screening libraries. *Nat. Chem. Biol.* **5**, pages479–483 (2009).
5. Smanski, M. J. *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nature Reviews Microbiology* **14**, 135–149 (2016).
6. Fürstenberg-Hägg, J., Zagrobelny, M. & Bak, S. Plant defense against insect herbivores. *International Journal of Molecular Sciences* **14**, 10242–10297 (2013).
7. Maier, M. E. Design and synthesis of analogues of natural products. *Organic and Biomolecular Chemistry* **13**, 5302–5343 (2015).
8. Chen, M. S. & White, M. C. A predictably selective aliphatic C-H oxidation reaction for complex molecule synthesis. *Science (80- )*. **318**, 783–787 (2007).
9. Cho, I., Jia, Z. J. & Arnold, F. H. Site-selective enzymatic C-H amidation for synthesis of diverse lactams. *Science (80- )*. **364**, 575–578 (2019).
10. Atanasov, A. G. *et al.* A Historical overview of natural products in drug discovery. *Metabolites* **33**, 1582–1614 (2012).
11. Paul, S. M. *et al.* How to improve RD productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery* **9**, 203–214 (2010).
12. Li, J. W. H. & Vederas, J. C. Drug discovery and natural products: End of era or an endless frontier? *Biomeditsinskaya Khimiya* **57**, 148–160 (2011).
13. Jensen, P. R., Chavarria, K. L., Fenical, W., Moore, B. S. & Ziemert, N. Challenges and triumphs to genomics-based natural product discovery. *Journal of Industrial Microbiology and Biotechnology* **41**, 203–209 (2014).
14. Medema, M. H. *et al.* AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339–W346 (2011).
15. Jensen, P. R. Natural Products and the Gene Cluster Revolution. *Trends Microbiol.* **24**, 968–977 (2016).
16. Yan, Y. *et al.* Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action. *Nature* **559**, 415–418 (2018).

17. Culp, E. J. *et al.* Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* **578**, 582–587 (2020).
18. Zhabinskii, V. N., Khripach, N. B. & Khripach, V. A. Steroid plant hormones: Effects outside plant kingdom. *Steroids* **97**, 87–97 (2015).
19. Li, Y. *et al.* Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3922–E3931 (2018).
20. Luo, X. *et al.* Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123–126 (2019).
21. He, R., Yu, Z., Zhang, R. & Zhang, Z. Protein tyrosine phosphatases as potential therapeutic targets. *Acta Pharmacol. Sin.* **35**, 1227–1246 (2014).
22. Paul, M. K. & Mukhopadhyay, A. K. Tyrosine kinase – Role and significance in Cancer. *Int. J. Med. Sci.* **1**, 101–115 (2012).
23. Ferguson, F. M. & Gray, N. S. Kinase inhibitors: The road ahead. *Nature Reviews Drug Discovery* **17**, 353–376 (2018).
24. Stanford, S. M. & Bottini, N. Targeting Tyrosine Phosphatases: Time to End the Stigma. *Trends in Pharmacological Sciences* **38**, 524–540 (2017).
25. Krishnan, N. *et al.* Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat. Chem. Biol.* **10**, 558–566 (2014).
26. Barr, A. J. *et al.* Large-Scale Structural Analysis of the Classical Human Protein Tyrosine Phosphatome. *Cell* **136**, 352–363 (2009).
27. Banno, R. *et al.* PTP1B and SHP2 in POMC neurons reciprocally regulate energy balance in mice. *J. Clin. Invest.* **120**, 720–734 (2010).
28. Zabolotny, J. M. *et al.* Protein-tyrosine phosphatase 1B expression is induced by inflammation in vivo. *J. Biol. Chem.* **283**, 14230–14241 (2008).
29. Matulka, K. *et al.* PTP1B is an effector of activin signaling and regulates neural specification of embryonic stem cells. *Cell Stem Cell* **13**, 706–719 (2013).
30. 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.* **17**, 1232–1240 (2010).
31. Antosch, J., Schaefer, F. & Gulder, T. A. M. Heterologous Reconstitution of Ikarugamycin Biosynthesis in *E. coli*. *Angew. Chemie Int. Ed.* **53**, 3011–3014 (2014).
32. Montalibet, J. & Kennedy, B. P. Using yeast to screen for inhibitors of protein tyrosine phosphatase 1B. *Biochem. Pharmacol.* **68**, 1807–1814 (2004).
33. Piserchio, A., Cowburn, D. & Ghose, R. Expression and purification of Src-family kinases for solution NMR studies. *Methods Mol. Biol.* **831**, 111–131 (2012).
34. Badran, A. H. *et al.* Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **533**, 58–63 (2016).
35. Kaneko, T. *et al.* Superbinder SH2 domains act as antagonists of cell signaling. *Sci. Signal.* **5**, (2012).

36. Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **117**, 11570–11648 (2017).
37. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products* **79**, 629–661 (2016).
38. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802 (2003).
39. Wiesmann, C. *et al.* Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat. Struct. Mol. Biol.* **11**, 730–737 (2004).
40. Hjortness, M. K. *et al.* Abietane-Type Diterpenoids Inhibit Protein Tyrosine Phosphatases by Stabilizing an Inactive Enzyme Conformation. *Biochemistry* **57**, 5886–5896 (2018).
41. Zhang, C., Chen, X., Stephanopoulos, G. & Too, H. P. Efflux transporter engineering markedly improves amorphadiene production in *Escherichia coli*. *Biotechnol. Bioeng.* **113**, 1755–1763 (2016).
42. 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* **16**, 55–95 (2017).
43. Keedy, D. A. *et al.* An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *Elife* **7**, doi: 10.7554/eLife.36307 (2018).
44. Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying ‘invisible’ excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012).
45. Amamuddy, O. S. *et al.* Integrated computational approaches and tools for allosteric drug discovery. *Int. J. Mol. Sci.* **21**, 847 (2020).
46. Menchon, G., Maveyraud, L. & Czaplicki, G. Molecular dynamics as a tool for virtual ligand screening. in *Methods in Molecular Biology* **1762**, 145–178 (Humana Press Inc., 2018).
47. Shimada, T. *et al.* Selectivity of Polycyclic Inhibitors for Human Cytochrome P450s 1A1, 1A2, and 1B1. *Chem. Res. Toxicol.* **11**, 1048–1056 (1998).
48. Manguso, R. T. *et al.* In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* **547**, 413–418 (2017).
49. Varone, A., Spano, D. & Corda, D. Shp1 in Solid Cancers and Their Therapy. *Frontiers in Oncology* **10**, 935 (2020).
50. Yang, C. F. *et al.* Targeting protein tyrosine phosphatase PTP-PEST (PTPN12) for therapeutic intervention in acute myocardial infarction. *Cardiovasc. Res.* **116**, 1032–1046 (2020).
51. Li, H. *et al.* Crystal Structure and Substrate Specificity of PTPN12. *Cell Rep.* **15**, 1345–1358 (2016).
52. 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.* **368**, 885–894 (2002).
53. Van Vliet, C. *et al.* Selective regulation of tumor necrosis factor-induced Erk signaling by Src family kinases and the T cell protein tyrosine phosphatase. *Nat. Immunol.* **6**, 253–260 (2005).

54. Zhang, S. & Zhang, Z. Y. PTP1B as a drug target: recent developments in PTP1B inhibitor discovery. *Drug Discov. Today* **12**, 373–381 (2007).
55. 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.* **138**, 14257–14263 (2016).
56. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology* **13**, 509–523 (2015).
57. Hartenfeller, M. & Schneider, G. De novo drug design. *Methods in molecular biology (Clifton, N.J.)* **672**, 299–323 (2011).
58. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
59. Johnston, C. W., Badran, A. H. & Collins, J. J. Continuous bioactivity-dependent evolution of an antibiotic biosynthetic pathway. *Nat. Commun.* **11**, (2020).
60. Chen, M. J., Dixon, J. E. & Manning, G. Genomics and evolution of protein phosphatases. *Sci. Signal.* **10**, 1–17 (2017).
61. 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.* **10**, 216–222 (2014).
62. Chen, X. *et al.* Statistical experimental design guided optimization of a one-pot biphasic multienzyme total synthesis of amorpha-4,11-diene. *PLoS One* **8**, e79650 (2013).
63. Edgar, S. *et al.* Mechanistic Insights into Taxadiene Epoxidation by Taxadiene-5 $\alpha$ -Hydroxylase. *ACS Chem. Biol.* **11**, 460–469 (2016).
64. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
65. 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.* **8**, 28–36 (2017).
66. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: a Practical Information-theoretic Approach, 2nd edn.* Springer-Verlag, New York. New York Springer **60**, (2002).
67. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
68. Afonine, P. V *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 352–67 (2012).
69. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
70. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB\_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).
71. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

72. Huang, J. *et al.* CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2016).
73. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
74. Vanommeslaeghe, K. *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.* **31**, 671–690 (2010).
75. 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.* **33**, 2451–2468 (2012).
76. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
77. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
78. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
79. Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182 (1981).