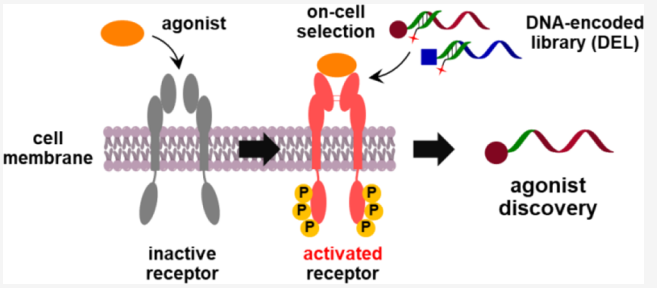
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Agonist Discovery for Membrane Proteins on Live Cells by Using DNA-encoded Libraries

Yiran Huang, Rui Hou, Fong Sang Lam, Yunxuan Jia, Yu Zhou, Xun He, Gang Li, Feng Xiong,* Yan Cao,* Dongyao Wang,* and Xiaoyu Li*

Cite This: J. Am. Chem. Soc. 2024, 146, 24638−24653

ABSTRACT: Identifying biologically active ligands for membrane proteins is an important task in chemical biology. We report an approach to directly identify small molecule agonists against membrane proteins by selecting DNA-encoded libraries (DELs) on live cells. This method connects extracellular ligand binding with intracellular biochemical transformation, thereby biasing the selection toward agonist identification. We have demonstrated the methodology with three membrane proteins: epidermal growth factor receptor (EGFR), thrombopoietin receptor (TPOR), and insulin receptor (INSR). A ∼30 million and a 1.033 billion-compound DEL were selected against these targets, and novel agonists with subnanomolar affinity and low micromolar cellular activities have been discovered. The INSR agonists activated the receptor by possibly binding to an allosteric site, exhibited clear synergistic effects with insulin, and activated the downstream signaling pathways. Notably, the agonists did not activate the insulin-like growth factor 1 receptor (IGF-1R), a highly homologous receptor whose activation may lead to tumor progression. Collectively, this work has developed an approach toward "functional" DEL selections on the cell surface and may provide a widely applicable method for agonist discovery for membrane proteins.



INTRODUCTION

Many human diseases are associated with aberrant membrane protein functions. Notable examples include G-protein coupled receptor proteins (GPCRs) and ion channels, which account for the targets of ∼33% and ∼18% of all approved small molecule drugs, respectively.[1] Given the importance, there has been great interest in developing small molecule ligands for membrane proteins, such as structure-based drug design,[2] virtual screening,[3] biophysical screening methods,[4,5] and sensor-based screening.[5] Phenotypic screening assays are also widely used.[6] However, ligand discovery for membrane proteins is highly challenging, as they are located in the hydrophobic lipid bilayer of the cell membrane, often making *in vitro* biochemical techniques such as expression, solubilization, purification, and crystallization very difficult.[7] In addition, purified membrane proteins may lose important biological features, such as post-translational modifications, cofactor binding, and protein complex formation. Thus, it is highly desirable to perform ligand discovery with membrane proteins directly on live cells.

The concept of DNA-encoded libraries (DELs or DECLs) was originally proposed by Brenner and Lerner in 1992 as a way to reduce library storage size and screening requirements that is needed with the on-resin one-bead, one-compound (OBOC) libraries.[8] This synthetic encoding concept was first demonstrated by Janda in 1993;[9] notably, a cleavable linker was introduced to allow library release after DEL synthesis, thereby realizing library characterization and selection in solution. In 2004, Neri,[10] Liu,[11] Harbury,[12] Winssinger,[13] and their respective co-workers independently reported four types of encoded libraries. In 2009, GSK published a seminal work on applying DELs in drug discovery.[14] Today, DEL has become a powerful screening platform (Figure 1a).[15−31] In a DEL, each library compound is conjugated with a unique DNA tag, which encodes its chemical structure and acts as an amplifiable tag for hit identification through next-generation sequencing (NGS). DELs allow the synthesis, processing, and selection of millions to billions of compounds simultaneously at a minute scale. DELs can be applied to a broad range of biological targets, including proteins, cell lysates, nucleic acids, live cells, whole bacteria, human sera, etc.[16,20,25] In the past decade, this technology has been widely adopted in drug discovery and also showed potential in academic research.[21]

Previously, DELs have been used to interrogate full-length membrane proteins stabilized with detergent,[32] nanodiscs,[33] or

Received: June 26, 2024
Revised: August 16, 2024
Accepted: August 16, 2024
Published: August 22, 2024

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

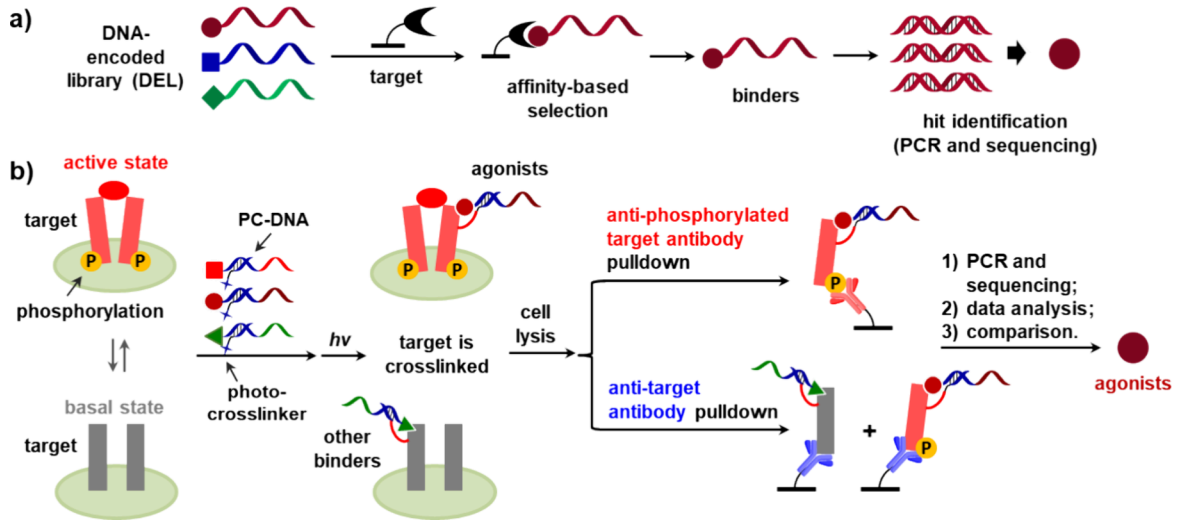


Figure 1. Schematic illustration of the selection methods. (a) Typical DEL selection with an immobilized protein target. (b) (This work) cell-based DEL selection for agonist identification. Parallel pulldowns with different antibodies are used to separate the agonists or all binders; a comparison of the pulldown results will identify the potential agonists; red oval: natural protein agonist of the target.

mutations.[34] The Bradley group pioneered PNA-encoded library screening for chemokine receptors and integrin on live cells.[35,36] GSK reported the first cell-based DEL selection, in which libraries were selected against the cells overexpressing receptor neurokinin-3.[37] The Krusemark group applied a covalent cross-linking method[38] to both cell-surface and intracellular targets.[39−41] The Neri group systematically optimized the experimental conditions for cell-surface selections.[42] Madsen et al. reported an intracellular DEL selection by using microinjection.[43] We reported a method that can selectively label the target with a DNA tag for selections against endogenous membrane proteins.[44] However, DEL selections only identify binders, and postselection assays are required to determine the biological activities of the hit compounds. Moreover, agonist discovery is even more difficult, as receptor agonism requires a specific switch of the protein conformation and/or multimer assembly. The "one-bead, one-compound" DEL (OBOC-DEL) method couples DEL selection with a functional readout and thus is amendable to identifying active compounds.[28] Recently, Krusemark and co-workers reported a selection method for GPCR agonist discovery, in which agonist binding triggers the intracellular biotinylation of the target via a split protein complementation approach.[41] However, the method was only tested with a small library, and the target protein needed to be genetically modified. It is still desirable to develop new methods that can select large DELs against native membrane proteins to identify ligands with functional relevance, especially agonists.

Herein, we report a cell-based selection method that connects ligand binding with the intracellular phosphorylation of the target (Figure 1b). The selected compounds favor binding to the activated receptor and thus are more likely to be agonists. We performed methodology development with three membrane proteins: insulin receptor (INSR), thrombopoietin receptor (TPOR), and epidermal growth factor receptor (EGFR). Further, a ∼30 million and a 1.033 billion-compound DEL were selected against these targets, and novel agonists with subnanomolar affinity and low micromolar cellular activities have been discovered. We show that the INSR agonists possibly bind to an allosteric site, are synergistic with insulin in inducing INSR phosphorylation, and activate the downstream signaling pathways. Notably, the agonists do not activate insulin-like growth factor 1 receptor (IGF-1R), a highly homologous receptor whose activation may lead to tumor progression. This work may provide a widely applicable method for agonist discovery for membrane proteins.

RESULTS

Design of the Selection Method. Using agonists to stabilize the active conformation of membrane proteins is an important technique for studying GPCRs and cytokine receptors in crystallography.[45,46] Ahn and co-workers and researchers at Nuevolution performed a DEL selection against purified full-length β2-adrenergic receptor (β2AR) in the presence of an orthosteric agonist, and they identified the first small molecule allosteric agonist of β2AR.[33] We reasoned that such a kind of "functional" selection may also be achieved on live cells. For membrane proteins, natural protein ligands can induce the conformational change and/or multimer assembly required for receptor activation,[47,48] and, for membrane kinases, lead to phosphorylation of the intracellular domain.[49] We reasoned that the phosphorylation sites could be exploited as a natural affinity handle to isolate the compounds binding to the activated receptor (Figure 1b). A photo-cross-linking DNA (PC-DNA) can hybridize at the primer-binding site (PBS) of the library.[38−41,44,50] Upon UV irradiation, PC-DNA covalently captures the target and stabilizes the binding. After cell lysis, the "protein−ligand−DNA" complex can be isolated with an antibody specific for the phosphorylated target,[38] and a parallel pulldown with a target-specific antibody will isolate all binders (Figure 1b). The pulldown results will be compared for hit identification: (i) agonists are more likely to be enriched by the antiphosphorylation antibody; (ii) although antibodies may cross-react with other proteins on the cell surface, target-specific binders should be enriched by both antibodies; and (iii) it will exclude the compounds that induce target phosphorylation through other indirect cellular pathways.

Initial Validation and Model Selection with INSR. INSR is a receptor tyrosine kinase (RTK) that plays crucial roles in many metabolic processes. Its natural agonist insulin activates INSR by stabilizing the receptor in a "stretched-out" manner, resulting in the phosphorylation of the cytoplasmic

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

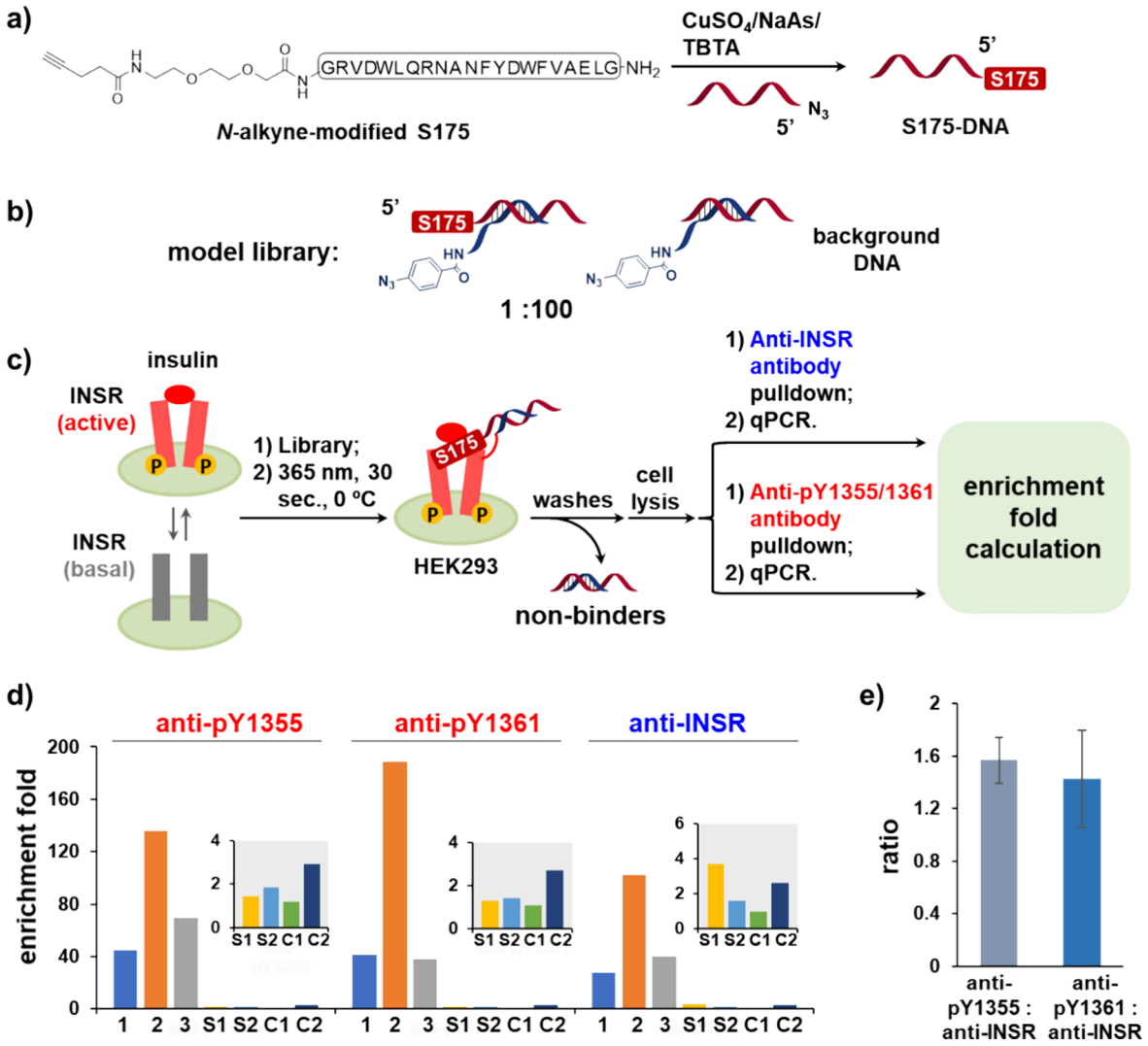


Figure 2. Model selections with INSR. (a) Synthesis of **S175-DNA**. NaAs: sodium ascorbate; TBTA: tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine. (b) A model library was prepared by mixing **S175-DNA** with background DNA and a photoreactive DNA. (c) Model selection scheme. (d) Column graphs showing enrichment folds from the selections with different antibodies. 1, 2, 3: triplicate selections; S1, S2: selections under starvation; C1, C2: selections with free **S175** (50 μM). The enlarged S1, S2, C1, and C2 data are shown in inlets. (e) Anti-pY1355/1361: anti-INSR EF ratios. See the Supporting Information for details.

domain and the activation of downstream signaling pathways.[51] Recently, Lerner and co-workers selected a natural product DEL (nDEL) against the ectodomain of INSR and identified Rutaecarpine as an activator.[52] Among the tyrosine phosphorylation sites, Y1355 and Y1361 at the *C*-terminal domain were chosen, since they are involved in the regulation of the kinase domain[53] and the *C*-terminal domain has a low degree of homology among RTKs;[54] therefore, the anti-phosphorylated tyrosine (anti-pY) antibodies are more specific to the target. The anti-pY antibodies for the tyrosines at the juxtamembrane and kinase domains often cross-react with IGF-1R, a homologous receptor, and may lead to nonspecific capture.[55] First, we measured the phosphorylation level of Y1355/1361 in transfected HEK293 cells overexpressing the INSR (Figure S1). In theory, a higher phosphorylation level should be better for compound enrichment; however, if the vast majority of the receptor is phosphorylated, there will be little difference between the parallel pulldowns. We reasoned that a "partially activated" receptor population may be more desirable. The pY1355/1361 levels in HEK293 cells were measured to be ~60% (pY1355) and ~67% (pY1361) for nonstarved cells, while the starved cells had only ~5−8% phosphorylation (Figure S2). The "whole-cell" INSR concentration was ~1.30 μM (Figure S2c), comparable to the target concentration of previous cell-based selections.[44] Since the receptors are concentrated on the cell surface, the actual molarity is much higher. The nonstarved cells provide a balanced target concentration and phosphorylation level for the selection.

S175 is a known INSR agonist (Figure 2a and Figure S3).[56] **S175** was conjugated to DNA, and the conjugate showed an EC_{50} of ~10−15 μM on increasing the pY1355/1361 levels (Figure 2a and Figure S4). **S175-DNA** was hybridized with a photoreactive DNA (capture probe, CP), and the DNA duplex was used to capture INSR on the cell surface (Figure S5a).[44] Since the spacer length between the cross-linker and the DNA is important,[44] a series of spacers ranging from 3 to 21 nt were screened. The CP with a 12 nt spacer gave higher efficiency and was used in the following studies (Figure S5b). Next, a model library was prepared by mixing **S175-DNA** with excess

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

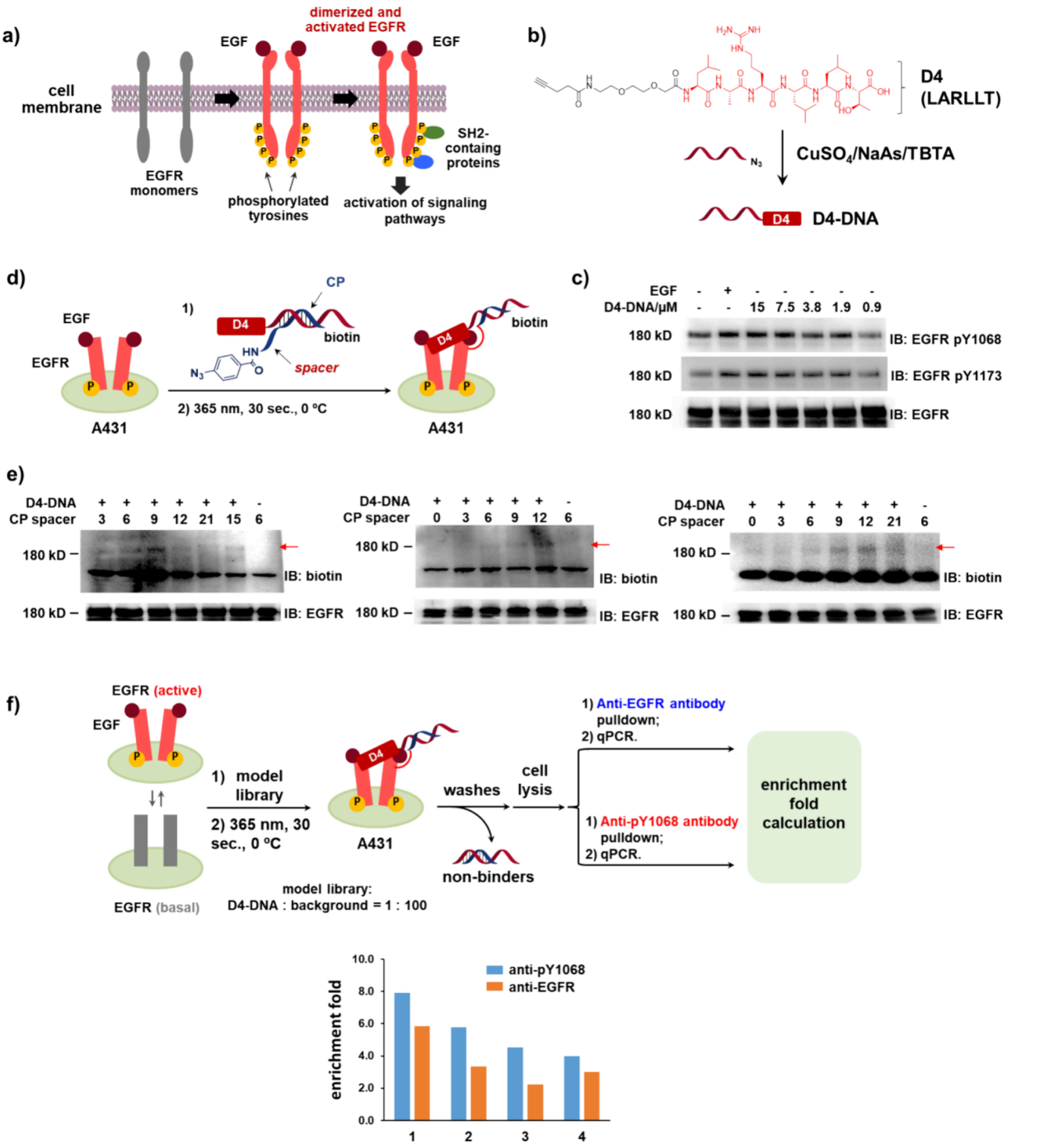


Figure 3. Model selections with EGFR. (a) Schematic illustration of EGFR activation by EGF. (b) Synthesis of the **D4-DNA** conjugate (DNA: 16 nt). (c) A431 cells treated with the **D4-DNA** or EGF (30 min at 37 °C) and the EGFR pY1068/1173 levels were analyzed with Western blot.[57−60] (d) DNA-templated labeling of EGFR on A431 cells.[44] (e) CPs with different spacer lengths were tested; the number indicates the number of nucleobases in the spacer; triplicate experiments were performed; red arrows indicate the labeling products. **D4-DNA**/CP, 5 μM; UV: 365 nm, 30 s., 0 °C. (f) Selection scheme; a library was prepared by mixing the **D4-DNA** with nonbinding background DNA (1:100), hybridized with a photoreactive CP. qPCR was used to quantify and calculate the enrichment folds. Column 1, 2, 3, 4: four replicates. See details and full data in the Supporting Information.

background DNA (1:4; Figure S6a). The library was incubated with INSR-transfected HEK293 cells (4 °C, 2.5 h) before UV irradiation (365 nm, 30 s, 0 °C). After washes, the cells were lysed under the condition of maintaining DNA hybridization.[38] The cell lysates were subjected to affinity pulldown using the anti-pY1355, pY1361, and INSR antibodies, respectively. By using quantitative PCR (qPCR), the enrichment fold (EF) of **S175-DNA** was calculated to be 6.14, 7.12, and 4.40, respectively (Figure S6). We changed the **S175**/background ratio to 1:100 and performed the pulldown experiments in triplicates. Significant enrichment of **S175** was observed in all replicates (Figure 2d and Figure S7). The EF values varied significantly, which may be due to multiple factors. First, different batches of the cells were used in the pulldowns; the cell batches may express different levels of INSR and the phosphorylation levels may vary, which would lead to

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

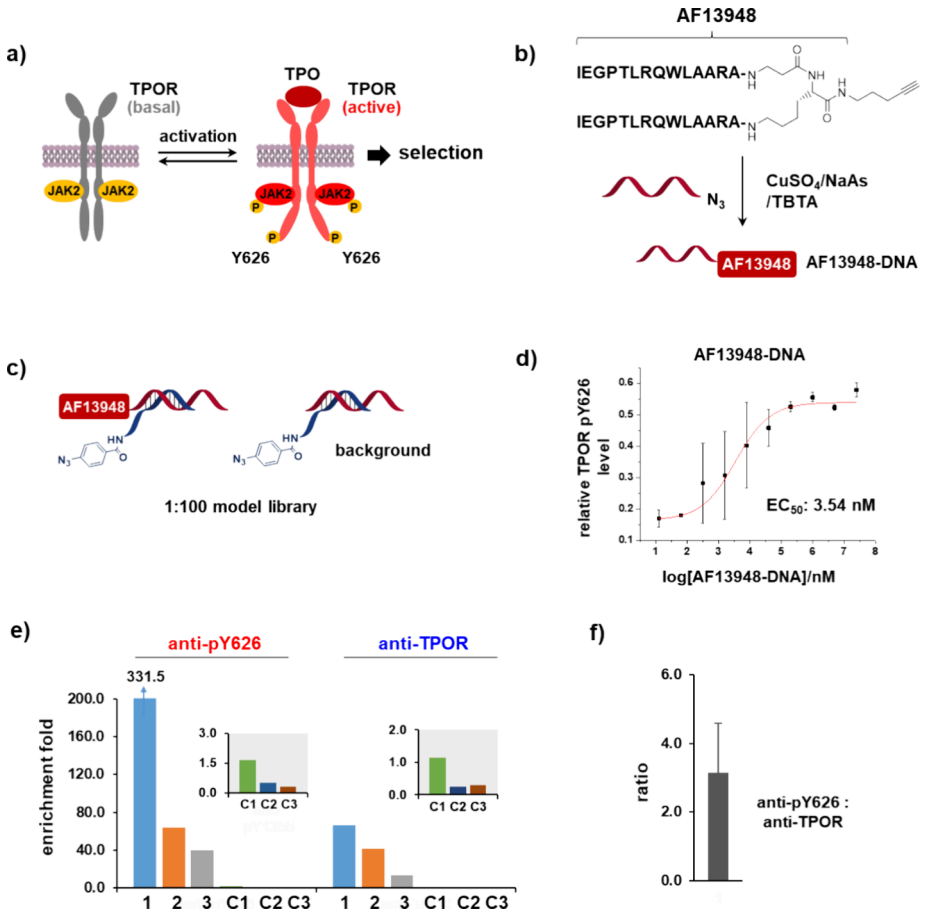


Figure 4. Model selections with TPOR. (a) TPO-TPOR binding leads to phosphorylation of JAK2, which further phosphorylates TPOR; Y626 was used as the affinity-handle. (b) Synthesis of **AF13948-DNA**. (c) The model library contained **AF13948-DNA**, 100-fold excess of the background DNA, and the CP. (d) **AF13948-DNA**'s effect on TPOR pY626 level. (e) Column graph showing the **AF13948** EF values from the selections with different antibodies. 1, 2, 3: triplicate selections; C1, C2, C3: triplicate selections with free **AF13948** (50 μM). The enlarged C1, C2, and C3 data are shown in the inlets. (f) Anti-pY626: anti-TPOR EF ratio. See Figure S14 for details.

variations in compound enrichment. Moreover, the cells were transfected each time before selection, and there may also be variations in transfection efficiency between the replicates. Second, batch variation in antibodies' efficiency/specificity in capturing INSR, the phosphorylated receptor, and other nonspecific proteins will also lead to variations in compound enrichment. Third, variations in the washing protocol, such as buffer conditions, washing times, etc., might be another factor. Using cells stably expressing the receptor and the same batch of antibodies and more stringent control of the washing protocol should minimize the variation among the replicates. The EF ratios of anti-pY/anti-INSR pulldowns were ~1.5−1.6 (Figure 2e), largely consistent with the theoretical model (Figure S8 and the Supporting Information worksheet). The model assumes that the anti-INSR pulldown enriches all binders, while the anti-pY pulldown preferably enriches agonistic binders. Calculation based on the model suggests that a lower phosphorylation level will give a higher anti-pY/anti-INSR ratio and be beneficial for agonist identification. Surprisingly, low enrichment was observed with the starved HEK293 cells (Figure 2d and Figure S7). We hypothesized that **S175** had a lower affinity to the unphosphorylated INSR[56] and the level of phosphorylated INSR was too low under starvation. No enrichment was observed with free **S175** competition. Collectively, the model selection supported that a "partially activated" receptor population is more beneficial and the agonists can be enriched by both antibodies but with a higher EF in the anti-pY pulldown.

Model Selections with EGFR and TPOR. EGFR is a well-studied RTK; upon binding with the native ligand EGF, EGFR dimerizes and autophosphorylates the tyrosine residues in the C-terminal domain (Figure 3a).[61] **D4** is a peptide that binds to the extracellular domain of EGFR (Figure 3b); it was mostly used in targeted delivery but its affinity and functions have not been reported.[62] **D4**-treated A431 cells, a cell line with high EGFR expression,[44] showed an increase of pY1068/1173 levels comparable to EGF (EC_{50}: ~15−20 μM; Figures S9 and S10). **D4** was conjugated to DNA (**D4-DNA**),[62] and the conjugate increased the pY1068 level but only had a slight effect on pY1173; thus, Y1068 was used as the affinity handle (Figure 3b,c). **D4-DNA** was hybridized with CPs with different spacer lengths and used to label EGFR. The CP with a 9 or 12 nt spacer gave a relatively higher efficiency (Figure 3d,e). Next, a model library containing **D4-DNA** and background DNA was selected against EGFR (Figure 3f and Figure S11), and all four selection replicates gave a higher EF with the anti-pY1068 antibody than that with the anti-EGFR antibody. The results showed that agonists with moderate activity, such as **D4**, may also be enriched by the selection.

The thrombopoietin receptor (TPOR) is a cytokine receptor. The binding of its native ligand thrombopoietin (TPO) leads to autophosphorylation of the receptor-associated

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

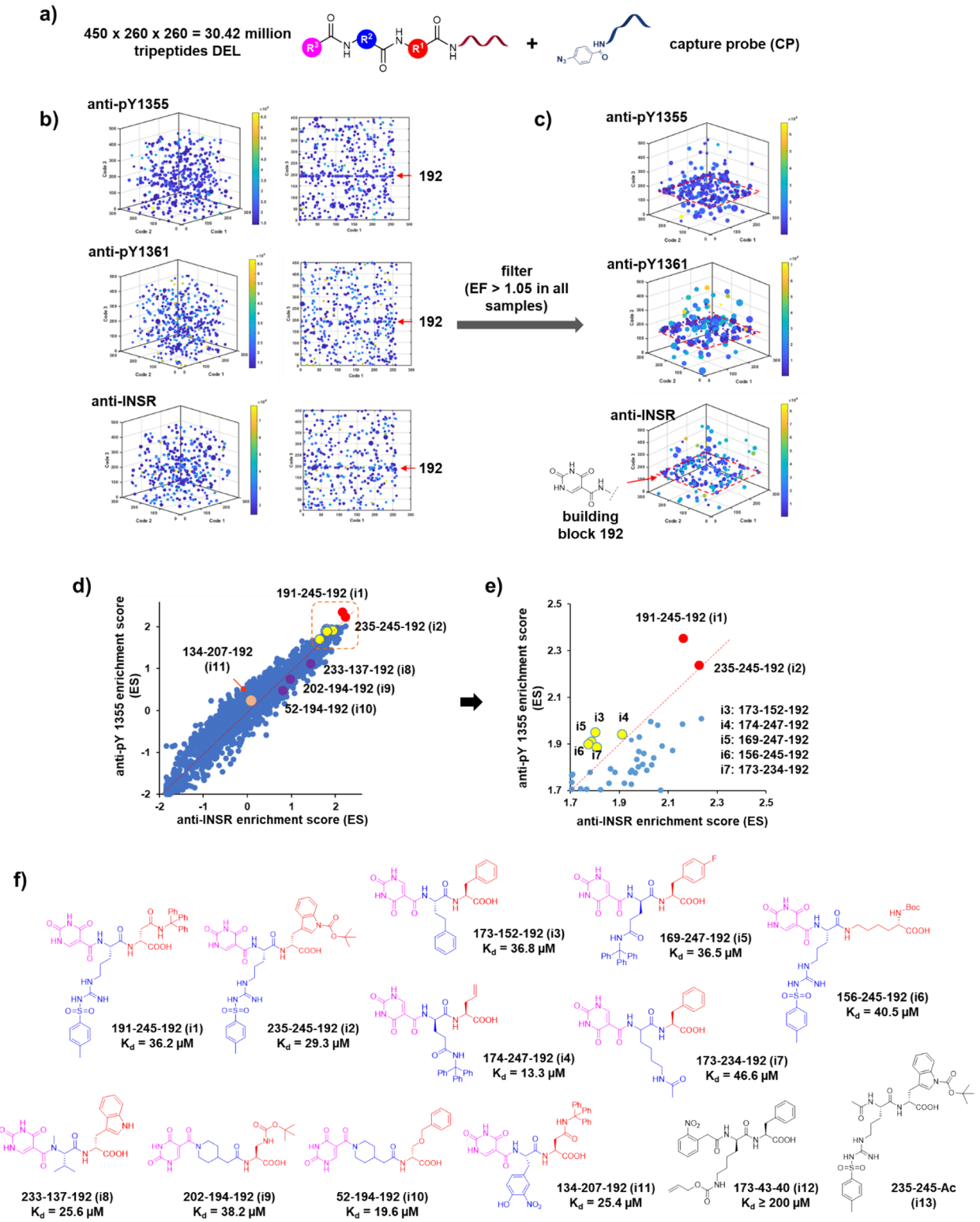


Figure 5. Selection of a 30.42 million DEL against INSR. (a) Library structure and BB diversity. (b) 3D plots of the selection results. The axes represent R^1, R^2, and R^3, respectively. The size of the data point represents the enrichment fold (EF); the color represents the postselection sequence count. For each 3D plot, a 2D projection along the code 2 (R^2) axis is shown to highlight the enriched BB **192**. (c) The 3D plots were filtered (EF > 1.05, present in all 9 samples), and the remaining compounds are shown. See Figures S16–S20 for the full data. (d) Plot of the ES values (anti-pY1355 vs anti-INSR); the BBs of the selected compounds are specified. (e) Zoom-in view of the upper right area of the plot in (d). (f) Structures of the hit compounds and their binding affinities measured by using SPR; two control compounds are shown in black.

JAK2, which further phosphorylates TPOR (Figure 4a).[63] The C-terminal tyrosine Y626 was chosen as the affinity handle. We screened several cell lines and found TPOR-transfected HEK293 cells expressed a high level of TPOR with a pY626 level of ∼55% (Figure S12). **AF13948** is a potent TPOR agonist (EC_{50} = 0.4 nM; Figure 4b);[64] the **AF13948-DNA**

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

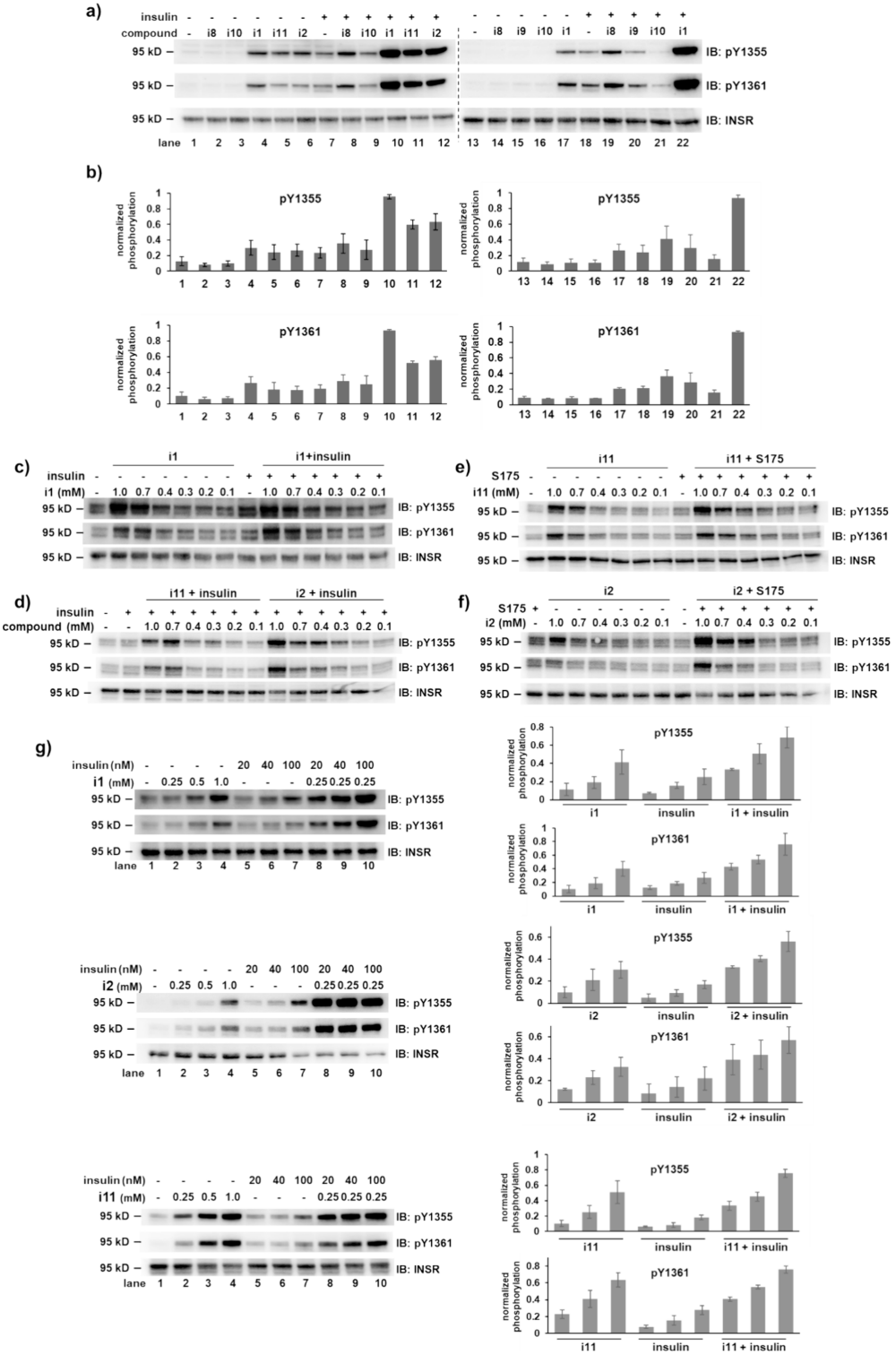


Figure 6. Effects of the hit compounds on INSR phosphorylation. HEK293 cells overexpressing INSR were starved and treated with the compounds (30 min, 37 °C) and the phosphorylation levels were analyzed with Western blot. (a) Cells treated with the individual compounds (1 mM) with and without insulin (20 nM); see Figure S23 for more replicates. (b) Column graphs summarizing the results; column 1−22 corresponds to lane 1−22 in (a). (c−f) Dose-dependency of the compounds' effects on pY1355/1361 levels, with and without insulin (20 nM) or **S175** (100 μM). (g) Study of the synergistic effect of the compounds with insulin. Left: a representative gel is shown for **i1**, **i2**, and **i11**, respectively; see Figure S25 for more replicates. Right: column graphs summarizing the results; the columns correspond to lane 2−10 in the gel images in (g); lane 1 was used as the normalization baseline. n = 3 biologically independent samples except for the right panel in (a) (n = 2); data are presented as mean values ± s.d.

conjugate exhibited a ∼3.54 nM EC_{50} activity on stimulating pY626 level (Figure 4d and Figure S12). Similar to INSR, the CP with a 12-nt spacer gave a higher labeling efficiency (Figure S13). Next, a model library containing **AF13948-DNA** and

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

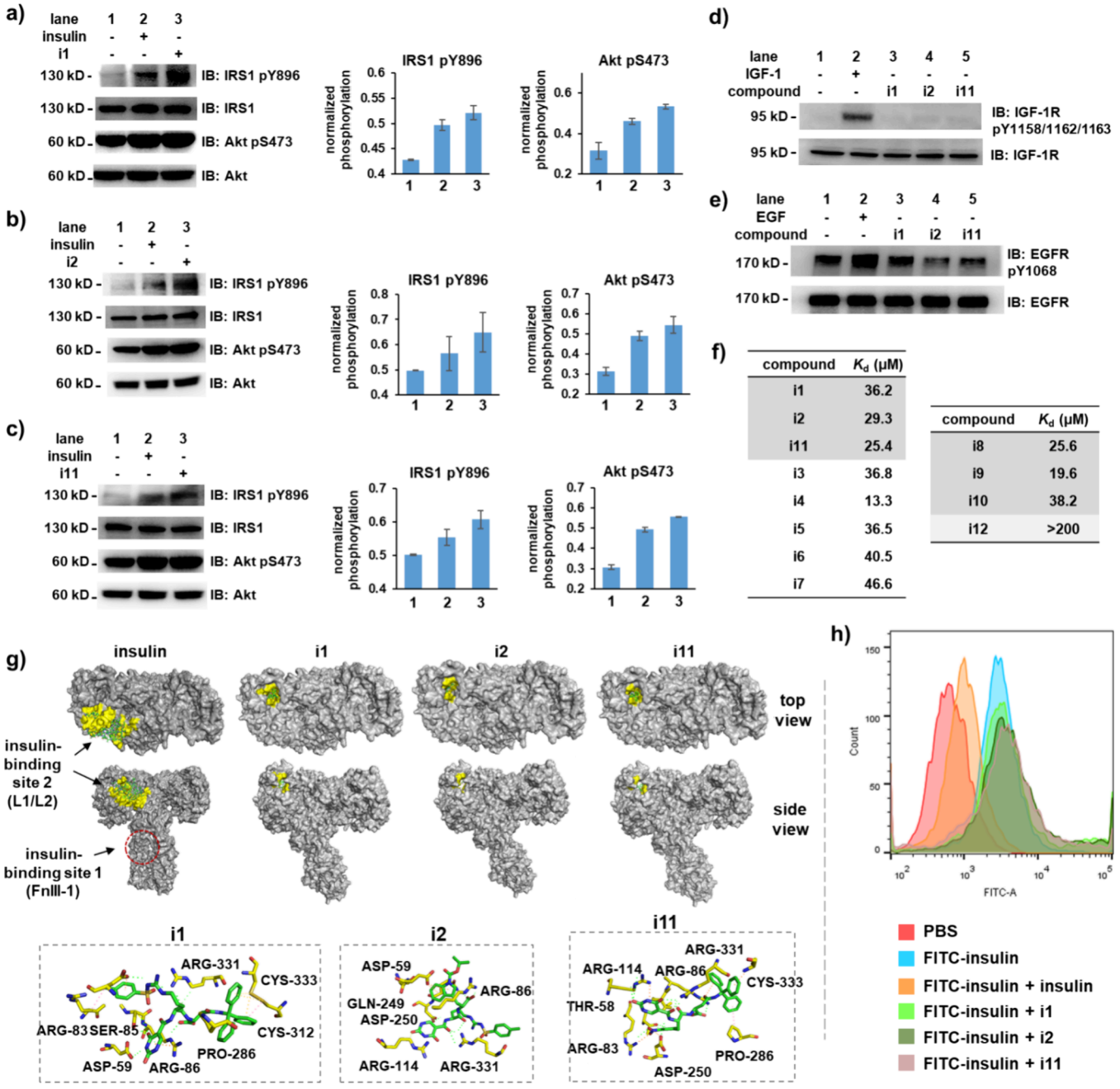


Figure 7. Effects of the selected compounds on the insulin-mediated signaling pathways and binding characterization. (a, c) HEK293 cells overexpressing INSR were starved and treated with insulin (20 nM) or the compounds (1 mM); the IRS1 pY896 and Akt pS473 levels were analyzed with Western blot. Left: a representative gel is shown for (a) **i1**, (b) **i2**, and (c) **i11**; see Figure S29 for replicates. Right: column graphs summarizing the results; columns 1−3 correspond to lane 1−3 in the gel images; $n = 2$ biologically independent samples; data are presented as mean values ± s.d. (d) HEK293 cells overexpressing IGF-1R were starved and treated with IGF-1 (20 nM) or the compound (1 mM); IGF-1R pY1158/1162/1163 levels were determined. (e) A431 cells were treated with EGF (20 nM) or the compound (1 mM); EGFR pY1068 level was determined. (f) SPR analysis of the compound's affinity; the sensorgrams and fitting curves are provided in Figure S30. (g) Molecular docking of **i1**, **i2**, and **i11** to the cryo-EM structure of the ectodomain of INSR (PDB: 6SOF);[73] the insulin-binding sites and the compound's possible binding sites are highlighted; detailed interaction maps are provided; small molecules are shown as sticks in green and the amino acid residues are shown yellow. Hydrogen bonds are shown as dashed lines. See Figure S31 for more details. (h) HEK293 cells overexpressing INSR were labeled with FITC-insulin in the presence of insulin or the compound, and the cell fluorescence was measured with flow cytometry.

excess background was selected against TPOR on HEK293 cells. Possibly due to its high affinity, significant enrichment of **AF13948** was observed (Figure 4e and Figure S14). The EFs showed large variations, but the pulldown experiments with the anti-TPOR pY626 antibody consistently gave higher EFs than the anti-TPOR antibody (ratio: ∼2.9; Figure 4f). Little enrichment was observed in the presence of the free **AF13948** competitor (Figure 4e and Figure S14). The results indicated that the selection could be applied to non-RTK membrane proteins.

Selection of a 30.42 Million DEL against INSR. Next, a 30.42 million tripeptide DEL (Figure 5a)[44] was selected

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

against INSR in triplicates following the same procedure. 500 pmol of the library (~10^7 copies/compound) was used to provide sufficient materials for the selection.[65,66] The DNA tags of the enriched compounds were amplified with NGS-compatible primers and submitted for sequencing (Figure S15). The sequencing data were processed with a Python script to quantitatively tally the codons of each compound and calculate the EF values (postselection% /preselection%),[44] which can generate 3D cubic plots for direct hit-picking or identifying structure−activity relationship (SAR) patterns. As expected for cell-based selections, the 3D plots exhibited a high noise level (Figure 5b and Figures S16−S18). The only recognizable pattern is a plane formed by the BB **192** in the third cycle (R^3), and there was only a small percentage of overlapping compounds among the top 5% of compounds (Figure S19). We reasoned that the true hit compounds should be enriched by all three antibodies in all three replicates; thus, the data sets were filtered by the following criteria: the compounds should be present in all nine samples with an EF > 1.05. In the end, ~2,000 compounds remained and **192** became much more recognizable (Figure 5c and Figure S20). Recently, machine learning (ML) has been shown to be effective in denoising DEL data sets.[67−69] Previously, we established a regression model using maximum aposteriori loss function, a probabilistic framework that can account for and quantify uncertainties of noisy data,[69] and showed that the model facilitated hit identification from cell-based selections.[70] Here, the model was trained with the selection sequencing counts in the form of disynthon aggregation,[71] and the trained model was then used to predict the selection outputs in the metric of a normalized enrichment score (ES). The ES metric diminished sample discrepancies and enabled comparison among the pY1355, pY1361, and INSR selections; it was calculated by converting the predicted EF values to the normalized *Z*-scores by standardization with the unit variance after removing the mean values (see details in the Supporting Information).[72] The ES values of the anti-pY1355 and anti-INSR pulldowns were plotted (Figure 5d) since the anti-pY1355 antibody gave more robust results in model studies. Two compounds, **i1** and **i2**, gave relatively high ES values, and they also had higher enrichment with the anti-pY1355 antibody in all three replicates (Figure S21). Besides **192**, **i1** and **i2** also share the same BB **245** at the R^2 position (Figure 5f). Five compounds with relatively high ES values and pY1355/INSR ES ratio >1.0 were chosen as the "2^{nd}-tier" hits (**i3**−**i7**; Figure 5e,f). Three compounds (**i8**−**i10**) with much lower ES values and pY1355/INSR ES ratio <1.0 were chosen for comparison. Since the ES metric is based on disynthon aggregation, we also picked a top compound (**i11**) from the original data set based on the average EF values and pY1355/INSR ratios of all three replicates (Figure S21). Finally, we chose a compound not present in the data set of Figure 5d (**i12**) as the "negative control".

Validation and Characterization of the Hit Compounds. The selected compounds were synthesized off-DNA and their effects on the pY1355/1361 levels were studied. First, we observed that at ~100 ng/mL insulin concentration, the pY1355/1361 levels of HEK293 cells showed high sensitivity to insulin (Figure S22). Next, we focused on **i1** and **i2**, the two compounds with high ES values, and **i11**, the top-ranked compound from the original data set. The cells were treated with the compounds at high concentrations (1 mM) to better observe the compounds' activity. **i1, i2**, and **i11** markedly increased the pY1355/1361 levels and exhibited a strong synergistic effect with insulin (Figure 6a,b, and Figure S23). In contrast, the compounds with lower ES values and pY1355/INSR ratios <1.0 (**i8**, **i9**, and **i10**) did not show activity. The randomly chosen compound **i12** also did not show any activity. We also synthesized an analogue of **i2**, in which **192** was replaced with an acetyl group (**235−245-Ac**, **i13**; Figure 5f), and it was inactive (Figure S24). Next, the cells were titrated with different compound concentrations, and the dose-dependency of **i1, i2,** and **i11** was observed (Figure 6c−6f). To further validate the synergistic effects, cells were treated with either the compound, insulin, or both at different concentrations in triplicate (Figure 6g and Figure S25). For simplicity, synergistic ratios (SRs) were calculated by dividing the average phosphorylation level of the "combo" treatment by the sum of the individuals. The SR values of the compounds are >1.0 with a slightly greater synergistic effect on pY1355 than pY1361, except **i1**, which has a higher effect with pY1361 (Figure S26). We also tested the compounds' effects on the phosphorylation of Y1158, Y1162, and Y1163, as these phosphorylation sites are the major determinants of kinase activation and important markers of INSR activation. Similar synergistic effects were also observed (Figure S27).

Next, the activities of the "2^{nd}-tier" hits (**i3**-**i7**) were tested. **i3** and **i5** appeared to be agonists; **i5** showed synergism with insulin, but **i3** only had a slight effect. **i4** slightly increased the pY1361 level, had no effect on pY1355, but showed a weak synergistic effect with insulin on pY1355 (Figure S28). **i6** and **i7** did not show agonistic effects; these two compounds might bind to the phosphorylated INSR at a site not correlated with activation. Insulin-induced tyrosine phosphorylation forms the binding scaffold for IRS proteins, which activates the PI3K/Akt pathway and leads to diverse cellular functions.[51] As shown in Figure 7a−7c and Figure S29, cells treated with **i1, i2,** and **i11** showed elevated phosphorylation at IRS-1 (Y896) and Akt (S473). IGF-1R is an RTK that plays important roles in mitogenic control; it is highly homologous to INSR and engages similar signaling pathways. Insulin also binds to IGF-1R with lower affinity.[74] Thus, a major concern in insulin replacement therapy is the collateral activation of IGF-1R, which may lead to tumor progression.[75] We anticipate nonorthosteric agonists may have better receptor selectivity. To verify this, HEK293 cells were treated with IGF-1, resulting in a significant increase in IGF-1R phosphorylation. In contrast, **i1, i2,** and **i11** had no effect on the IGF-1R, despite the high concentration (1 mM) used (Figure 7d and Figure S29). Moreover, the extracellular domains of INSR and EGFR share a common module with two ligand-binding sites connected by a cysteine-rich domain; therefore, the compounds' effect on EGFR was also tested, and no increase in the EGFR phosphorylation level was observed (Figure 7e). Collectively, these results indicated that the hit compounds are selective INSR agonists. Although with moderate activity, the compounds are synergistic with insulin in enhancing INSR phosphorylation and activating the downstream signaling pathways without the collateral activation of IGF-1R and EGFR.

Surface plasmon resonance (SPR) analysis showed that the compounds had moderate affinities (K_d: ~13−40 μM) to INSR, including the agonists and the inactive ones (Figure 7f, and Figure S30). The "non-hit" **i12** did not show detectable affinity. Since the activity test results indicated that the compounds were not competitive with insulin, molecular

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

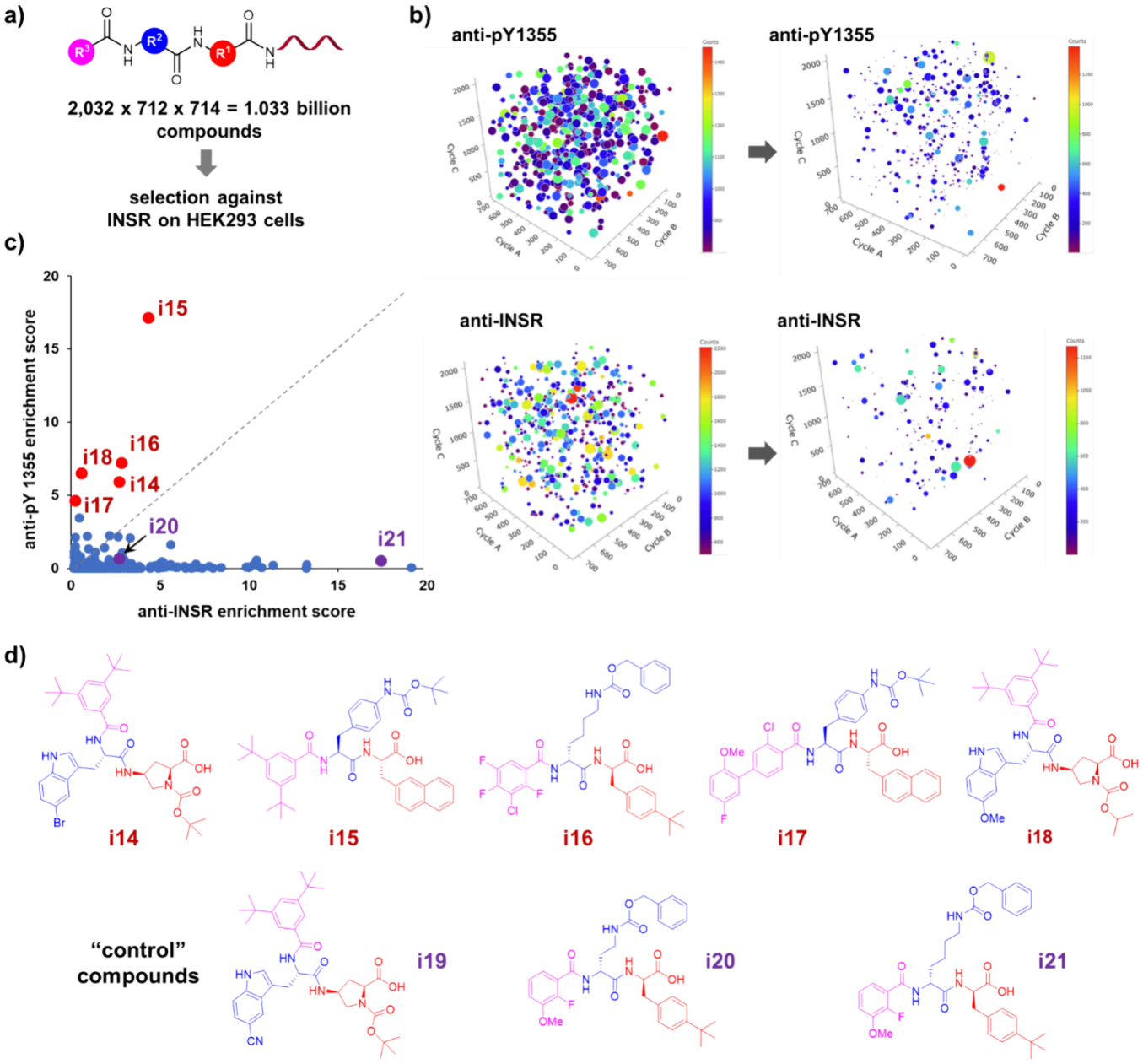


Figure 8. Selection of a 1.033 billion DEL against INSR. (a) Library structure and BB diversity. (b) 3D plots of the selection results. The axes represent R^1, R^2, and R^3, respectively; the size of the data point represents the EF value; the color represents the postselection sequence count; the 3D plots were filtered and the compounds enriched in all replicates are shown on the right. (c) Plot of the normalized ES values of the remaining compounds in (b) (anti-pY1355 vs anti-INSR); the selected hits and the control compounds are highlighted. (d) Structures of the selected compounds.

docking was used to probe the possible binding site. **i1, i2,** and **i11** were docked to the Cryo-EM structure of the ectodomain of INSR (PDB: 6SOF).[73,76] The results showed that the compounds possibly bind to a site in the CR domain, which is distal to the insulin-binding site 1 and proximal to the insulin-binding site 2 (Figure 7g).[73] The BB **192** may form hydrogen bonds within multiple hydrophilic residues. Finally, to test whether the compounds bind to different sites from insulin, FITC-labeled insulin was incubated with HEK293 cells in the presence of insulin or the test compound; the cells were washed and analyzed with flow cytometry. As shown in Figure 7h, unlike insulin, **i1, i2,** and **i11** did not reduce the fluorescence arising from the FITC-insulin labeling, suggesting that the compounds are not competitive with insulin. Collectively, these results have supported that the compounds may possibly bind to an allosteric site of INSR.

Selection of a 1.033 Billion DEL against INSR. Although the selections identified several INSR agonists, they have weak affinity and activities. To further validate the selection method, a larger library of ~1.033 billion compounds was prepared (Figure 8a). The library was selected against INSR on HEK293 cells in triplicates. To identify more potent compounds, more stringent criteria were applied in data analysis: the compounds needed to be enriched in the three replicates with a greater enrichment in the anti-pY1355 selection. As shown in Figure 8b, significantly fewer compounds remained after only keeping the compounds enriched in all replicates, which are further plotted in a 2D

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

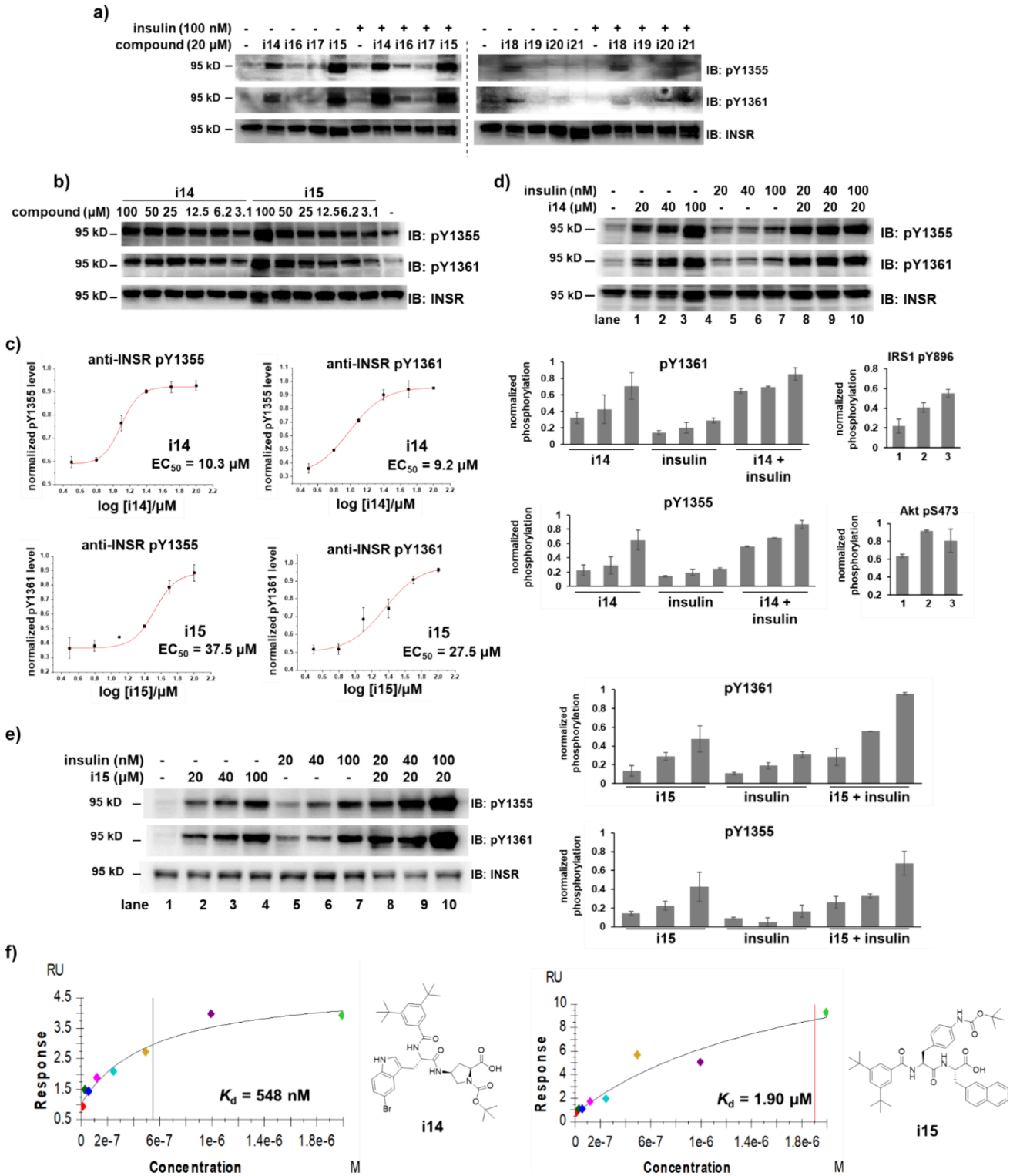


Figure 9. Characterization of the hit compounds. (a) HEK293 cells overexpressing INSR were starved and treated with the compounds (30 min, 37 °C; 20 μM) with and without insulin (20 nM); the phosphorylation levels were analyzed. (b) Dose dependency of the compounds' effect; see Figure S32 for more replicate. (c) Titration curves to determine the EC_{50} values of **i14** and **i15**. $n = 2$ biologically independent samples; data are presented as mean values ± s.d. (d) Synergistic effect of **i14** with insulin and its effect on IRS1 pY896 and Akt pS473 levels; column 1: without insulin and **i14**; column 2: with **i14**; column 3: with insulin. (e) Synergistic effect of **i15** with insulin; the columns correspond to lane 2−10 in the gel images in (d) and (e); lane 1 was used as the normalization baseline. (f) SPR analysis of the binding affinity of **i14** and **i15** to INSR.

scatter plot comparing the normalized ES values in the anti-pY1355 vs anti-INSR selections (Figure 8c). Five compounds that are more significantly enriched by the anti-pY1355 antibody were chosen as the potential hits (**i14, i15, i16, i17**, and **i18**; Figure 8c,d). Three additional compounds were chosen as the "negative controls" (**i19, i20**, and **i21**), which contain similar BB structures as the hit compounds; however, **i19** was not enriched by both antibodies, **i20** has a low

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

enrichment fold in both pulldowns, and **i21** was more significantly enriched in the anti-INSR selection (Figure 8c). First, these compounds were screened for their effects on INSR pY1355/1361 levels in the cellular assay at a much lower compound concentration (20 μM). As shown in Figure 9a, **i14** and **i15** significantly increased the INSR pY1355/1361 levels, **i16**, **i17**, and **i18** showed modest activities, while the "negative controls" had no effect. The screening result is largely consistent with the compounds' enrichment profile. Dose-dependency experiments showed that **i14** and **i15** increased the INSR pY1355/1361 levels at low μM concentrations (Figure 9b and Figure S32). The EC_{50} values of **i14** are ∼10.3 μM (pY1355)/∼9.2 μM (pY1361), and **i15** gave slightly higher EC_{50} values (Figure 9c), which are significantly more potent than the hits from the previous selection. **i14** and **i15** showed a synergistic effect with insulin, and **i14** also increased the IRS pY896 and Akt pS473 levels (Figure 9d,e). Finally, we measured the binding affinity of **i14** and **i15** with SPR. **i14** gave sub-μM binding affinity (∼548 nM) and **i15** showed single-digit μM affinity (∼1.90 μM) (Figure 9f). Collectively, these results have demonstrated that the selection with a larger library may identify hit compounds with significantly improved binding affinity and agonistic activity.

Selection against TPOR. To demonstrate the method with another target, the library was also selected against TPOR. The effective concentration of TPOR on the cell was ∼1.34 μM (Figure S33), similar to INSR. After the selection, the enrichment profile was processed and presented in 3D plots (Figure S34), which gave few recognizable SAR patterns. A large difference between the anti-pY626 and anti-TPOR pulldowns was observed (Figure S35a). We reason that, since the anti-TPOR antibody captures all binders to the receptor, it enriched a greater number of compounds. The anti-pY626 antibody can pull down only the binders to the activated receptor and enrich fewer compounds. Thus, the compounds from the anti-TPOR pulldown may show lower enrichment fold values because the greater number of compounds will "even out" the percentage of each compound, while the smaller number of compounds in the anti-pY626 pulldown led to higher enrichment fold values. Also, the anti-TPOR pY626 antibody may be more efficient (higher binding affinity and/or less nonspecific binding), which may lead to a higher enrichment fold. It is more desirable to use antibodies with comparable pull-down efficiencies for the selections. Hit compounds were directly chosen from the 2D plot. Two compounds (**T1**, **T2**) showing high enrichment in the anti-pY626 pulldown and some enrichment with anti-TPOR antibody were selected. Three **192**-containing compounds (**T3**-**T5**) were chosen as controls: they were either enriched only by the anti-TPOR antibody (**T4, T5**) or enriched by neither antibody (**T3**); they can also be used to check whether BB **192** has any effect on TPOR phosphorylation. HEK293 cells were treated with the compounds, respectively. Western blots showed that **T1** and **T2** increased the pY626 level, albeit without a synergistic effect with the natural agonist TPO. **T3**, **T4**, and **T5** did not show any agonistic activity on TPOR (Figure S35c). SPR analysis showed that **T1** and **T2** bind to TPOR at 18.2 and 32.1 μM, respectively, **T3** and **T4** are weaker binders, and **T5** did not bind to TPOR (Figure S36).

Test of the Amide Derivatives of the Hit Compounds. The off-DNA hit compounds were synthesized and tested as carboxylic acids, while the library compounds were connected to the DNA tag via an amide linkage. A recent report showed the "DNA linker scar" may affect the binding properties of DEL hits.[77] Thus, we selected four hit compounds and prepared the methylamide derivatives, including three INSR hits (**i2-amide, i14-amide**, and **i15-amide**) and one TPOR hit (**T1-amide**) (Figure S37a). SPR analysis showed that, except **i2**, the methyl amide derivatives have similar binding affinities to the carboxylic acids Figure S37b), and **i2-amide** gave higher affinity than the carboxylic derivative. The results have shown that the methylamide form of the compounds also binds to the target receptor. Furthermore, the methylamide derivatives also increased the INSR pY1355/1361 levels significantly (Figure S37c).

CONCLUSIONS AND DISCUSSION

We developed a "functional" cell-based DEL selection method for agonist discovery with membrane proteins. This method leverages the property that agonists may preferentially bind to the activated receptor,[33] thereby biasing the selection toward agonist identification. The theoretical model (Figure S8) and our data suggested that a "partially activated" receptor population may be beneficial: a high phosphorylation level of the target makes it difficult to compare the pulldowns of the different antibodies, which is important for comparative hit identification, while a very low level of phosphorylation leads to low enrichment. Measurement of the target phosphorylation level is an important step prior to selection, and ∼50−70% phosphorylation level appeared to be optimal.

Recently, the Krusemark group reported a cell-based DEL selection method to identify GPCR agonists.[41] Instead of using the activated receptor to drive the binding equilibrium, their method relied on the binding of active library compounds to induce biotinylation of the target via a split-TurboID technique. The authors tested this method with only small focused libraries. In large DELs, the copy numbers of the active compounds (e.g., agonists) are extremely small and may not be sufficient to induce a detectable signal readout. Although the enzymatic turnover of TurboID may alleviate the issue, its applicability has yet to be validated with large libraries. Moreover, our method does not require genetic tagging of the target, and the endogenous phosphorylation is used for agonist enrichment. The two methods may complement each other in agonist discovery for membrane proteins, and the suitability may depend on the size of the library, the nature of the targets, and the activation mechanism of the receptor.

Several points need to be noted. First, the method is more likely to identify allosteric agonists as the orthosteric site may be occupied by the endogenous ligand in the serum. This can be leveraged for identifying allosteric agonists. Second, the antibody's specificity and affinity are important. The cross-reactivity of the antibody leads to false positives. Comparison of the parallel pulldowns with the anti-target and anti-phosphorylated target antibodies could help identify target-specific agonists and exclude the compounds that induce target phosphorylation through indirect cellular pathways. The affinities of the pulldown antibodies are mostly in the 10^{-10}−10^{-11} M (K_d) range,[78] which are sufficient for binder recovery. Moreover, the antibody's affinity will affect the overall library recovery rather than altering the relative enrichment fold of individual compounds; thus, it may be beneficial to measure the library recovery of the antibodies and then calibrate the affinity difference in data analysis. Third, this study assumed that the compounds preferentially binding to the phosphorylated receptor are more likely to be agonists;

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653

however, such binders may just have a higher affinity for the activated receptor but are not agonists. More selections with additional types of membrane proteins are needed to understand the scope of this approach. Fourth, cell-based DEL selections are very noisy. The comparisons of not only different antibodies but also multiple replicates are essential for reliable hit identification. Finally, the quality, size, and amount of the library are also important due to the target complexity. High-quality DELs built with robust chemistry are more desirable as they will reduce the noise level.[79] In addition, DEL selections require 10^3−10^5 DNA copies for each compound.[65,66] Here, $10^6 \sim 10^7$ copies were used to provide a "safety margin", since the cellular environment results in more significant nonspecific interactions. For very large DELs, it will also consume a large quantity of library; therefore, DELs with moderate size (e.g., a few hundred million compounds) may be more beneficial.

The selection identified several INSR agonists with sub/low nanomolar affinity/potency. The focused library may be designed based on the most represented BBs and selected against INSR to identify better compounds for further medicinal chemistry optimization. The analysis also showed that the hit compounds have less favorable physicochemical properties (Figure S38). Since the SAR suggested R^1 is less important, compounds with smaller or omitted R^1 may be tested to improve the physiochemical properties.

EXPERIMENTAL SECTION

Library Selection. HEK293 cells were transfected with corresponding plasmids and cultured at 37 °C for 24 h. Before selection, the cells were harvested and washed three times with 1x PBS buffer (supplemented with Phosphatase Inhibitor Cocktail IV) by gentle resuspension and centrifugation at 500× g for 5 min. The cells were resuspended in 1x PBS buffer (supplemented with Phosphatase Inhibitor Cocktail IV) containing a DNA-encoded library and corresponding CP. Cells were incubated at 4 °C for 2.5 h before UV irradiation on ice at 365 nm for 30 s by a Uvata UV LED point light source. After irradiation, the cells were washed three times with 1x PBS buffer supplemented with Phosphatase Inhibitor Cocktail IV to remove unbound molecules by gentle resuspension and centrifugation at 500× g for 5 min.

The postselection cells were lysed with modified RIPA buffer (supplemented with Pierce Universal nuclease, Roche EDTA-free protease inhibitor cocktail, and phosphatase inhibitor cocktail IV) at 4 °C for 15 min. Cell lysates were obtained by collecting the supernatant after centrifugation at 13,300 rpm under 4 °C for 10 min to remove insoluble cell debris. The lysate protein concentration was determined with a BCA Protein Assay Kit. The cell lysates were diluted to a final protein concentration of 2 mg/mL and incubated at 4 °C for 1 h with corresponding anti-phospho-target antibodies and anti-target antibodies, respectively. The mixture was then incubated with Protein A/G PLUS Agarose (Santa Cruz; catalog number: sc-2003) at 4 °C overnight. The resin was then washed with ice-cold 1× PBS buffer supplemented with Roche EDTA-free Protease Inhibitor Cocktail and Phosphatase Inhibitor Cocktail IV for 4 times. The enriched proteins were eluted with 25 μL H_2O at 95 °C for 10 min. The supernatant that contains the bound library members was collected and subjected to qPCR quantification and PCR amplification.

Selection Data Analysis. After Illumina sequencing, raw data (fastq files) were exported for processing with a custom Python script (see supplement files). The sequence counts for each library member before and after the selection were tallied to calculate the enrichment fold for each compound, which further generated a 3D interactive cubic stereogram. The sequence counts of library compounds were also used to generate a Venn diagram and one-hot encoded regression model.

Cell-Based Agonistic Activity Assays. HEK293 cells were transfected with corresponding plasmids and cultured at 37 °C for 24 h. For serum starvation, transfected HEK293 cells were washed with DMEM twice and cultured in DMEM supplemented with 100 unit/mL penicillin and 100 μg/mL streptomycin for 24 h. After serum starvation, the cells were treated with natural agonists and/or hit compounds diluted in DMEM as indicated at 37 °C for 30 min. The cells were harvested and then lysed with modified RIPA buffer (25 mM HEPES, pH = 7.5, 150 mM NaCl, 0.1% SDS, 1% NP-40, 1% triton, 1% sodium deoxycholate, 1 mM EDTA, supplemented with Pierce universal nuclease, Roche EDTA-free protease inhibitor cocktail, and phosphatase inhibitor cocktail IV) at 4 °C for 15 min. Cell lysates were obtained by collecting the supernatant after centrifugation at 13,300 rpm under 4 °C for 10 min to remove insoluble cell debris. The lysate protein concentration was determined with a BCA Protein Assay Kit. The cell lysates were then resolved with SDS-PAGE and electrotransferred onto immune-blot PVDF membranes. The membrane was blocked with 5% nonfat milk in TBST buffer, incubated with corresponding anti-phospho-target antibodies and anti-target antibodies as indicated, followed by incubation with HRP conjugated goat anti-rabbit IgG antibody. The membranes were developed with Clarity Western ECL substrate. The relative band intensities were quantified with the ImageJ software, and the phosphorylated protein level was calculated based on the quantification.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.4c08624.

Synthetic procedures, characterization data, library information, selection protocols and data analysis methods, and other experimental details (PDF)

Model selection calculation worksheet (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Feng Xiong − *Shenzhen NewDEL Biotech Co., Ltd., Shenzhen 518110, China*; Email: xiongfeng@newdel.com.cn

Yan Cao − *School of Pharmacy, Naval Medical University, Shanghai 200433, China*; orcid.org/0000-0001-5510-9773; Email: caoyan@smmu.edu.cn

Dongyao Wang − *School of Pharmacy, Naval Medical University, Shanghai 200433, China*; orcid.org/0000-0002-5852-2640; Email: wangdongyao@smmu.edu.cn

Xiaoyu Li − *Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong SAR 999077, China; Laboratory for Synthetic Chemistry and Chemical Biology Limited, Health@InnoHK, Innovation and Technology Commission, Hong Kong SAR 999077, China*; orcid.org/0000-0002-8907-6727; Email: xiaoyuli@hku.hk

Authors

Yiran Huang − *Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong SAR 999077, China*

Rui Hou − *Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong SAR 999077, China; Laboratory for Synthetic Chemistry and Chemical Biology Limited, Health@InnoHK, Innovation and Technology Commission, Hong Kong SAR 999077, China*

Fong Sang Lam − *Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong SAR 999077, China*

Yunxuan Jia − *Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong SAR 999077, China*

Yu Zhou − *Department of Chemistry and State Key Laboratory of Synthetic Chemistry, The University of Hong Kong, Hong Kong SAR 999077, China; Laboratory for Synthetic Chemistry and Chemical Biology Limited, Health@InnoHK, Innovation and Technology Commission, Hong Kong SAR 999077, China;* Present Address: Institute of Translational Medicine and School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, China, 211198 (Y.Z.); orcid.org/0000-0003-1497-8264

Xun He − *Shenzhen NewDEL Biotech Co., Ltd., Shenzhen 518110, China*

Gang Li − *Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518000, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.4c08624

Author Contributions

The manuscript was written through contributions of all authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the Shenzhen Bay Laboratory Open Fund SZBL2020090501008 (XL, GL) National Natural Science Foundation of China 21877093, 91953119 (XL), grants from National Natural Science Foundation of China 82003711 (DYW), grants from Shanghai Sailing Programme 19YF1459400 (DYW), grants from Research Grants Council of Hong Kong SAR, China AoE/P-705/16, 17301118, 17111319, 17303220, 17300321, 17300423, C7005-20G, and C7016-22G (XL), and The GuangDong Basic and Applied Basic Research Foundation General Program 2023A1515010711 (XL). We acknowledge the support from "Laboratory for Synthetic Chemistry and Chemical Biology" under the Health@InnoHK Program and State Key Laboratory of Synthetic Chemistry by Innovation and Technology Commission, Hong Kong SAR, China.

■ REFERENCES

(1) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16* (1), 19−34.

(2) Jazayeri, A.; Dias, J. M.; Marshall, F. H. From G Protein-coupled Receptor Structure Resolution to Rational Drug Design. *J. Biol. Chem.* **2015**, *290* (32), 19489−19495.

(3) Lee, Y.; Basith, S.; Choi, S. Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. *J. Med. Chem.* **2018**, *61* (1), 1−46.

(4) Holdgate, G.; Geschwindner, S.; Breeze, A.; Davies, G.; Colclough, N.; Temesi, D.; Ward, L. Biophysical methods in drug discovery from small molecule to pharmaceutical. *Methods Mol. Biol.* **2013**, *1008*, 327−355.

(5) Congreve, M.; Rich, R. L.; Myszka, D. G.; Figaroa, F.; Siegal, G.; Marshall, F. H. Fragment screening of stabilized G-protein-coupled receptors using biophysical methods. *Methods Enzymol.* **2011**, *493*, 115−136.

(6) Takakura, H.; Hattori, M.; Tanaka, M.; Ozawa, T. Cell-based assays and animal models for GPCR drug screening. *Methods Mol. Biol.* **2015**, *1272*, 257−270.

(7) Rawlings, A. E. Membrane proteins: always an insoluble problem? *Biochem. Soc. Trans.* **2016**, *44*, 790−795.

(8) Brenner, S.; Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89* (12), 5381−5383.

(9) Nielsen, J.; Brenner, S.; Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **1993**, *115*, 9812−9813.

(10) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **2004**, *22* (5), 568−574.

(11) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **2004**, *305* (5690), 1601−1605.

(12) Halpin, D. R.; Harbury, P. B.; Joyce, G. DNA display I. Sequence-encoded routing of DNA populations. *PLoS Biol.* **2004**, *2* (7), No. e173.

(13) Debaene, F.; Mejias, L.; Harris, J. L.; Winssinger, N. Synthesis of a PNA-encoded cysteine protease inhibitor library. *Tetrahedron* **2004**, *60* (39), 8677−8690.

(14) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **2009**, *5* (9), 647−654.

(15) Conole, D.; H Hunter, J.; J Waring, M. The maturation of DNA encoded libraries: opportunities for new users. *Future Med. Chem.* **2021**, *13* (2), 173−191.

(16) Satz, A. L.; Kuai, L.; Peng, X. Selections and screenings of DNA-encoded chemical libraries against enzyme and cellular targets. *Bioorg. Med. Chem. Lett.* **2021**, *39*, No. 127851.

(17) Song, M.; Hwang, G. T. DNA-Encoded Library Screening as a Core Platform Technology in Drug Discovery. Its Synthetic Method Development and Applications in DEL Synthesis. *J. Med. Chem.* **2020**, *63* (13), 6578−6599.

(18) Kunig, V. B. K.; Potowski, M.; Klika Skopic, M.; Brunschweiger, A. Scanning Protein Surfaces with DNA-Encoded Libraries. *ChemMedChem.* **2021**, *16* (7), 1048−1062.

(19) Fitzgerald, P. R.; Paegel, B. M. DNA-Encoded Chemistry: Drug Discovery from a Few Good Reactions. *Chem. Rev.* **2021**, *121* (12), 7155−7177.

(20) Kodadek, T.; Paciaroni, N. G.; Balzarini, M.; Dickson, P. Beyond protein binding: recent advances in screening DNA-encoded libraries. *Chem. Commun.* **2019**, *55* (89), 13330−13341.

(21) Satz, A. L.; Brunschweiger, A.; Flanagan, M. E.; Gloger, A.; Hansen, N. J. V.; Kuai, L.; Kunig, V. B. K.; Lu, X.; Madsen, D.; Marcaurelle, L. A.; Mulrooney, C.; O'Donovan, G.; Sakata, S.; Scheuermann, J.; et al. DNA-encoded chemical libraries. *Nat. Rev. Methods Primers* **2022**, *2* (1), 3.

(22) Plais, L.; Scheuermann, J. Macrocyclic DNA-encoded chemical libraries: a historical perspective. *RSC Chem. Biol.* **2022**, *3* (1), 7−17.

(23) Neri, D.; Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based On Endowing Organic Compounds with Amplifiable Information. *Annu. Rev. Biochem.* **2018**, *87*, 479−502.

(24) Reddavide, F. V.; Thompson, M.; Mannocci, L.; Zhang, Y. X. DNA-Encoded Fragment Libraries: Dynamic Assembly, Single-Molecule Detection, and High-Throughput Hit Validation. *Aldrichim. Acta* **2019**, *52* (3), 63−74.

(25) Huang, Y. R.; Li, Y. Z.; Li, X. Y. Strategies for developing DNA-encoded libraries beyond binding assays. *Nat. Chem.* **2022**, *14* (2), 129−140.

(26) Dockerill, M.; Winssinger, N. DNA-Encoded Libraries: Towards Harnessing their Full Power with Darwinian Evolution. *Angew. Chem., Int. Ed.* **2023**, *62* (9), No. e202215542.

(27) Yuen, L. H.; Franzini, R. M. Achievements, Challenges, and Opportunities in DNA-Encoded Library Research: An Academic Point of View. *Chembiochem* **2017**, *18* (9), 829−836.
(28) Dixit, A.; Barhoosh, H.; Paegel, B. M. Translating the Genome into Drugs. *Acc. Chem. Res.* **2023**, *56* (4), 489−499.
(29) Matsuo, B.; Granados, A.; Levitre, G.; Molander, G. A. Photochemical Methods Applied to DNA Encoded Library (DEL) Synthesis. *Acc. Chem. Res.* **2023**, *56* (3), 385−401.
(30) Sunkari, Y. K.; Siripuram, V. K.; Nguyen, T. L.; Flajolet, M. High-power screening (HPS) empowered by DNA-encoded libraries. *Trends Pharmacol. Sci.* **2022**, *43* (1), 4−15.
(31) Peterson, A. A.; Liu, D. R. Small-molecule discovery through DNA-encoded libraries. *Nat. Rev. Drug Discovery* **2023**, *22*, 699−722.
(32) Ahn, S.; Kahsai, A. W.; Pani, B.; Wang, Q. T.; Zhao, S.; Wall, A. L.; Strachan, R. T.; Staus, D. P.; Wingler, L. M.; Sun, L. D.; et al. Allosteric "beta-blocker" isolated from a DNA-encoded small molecule library. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (7), 1708−1713.
(33) Ahn, S.; Pani, B.; Kahsai, A. W.; Olsen, E. K.; Husemoen, G.; Vestergaard, M.; Jin, L.; Zhao, S.; Wingler, L. M.; Rambarat, P. K.; et al. Small-Molecule Positive Allosteric Modulators of the beta2-Adrenoceptor Isolated from DNA-Encoded Libraries. *Mol. Pharmacol.* **2018**, *94* (2), 850−861.
(34) Brown, D. G.; Brown, G. A.; Centrella, P.; Certel, K.; Cooke, R. M.; Cuozzo, J. W.; Dekker, N.; Dumelin, C. E.; Ferguson, A.; Fiez-Vandal, C.; et al. Agonists and Antagonists of Protease-Activated Receptor 2 Discovered within a DNA-Encoded Chemical Library Using Mutational Stabilization of the Target. *SLAS Discovery* **2018**, *23* (5), 429−436.
(35) Svensen, N.; Diaz-Mochon, J. J.; Bradley, M. Decoding a PNA encoded peptide library by PCR: the discovery of new cell surface receptor ligands. *Chem. Biol.* **2011**, *18* (10), 1284−1289.
(36) Svensen, N.; Diaz-Mochon, J. J.; Bradley, M. Encoded peptide libraries and the discovery of new cell binding ligands. *Chem. Commun.* **2011**, *47* (27), 7638−7640.
(37) Wu, Z.; Graybill, T. L.; Zeng, X.; Platchek, M.; Zhang, J.; Bodmer, V. Q.; Wisnoski, D. D.; Deng, J.; Coppo, F. T.; Yao, G.; et al. Cell-Based Selection Expands the Utility of DNA-Encoded Small-Molecule Library Technology to Cell Surface Drug Targets: Identification of Novel Antagonists of the NK3 Tachykinin Receptor. *ACS Comb. Sci.* **2015**, *17* (12), 722−731.
(38) Denton, K. E.; Krusemark, C. J. Crosslinking of DNA-linked ligands to target proteins for enrichment from DNA-encoded libraries. *Medchemcomm* **2016**, *7* (10), 2020−2027.
(39) Cai, B.; Kim, D.; Akhand, S.; Sun, Y.; Cassell, R. J.; Alpsoy, A.; Dykhuizen, E. C.; Van Rijn, R. M.; Wendt, M. K.; Krusemark, C. J. Selection of DNA-Encoded Libraries to Protein Targets within and on Living Cells. *J. Am. Chem. Soc.* **2019**, *141* (43), 17057−17061.
(40) Cai, B.; Mhetre, A. B.; Krusemark, C. J. Selection methods for proximity-dependent enrichment of ligands from DNA-encoded libraries using enzymatic fusion proteins. *Chem. Sci.* **2023**, *14* (2), 245−250.
(41) Cai, B.; El Daibani, A.; Bai, Y.; Che, T.; Krusemark, C. J. Direct Selection of DNA-Encoded Libraries for Biased Agonists of GPCRs on Live Cells. *JACS Au* **2023**, *3* (4), 1076−1088.
(42) Oehler, S.; Catalano, M.; Scapozza, I.; Bigatti, M.; Bassi, G.; Favalli, N.; Mortensen, M. R.; Samain, F.; Scheuermann, J.; Neri, D. Affinity Selections of DNA-Encoded Chemical Libraries on Carbonic Anhydrase IX-Expressing Tumor Cells Reveal a Dependence on Ligand Valence. *Chem. - Eur. J.* **2021**, *27* (35), 8985−8993.
(43) Petersen, L. K.; Christensen, A. B.; Andersen, J.; Folkesson, C. G.; Kristensen, O.; Andersen, C.; Alzu, A.; Slok, F. A.; Blakskjaer, P.; Madsen, D.; et al. Screening of DNA-Encoded Small Molecule Libraries inside a Living Cell. *J. Am. Chem. Soc.* **2021**, *143* (7), 2751−2756.
(44) Huang, Y.; Meng, L.; Nie, Q.; Zhou, Y.; Chen, L.; Yang, S.; Fung, Y. M. E.; Li, X.; Huang, C.; Cao, Y.; et al. Selection of DNA-encoded chemical libraries against endogenous membrane proteins on live cells. *Nat. Chem.* **2021**, *13* (1), 77−88.
(45) Rasmussen, S. G.; Choi, H. J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; Devree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; et al. Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **2011**, *469* (7329), 175−180.
(46) Boger, D. L.; Goldberg, J. Cytokine receptor dimerization and activation: prospects for small molecule agonists. *Bioorg. Med. Chem.* **2001**, *9* (3), 557−562.
(47) von Heijne, G. Membrane-protein topology. *Nat. Rev. Mol. Cell Bio.* **2006**, *7* (12), 909−918.
(48) Engel, A.; Gaub, H. E. Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.* **2008**, *77*, 127−148.
(49) Lemmon, M. A.; Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2010**, *141* (7), 1117−1134.
(50) Zhao, P.; Chen, Z.; Li, Y.; Sun, D.; Gao, Y.; Huang, Y.; Li, X. Selection of DNA-encoded small molecule libraries against unmodified and non-immobilized protein targets. *Angew. Chem., Int. Ed.* **2014**, *53* (38), 10056−10059.
(51) Haeusler, R. A.; McGraw, T. E.; Accili, D. Biochemical and cellular properties of insulin receptor signalling. *Na.t Rev. Mol. Cell Bio.* **2018**, *19* (1), 31−44.
(52) Xie, J.; Wang, S.; Ma, P.; Ma, F.; Li, J.; Wang, W.; Lu, F.; Xiong, H.; Gu, Y.; Zhang, S.; et al. Selection of Small Molecules that Bind to and Activate the Insulin Receptor from a DNA-Encoded Library of Natural Products. *iScience* **2020**, *23* (6), No. 101197.
(53) Tennagels, N.; Bergschneider, E.; Al-Hasani, H.; Klein, H. W. Autophosphorylation of the two C-terminal tyrosine residues Tyr1316 and Tyr1322 modulates the activity of the insulin receptor kinase in vitro. *FEBS Lett.* **2000**, *479* (1−2), 67−71.
(54) Yarden, Y.; Ullrich, A. Growth factor receptor tyrosine kinases. *Annu. Rev. Biochem.* **1988**, *57*, 443−478.
(55) Yee, D. A tale of two receptors: insulin and insulin-like growth factor signaling in cancer. *Clin. Cancer. Res.* **2015**, *21* (4), 667−669.
(56) Pillutla, R. C.; Hsiao, K. C.; Beasley, J. R.; Brandt, J.; Ostergaard, S.; Hansen, P. H.; Spetzler, J. C.; Danielsen, G. M.; Andersen, A. S.; Brissette, R. E.; et al. Peptides identify the critical hotspots involved in the biological activation of the insulin receptor. *J. Biol. Chem.* **2002**, *277* (25), 22590−22594.
(57) Yang, W.; Xia, Y.; Ji, H.; Zheng, Y.; Liang, J.; Huang, W.; Gao, X.; Aldape, K.; Lu, Z. Nuclear PKM2 regulates beta-catenin transactivation upon EGFR activation. *Nature* **2011**, *480* (7375), 118−122.
(58) Yang, W.; Xia, Y.; Hawke, D.; Li, X.; Liang, J.; Xing, D.; Aldape, K.; Hunter, T.; Alfred Yung, W. K.; Lu, Z. PKM2 phosphorylates histone H3 and promotes gene transcription and tumorigenesis. *Cell* **2012**, *150* (4), 685−696.
(59) Hamabe, A.; Konno, M.; Tanuma, N.; Shima, H.; Tsunekuni, K.; Kawamoto, K.; Nishida, N.; Koseki, J.; Mimori, K.; Gotoh, N.; et al. Role of pyruvate kinase M2 in transcriptional regulation leading to epithelial-mesenchymal transition. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (43), 15526−15531.
(60) Liang, S. I.; van Lengerich, B.; Eichel, K.; Cha, M.; Patterson, D. M.; Yoon, T. Y.; von Zastrow, M.; Jura, N.; Gartner, Z. J. Phosphorylated EGFR Dimers Are Not Sufficient to Activate Ras. *Cell Rep.* **2018**, *22* (10), 2593−2600.
(61) Sharma, S. V.; Bell, D. W.; Settleman, J.; Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **2007**, *7* (3), 169−181.
(62) Song, S.; Liu, D.; Peng, J.; Deng, H.; Guo, Y.; Xu, L. X.; Miller, A. D.; Xu, Y. Novel peptide ligand directs liposomes toward EGF-R high-expressing cancer cells in vitro and in vivo. *FASEB J.* **2009**, *23* (5), 1396−1404.
(63) Varghese, L. N.; Defour, J. P.; Pecquet, C.; Constantinescu, S. N. The Thrombopoietin Receptor: Structural Basis of Traffic and Activation by Ligand, Mutations, Agonists, and Mutated Calreticulin. *Front. Endocrinol.* **2017**, *8*, 59.
(64) Cwirla, S. E.; Balasubramanian, P.; Duffin, D. J.; Wagstrom, C. R.; Gates, C. M.; Singer, S. C.; Davis, A. M.; Tansik, R. L.; Mattheakis, L. C.; Boytos, C. M.; et al. Peptide agonist of the thrombopoietin

receptor as potent as the natural cytokine. *Science* **1997**, *276* (5319), 1696−1699.
(65) Chen, Q.; Cheng, X.; Zhang, L.; Li, X.; Chen, P.; Liu, J.; Zhang, L.; Wei, H.; Li, Z.; Dou, D. Exploring the Lower Limit of Individual DNA-Encoded Library Molecules in Selection. *SLAS Discovery* **2020**, *25* (5), 523−529.
(66) Sannino, A.; Gabriele, E.; Bigatti, M.; Mulatto, S.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. Quantitative Assessment of Affinity Selection Performance by Using DNA-Encoded Chemical Libraries. *Chembiochem* **2019**, *20* (7), 955−962.
(67) Shmilovich, K.; Chen, B.; Karaletsos, T.; Sultan, M. M. DEL-Dock: Molecular Docking-Enabled Modeling of DNA-Encoded Libraries. *J. Chem. Inf. Model* **2023**, *63* (9), 2719−2727.
(68) Komar, P.; Kalinic, M. Denoising DNA Encoded Library Screens with Sparse Learning. *ACS Comb. Sci.* **2020**, *22* (8), 410−421.
(69) Lim, K. S.; Reidenbach, A. G.; Hua, B. K.; Mason, J. W.; Gerry, C. J.; Clemons, P. A.; Coley, C. W. Machine Learning on DNA-Encoded Library Count Data Using an Uncertainty-Aware Probabilistic Loss Function. *J. Chem. Inf. Model* **2022**, *62* (10), 2316−2331.
(70) Hou, R.; Xie, C.; Gui, Y.; Li, G.; Li, X. Machine-Learning-Based Data Analysis Method for Cell-Based Selection of DNA-Encoded Libraries. *ACS Omega* **2023**, *8* (21), 19057−19071.
(71) McCloskey, K.; Sigel, E. A.; Kearnes, S.; Xue, L.; Tian, X.; Moccia, D.; Gikunju, D.; Bazzaz, S.; Chan, B.; Clark, M. A.; et al. Machine Learning on DNA-Encoded Libraries: A New Paradigm for Hit Finding. *J. Med. Chem.* **2020**, *63* (16), 8857−8866.
(72) Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media: USA, 2 nd ed., 2019, ISBN-10:1492032646.
(73) Gutmann, T.; Schafer, I. B.; Poojari, C.; Brankatschk, B.; Vattulainen, I.; Strauss, M.; Coskun, U. Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain. *J. Cell Biol.* **2020**, *219* (1), No. e201907210.
(74) Xu, Y.; Kong, G. K. W.; Menting, J. G.; Margetts, M. B.; Delaine, C. A.; Jenkin, L. M.; Kiselyov, V. V.; De Meyts, P.; Forbes, B. E.; Lawrence, M. C. How ligand binds to the type 1 insulin-like growth factor receptor. *Nat. Commun.* **2018**, *9* (1), 821.
(75) De Meyts, P.; Whittaker, J. Structural biology of insulin and IGF1 receptors: implications for drug design. *Nat. Rev. Drug Discovery* **2002**, *1* (10), 769−783.
(76) Gagnon, J. K.; Law, S. M.; Brooks, C. L., 3rd Flexible CDOCKER: Development and application of a pseudo-explicit structure-based docking method within CHARMM. *J. Comput. Chem.* **2016**, *37* (8), 753−762.
(77) Montoya, A. L.; Glavatskikh, M.; Halverson, B. J.; Yuen, L. H.; Schüler, H.; Kireev, D.; Franzini, R. M. Combining pharmacophore models derived from DNA-encoded chemical libraries with structure-based exploration to predict Tankyrase 1 inhibitors. *Eur. J. Med. Chem.* **2023**, *246*, No. 114980.
(78) Landry, J. P.; Ke, Y.; Yu, G. L.; Zhu, X. D. Measuring affinity constants of 1450 monoclonal antibodies to peptide targets with a microarray-based label-free assay platform. *J. Immunol. Methods* **2015**, *417*, 86−96.
(79) Keller, M.; Petrov, D.; Gloger, A.; Dietschi, B.; Jobin, K.; Gradinger, T.; Martinelli, A.; Plais, L.; Onda, Y.; Neri, D.; Scheuermann, J. Highly pure DNA-encoded chemical libraries by dual-linker solid-phase synthesis. *Science* **2024**, *384* (6701), 1259−1265.

https://doi.org/10.1021/jacs.4c08624
J. Am. Chem. Soc. 2024, 146, 24638−24653