Artificial Intelligence uncovers Evolutionarily Conserved Intracellular Allosteric Modulators of GPCR-Gα interface.

Sanjay Kumar Mohanty[1,†], Aayushi Mittal[1,†], Aakash Gaur[2,3], Saveena Solanki[1], Suvendu Kumar[1], Subhadeep Duari[1], Sakshi Arora[1], Vishakha Gautam[1], Nilesh Kumar Dixit[1], Tarini Shankar Ghosh[1], Debarka Sengupta[1], Shashi Kumar Gupta[2,3], Natarajan Arul Murugan[1], and Gaurav Ahuja[1]

[1]Department of Computational Biology, Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi), Okhla, Phase III, New Delhi, 110020, India.

[2]Pharmacology Division, CSIR-Central Drug Research Institute, Lucknow, India

[3]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India.

[†]Equal contributions

Correspondence: Gaurav Ahuja; Email: gaurav.ahuja@iiitd.ac.in.

Running title: Modulation of GPCR-Gα binding via intracellular metabolites

Keywords: apoptosis, Deep Learning, GPCR, hypertrophy, mating, yeast

ABSTRACT

Traditionally, the exogenous allosteric modulators of G protein-coupled receptors (GPCRs) have been extensively investigated due to their pharmacological significance. However, to date, only a handful of endogenous intracellular allosteric modulators are known, that too with inconclusive binding information and their associated phenotypes. This limitation primarily stems from the non-availability of robust computational techniques that entails unbiased cavity identification across GPCR protein topology, cavity-specific ligand design, their synthesis, and cross-verification. Here, we introduce Gcoupler, which leverages an integrative approach combining *de novo* ligand design, statistical methods, and Graph Neural Networks for rational prediction of high-affinity ligands. Gcoupler offers an efficient and comparatively faster route to explore endogenous allosteric sites of GPCRs, including the GPCR-Gα interface. We tested and validated the applicability of Gcoupler in decrypting the cellular metabolites that could intrinsically but directly modulate the Ste2 (GPCR)-mediated pheromone-induced cell death in yeast. Our rigorous interrogation using Gcoupler and experimental approaches, including yeast genetic screening, high-resolution metabolomics, and functional assays identified endogenous hydrophobic metabolites as intracellular allosteric inhibitors of Ste2p signaling. Elevated intracellular levels of these metabolites, either naturally, through genetic alterations, or exogenous treatment, rescue the pheromone-induced programmed cell death. Mechanistic insights revealed that metabolites harbor high-binding affinity to the conserved GPCR-Gα interface and trigger a cohesive response that potentially obstructs downstream signaling. Finally, by utilizing isoproterenol-induced, GPCR-mediated human and primary neonatal rat hypertrophy models, we observed that elevated metabolite levels attenuate hypertrophic response, reinforcing the functional and evolutionary relevance of this allosteric modulation mechanism. In summary, our study reports a robust computational method that uncovered a novel, evolutionary conserved, and metabolite-driven regulatory mechanism of GPCR signaling.

INTRODUCTION

G Protein-Coupled Receptors (GPCRs) are one of the most preferred drug targets due to their functional importance in regulating vital cellular functions such as homeostasis, growth, metabolism, and death [1–3]. Traditional therapeutic approaches devised to target GPCR activity are centered around the orthosteric or the ligand binding sites [4,5]. Recent advancements in the structural and functional information about the GPCRs led to the identification of novel, albeit spatially distinct allosteric sites that could potentially modulate the receptor activity response [5–7]. These allosteric sites can be exploited using chemical probes to alter (positively or negatively) the GPCR-induced physiological responses. While many exogenous synthetic compounds with allosteric characteristics are known in the literature, limited information is available about intracellular, allosteric, small molecule modulators, their mode of regulation, and their physiological relevance [8–13]. The practical design or identification of highly-specific and high-affinity endogenous allosteric compounds demand efficient approaches leveraging structure-based drug design, artificial intelligence (AI), and biochemical assays, each possessing defined limitations. For example, classical computational SAR (structure-activity relationships) techniques demand a substantial amount of experimentally validated compounds for the designated cavity for predictive modeling, which, unfortunately, at least in the case of GPCRs, is the limiting factor [14,15]. Similarly, experimental approaches such as Förster resonance energy transfer (FRET), bioluminescence resonance energy transfer (BRET), and single-molecule detection fluorescence can be employed to validate potential allosteric compounds; however, their implementation for the hunt of novel intracellular allosteric modulators via high-throughput screening is technically challenging [16–18]. Identification and subsequent characterization of novel endogenous GPCR intracellular allosteric modulators harbor multiple challenges due to the lack of accurate topology information for multiple GPCRs [19], methods required to detect them, large-scale screening to evaluate their modulatory function, astronomically larger chemical space among others [1,20]. This necessitates the implementation of a hybrid computational approach that could leverage the power of the predictive structure approaches for unbiased detection of allosteric sites (cavities), *de novo* ligands synthesis based on cavity topological features, and finally, to accelerate the screening procedure, a helping hand from statistical methods and state-of-the-art deep learning approaches. While multiple computational tools have been developed over the years that allow *de novo* drug design, they largely lack practical suitability for discovering GPCR endogenous intracellular allosteric modulators due to computational constraints and excessive demand for domain knowledge and technical skills [21–27]. To overcome such limitations, we developed a python package, Gcoupler, that leverages computational structural biology, statistical methods, and

deep learning approaches and offers a novel, computationally inexpensive, and highly efficient approach to decrypt the endogenous intracellular allosteric modulators of GPCRs systematically.

We demonstrated the usability and applicability of Gcoupler in identifying novel endogenous modulators of GPCRs by exploiting the α-pheromone (α-factor)-induced mating or programmed cell death pathway of *S. cerevisiae*. Ste2p (GPCR) is the hub protein of the yeast mating response, where the pheromone-bound Ste2p triggers the exchange of GDP for GTP in a heterotrimeric G protein (Gpa1p) and initiates mitogen-activated protein kinase (MAPK) signal transduction cascade, resulting in the expression of mating-related genes [28–30]. Notably, it has been previously reported that exposure to low doses of mating pheromones, either during the failed mating event or in the absence of a suitable partner, triggers programmed cell death (PCD) in the haploid cells [31–34]. Mechanistically, pheromone-induced PCD induction activates the Ste2/Ste3-mediated MAP-kinase signaling pathway, which triggers an increase in intracellular Ca^{2+} levels, and mitochondrial activity, leading to the release of cytochrome c [31–34]. While the pheromone-induced PCD favors the diploid state, which is comparatively more advantageous than the haploid one, it is arguable that evolution only unilaterally supported one state over the other without imparting an innate defense mechanism to nullify accidental or intentional pheromone-induced PCD induction. Notably, it is well documented that the highly elevated, non-physiological levels of α-factor trigger PCD in less than half of the MATa population [33]. Moreover, an equivalent concentration of α-factor triggers distinct PCD kinetics across distinct laboratory strains; for example, BY4741 is more resistant than the W303 strain [35]. Earlier studies utilizing the genetic screening approach identified Yeast Suicidal Protein 1 (LAM1) and Yeast Suicidal Protein 2 (LAM2) as the positive regulators of α-factor-induced PCD [35]. While the exact mechanisms that facilitate resistive or sensitive response towards pheromone-induced PCD are still elusive, the evidence clearly suggests the involvement of a previously undiscovered innate rescue mechanism that imparts resistance to MATa upon α-factor exposure.

We hypothesized that a subset of pheromone-resistant cells might regulate the Ste2p-mediated mating pathway by the endogenous intracellular metabolites at the Ste2-Gα binding interface. By using Gcoupler, we identified a subset of intracellular metabolites that could potentially bind to Ste2p (GPCR) at the Gpa1 (Gα) binding interface and regulate the downstream signaling. Our results suggest that hydrophobic ligands such as sterols strengthen the Gpa1p-Ste2p binding and might trigger a cohesive response that potentially obstructs downstream signaling. Experimental evidence further supports these findings that the elevated intracellular levels of these metabolites rescue the pheromone-induced

programmed cell death. To evaluate the evolutionary conservation and possible clinically-relevant translation of this mechanism, we tested these metabolites on the human isoproterenol-induced, GPCR-mediated hypertrophy model system and observed attenuated hypertrophic response in the cardiomyocytes pretreated with GPCR-Gα-protein interface modulating metabolites. This study presents Gcoupler, a computational workflow as a Python package that leverages structure-based *de novo* ligand design and Graph-Neural Network-based identification of high-affinity ligands. Moreover, by leveraging Gcoupler, we underpin the intracellular allosteric modulating mechanism by cellular metabolites and showed their functional importance in regulating α-factor induced PCD in yeast, and isoproterenol-mediated hypertrophy models.

RESULTS

Structure-based ligand design and validation using Gcoupler

Identifying hit compounds for any given protein is a daunting task, primarily due to the diversity and complexity of the chemical space [20,36]. G-protein coupled receptors (GPCR) are one of the most pharmaceutically important targets because they regulate many essential human physiological functions, including growth, metabolism, and homeostasis [2,37,38]. Presently, both the computationally exhaustive *in silico* methods such as docking, and complex, time-consuming, and expertise-demanding experimental approaches such as structure elucidation and biochemical assays are used to identify and validate hit molecules. Therefore, to ease and accelerate the ligand designing for distinct cavities of the topological surface of GPCRs, we developed Gcoupler Python Package. Gcoupler adopts an integrative approach utilizing structure-based *de novo* ligand design, robust statistical method, and highly powerful Graph Neural Networks. It is primarily designed to identify ligands that could potentially bind to the GPCR-Gα-protein interface and modulate Gα-protein coupling, and hence the downstream signaling.

The Gcoupler package possesses three significant modules, i.e., Synthesizer, Authenticator, and Generator, that collectively impart a smoother, user-friendly, and minimalistic experience for the *de novo* ligand design. Due to the implementation of Graph Neural Network algorithm-based models, Gcoupler provides a rapid solution for large-scale screening of any chemical libraries (such as ZINC15 [39] or Enamine Real database [40]) or user query compound. Synthesizer, the first module of Gcoupler, takes a protein structure as input in PDB format, identifies putative cavities across the topological/entire protein surface, and subsequently selects the user-preferred cavity for the downstream analysis. Post selection of the target cavity, the Synthesizer module allows structure-based *de novo* ligand designing that generates potential ligands and outputs the Simplified Molecular Input Line Entry System (SMILES) information of the synthesized compounds, three-dimensional cavity coordinates and other files required by the succeeding Gcoupler modules **(Figure 1a)**. The Synthesizer module of Gcoupler employs LigBuilder V3 [41], which utilizes the genetic algorithm (GA) for the ligand synthesis. Authenticator, the next module of Gcoupler, further segregates the synthesized compounds into high-affinity binders (HAB) and low-affinity binders (LAB) by leveraging a structure-based virtual screening approach and statistically-backed hypothesis testing for distribution comparisons **(Figure 1a)**. The Authenticator module outputs the free binding energies of all the generated compounds. The statistical submodule further segregates the compounds into HAB and LAB while ensuring

the optimal binding energy threshold and class balance. Of note, Authenticator leverages the Empirical Cumulative Distribution Function (ECDF) for binding energy distribution comparisons of HABs and LABs and performs Kolmogorov–Smirnov test [42], Epps-Singleton test [43], and Anderson-Darling Test [44] for hypothesis testing. Moreover, it also follows a strict criterion to ensure adequate representation of HABs and LABs to avoid class imbalance issues for the downstream analysis. Importantly, in case the Authenticator module fails to generate optimal LABs, Gcoupler also allows users to generate or submit external decoys that might serve as a negative dataset for model building in the next step [45]. Generator, the last module of the Gcoupler, takes these outputs and applies state-of-the-art Graph Neural Network algorithms such as Graph Convolution Model, Graph Convolution Network, Attentive FP, and Graph Attention Network to generate classification models [46]. Noteworthily, the hyperparameter tuning and cross-validation submodules are already automated and implemented in the Generator module. By default, Gcoupler builds four distinct base models using the four aforementioned algorithms. Moreover, it also supports the augmentation of the best base model performance metrics by facilitating hyperparameter tuning along with parameter cross-validation (default: 3 folds). Finally, it allows users to predict the binding probability of user-supplied query compounds **(Figure 1a)**. Taken together, Gcoupler overcomes the previously mentioned prerequisite limitations and offers a hassle-free and highly efficient method that leverages *de novo* ligand design and deep learning algorithms.

To test and compare the performance of Gcoupler with a state-of-the-art molecular docking algorithm (AutoDock) [47], we selected alpha-1A adrenergic receptor (ADRA1A). We selected this receptor due to the availability of information about its experimentally validated agonists in the ChEMBL database [48]. For better comparison, we followed the conventional steps of AutoDock Tools for molecular docking while keeping track of execution time for each step during the whole process till completion **(Figure 1b, d, Supplementary Figure 1a-b)**. We also followed the same timestamp procedure for the Gcoupler, including its separate module sub-functions **(Figure 1c, d, Supplementary Figure 1a-d)**. The processing time consumption comparison between the two approaches revealed that Gcoupler takes 13.5 folds less time than AutoDock while preserving the prediction accuracy **(Figure 1d, Supplementary Figure 1e)**. Both approaches produced similar predictions, i.e., low binding energy and high predicted probability for all the 993 known agonists **(Supplementary Figure 1e)**. These results clearly suggest that Gcoupler offers a novel and highly rapid alternative approach to ligand design.

To gain deeper insights into the GPCR-Gα-protein interface from the sequence, structural and functional perspectives and to investigate whether the intracellular metabolites could

target the exposed, intracellular GPCR-Gα-protein binding interface, we first downloaded all the experimentally deduced structures of the human GPCR-Gα complexes and systematically investigated and extracted the GPCR-Gα interacting interface residues **(Figure 1e, Supplementary Table 1)**. Next, by interrogating the sequence level amino acid conservation at the GPCR-Gα-protein interface, we identified multiple conserved motifs, including the canonical DRY, CWxL, and NPxxY motifs [3] **(Figure 1f)**. To determine the topological similarity of the GPCR-Gα-protein interface, we identified and extracted the cavities across all GPCR-Gα-protein complexes and observed structural conservation as indicated by normalized Root Mean Square Deviation (RMSD) values (mean: 1.47 Å; median: 0.86 Å) **(Figure 1g-i, Supplementary Figure 1f-g, Supplementary Table 2)**. Finally, to test whether this topological and sequence conservation also impacts the ligand profiles that could potentially bind to this interface, we performed the Gcoupler workflow on all 66 GPCRs and synthesized ~50 unique ligands per GPCR **(Figure 1g).** We next computed and compared the physicochemical properties of these synthesized ligands and observed higher cosine similarity indicating functional conservation of the ligand profiles that could bind to these interfaces **(Figure 1g, j)**. In summary, we used Gcoupler to systematically evaluate and analyze the ligand profiles of the GPCR-Gα-protein interface and observed a higher degree of sequence, topological and functional conservation.

Gcoupler reveals endogenous, intracellular Ste2p allosteric modulators.

Next, we used Gcoupler to investigate the α-factor-mediated programmed cell death/apoptosis in yeast to identify novel (hitherto uncharacterized) intracellular allosteric modulating metabolites that impart natural resistance to PCD induction. The yeast mating pathway is one of the most extensively studied cellular pathways to date [28,29,33,34]. To delineate the identity and mode-of-action of endogenous metabolites that modulate the GPCR-G-protein interface, we used the yeast mating receptor Ste2p, a pheromone-activated G protein-coupled receptor, as a model system. Importantly, our central aim is to pin down the intracellular metabolites that can bind to the Ste2p-Gpa1p interface and modulate the signaling cascade **(Figure 2a)**. To achieve this, we used Gcoupler to screen for such metabolites against the Yeast Metabolome Database (YMDB) [49]. We utilized the cryo-EM structure of the Ste2 protein[50] and performed a small-scale molecular dynamic simulation within a lipid bilayer environment, and subjected the simulated stable structure to the Gcoupler workflow **(Figure 2b, Supplementary Figure 2a-g)**. The Synthesizer module of the Gcoupler package identified 17 potential surface cavities spanned across the extracellular, intracellular, and transmembrane regions of the Ste2 protein **(Figure 2c, Supplementary Figure 2h-k)**. Careful interrogation of the structurally-supported

Ste2p-Gpa1p interface revealed two distinct predicted cavities (annotated as IC4 (Intracellular Cavity 4) and IC5 (Intracellular Cavity 5)), collectively capturing >90% of the interface regions **(Figure 2d, Supplementary Figure 2l, m)**. We also measured the residue level conservation of the Ste2 protein by leveraging Ste2 protein sequences from 14 other closely related yeast species. Residue level conservation scores revealed a higher degree of conservation across IC4 and IC5 **(Supplementary Figure 2l, m)**. We next synthesized ~500 synthetic compounds for both IC4 and IC5 each by leveraging the Synthesizer module of Gcoupler. Finally, these compounds were segregated into high-affinity binders (HABs) and low-affinity binders (LABs) by the Authenticator module of Gcoupler. Notably, the estimated binding energy threshold was set at -7 kcal/mol, a widely accepted cutoff in virtual screening **(Figure 2e,f)**. A deep investigation into these classified synthetic compounds showed a comparable similarity between the HABs and LABs of the aforementioned target cavities, respectively **(Supplementary Figure 2n, Supplementary Figure 3a-b, Supplementary Table 3)**. Finally, the Generator module built classification models by implementing four distinct Graph Neural Network algorithms. Comparing the model performance metrics suggests that Attentive FP outperformed other algorithms for both cavities **(Figure 2g, Supplementary Figure 3c)**. Finally, we screened all the known yeast metabolites, compiled from YMDB, on the best-performing model (hyperparameter tuned Attentive FP model) and predicted metabolites that could potentially bind to the IC4 and IC5 of the Ste2p-Gpa1p interface with a binding probability cutoff of > 0.5 **(Figure 2e)**. To further optimize the lead metabolite list, we parallelly performed the standard molecular docking using AutoDock with YMDB metabolites against IC4 and IC5 of Ste2p, respectively, and finally selected the consensus metabolites (binding energy < -7 kcal/mol and binding probability > 0.5) for the downstream analysis **(Figure 2e, Supplementary Figure 2o, Supplementary Table 4)**. Of note, the consensus metabolites list was further segregated into HAM (High-Affinity Metabolites) and LAM (Low-Affinity Metabolites) based on the binding prediction probability (cutoff = 0.5). Comparative analysis of the binding prediction probabilities of Gcoupler and binding energies from Autodock of HAB and LAB revealed a significant negative correlation that further validates our novel approach's authenticity **(Figure 2h).** Of note, as expected, HAMs and LAMs possess distinct atomic fingerprints as indicated by the Principal Component Analysis (**Supplementary Figure 3d)**. Furthermore, the functional group enrichment analysis of these metabolites suggests enrichment of R2NH, R3N, ROPO3, ROH, and ROR functional groups in the high-affinity binding metabolites for both cavities **(Figure 2i, Supplementary Figure 3e)**. To gain the pathway-level information about these putative endogenous intracellular allosteric modulators of Ste2p, we performed Pathway level Over Representation Analysis (ORA)[51] and observed the selective enrichment of metabolites involved in the steroid, sesquiterpenoid and triterpenoid biosynthesis and one

carbon pool by folate pathways **(Figure 2j)**. Taken together, by using Gcoupler, we narrowed down cellular metabolites that could potentially bind to the Ste2p-Gpa1p interface and modulate pheromone-induced PCD or mating response.

Ste2p-Gpa1p interface regulatory metabolites inhibit pheromone signaling.

Gcoupler revealed a subset of intracellular metabolites that could potentially bind to the Ste2p-Gpa1p interface and regulate the downstream signaling. To test the impact of these metabolites on the Ste2-mediated signaling cascade, we performed a genetic screening of the metabolic mutants. We first mapped the predicted allosteric-modulating metabolites to the biochemical pathway space using KEGG [52] and MetaCyc [53], and identified the enzymes responsible for processing these metabolites **(Figure 3a,b, Supplementary Table 5)**. Notably, among the 53 single metabolic mutants (+ *ste2Δ*) used for screening, only Ste2 was reported in the KEGG pathway (ID: sce04011) for an altered mating response **(Supplementary Figure 4a)**. Second, we performed the large-scale activity screening of single metabolic mutants, assuming that their loss leads to the intracellular accumulation of these selected metabolites and, therefore, be used as a proxy to evaluate the crosstalk between metabolism (metabolites) and GPCR modulation **(Figure 3a)**. The signaling cascade activity was estimated using the α-factor-induced PCD assay using Propidium Iodide (PI). Briefly, selected single metabolic mutants (MATa) were grown under optimal growth conditions until the late log phase (16 hours). Noteworthily, intercomparison of the growth profiles of these single metabolic mutants revealed varied growth responses with respect to the wild-type, with the majority of the mutants harboring delayed growth response **(Supplementary Figure 4b)**. The late log phase cells were treated with the α-factor, and the PCD induction was measured using the PI-based cell viability assay **(Figure 3c)**. Heat-killed cells were used as the positive control (100% cell death). As expected, we observed a significant increase in the PI fluorescence in the wild-type BY4741 yeast strains treated with α-factor, indicating pheromone-induced PCD, whereas no significant death was observed in the *STE2* loss of function mutants. Interestingly, most single metabolic mutants (94.4%) resisted the α-factor-induced cell death, while a subset of them also displayed accelerated growth in the presence of α-factor, indicating a cross-talk between the central metabolism (or metabolites) and the Ste2 signaling. Our screening procedure also revealed some metabolic mutants that harbor enhanced sensitivity to the α-factor **(Figure 3d)**. These results indicate that a significant proportion of Gcoupler-predicted metabolites could directly or indirectly influence the Ste2 signaling pathway, establishing a link between metabolism and Ste2 signaling.

We next performed a high-resolution metabolomics approach to gain further insight into the metabolic pathways or metabolomes imparting α-factor-induced PCD resistance. To achieve this, we first selectively enriched the cells that tolerated the α-factor-induced PCD at different pheromone concentrations, termed as survivors (**Supplementary Figure 4c**), and subsequently performed high-resolution metabolomics to uncover their metabolic profiles **(Figure 3e)**. Of note, we exposed yeast cells to the increasing concentrations of α-factor to narrow down on differentially enriched metabolites in the surviving population. In addition to the wild type, we profiled the Ste2 knockout. Unbiased high-resolution metabolome analysis revealed a subset of metabolites that were differentially and significantly enriched in the survived population **(Figure 3f, Supplementary Figure 4d-f, Supplementary Table 6)**. In a cross-comparison analysis, we observed an overlap of seven metabolites between the Gcoupler prediction of Ste2p intracellular allosteric modulators and differentially-enriched metabolites from the survived cells, and among these, Ubiquinone 6 and Zymosterol showed prominent enrichment in the survived cells at all tested concentrations **(Figure 3g-i, Supplementary Figure 4g)**. Moreover, Over Representation Analysis (ORA) of the differentially-enriched metabolites suggests their involvement in glyoxylate and dicarboxylate metabolism, purine metabolism, vitamin B6 metabolism, etc. **(Supplementary Figure 4h)**. Taken together, the results from genetic screening of the metabolic mutants and untargeted metabolomics from the PCD-survived mutants hint towards the interplay between the central metabolism and Ste2 signaling. Notably, these results also pinpointed two key metabolites that could potentially impart resistance against α-factor-induced PCD.

Elevated metabolic levels selectively inhibit GPCR signaling.

We next asked whether the observed resistance toward α-factor-induced PCD in single metabolic mutants might be the direct consequence of the identified metabolites or a pleiotropic response due to an altered genome. To answer this, we exogenously elevated zymosterol (ZST), ubiquinone 6 (CoQ6), and lanosterol (LST) levels in the wild-type strain and measured α-factor-induced PCD by three different techniques. Notably, lanosterol was predicted as a high-affinity binder to the Ste2p-Gpa1p interface by Gcoupler (prediction probability of 0.99 for IC4 and IC5 each) and was also supported by the docking technique; however, it was not detected in our untargeted metabolomics screening. Methodologically, we first pre-loaded the yeast cells with these metabolites for 24 hours. Notably, in almost all the cases, except for 10μM concentration of lanosterol, we did not observe any significant alterations in the yeast growth profiles, indicating the physiological relevance of the selected doses **(Supplementary Figure 5a)**. Post-loading the cells with metabolites, the cells were screened for alterations in the α-factor-induced PCD induction by measuring their growth

profiles, cellular viability measurement using PI, and cellular viability and vitality measurement using FUN™1 staining [54,55]. In the case of growth kinetics-based PCD screening, we measured and compared the area under the growth profile's curve (refer to methods section). As expected, we observed a significant decrease in AUC values in the wild-type strain treated with α-factor, while no significant differences were observed in the metabolite pre-loaded cells as well as in the *ste2Δ*, indicating the direct role of these metabolites in modulating Ste2p signaling **(Figure 4a, Supplementary Figure 5b)**. A similar rescue phenotype was observed in the pre-loaded cells using the PI-based fluorometric measurement for cell viability. In this case, we also used heat-killed cells as a positive control **(Figure 4b, c, Supplementary Figure 5c)**. Finally, we used FUN™ 1 staining to estimate and compare the cell vitality and viability using fluorescent microscopy. In the case of the FUN™ 1-based assay [54], we also used acetic acid, an alternative exogenous stimulus, to induce PCD in yeast [56,57]. The quantification of metabolically active and viable (faint green with red spots) cells revealed α-factor-induced PCD rescue response in the case of cells pre-loaded with zymosterol (ZST), ubiquinone 6 (CoQ6), and lanosterol (LST) compared to the non-pre-loaded wild-type cells. As expected, we did not observe α-factor-induced PCD in the *ste2Δ* cells. Interestingly, the observed PCD rescue response in the metabolite-pre-loaded cells was only valid for α-factor but not in the acetic acid treatment **(Figure 4d, e)**. These results suggest that the observed rescue response is exclusive to α-factor-induced PCD (and not a consequence of an altered response) and likely via modulation of the Ste2p-Gpa1p interface.

We next probed whether the observed modulation of Ste2 signaling is only applicable to the exogenous pheromone exposure-based assays (as discussed above) or if it also modulates the natural yeast mating behavior, an alternate but natural proxy to evaluate the Ste2 signaling pathway [58]. We investigated this by performing a mating assay with MATa and MATα strains. Of note, we used *ste2Δ* strain as a negative control in the mating assay [59]. In line with the aforementioned exogenous pheromone exposure-based assays, the mating assay also revealed a decline in Ste2 signaling in metabolite-pre-loaded cells, suggesting the interlink between these metabolites and Ste2p signaling. Notably, in the case of ubiquinone 6 (CoQ6), we failed to observe any significant decline in the mating response **(Figure 4f, Supplementary Figure 5d),** which might be due to the naturally lower levels of α-factor in contrast to the artificially elevated conditions used for cellular pre-loading in the aforementioned assays.

We next investigated whether the interaction of these metabolites at the Ste2p-Gpa1p interface is stable or transient. To achieve this, we performed molecular dynamics (MD)

simulations of the Ste2p-metabolite complex **(Figure 4g-h, Supplementary Figure 6a)**. MD simulation results suggest that the interactions between the metabolites and Ste2p at the Ste2p-Gpa1p interface are thermodynamically stable in almost all the cases across both the cavities (IC4 and IC5), except for the ubiquinone 6 (CoQ6), which harbored a fluctuating Root Mean Square Deviation (RMSD) over simulation timeframe **(Figure 4i)**. Notably, fluctuating RMSD is observed only in the case of IC5, while it is within the permissive range for IC4 **(Figure 4i, Supplementary Figure 6b)**. These simulation results could potentially explain the non-significant defects in the mating efficiency of the MATa pre-loaded with ubiquinone 6 (CoQ6) **(Figure 4f)**. To gain deeper insight into the mode of action of these metabolites in inhibiting Ste2p signaling, we first analyzed their impact on the α-factor binding. We performed protein-peptide docking of Ste2p and α-factor and observed that metabolite-binding at the Ste2p-Gpa1p interface favors the α-factor interaction, as inferred from the binding free energies ΔG (kcal/mol) **(Supplementary Figure 6c-d)**. We also analyzed protein-protein interaction between the Ste2p (GPCR) and miniGpa1-protein [50] by selecting GPCR configurations with and without metabolite-induced altered cavity topologies (IC4 and IC5), respectively (frames output from the aforementioned Ste2p-metabolite complex) and computed the dissociation constant (K_d), binding affinity (ΔG) and the structural changes in the overall Ste2p topology **(Figure 4j-k, Supplementary Figure 6c-e)**. These computational analyses revealed that in contrast to metabolite-free Ste2p-Gpa1p interaction, referred to as wild-type condition, the interaction K_d value is multi-fold lower in the presence of metabolite, indicating a cohesive response induced by these metabolites **(Figure 4k)**. A multi-fold lower K_d value further indicates and potentially explains that the metabolites binding favors Ste2p (GPCR) and miniGpa1-protein interaction and enables the establishment of a stable complex that might influence the shielding of the effector-regulating domains of the Gpa1 protein or influence its binding with the Ste4p (Gβ)-Ste18p (Gγ) complex. This could be the possible mechanism that might explain the obstruction of the downstream signaling. While the mechanistic model is highly speculative and based on computational findings, it demands further experimental evidence using Förster resonance energy transfer (FRET) or other protein-protein interaction assays for validation [16]. In summary, the combination of experimental and computational techniques resulted in a working model that could potentially explain the blockage of α-factor-induced Ste2 signaling via these metabolites.

Since our intensive computational interrogation of all the available human GPCR-Gα complexes revealed a higher degree of functional conservation **(Figure 1i, j)**, we next asked whether zymosterol (ZST), lanosterol (LST), and ubiquinone 6 (CoQ6) could also modulate human GPCR signaling. In other words, we intended to investigate whether sterols *per se*

could intrinsically and intracellularly modulate the GPCR activity in higher vertebrates as well. To test this hypothesis, we first performed the computational analysis on the human and rat beta1/2-adrenergic receptors. Briefly, by using Gcoupler, we identified the putative GPCR-Gα interface and performed molecular docking with zymosterol (ZST), lanosterol (LST), and ubiquinone 6 (CoQ6) **(Supplementary Figure 6f)**. Docking results revealed a high binding affinity of these selected metabolites at the GPCR-Gα interface of adrenergic receptors, reminiscent of Ste2p-metabolite interactions **(Supplementary Figure 6g-j)**. Finally, to elucidate the functional relevance of this predicted interaction, we selected isoproterenol-induced, adrenergic receptor-mediated cardiac hypertrophy models in cultured human cardiac cell line (AC16) and primary neonatal rat cardiomyocytes **(Figure 4l)**. Our results suggest that both the AC16 and neonatal cardiomyocytes preloaded with these metabolites attenuate the isoproterenol-induced hypertrophic response as inferred from the quantitative analysis of the single cardiomyocytes' surface area **(Figure 4m, n)**. Taken together, these results collectively reinforce that modulation of GPCR activity by endogenous intracellular metabolites is an evolutionarily conserved phenomenon.

DISCUSSION

Over the last few decades, a tremendous amount of research has been conducted to identify the allosteric modulators of GPCRs [7,20]. Due to their relevance in drug discovery, most of the examples belong to the exogenous allosteric modulators [3,4,6–8,20]. The intracellular side of the GPCRs is exposed to the cytosol and therefore allows binding to the signaling proteins such as trimeric G-proteins and arrestins [1,5,8]. In addition to these ubiquitous binders, the GPCR-Gα interface or other adjoining intracellular sites can also be targeted by chemically heterogeneous modulators, for example, GPCR-directed autoantibodies, Pepducins, sodium ions, etc.; however, the underlying mechanism of their mode of action is still elusive [4–6,8,9,11–13,60]. Systematic interrogation of allosteric sites, including the GPCR-Gα interface or nearby regions, demands multiple resource-intensive computational tools, in-depth technical skills, and domain knowledge [14,19,61]. While methods such as conventional computational structural biology or empirical machine learning approaches can be used to identify novel intracellular allosteric modulators of GPCRs, unfortunately, due to the lack of information about compounds regulating novel allosteric sites, implementation of these approaches is mainly unfeasible [62–65]. Here, we developed Gcoupler, a computational framework that harnesses the power of *de novo* cavity identification on GPCR topological surface and subsequent ligand synthesis, coupled with rigorous statistical methods for segregating synthesized compounds into high-affinity and low-affinity binders. Finally, the deep learning (Graph Neural Networks) algorithms in Gcoupler empower predictive modeling and allow large-scale screening of the user-queried compounds. Interestingly, Gcoupler also offers a novel approach that not only alleviates the need for experimentally validated compounds for the given cavity (allosteric site); but also offers an efficient way for the large-scale screening of the user-specified compounds in a hassle-free manner.

We used Gcoupler to discover the molecular basis of innate resistance towards α-factor-induced programmed cell death (PCD) in yeast [31,33,34,55]. Unlike the human genome that encodes ~800 GPCRs, the yeast genome possesses only two GPCR systems that allow pheromone and glucose sensing [66]. The pheromone-sensing pathway in yeast has been significantly interrogated over the last few decades leading to the identification of feedback inhibition systems involved in this pathway; however, to the best of our knowledge, the majority of the known regulatory mechanisms operate downstream of the signaling cascade [29,34]. This further suggests that the rescue mechanism that imparts innate resistance towards α-factor-induced PCD might operate much upstream of the signaling cascade via a direct, physical interaction-based regulatory mechanism, in contrast to transcriptome/epigenome-mediated feedback mechanisms that might be comparatively

slower. One such case could be the regulation at the receptor level via the endogenous metabolites that can intracellularly induce allosteric modulation of the Ste2 protein. This could be feasible since a handful of biochemically diverse, endogenous intracellular allosteric modulators for different GPCRs are reported in the literature [8–10]. Gcoupler enabled pinpointing a subset of cellular metabolites that could physically bind to the Ste2p-miniGpa1 interface and regulate the downstream signaling pathway [50]. Experimental evidence further strengthens the link between these selective metabolites and Ste2p-mediated signaling.

While the present evidence from this work clearly demonstrates this interlink; however, it possesses certain limitations. First, while genetic screening for single metabolic mutants hints towards this interlink, it is still unknown whether the observed deviations in the α-factor mediated PCD are metabolite mediated or have emerged from the pleiotropic effect in these mutants. Second, the concentrations of the metabolites used in the exogenous treatment experiments might not be the same as those present in natural conditions. Third, the interaction between metabolites (ZST, CoQ6, and LST) and Ste2p was only observed *in silico* and demands further experimental proof involving FRET or BRET-based assays [16]. Fourth, the metabolites-protein interaction predictions can be substantiated by the mutagenesis experiments of the IC4 and IC5 residues; however, such experiments might further complicate the interpretation and pose challenges to elucidate whether the speculative mutation-based activity loss is due to interference with trimeric G-protein binding or metabolites-induced binding enhancement or hindrance. Of note, previous mutagenesis experiments revealed multiple critical amino acid residues that overlap with IC4 and IC5, suggesting their functional relevance in Ste2p downstream signaling **(Supplementary Table 7).** Finally, it is technically challenging to assess the specificity of the observed mechanism by using another yeast GPCR system (Gpr1p) due to the high level of sequence, structural and functional conservation of the GPCR-Gα interface.

Functional characterization of sterol biosynthesis intermediate (*ergΔ*) mutants revealed impaired mating response. Mechanistic insights further revealed that the underlying phenotypic defects are highly heterogeneous and range from defective schmoo morphology, reduced sterol accumulation at the mating projection tip, impaired membrane fusion, and lower expression levels of *FUS1* [67–71]. The present study uncovers a new molecular mechanism that provides a novel function of cellular sterols and also explains the underlying reason for the reduced efficiency of mating-dependent yeast microbial cell factories or biosensors in mutants with altered sterol levels [72,73]. Notably, in addition to the Ste2p regulation by endogenous intracellular sterols, their dysregulation also imparts tolerance to

mycotoxic compounds, ethanol, and high temperatures; however, the underlying mechanism is largely elusive [74].

Since sterols facilitate many distinct biological functions, their *in vivo* functional characterization, such as co-measurement of their interactions with cellular proteins, and induced phenotypic changes are technically challenging. The present study utilizes both the computational and experimental approaches and showed that sterols can target the conserved Gα binding sites of GPCRs and modulate their activity. The human and rat hypertrophy experiments further reinforce the specificity of this evolutionary conserved regulatory mechanism. In summary, in addition to the uncovering of a novel GPCR regulatory mechanism by endogenous intracellular metabolites, the present study also proposes a novel computational approach that could accelerate and maximize our understanding of multiple unexplored allosteric sites of GPCRs.

MATERIAL AND METHODS

Backend code for the Gcoupler

The back-end code for the Gcoupler is written entirely in Python (3.8). The Gcoupler workflow consists of three distinct modules, namely, Synthesizer, Authenticator, and Generator. The Synthesizer module leverages the LigBuilder V3.0 [41] to identify the putative cavities on the protein surface with the potential to be an active or an allosteric site and perform *de novo* drug synthesis within the target cavity by following a hybrid approach of GROW and LINK method using Genetic Algorithm. Notably, among all the predicted cavities, the Synthesizer module pinpoints one cavity as the target for the subsequent ligand synthesis process. The decision is autonomous in nature and is taken based on user-supplied residue-position pairs of interest. Synthesizer performs *de novo* synthesis of the active compounds within the target cavity, complementing the pharmacophores in the cartesian space based on the target cavity. In summary, Synthesizer, the first module generates and outputs the simplified molecular-input line-entry system (SMILES) and PDBQT files of the synthesized ligands along with cavity information in the form of grid coordinates, which is required as input by the next module.

Authenticator, the next module in Gcoupler takes the PDBQT files and the cavity grid information generated using Synthesizer as input and further allows the segregation of these synthetic compounds based on their actual interaction at the molecular level. Authenticator leverages Vina (1.2.3) python library for virtual screening [75]. The binding energies produced by the Authenticator module are further used to segregate the synthetic compounds into High-Affinity Binders (HABs) and Low-Affinity Binders (LABs) based on the distribution while preserving the class balance for the downstream deep learning steps. Of note, the default value of binding energy for the distribution split is set to -7 kcal/mol; however, the module provides enough flexibility to the users to try and test different cutoffs (loose to stringent) and can visualize the change in distribution and statistically compare the resulting distributions within the Gcoupler workflow. The Authenticator uses the Kolmogorov-Smirnov test [42], Anderson-Darling test [44], and Epps-Singleton test [43] for hypothesis testing for the distribution comparison.

(1) Two Sample Kolmogorov-Smirnov test:

$$D_{m,n} = max_x |F(x) - G(x)|,$$

where,

$F(x)$ = observed cumulative distribution function of the first sample of size m

$G(x)$ = observed cumulative distribution function of the second sample of size n

(2) Anderson-Darling test:

$$ADK = \frac{n-1}{n^2(k-1)} \sum_{i=1}^{k} \left[\frac{1}{n_i} \sum_{j=1}^{L} h_j \frac{(nF_{ij} - n_i H_j)^2}{H_j(n-H_j) - \frac{nh_j}{4}} \right],$$

where,

h_j = the number of values in the combined samples equal to Z_j

H_j = the number of values in the combined samples less than Z_j plus one half the number of values in the combined samples equal to Z_j

F_{ij} = the number of values in the i-th group which are less than Z_j plus one half the number of values in this group which are equal to Z_j

k = the number of samples (groups)

n_i = the number of observations in group i

x_{ij} = the jth observation in the ith group, and

Z_j = Z_1, Z_2 ..., Z_L are the distinct values in the combined data set ordered from smallest to largest (L is less than n if there are tied observations)

(3) Epps-Singleton test:

Epps-Singleton test was used to compare the binding energies distributions of HAB and LAB. It returns a p-value that signifies the probability of falsely rejecting the null hypothesis (H_o) that both HAB and LAB are taken from the same population.

The Authenticator module supports the graphical plots, such as overlapping density plots and Empirical Cumulative Distribution Function (ECDF), to visualize the distributions. In case either the Authenticator module fails to recognize the default cutoff as an optimal threshold for the distribution split or the user decides that the resulting distributions fail to obtain the statistical inference under any threshold of their choice, the module is also incorporated with the option for synthesizing decoys for the generation of negative class. Authenticator supports decoys generation by using the Chem module of RDKit Python package [45].

The last module of Gcoupler, known as Generator, leverages the DeepChem (2.6.1) [46] python library to build the Graph-Neural Networks supported classification models. The Generator takes the segregated HAB and LAB/decoys from Authenticator module and uses

these synthetic compounds to build prediction models. By default, the Generator module supports four different graph-based models, i.e., Graph Convolution Model (GCM), Graph Convolution Network (GCN) [76], Attentive FP (AFP) [77], and Graph Attention Network (GAT) [78], to generate classification models. Before the model-building step, the Generator module evaluates the class imbalance in the synthetic dataset and implements the upsampling techniques to counter it, if required. Generator by default, tests the pre-processed synthetic data on all four deep learning-based models under the default hyperparameters and returns models' performance parameter list, allowing users to select the best-performing model for the downstream refinement process. Users can either manually set up the hyperparameter tuning grid parameters or opt for Generator recommended grid range along with the fold count for k-fold cross-validation to tune the selected model around the best hyperparameters and also evaluate the stability of the said optimized parameters. In the final step, Generator builds the resultant model with the best-performing hyperparameters by utilizing the complete synthetic compound library (HAB + LAB/decoys) for training, allowing large-scale screening of the user query compounds by using SMILES information.

Runtime Analysis

To showcase the speed and efficiency of our *de novo* coupled Deep Learning approach against conventional docking, we performed a comparative study involving analysis of the same query compound dataset against a single target of interest through both approaches. A system with 125 GB RAM, 16 Threads, and 8 CPU cores with no upgradation to the software or hardware within the duration of this analysis was used as the platform for this comparison. To compare the time complexity between Gcoupler and classical docking procedures using AutoDock [47], we selected human alpha-1A adrenergic receptors (ADRA1A) as our GPCR of interest. Since the experimentally elucidated structure of ADRA1A is unknown, we used AlphaFold [79] predicted structure as our starting point. The ligand binding cavity at the extracellular site was determined using the LigBuilder V3.0 [41] cavity function. Moreover, we obtained the ADRA1A ligand information from the ChEMBL database corresponding to the accession ID: CHEMBL229 [48]. We obtained a total of 3684 bioactive compounds against ADRA1A; however, to ensure uniformity, we only selected those ligands (#933) that were annotated as “single protein format experiments”. Files preparation for AutoDock-based docking includes conversion of ligand SMILES to MOL2 using OpenBabel (2.4.1) [80], followed by conversion of MOL2 to PDBQT via MGLtools (1.5.6) [81], and receptor PDB to PDBQT using OpenBabel (2.4.1) [80]. Noteworthily, the grid parameter calculations were manually performed, and docking was performed using AutoDock (4.2.6).

The Gcoupler workflow was performed as per the aforementioned steps. To ensure uniformity of the target cavity between both approaches, the cavity information along with the grid coordinate obtained during the initial steps of the AutoDock approach [47] was reused by the Synthesizer and the Authenticator module for the synthesis of the *de novo* ligands and their molecular interaction based classification respectively. The Authenticator module selected an optimal binding energy cutoff of -8 kcal/mol; however, we opted for decoys for the negative dataset generation due to severe class imbalance. Under default hyperparameters, AttentiveFP showed better overall performance scores among the four deep learning-based models tested. We used the AttentiveFP model with the following hyperparameters: {'num_layers': 2, 'num_timesteps': 2, 'graph_feat_size': 200, 'dropout': 0} after assessment with 10-fold cross-validation. The final model was generated using the complete HAB compound library as the positive and the respective decoys dataset as a negative class for the model learning. Post-model generation, all 933 ligands were screened for their binding prediction probabilities.

Sequence-Structural-Functional level analysis

All the available, non-redundant sixty-six human GPCR-Gα complexes were downloaded from the RCSB PDB database (https://www.rcsb.org/). The sequence level information for each of these GPCRs was obtained from the UniProt database (https://www.uniprot.org). Multiple Sequence Alignment was performed to study the sequence level conservation, using the FASTA sequences of the GPCRs as input for the MUltiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm [82]. Of note, MUSCLE uses PAM substitution matrix and sum-of-pairs (SP) score for alignment scoring. The MUSCLE alignment output in CLW format was then subsequently used to compute amino acid level conservation using the WebLogo tool (https://weblogo.berkeley.edu/logo.cgi). To identify conserved motifs native to the GPCR-Gα interface, residues at the GPCR-Gα interface for each GPCR-Gα complex were retrieved using EMBL-EBI developed PDBePISA tool and then subsequently mapped on the WebLogos [83]. To test the structural level conservation, the pairwise similarity of the PDBePISA provided GPCR-Gα interfaces (cavity) of all 66 GPCRs was achieved using the normalized Root Mean Square Deviation (RMSD), calculated using the PyMOL software (https://pymol.org/2/). The RMSD values were normalized by dividing the cavity RMSD with their respective whole protein RMSD. Finally, to analyze the functional level conservation of the GPCR-Gα interfaces, we used the Gcoupler's Synthesizer module to compute 50 synthetic ligands for each GPCR-Gα interface (cavity) and calculated their physicochemical properties (descriptors) using Mordred [84]. For each GPCR-Gα-protein interface, we computed a single vector representing the aggregated physicochemical property of all the

synthesized compounds for the given GPCR cavity and measured their similarities using the Cosine similarity.

$$cosine\ similarity = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

where, A_i and B_i are components of vectors A and B of size n respectively.

Molecular Dynamics Simulation

The Molecular Dynamics simulation was performed using either GROMACS [85] or AMBER [86] software suite. For Ste2p, the experimentally elucidated structure with PDB ID: 7ad3 was downloaded from the RCSB PDB database [50]. Since the available structure was provided for the dimer, PDB preprocessing steps were used to isolate a single chain (chain B) that refers to the GPCR. Embedding in the lipid bilayer was performed using the CHARM-GUI [87], followed by the simulations using GROMACS [85]. Notably, we performed six rounds of equilibration with the membrane system and nine rounds of production runs of 5 ns each. We finally selected the last frame of the simulation step from the production run for further analysis. Protein-ligand complexes were prepared for the molecular dynamics simulations using AMBER software [86], as discussed above. Notably, we performed all analyses using the Ste2p monomers. The grid box center and several grids were chosen appropriately to restrict the conformational sampling within the specific binding site. By default, the software estimates docking energies for 2,50,000 conformations of the ligands within the binding sites and ranks them to find the top 10 binding modes. The best binding modes for the ligands in these two sites were analyzed. The subsequent molecular dynamics simulations were performed for the ligands in these two binding sites, and the input configuration is based on the aforementioned molecular docking study results. The charges for the ligands were computed using density functional theory calculations at the B3LYP/6-31+G* level of theory. We used a general amber force field (GAFF) to describe the molecular interactions [88]. For protein, we employed FF19SB force-field [89], and for water, TIP3P force-field was used [90]. The Ste2p-ligand complexes were embedded in about 23,000 water molecules. The minimization run, simulation in the constant volume ensemble, and simulation in the isothermal-isobaric ensemble were carried out. The time step for integrating the equation of motion was 2 fs. Followed by the equilibration run, a long time scale simulation for a 100 ns time scale was carried out. All the Ste2p-ligand complex simulations were performed using Amber20 software [86]. Various structural properties, such as RMSD and RMSF were computed and analyzed. The binding free energies using molecular mechanics-Generalizer

Born surface area approach (MM-GBSA) were computed for 1000 configurations picked up from the last 10 ns trajectory.

Molecular Docking (AutoDock)

YMDB metabolites were downloaded from the YMDB database [49]. The file preparation for the docking protocol involved converting the YMDB SMILES (1D) to MOL2 (3D) format. This was achieved using the OpenBabel software (2.4.1) [80]. Further, the compounds from the MOL2 format were converted into dockable PDBQT format by following the MGLtools (1.5.6) ligand preparation procedure [81]. Of note, the sub-processing steps include detecting the ligand center, selecting single bonds as torsion/rotatable bonds, etc. As discussed above, the receptor PDBs were also converted into PDBQT format using the OpenBabel software [80]. AutoDock (4.2.6) [47] was used for molecular docking by implementing Genetic algorithm with the following parameters (population size = 150, maximum number of evaluations = 2,50,00,000, and maximum number of generations = 27,000). Notably, ten complexes were generated per YMDB compound. The docking results, i.e., extracting the lowest binding energy from each docked protein-ligand complex, were analyzed using the MGLtools (1.5.6) [81].

Functional Enrichment Analysis

Functional enrichment analysis of the differentially enriched metabolites or the metabolites of interest was performed using MetaboAnalyst (https://www.metaboanalyst.ca/) web server [51]. In the case of the YMDB compound library, the union list of the lead compounds for each target cavity (IC4 & IC5) was used as input. In the case of the untargeted metabolomics data, the list of differentially enriched metabolites between the α-factor treated and untreated conditions were used as input. The hypergeometric test was used for hypothesis testing and for providing subsequent p-values for each resultant enrichment pathway.

Protein-Protein Docking

To determine the impact of metabolite binding on GPCR-Gα and GPCR-α-factor interaction, we performed Protein-Protein docking using the HADDOCK (High Ambiguity Driven protein-protein DOCKing) webserver [91]. In the case of Ste2p-miniGpa1p or Ste2p–α-factor docking, the PDB chains of both proteins were extracted from the PDB complex (accession id: 7ad3) [50]. Ste2p in docked complex with the respective metabolites were simulated via the AMBER software suite [86]. Post simulation, the most stable conformation of protein-metabolite was taken, and the metabolite was manually removed to obtain a metabolite-influenced stable Ste2p structure. HADDOCK was subsequently used for the

protein-protein docking between wild Ste2p or metabolite-influenced bound stable Ste2p against miniGpa1-protein or α-factor (extracted from PDB ID: 7ad3). The optimal representative complexes were obtained from the HADDOCK results [91] and were used as input in PRODIGY (PROtein binDIng enerGY prediction) [92] for K_d and ΔG calculations.

Yeast strains

BY4741 *(MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*), and *ste2Δ* (*MATa ste2Δ::kanMX his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*) strains of *Saccharomyces cerevisiae* were used in all the experiments. For the mating assay, BY4742 strain *(MAT*α *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) was also used. All the knockouts were obtained from Yeast Deletion Collection (*MATa xxx::kanMX his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*). Unless mentioned, the yeast was grown at 30°C at 200 rpm in yeast extract–peptone–dextrose (YPD) (1% yeast extract, 2% peptone, and 2% dextrose). Agar (1.5%) was additionally added to YPD to prepare plates.

Metabolomics

Wildtype (BY4741) and *ste2Δ* yeast strains were inoculated in the primary culture, followed by secondary culture, in YPD medium at 30°C at 150 rpm for 16 hours each. 1.5 mL of the secondary cultures were aliquoted into a 96-well deep well plate to ensure an equal number of cells for both strains. α-factor (T6901, Sigma-Aldrich) was added to the wells to attain final concentrations of 10 μM, 20 μM, 30 μM, 40 μM and 50 μM, with eight biological replicates each. DMSO (28580, SRL) (volume equivalent to the 30 μM α-factor sample) was added to the solvent control. DMSO or α-factor was not added to untreated (WT) and *ste2Δ* conditions. The plate was covered with a breathe-easy membrane (Z380059, Sigma-Aldrich) and incubated at 30°C at 150 rpm for 4 hours. After incubation, an aliquot of 50 μL was used to perform Propidium Iodide (PI) fluorometric assay as described in the later section.

Post Propidium Iodide-based cell viability assay, four biological replicates were created by merging two biological replicates. These were transferred to 1.5 mL microcentrifuge tubes and pelleted down at 6000 rpm for 5 minutes at room temperature (RT). The cell pellet was treated with zymolyase enzyme (L2524, Sigma-Aldrich) at a final concentration of 40 units/mL in 1X PBS and incubated at 30°C for 1 hour. Following this, the cells were washed with 1X PBS, and the cell pellet was stored at -80°C before metabolomics.

Metabolomics was performed on 1290 Infinity HPLC coupled with 6545 QTOF (Agilent Technologies, Santa Clara, CA, USA) equipped with dual ESI Source. The internal controls used in metabolomics include Jasmonic acid, Gibberellic acid, zeatin, Tryptophan 15N,

Estrone D4, Arginine 15N, Thiamine D4, 15N5-8-OH-2-dG, 13C12 dityrosine, and 8-PGF2-D4. For the sample preparation, the protein precipitation method was used. Briefly, 10 µL of a cocktail containing commercial stable isotopes internal standards were added, followed by the addition of chilled 750 µL methanol: water (4:1) and probe sonicated. Then, 450µL of chilled chloroform was added and vortexed for 5 min, followed by adding 150 µL of chilled water, vortexed, and kept the sample at -20 °C for 30 min. Next, samples were centrifuged at 5000 rpm at 4°C for 10 min to separate the methanol and chloroform layers. Both layers were collected and dried at 30 °C using a vacuum concentrator and reconstituted in 100 µL methanol: water (1:1). Quality Control (QC) samples were injected between every five samples to check the drift in the instrument and to evaluate the signal repeatability. Pearson's rank order showed that QCs (n=6) were strongly correlated($r>0.9$) with each other, suggesting variations were insignificant.

Data analysis of the metabolomics data includes peak normalization. Furthermore, the normalized peak intensities of all the samples' replicates were merged, and the metabolites with a constant value or with more than 50% missing values were omitted. The missing values were imputed using the feature-based k nearest neighbors (kNN) via Metaboanalyst webserver. The data were filtered based on the interquartile range (IQR). To find the list of Differentially Enriched Metabolites (DEMs), $\log_2$ fold change ($\log_2$FC) values were calculated for each sample with respect to the solvent control, along with its statistical significance student's t-test. The metabolites with $|\log_2 FC \geq 1|$ and p-value < 0.05 were considered DEMs. Pathway Over Representation Analysis (ORA) was performed using MetaboAnalyst [51].

Genetic Screening

Fifty-three knockout strains from Yeast Deletion Collection, along with WT and *ste2Δ* strains, were grown in primary culture, followed by secondary culture, in YPD medium at 30°C at 150 rpm for 16 hours each. Ensuring an equal number of cells in all the strains, 5 µL of secondary culture was inoculated into 96-well plates containing 145 µL of YPD and grown for 16 hours at 30°C at 200 rpm with eight biological replicates. Afterward, the α-factor was added to half the replicates at a final concentration of 30 µM. An equal volume of DMSO was added to the other half to serve as solvent control. The plates were incubated for another 4 hours. After incubation, a 50 µL aliquot was mixed with 50 µL of staining solution (Propidium Iodide (PI; 11195, SRL) in 1x PBS, freshly prepared) to attain a final PI concentration of 5 µg/mL in 96-well black plates. The plates were incubated in the dark for 15 minutes. Heat-killed (HK) cells were used as a positive control. The fluorescence was measured using

Biotek Synergy HTX multi-mode reader at excitation and emission wavelengths of 530/25 nm and 590/25 nm, respectively. Another 50 µL aliquot from the original plates was used to measure OD_{600} after 1:2 dilution with 1X PBS. The fluorescence values were first normalized for data analysis with the blank normalized OD_{600} values of the respective well. These normalized fluorescence values were subjected to two more rounds of normalization, first with unstained cells followed by heat-killed cells. Following this, percentage fold change was calculated for the treated group with respect to the untreated group. The p-value was computed using a one-sample student's t-test.

Cell viability Assay using Growth Kinetics

Wildtype and *ste2Δ* yeast strains were grown in primary culture, followed by secondary culture, in YPD medium at 30°C at 200 rpm for 16 hours each. Ensuring an equal number of cells for both strains, 5 µL of secondary culture was inoculated into a 96-well plate containing 145 µL of YPD along with respective metabolites (CoQ6: Coenzyme Q6, 900150O, Avanti® Polar Lipids, Sigma-Aldrich; ZST: Zymosterol, 700068P, Avanti® Polar Lipids, Sigma-Aldrich; and LST: Lanosterol, L5768, Sigma-Aldrich) at concentrations of 0.1 µM, 1 µM, and 10 µM each and grown for 24 hours at 30°C at 200 rpm with multiple biological replicates. An equal volume of ethanol was added to the control group to negate the effects of the solvent. After incubation, 5 µL cultures from this plate were used as an inoculum for another 96-well plate containing 145 µL of YPD along with respective metabolites (CoQ6, ZST, and LST) at concentrations of 0.1 µM, 1 µM, and 10 µM each in the presence of 10 µM α-factor (T6901, Sigma-Aldrich) in the treated group and DMSO (28580, SRL) in the untreated group with multiple biological replicates. The plates were sealed with breathe-easy membranes to prevent evaporation. The growth kinetics was performed in Biotek Synergy HTX multi-mode reader for 16.5 hours at 30°C under continuous orbital shaking at 559 CPM. Absorbance was measured at 600 nm every 30 minutes. The area under the curve (AUC) was calculated for each sample using the bayestestR package in R with default parameters and normalized using the OD_{600} value at time t=0.

Cell viability Assay using Propidium Iodide

Wildtype and *ste2Δ* yeast strains were grown in primary culture, followed by secondary culture, in YPD medium at 30°C at 200 rpm for 16 hours each. Ensuring an equal number of cells for both strains, 5 µL of secondary culture was inoculated into a 96-well plate containing 145 µL of YPD along with respective metabolites (CoQ6, ZST, and LST) at concentrations of 0.1 µM, 1 µM, and 10 µM each and grown for 24 hours at 30°C at 200 rpm with multiple biological replicates. An equal volume of ethanol was added to the control group to negate

the effects of the solvent. Following this, the α-factor was added to half the wells at a final concentration of 30 μM. An equal volume of DMSO was added to the other half to serve as solvent control. The plates were incubated for another 4 hours. After incubation, a 50 μL aliquot was mixed with 50 μL of staining solution (Propidium Iodide (PI) in 1X PBS, freshly prepared) to attain a final PI concentration of 5 μg/mL in 96 well-black plates. The plates were incubated in the dark for 15 minutes. Heat-killed (HK) cells were used as a positive control. The fluorescence was measured using Biotek Synergy HTX multi-mode reader at excitation and emission wavelengths of 530/25nm and 590/25nm, respectively. Another 50 μL aliquot from the original plate was used to measure OD_{600} after 1:2 dilution with 1X PBS.

The fluorescence values were first normalized for data analysis with the blank normalized OD_{600} values of the respective well. Cellular background fluorescence values of the unstained cells were subtracted from these normalized fluorescence values. Post this, fold change between the untreated solvent control and the α-factor treated samples was computed by dividing the latter with the former. Mann Whitney U-test was used to measure the statistical significance.

Cell viability Assay using FUN™1 staining

Wildtype and *ste2Δ* yeast strains were grown in primary culture, followed by secondary culture, in YPD medium at 30°C at 150 rpm for 16 hours each. Ensuring an equal number of cells for both strains, 5 μL of secondary culture was inoculated into a 96-well plate containing 145 μL of YPD along with respective metabolites at concentrations of 1 μM (ZST and LST), and 10 μM (CoQ6) and grown for 24 hours at 30°C at 150 rpm with nine biological replicates. An equal volume of ethanol was added to the control group to negate the effects of the solvent. Following this, α-factor (T6901, Sigma-Aldrich) was added to a third of the wells at a final concentration of 30 μM, acetic acid (85801, SRL) to another third at a final concentration of 199 mM, and DMSO to the remaining third to serve as solvent control. The plates were incubated for another 4 hours. Post incubation, the cells were pelleted down and washed with 1X PBS to remove any media traces at 6000 rpm for 5 minutes at RT. The staining solution was freshly prepared in 1X GH solution (10 mM HEPES buffer, pH 7.2, 2% glucose) with FUN™ 1 Cell Stain (F7030, Invitrogen) at a final concentration of 5 μM. The samples were incubated in the dark for 30 minutes, centrifuged at 6000 rpm for 5 minutes, and the pellet resuspended in 25 μL of 1x GH solution. 5 μL of the sample was used for fluorescence microscopy (Nikon Eclipse Ci-L Fluorescence Microscope with Nikon DS-Qi2 camera). FITC and TRITC filters were used for measuring green and red fluorescence, respectively, with an exposure time of 1 second each. For each sample, the number of cells

with red and green fluorescence, indicating cellular viability and vitality in each image, were counted using ImageJ and normalized with the total cell counts.

Mating Assay

BY4741 (*MATa*), BY4742 (*MATα*), and *ste2Δ* yeast strains were grown in primary culture, followed by secondary culture, in YPD medium at 30°C at 200 rpm for 16 hours each. Ensuring an equal number of cells for both strains, 5 µL of secondary culture was inoculated into a 96-well plate containing 145 µL of YPD along with respective metabolites at concentrations of 1 µM (ZST and LST) and 10 µM (CoQ6) and grown for 24 hours at 30°C at 200 rpm with 3 biological replicates. An equal volume of ethanol was added to the control group to negate the effects of the solvent. After 24 hours, the cells were harvested and washed with fresh YPD at 6000 rpm for 5 minutes at RT to remove all traces of metabolites. The pellet was resuspended in fresh YPD. After ensuring the number of cells is equal in all the samples, 100 µL of MATa cells were mixed with 100 µL of MATα in a 96-well plate, mixed thoroughly, covered with breathe easy membrane and incubated at 30°C at 200 rpm for 4 hours for mating. After mating, the cells were harvested and washed with distilled water at 6000 rpm for 5 minutes at RT to remove all media traces. Following this, the cells were resuspended in 200 µL of distilled water, serial dilutions of 1:10 were prepared in distilled water, and 5 µL culture was spotted on the SC-Met (His, Leu, Ade, Ura, Lys), SC-Lys (His, Leu, Ade, Ura, Met), and SC-Met-Lys (His, Leu, Ade, Ura) agar plates each in 3 technical replicates. Composition of the SC agar plates: 2% glucose, 2% agar, 20 mg/L histidine, 120 mg/L leucine, 60 mg/L lysine, 20 mg/L methionine, 20 mg/L uracil, and 20 mg/L adenine. The plates were incubated at 30°C for two days.

After incubation, the colonies were counted in ImageJ in the 10^4 dilutions, which had the best resolution of colonies. Mating efficiency was calculated using the following formula:

$$Mating\ Efficiency = \frac{colonies\ in\ DD \times Dilution\ Factor}{colonies\ in\ SD \times Dilution\ Factor} \times 100$$

where, SD is single dropout media (SC-Met for MATα, and SC-Lys for MATa), and DD is double dropout (SC-Met-Lys).

Ethical statement

The local IAEC (Institutional Animal Ethics Committee) committee at CSIR-Central Drug Research Institute approved all the animal experiments (IAEC/2020/38) following the

guidelines of the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), New Delhi, Government of India.

Cardiomyocytes Hypertrophy Models

Human cardiomyocyte cell line AC16 was cultured in DMEM-F12 (Thermo Scientific) with 12.5% fetal bovine serum (Thermo Scientific) at 37°C in the presence of 5% CO_2. AC16 cells were seeded in a 24-well cell culture plate for cell size measurements. After 24 hours, cells were treated with metabolites (CoQ6, ZST, LST) at the final concentration of 2.5 μM in the presence of 1% fetal bovine serum (FBS) and incubated overnight. The next day, the medium was changed, and fresh metabolites were again added at 2.5 μM final concentration together with isoproterenol treatment at 25 μM concentration and incubated for 48 hours in the presence of 1% fetal bovine serum. Next, cells were washed with 1X PBS, fixed with 4% paraformaldehyde, and stained with wheat germ agglutinin (Thermo Scientific) and DAPI. Images were taken by Leica DMI 6000 B (Leica Microsystems, USA) at 20X objective. Cell area was calculated by using ImageJ software.

Neonatal rat cardiomyocytes were isolated from one to three-day-old SD rat pups using Collagenase Type II. Briefly, pups were sacrificed, hearts were explanted, washed in 1X PBS, and chopped into small pieces. Chopped heart pieces were digested with Collagenase Type II at 1 mg/ml (Thermo Scientific, USA) at 37°C. Digested cells are then centrifuged, resuspended in DMEM medium, and pre-plated for 90 minutes to remove fibroblasts. Cardiomyocytes containing supernatant were then collected and seeded in a gelatin-coated 24-well cell culture plate coated. After 48 hours later, cells were washed with 1X PBS and incubated with metabolites at a concentration of 2.5 μM overnight in the presence of 1% FBS. The next day, the medium was changed, and cells were treated with metabolites (2.5 μM) and isoproterenol (10 μM) in the presence of 1% FBS. After 72 hours, cells were fixed and stained with alpha-sarcomeric actinin (#MA1- 22863 Thermo Scientific, USA) and DAPI. Images were taken by Leica DMI 6000 B (Leica Microsystems, USA) at 20X objective. Cell area was calculated by using ImageJ software.

Statistical Analysis

All statistical analyses were performed using either the Past4 software [93] or R-Programming [94]. Mann-Whitney U test was used to compare the medians of the two distributions (non-parametric), while Student's t-test was used for comparing the means (pairwise comparisons). In case of p-value correction, where required, Bonferroni method of p-value

adjustment is used. The p-value cutoff used in this study is 0.05. *, **, ***, and **** refer to p-values <0.05, <0.01, <0.001, and <0.0001, respectively.

Data Availability

The processed untargeted metabolomics data is provided as Supplementary Information. The raw data files are available from the lead author upon request.

Code Availability

A Python package for Gcoupler is provided via pip https://test.pypi.org/project/Gcoupler/ or from the project GitHub page: https://github.com/the-ahuja-lab/Gcoupler. The project's GitHub page provides the source code used to build Gcoupler.

Acknowledgments

The authors thank the IT-HelpDesk team of IIIT-Delhi for providing assistance with the computational resources. We thank all the members of the Ahuja lab for their intellectual contributions at various stages of this project. We thank Prof. G.P.S Raghava for providing critical comments on our manuscript. We also thank Dr. Martin Graef and Dr. Kaushik Chakraborty for sharing yeast strains and the NIPER Guwahati central facility for helping us with high-resolution metabolomics. The Ahuja lab is supported by the Ramalingaswami Re-entry Fellowship (BT/HRD/35/02/2006), a re-entry scheme of the Department of Biotechnology, Ministry of Science & Technology, Government of India, Start-Up Research Grant (SRG/2020/000232) from the Science and Engineering Research Board, and a research grant from IHUB Anubhuti (Project Grant/23) and an intramural Start-up grant from Indraprastha Institute of Information Technology-Delhi. The INSPIRE faculty grant from the Department of Science & Technology, India funds the Sengupta lab. Gupta Lab is funded by Ramalingaswami Re-entry Fellowship (BT/RLF/re-entry/14/2019) from the Department of Biotechnology, Government of India.

Author Contributions

The study was conceived by G.A. Computational analysis workflows were designed by G.A. Statistical guidance was provided by D.S. Docking, and Molecular Simulations were supervised by N.A.M and conducted by N.A.M, and S.K.M. Data analysis was performed by S.K.M., A.M., S.S., S.K., S.D., S.A., V.G., T.S.G., and N.K.D. S.K.M build the Gcoupler Python package. G.A. and A.M. designed yeast experimental workflows and performed by A.M. Human cell culture-based experiments were performed and analyzed by A.G., and

S.K.G. Illustrations were drafted by A.M., S.K.M., and G.A. G.A. wrote the paper. All authors have read and approved the manuscript.

FIGURE LEGENDS

Figure 1: Development of the Gcoupler computational package to decode GPCR-Gα-protein interface topology.

(a) Schematic workflow depicting the main modules implemented in the Gcoupler package. Of note, Gcoupler possesses three major modules, i.e., Synthesizer, Authenticator, and Generator. Synthesizer takes the information about the GPCR of interest from the users in the PDB format. It also takes the key residues which can be used for cavity annotation. Synthesizer allows generation of *de novo* design of the synthetic ligands based on the cavity topology. Authenticator, the second module, allows the screening of these synthetic ligands via a virtual screening approach. It also supports inferential statistics and graphical methods for statistics and data visualization, respectively. Generator, the last module, allows building the classification models for large-scale screening of the user-provided compounds. **(b)** Waterfall chart depicting the timestamp information about the key steps involved in AutoDock. Of note, time consumed at the indicated steps is mentioned in seconds and hours along the y-axis and above bars, respectively. **(c)** Waterfall chart depicting the timestamp information about the key steps involved in Gcoupler. Of note, time consumed at the indicated steps is mentioned in seconds and hours along the y-axis and above bars, respectively. **(d)** Bar plot depicting the comparison between the total time consumed by the AutoDock and Gcoupler to predict the binding of experimentally validated ligands for ADRA1A (AF-P35348). **(e)** Schematic workflow depicting the major steps in determining the sequence, structural and functional conservation of the human GPCRs-Gα-protein interfaces. **(f)** Snakeplot depicting the standard human GPCR two-dimensional sequence level information. Conserved motifs of the GPCR-Gα-protein interfaces are depicted as WebLogo. Asterisks represent residues of conserved motifs present in the GPCRs-Gα-protein interfaces. Of note, the location of the motifs indicated in the exemplary GPCR snake plot is approximated. **(g)** Schematic workflow illustrating the steps involved in measuring and comparing the structural conservation of the GPCR-Gα-protein interfaces across human GPCRs. **(h)** Representative structures of the proteins depicting highly conserved (low RMSD) and highly divergent (high RMSD) GPCR-Gα-protein interfaces. PDB accession numbers are indicated at the bottom. **(i)** Heatmap depicting the RMSD values obtained by the pairwise comparison of all the GPCR-Gα-protein interfaces of the available human GPCRs from the protein databank. Of note, the RMSD of the Gα-protein cavity was normalized with the RMSDs of the respective whole proteins across all pairwise comparisons. **(j)** Heatmap depicting the pairwise cosine similarities between the synthesized ligands at the GPCR-Gα-protein interfaces of the available human GPCRs using Gcoupler.

Figure 2: Gcoupler predicted endogenous metabolic modulators of Ste2p.

(a) Schematic diagram depicting the hypothesis that the intracellular yeast metabolites could allosterically modulate the GPCR-Gα-protein (Ste2p-Gpa1p) interface and, therefore, the programmed cell death pathway in yeast. **(b)** Overlapping Ste2p (GPCR) ribbon diagram depicting the structural similarity between the experimentally elucidated Ste2p structure before and after molecular dynamics simulation using GROMCS software. RMSD value is depicted at the bottom. **(c)** Schematic diagram depicting the topology of all the cavities identified using the Synthesizer module of the Gcoupler python package. Of note, the cavity nomenclature includes the cavity location, i.e., EC (extracellular), IC (intracellular), and TM (transmembrane), succeeded by numerical number. **(d)** Diagram depicting the three-dimensional view of the Ste2 protein, with highlighted Gα-protein binding site (Gpa1) and the Gcoupler intracellular cavities (IC4 and IC5). The bar plot at the bottom depicting the percentage overlap at the amino acid levels between the Gα-binding site and predicted IC4 and IC5. **(e)** Schematic representation of the overall workflow used to predict the endogenous intracellular allosteric modulators of Ste2 receptor using Gcoupler and molecular docking technique. Of note, Yeast Metabolome DataBase (YMDB) metabolites were used as query compounds. **(f)** Overlapping density plots depicting and comparing the distributions of synthetic compounds predicted to target the IC4 and IC5 of the Ste2 receptor using the Gcoupler package. Of note, the Authenticator module of Gcoupler segregated the synthesized compound for each cavity (IC4 or IC5) into High-Affinity Binders (HAB) and Low-Affinity Binders (LAB). **(g)** AUC (Area under the curve) plots representing the performance of the indicated models. Notably, the models were trained using the cavity-specific synthetic compounds generated using the Gcoupler package. **(h)** Scatter plots depicting the relationship (correlation) between the binding prediction probabilities using Gcoupler and binding free energies computed using molecular docking (AutoDock). **(i)** Heatmap illustrating the relative enrichment of the indicated functional groups (RNH2: primary amine, R2NH: secondary amine, R3N: tertiary amine, ROPO3: monophosphate, ROH: alcohol, RCHO: aldehyde, RCOR: ketone, RCOOH: carboxylic acid, RCOOR: ester, ROR: ether, RCCH: terminal alkyne, RCN: nitrile) among the endogenous intracellular allosteric modulators (metabolites) of Ste2 receptor across IC4 and IC5. **(j)** Scatterplot depicting the Pathway Over Representation Analysis (ORA) results of the endogenous metabolites that were predicted to bind to the GPCR-Gα-protein (Ste2p-Gpa1p) interface using both Gcoupler and molecular docking. Structures of the top enriched endogenous metabolites are depicted on the right.

Figure 3: Metabolomics and metabolic mutant screening unveil the link between metabolism and programmed cell death.

(a) Schematic diagram depicting the workflow opted to narrow down on the single metabolic gene mutants. **(b)** Screenshot of the central yeast metabolism from the Kyoto Encyclopedia of Genes and Genomes (KEGG) indicating the key metabolic pathways (highlighted in red) harboring metabolic intermediates predicted as high-affinity binders of GPCR-Gα-protein (Ste2p-Gpa1p) interface by Gcoupler. **(c)** Schematic diagram depicting the experimental design used to screen single metabolic gene mutants for α-factor-induced programmed cell death. Cell viability was assessed using Propidium iodide-based fluorometric assay. **(d)** Scatter plot depicting the impact of α-factor stimuli on cellular viability, assessed using Propidium iodide-based fluorometric assay. The y-axis represents $-\log_{10}$(p-value) of the one-sample student's t-test between the normalized PI fluorescence of untreated and treated conditions. The x-axis represents the percentage inhibition or increase in cellular viability, estimated using Propidium Iodide-based assay. The mutants reported to be involved in mating, apoptosis, or both are indicated in orange, green, and blue, respectively. **(e)** Schematic diagram depicting the experimental workflow of the untargeted metabolomics experiment. Notably, the experiment involves the usage of increasing doses of α-factor to induce programmed cell death and to enrich the survival cells selectively. **(f)** Heatmap depicting the relative enrichment/de-enrichment of differentially enriched metabolites in the indicated conditions. Of note, four biological replicates per condition were used in the untargeted metabolomics. **(g)** Venn diagram depicting the overlap between the predicted endogenous intracellular allosteric modulators of Ste2p and differentially enriched metabolites (DEMs) identified using untargeted metabolomics. **(h)** Mean-whisker plot depicting the relative abundance of ubiquinone 6 (CoQ6) in the indicated conditions. Student's t-test was used to compute statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. **(i)** Mean-whisker plot depicting the relative abundance of zymosterol (ZST) in the indicated conditions. Student's t-test was used to compute statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance.

Figure 4: Exogenous metabolites treatment resists α-factor-induced programmed cell death.

(a) Schematic diagram (left) depicting the workflow opted to monitor α-factor-induced programmed cell death in the wild-type yeast cells pre-loaded for 24 hours with ubiquinone 6 (CoQ6), zymosterol (ZST), and lanosterol (LST) at three different concentrations. Changes

in the programmed cell death were quantified using the area under the curve (AUC) of the growth kinetics with respect to their respective controls. Box plot on the right depicting the rescue from the α-factor-induced programmed cell death in the indicated conditions (n=8 or 9 biological replicates). The y-axis represents the fold change of the AUC values with respect to their respective controls. Mann Whitney U test was used to calculate statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. **(b)** The schematic diagram depicting the experimental workflow used to quantify α-factor-induced programmed cell death using a propidium iodide-based cell viability fluorometric assay. **(c)** Box plot depicting the rescue from the α-factor-induced programmed cell death in the indicated conditions as inferred using propidium iodide-based cell viability fluorometric assay (n=9 or 10 biological replicates; Heatkilled = 2). The y-axis represents the fold change of the propidium iodide fluorescence values with respect to their respective controls. Mann Whitney U test was used to calculate statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. **(d)** Schematic representation of the experimental approach used to measure cell vitality and viability using microscopy-based FUN™1 staining. Scale 10 μm. **(e)** Mean-whisker plot depicting the relative proportion of the vital and viable yeast cells observed using FUN™1 staining in the indicated conditions (n = 3 biological replicates). A Student's t-test was used to compute statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. Error bars represent the standard error of mean (SEM). **(f)** Schematic representation (left) of the experimental design for the mating assay (n = 3 biological replicates, each with three technical replicates). Notably, MATa yeast was pre-loaded with the metabolites and then mated with MATα cells to evaluate the mating efficiency. Representative micrographs in the middle qualitatively depicting the mating efficiency in the indicated conditions. Bar plots on the right depicting the mating efficiency (mean ± SEM) in the indicated conditions. Student's t-test was used to compute statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. **(g)** Workflow depicting the steps involved in the molecular dynamics simulation of the Ste2p-metabolite docked complex using the AMBER software suite. **(h)** Barplots depicting the binding energies obtained by the docking of Ste2p and indicated metabolites across IC4 and IC5. **(i)** Line plot depicting the changes in the Root Mean Square Deviation (RMSD) over simulation timeframes. Notably, RMSD is provided in Angstroms (Å), whereas the simulation time is in nanoseconds (ns). **(j)** Workflow depicting the steps involved in Ste2p-miniG-protein docking using HADDOCK and PRODIGY web servers. **(k)** Barplots depicting the fold change of the dissociation constant (K_d) in the indicated conditions. Notably, fold change was computed with respect to the wild-type condition (Ste2p-miniG-protein). **(l)** Schematic representation of the experimental workflow followed to deduce the impact of indicated

metabolites treatment on isoproterenol (ISO)induced, GPCR-mediated hypertrophy response in human (AC16) and neonatal rat cardiomyocytes. Notably, in the case of AC16 cells, Wheat germ agglutinin (WGA) was used to stain the cardiomyocytes, whereas for neonatal cardiomyocytes alpha-sarcomeric actinin staining was used. **(m)** Micrographs depicting the human (above; green colored) and neonatal rat (below; red colored) cardiomyocytes in the indicated conditions. Scale 50 μm. **(n)** Box plot depicting the surface area of human (AC16) and neonatal rat cardiomyocytes in the indicated conditions. Statistical significance of indicated metabolites with untreated control and isoproterenol-treated conditions are indicated in green and grey text respectively. Mann Whitney U test with Bonferroni corrected p-values was used to compute statistical significance.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Development of Gcoupler package.

(a) Flowchart depicting the entire workflow used to compute and compare the runtime using AutoDock and Gcoupler. Of note, the three major modules of Gcoupler, i.e., Synthesizer, Authenticator, and Generator, are indicated alongside their key processes. **(b)** Ribbon diagram depicting the structure of ADRA1A predicted structure using Alphafold (ID: AF-P35348). **(c)** Boxplot depicting the performance metrics of the 10-fold cross-validation of the training data obtained using Gcoupler. **(d)** Boxplot depicting the performance metrics of the 10-fold cross-validation of the testing data obtained using Gcoupler. **(e)** Scatterplot depicting the relationship between AutoDock computed binding energies and Gcoupler computed prediction probabilities of the experimentally validated ligands of ADRA1A protein. **(f)** Heatmap depicting the RMSD values obtained by comparing all the GPCR-Gα-protein interfaces of the available human GPCRs from the protein databank. **(g)** Heatmap depicting the RMSD values obtained by comparing the selected human GPCRs available in the protein databank.

Figure S2: Ste2 protein cavities prediction using Gcoupler

(a) Flowchart depicting the key steps used for the molecular dynamics simulation of the Ste2p using CHARMM-GUI and GROMACS software. **(b)** Line plot depicting the changes in the density (kg/m³) along the z-axis of the three-dimensional Ste2p structure. **(c)** Line plot depicting the decline in the potential (kJ/mol) over molecular dynamics simulation time in picoseconds (ps). **(d)** Line plot depicting the changes in the density (kg/m³) of the overall system over molecular dynamics simulation time in picoseconds (ps). **(e)** Line plots depicting the changes in the pressure (in bars) and temperature (in Kelvin) of the overall system during molecular dynamics simulation of Ste2 protein. **(f)** Line plot depicting the Root Mean Square Deviation (RMSD) changes of the Ste2 protein during molecular dynamics simulation. Note the RMSD values are indicated in nanometers (nm). **(g)** Ramachandran plot depicting the location of simulated Ste2 protein residues in the favorable and unfavorable regions. **(h)** Barplot depicting the predicted cavities' volumes (Å). **(i)** Scatterplot depicting the relationship between the cavity volume and surface area in the indicated cavities of the Ste2 protein. **(j)** Heatmap depicting the percentage of amino acids of the indicated cavities residing in the EC (extracellular), IC (intracellular), and TM (transmembrane) region of the Ste2 protein. **(k)** Scatterplot depicting the predicted drug score of the indicated cavities of the Ste2 protein. **(l)** Snake plot depicting the conservation of Ste2 protein at the amino acid

levels. Of note, Ste2 proteins from 15-related yeast species were used for computing the conservation score. **(m)** Snake plot depicting the key amino acids of the Ste2 protein alongside their location within the indicated functional cavities. **(n)** Heatmap illustrating the relative enrichment of the indicated functional groups (RNH2: primary amine, R2NH: secondary amine, R3N: tertiary amine, ROPO3: monophosphate, ROH: alcohol, RCHO: aldehyde, RCOR: ketone, RCOOH: carboxylic acid, RCOOR: ester, ROR: ether, RCCH: terminal alkyne, RCN: nitrile) among the Gcoupler-based *de novo* synthesized binders of Ste2p receptor across IC4 and IC5. **(o)** Venn diagrams depicting the number of yeast metabolites from YMDB predicted to bind to the Ste2p-Gpa1p interface, predicted using Gcoupler and AutoDock. Of note, IC4 and IC5 represent intracellular cavities 4 and 5 of the Ste2 protein, respectively.

Figure S3: Gcoupler predicted the high-affinity endogenous intracellular allosteric modulators of Ste2 protein.

(a) Principal Component Analysis depicting chemical heterogeneity of synthetic compounds synthesized using Gcoupler across intracellular cavities (IC4 and IC5). Notably, the compounds were segregated into High-Affinity Binders (HAB) and Low-Affinity Binders (LAB) by the Authenticator module of the Gcoupler. **(b)** Heatmap depicting the chemical similarity of the *de novo* synthesized HAB (dark green) and LAB (light green) across intracellular cavities (IC4 and IC5) of the Ste2 protein. Tanimoto similarities were computed using Atomic fingerprints. **(c)** Box plots depicting the model's performances across the 10-fold cross-validation. The upper and lower box plots depicting the performance metrics of models trained using the *de novo* ligands for intracellular cavities (IC4 and IC5), respectively of the Ste2 protein. **(d)** Principal Component Analysis depicting chemical heterogeneity of endogenous yeast metabolites predicted to bind to the intracellular cavities (IC4 and IC5) of the Ste2 protein. Notably, the compounds were segregated into High-Affinity Metabolites (HAM) and Low-Affinity Metabolites (LAM) by the Generator module of the Gcoupler. **(e)** Percentage stacked barplot depicting the relative enrichment of the indicated functional groups of the High-Affinity Metabolites (HAM) and Low-Affinity Metabolites (LAM) across intracellular cavities (IC4 and IC5) of the Ste2 protein.

Figure S4: Untargeted metabolomics and genetic screening suggest an interlink between metabolites and Ste2p-mediated Programmed Cell Death.

(a) Venn diagram depicting the percentage overlaps of the mating-related genes reported in the literature and the ones used in the screening. The *STE2* was the only common gene. **(b)**

Heatmap depicting the growth curve profiles of the wild-type, single metabolic mutants, and *ste2Δ* under optimal growth conditions. **(c)** Boxplot depicting the increase in the propidium iodide fluorescence upon the increasing concentration of α-factor treatment. Heat-Killed (HK), untreated, and vehicle (DMSO-treated) were used as controls, whereas the increasing concentration of α-factor survived wild-type cells were used as test conditions. Mann Whitney U test was used to calculate statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. **(d)** Principal Component Analysis depicting the segregation of indicated samples based on their metabolomics profiles along the PC1, PC2, and PC3. **(e)** Corrplot depicting Pearson's correlation of the metabolome of the indicated samples. **(f)** Heatmap depicting the relative enrichment of all the detected metabolites in the indicated condition. Heat-Killed (HK), untreated, and vehicle (DMSO-treated) were used as controls, whereas the increasing concentration of α-factor survived wild-type cells were used as test conditions. **(g)** Venn diagram indicating that 38 predicted intracellular allosteric modulators of Ste2p were identified in the untargeted metabolomics profiling. **(h)** Scatterplot depicting the Pathway Over Representation Analysis (ORA) results of the differentially enriched metabolites (treated vs vehicle control) identified using untargeted metabolomics.

Figure S5: Exogenous supplementation of selective metabolites rescues α-factor-mediated Programmed Cell Death.

(a) Growth curve profiles of the wild-type (control), *ste2Δ,* and wild-type yeast pre-treated with the indicated metabolites. **(b)** Beanplots depicting the changes in the α-factor-mediated programmed cell death in the indicated conditions. Notably, cell viability was assessed by computing the area under the curve (AUC) of the growth kinetics. Mann Whitney U test was used to calculate statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. **(c)** Beanplot depicting the fold change in the α-factor-mediated programmed cell death, assessed using the propidium iodide-mediated programmed cell death in the indicated conditions. Mann Whitney U test was used to calculate statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance. **(d)** Representative micrographs (left) depicting the growth of the cells pretreated with indicated metabolites in a single drop-out medium (SC-Lys and SC-Met). Quantifications of the mating response are indicated as a bar graph (mean ± SEM) on the right (n = 3 biological replicates, each with three technical replicates). Student's t-test was used to compute statistical significance. Asterisks indicate statistical significance, whereas ns represents non-significance.

Figure S6: Docking and Molecular Dynamics simulations suggest the stability of the Ste2-metabolite complex.

(a) Ligplots depicting the interacting amino acid residues and atoms of the Ste2 protein and indicated metabolites, for the intracellular cavities (IC4 and IC5). **(b)** Table indicating the Van der Waals (ΔE_{vdw}), electrostatic (ΔE_{elec}), polar solvation (ΔE_{psolv}), and non-polar solvation (ΔE_{npsolv}) energies along with net binding free-energy (ΔG_{bind}) in kcal/mol of the indicated metabolites across intracellular cavities IC4 and IC5. **(c)** Micrographs depicting the structures of the Ste2 protein (blue) and the miniGpa1-protein (cyan), with the highlighted interface residues (red and green) in the indicated conditions. **(d)** Scatterplot depicting the net binding free-energy ΔG (kcal/mol) of the Ste2p (with and without indicated metabolites) with miniG-protein (y-axis) and α-factor (x-axis). **(e)** Barplot depicting RMSD values of the Ste2 protein with indicated metabolite across intracellular cavities IC4 and IC5. **(f)** Micrographs depicting the topology of human and rat ADRB1 and ADRB2 receptors. GPCR-Gα interface cavities are color-coded at the intracellular sites. Of note, in the case of human ADRB1 the nomenclature of the two cavities detected at the interface includes 'Cv' (cavity) succeeded by the numerical number. **(g-j)** Barplots depicting the binding energies obtained by the docking of Human ADRB1 (7BVQ), Human ADRB2 (5X7D), Rat ADRB1 (P18090), and Rat ADRB2 (P10608) with the indicated metabolites at the GPCR-Gα interfaces.

SUPPLEMENTARY DATA

Supplementary Table 1: Table contains information about the structure information about the human GPCR-Gα-Protein complexes downloaded from Protein Data Bank (PDB). It also contains information about the PDB IDs, Protein name, species information, whether it is used in this study or not, UniProt IDs, and other information.

Supplementary Table 2: The table contains information about the normalized Root Mean Square Deviation (RMSD) of the pairwise comparison of the GPCR-Gα-Protein interface (cavities) for the indicated GPCRs.

Supplementary Table 3: The table contains information about the chemical similarities of the *de novo* synthesized synthetic compounds of the indicated GPCR-Gα-Protein interface (cavities). Of note, the chemical similarity was computed using the Taniomoto coefficient on the Atomic fingerprints.

Supplementary Table 4: The table contains information about the YMDB metabolites predicted to bind at the intracellular cavities IC4 and IC5 of the Ste2 protein. Of note, predictions were made using molecular docking and Gcoupler. It contains information about the YMDB ID, Metabolite name, SMILES (Simplified Molecular Input Line Entry System), and cavities information. Of note, BE refers to binding energies, whereas probabilities were the outcome of Gcoupler.

Supplementary Table 5: The table contains information about the list of metabolite-gene pairs from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway that is used in this study.

Supplementary Table 6: The table contains the results of the untargeted metabolomics. Of note, the values represent normalized and imputed peak intensities of the indicated metabolites.

Supplementary Table 7: The table contains information about the known mutants from the literature that could modulate Ste2 signaling. Of note, the table also contains information about the site of mutation, the nature of the mutation, its position within the intracellular cavities IC4 and IC5, their impact on the Ste2 signaling pathway, and references.

REFERENCES

1. Yang, D. *et al.* G protein-coupled receptors: structure- and function-based drug discovery. *Signal Transduction and Targeted Therapy* **6**, 1–27 (2021).
2. Alhosaini, K., Azhar, A., Alonazi, A. & Al-Zoghaibi, F. GPCRs: The most promiscuous druggable receptor of the mankind. *Saudi Pharmaceutical Journal* **29**, 539–551 (2021).
3. Calebiro, D., Koszegi, Z., Lanoiselée, Y., Miljus, T. & O'Brien, S. G protein-coupled receptor-G protein interactions: a single-molecule perspective. *Physiol. Rev.* **101**, 857–906 (2021).
4. Wold, E. A. & Zhou, J. GPCR Allosteric Modulators: Mechanistic Advantages and Therapeutic Applications. *Curr. Top. Med. Chem.* **18**, 2002–2006 (2018).
5. Bourque, K. *et al.* Exploring the use of intracellular and extracellular allosteric modulators to understand GPCR signaling. *Allosteric Modulation of G Protein-Coupled Receptors* 135–160 Preprint at https://doi.org/10.1016/b978-0-12-819771-4.00008-7 (2022).
6. Rees, S., Morrow, D. & Kenakin, T. GPCR Drug Discovery Through the Exploitation of Allosteric Drug Binding Sites. *Receptors and Channels* vol. 8 261–268 Preprint at https://doi.org/10.3109/10606820214640 (2002).
7. Leach, K., Sexton, P. M. & Christopoulos, A. Allosteric GPCR modulators: taking advantage of permissive receptor pharmacology. *Trends Pharmacol. Sci.* **28**, 382–389 (2007).
8. van der Westhuizen, E. T., Valant, C., Sexton, P. M. & Christopoulos, A. Endogenous Allosteric Modulators of G Protein–Coupled Receptors. *J. Pharmacol. Exp. Ther.* **353**, 246–260 (2015).
9. Stornaiuolo, M. *et al.* Endogenous vs Exogenous Allosteric Modulators in GPCRs: A dispute for shuttling CB1 among different membrane microenvironments. *Sci. Rep.* **5**, 15453 (2015).

10. Doller, D. Endogenous Allosteric Modulators of G Protein-Coupled Receptors: Implications in Drug Design. *2017 Medicinal Chemistry Reviews* 343–360 Preprint at https://doi.org/10.29200/acsmedchemrev-v52.ch18 (2017).
11. Reyes-Alcaraz, A., Lucero Garcia-Rojas, E. Y., Bond, R. A. & McConnell, B. K. Allosteric Modulators for GPCRs as a Therapeutic Alternative with High Potential in Drug Discovery. *Molecular Pharmacology* Preprint at https://doi.org/10.5772/intechopen.91838 (2020).
12. Zhang, P., Covic, L. & Kuliopulos, A. Pepducins and Other Lipidated Peptides as Mechanistic Probes and Therapeutics. *Methods Mol. Biol.* **1324**, 191–203 (2015).
13. O'Callaghan, K., Kuliopulos, A. & Covic, L. Turning receptors on and off with intracellular pepducins: new insights into G-protein-coupled receptor drug development. *J. Biol. Chem.* **287**, 12787–12796 (2012).
14. Basith, S. *et al.* Exploring G Protein-Coupled Receptors (GPCRs) Ligand Space via Cheminformatics Approaches: Impact on Rational Drug Design. *Front. Pharmacol.* **9**, 128 (2018).
15. Gupta, A. *et al.* Machine-OlF-Action: A unified framework for developing and interpreting machine-learning models for chemosensory research. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btaa1104.
16. Zhou, Y., Meng, J., Xu, C. & Liu, J. Multiple GPCR Functional Assays Based on Resonance Energy Transfer Sensors. *Front Cell Dev Biol* **9**, 611443 (2021).
17. Hoffmann, C. & Bünemann, M. Fluorescence and Resonance Energy Transfer Shine New Light on GPCR Function. *GPCR Molecular Pharmacology and Drug Targeting* 226–251 Preprint at https://doi.org/10.1002/9780470627327.ch10 (2010).
18. Jaeger, W. C., Pfleger, K. D. G. & Eidne, K. A. Monitoring GPCR-Protein Complexes Using Bioluminescence Resonance Energy Transfer. *G Protein-Coupled Receptors* 111–132 Preprint at https://doi.org/10.1002/9780470749210.ch6.
19. Congreve, M., de Graaf, C., Swain, N. A. & Tate, C. G. Impact of GPCR Structures on Drug Discovery. *Cell* **181**, 81–91 (2020).

20. Topiol, S. Current and Future Challenges in GPCR Drug Discovery. *Methods Mol. Biol.* **1705**, 1–21 (2018).
21. Spiegel, J. O. & Durrant, J. D. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. *J. Cheminform.* **12**, 25 (2020).
22. Sicho, M., Liu, X., Svozil, D. & van Westen, G. J. P. GenUI: interactive and extensible open source software platform for de novo molecular generation and cheminformatics. *J. Cheminform.* **13**, 73 (2021).
23. Li, Y., Pei, J. & Lai, L. Structure-based de novo drug design using 3D deep generative models. *Chem. Sci.* **12**, 13664–13675 (2021).
24. Böhm, H. J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.* **6**, 61–78 (1992).
25. Nishibata, Y. & Itai, A. Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron* **47**, 8985–8990 (1991).
26. Pearlman, D. A. & Murcko, M. A. CONCEPTS: New dynamic algorithm forde novo drug suggestion. *J. Comput. Chem.* **14**, 1184–1193 (1993).
27. Yuan, Y., Pei, J. & Lai, L. LigBuilder 2: a practical de novo drug design approach. *J. Chem. Inf. Model.* **51**, 1083–1091 (2011).
28. Nakayama, N., Kaziro, Y., Arai, K. & Matsumoto, K. Role of STE genes in the mating factor signaling pathway mediated by GPA1 in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **8**, 3777–3783 (1988).
29. Alvaro, C. G. & Thorner, J. Heterotrimeric G Protein-coupled Receptor Signaling in Yeast Mating Pheromone Response. *J. Biol. Chem.* **291**, 7788–7795 (2016).
30. Blumer, K. J., Reneke, J. E. & Thorner, J. The STE2 gene product is the ligand-binding component of the alpha-factor receptor of Saccharomyces cerevisiae. *J. Biol. Chem.* **263**, 10836–10842 (1988).
31. Büttner, S. *et al.* Why yeast cells can undergo apoptosis: death in times of peace, love, and war. *J. Cell Biol.* **175**, 521–525 (2006).

32. Severin, F. F. & Hyman, A. A. Pheromone induces programmed cell death in S. cerevisiae. *Curr. Biol.* **12**, R233–5 (2002).
33. Zhang, N.-N. *et al.* Multiple signaling pathways regulate yeast cell death during the response to mating pheromones. *Mol. Biol. Cell* **17**, 3409–3422 (2006).
34. Carmona-Gutierrez, D. *et al.* Apoptosis in yeast: triggers, pathways, subroutines. *Cell Death Differ.* **17**, 763–773 (2010).
35. Sokolov, S. S., Galkina, K. V., Litvinova, E. A., Knorre, D. A. & Severin, F. F. The Role of LAM Genes in the Pheromone-Induced Cell Death of S. cerevisiae Yeast. *Biochemistry* **85**, 300–309 (2020).
36. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162**, 1239–1249 (2011).
37. Schöneberg, T., Hofreiter, M., Schulz, A. & Römpler, H. Learning from the past: evolution of GPCR functions. *Trends Pharmacol. Sci.* **28**, 117–121 (2007).
38. Magalhaes, A. C., Dunn, H. & Ferguson, S. S. Regulation of GPCR activity, trafficking and localization by GPCR-interacting proteins. *Br. J. Pharmacol.* **165**, 1717–1736 (2012).
39. Irwin, J. J. & Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
40. Shivanyuk, A. N., Ryabukhin, S. V. & Tolmachev, A. Enamine real database: Making chemical diversity real. *Chemistry today* (2007).
41. Yuan, Y., Pei, J. & Lai, L. LigBuilder V3: A Multi-Target de novo Drug Design Approach. *Front Chem* **8**, 142 (2020).
42. Berger, V. W. & Zhou, Y. Kolmogorov–Smirnov Test: Overview. *Wiley StatsRef: Statistics Reference Online* Preprint at https://doi.org/10.1002/9781118445112.stat06558 (2014).
43. Goerg, S. J. & Kaiser, J. Nonparametric Testing of Distributions—the Epps–Singleton Two-Sample Test using the Empirical Characteristic Function. *Stata J.* **9**, 454–465 (2009).
44. Engmann, S. & Cousineau, D. Comparing distributions: the two-sample

Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. *Journal of applied quantitative methods* **6**, (2011).

45. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
46. Ramsundar, B. *et al.* DeepChem: Democratizing Deep-Learning for Drug Discovery, Quantum Chemistry. *Materials Science and Biology. https://github. com/deepchem/deepchem (accessed Aug 8, 2017)* (2019).
47. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
48. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–7 (2012).
49. Jewison, T. *et al.* YMDB: the Yeast Metabolome Database. *Nucleic Acids Res.* **40**, D815–20 (2012).
50. Velazhahan, V. *et al.* Structure of the class D GPCR Ste2 dimer coupled to two G proteins. *Nature* **589**, 148–153 (2021).
51. Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).
52. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
53. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **44**, D471–80 (2016).
54. Van Zandycke, S. M., Simal, O., Gualdoni, S. & Smart, K. A. Determination of Yeast Viability Using Fluorophores. *J. Am. Soc. Brew. Chem.* **61**, 15–22 (2003).
55. Teng, X. & Hardwick, J. M. Reliable method for detection of programmed cell death in yeast. *Methods Mol. Biol.* **559**, 335–342 (2009).
56. Guedes, J. P. *et al.* Acetic acid triggers cytochrome c release in yeast heterologously

expressing human Bax. *Apoptosis* **27**, 368–381 (2022).

57. Martins, V. M. *et al.* Contacts in Death: The Role of the ER-Mitochondria Axis in Acetic Acid-Induced Apoptosis in Yeast. *J. Mol. Biol.* **431**, 273–288 (2019).
58. Bardwell, L. A walk-through of the yeast mating pheromone response pathway. *Peptides* **26**, 339–350 (2005).
59. Hartwell, L. H. Mutants of Saccharomyces cerevisiae unresponsive to cell division control by polypeptide mating hormone. *J. Cell Biol.* **85**, 811–822 (1980).
60. Nachtergaele, S. *et al.* Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nat. Chem. Biol.* **8**, 211–220 (2012).
61. Hedderich, J. B. *et al.* The pocketome of G-protein-coupled receptors reveals previously untargeted allosteric sites. *Nat. Commun.* **13**, 2567 (2022).
62. Chatzigoulas, A. & Cournia, Z. Rational design of allosteric modulators: Challenges and successes. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**, (2021).
63. Bartuzi, D., Kaczor, A. A. & Matosiuk, D. Opportunities and Challenges in the Discovery of Allosteric Modulators of GPCRs. *Methods Mol. Biol.* **1705**, 297–319 (2018).
64. Hou, T., Bian, Y., McGuire, T. & Xie, X.-Q. Integrated Multi-Class Classification and Prediction of GPCR Allosteric Modulators by Machine Learning Intelligence. *Biomolecules* **11**, (2021).
65. Topiol, S. New opportunities for GPCR allosteric modulators. *Future Medicinal Chemistry* vol. 10 707–710 Preprint at https://doi.org/10.4155/fmc-2017-0313 (2018).
66. Horn, F. *et al.* GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.* **31**, 294–297 (2003).
67. Bagnat, M. & Simons, K. Cell surface polarization during yeast mating. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14183–14188 (2002).
68. Jin, H., McCaffery, J. M. & Grote, E. Ergosterol promotes pheromone signaling and plasma membrane fusion in mating yeast. *J. Cell Biol.* **180**, 813–826 (2008).
69. Heese-Peck, A. *et al.* Multiple functions of sterols in yeast endocytosis. *Mol. Biol. Cell* **13**, 2664–2680 (2002).

70. Aguilar, P. S. *et al.* Structure of sterol aliphatic chains affects yeast cell shape and cell fusion during mating. *Proceedings of the National Academy of Sciences* **107**, 4170–4175 (2010).
71. Tiedje, C., Holland, D. G., Just, U. & Höfken, T. Proteins involved in sterol synthesis interact with Ste20 and regulate cell polarity. *J. Cell Sci.* **120**, 3613–3624 (2007).
72. Shaw, W. M. *et al.* Engineering a Model Cell for Rational Tuning of GPCR Signaling. *Cell* **177**, 782–796.e27 (2019).
73. Ostrov, N. *et al.* A modular yeast biosensor for low-cost point-of-care pathogen detection. *Sci Adv* **3**, e1603221 (2017).
74. Johnston, E. J., Moses, T. & Rosser, S. J. The wide-ranging phenotypes of ergosterol biosynthesis mutants, and implications for microbial cell factories. *Yeast* **37**, 27–44 (2020).
75. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
76. Kipf, T. N. & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv [cs.LG]* (2016).
77. Xiong, Z. *et al.* Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *J. Med. Chem.* **63**, 8749–8760 (2020).
78. Veličković, P. *et al.* Graph Attention Networks. *https://openreview.net › forumhttps://openreview.net › forum* (2018).
79. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
80. O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
81. Dallakyan, S. MGLTools. *Reference Source*.
82. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

83. Battle, G. M. PDBePISA: Identifying and interpreting the likely biological assemblies of a protein structure. *Chemistry*.

84. Moriwaki, H., Tian, Y.-S., Kawashita, N. & Takagi, T. Mordred: a molecular descriptor calculator. *J. Cheminform.* **10**, 4 (2018).

85. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).

86. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).

87. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

88. Sprenger, K. G., Jaeger, V. W. & Pfaendtner, J. The general AMBER force field (GAFF) can accurately predict thermodynamic and transport properties of many ionic liquids. *J. Phys. Chem. B* **119**, 5882–5895 (2015).

89. Tian, C. *et al.* ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **16**, 528–552 (2020).

90. Wang, H. & Yang, W. Force Field for Water Based on Neural Network. *J. Phys. Chem. Lett.* **9**, 3232–3240 (2018).

91. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).

92. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).

93. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electronica* (2001).

94. Core Team, R. R: A language and environment for statistical computing. Version 3.6. 0. Vienna, Austria. */ra-language-and-environment-forstatistical-computing*.

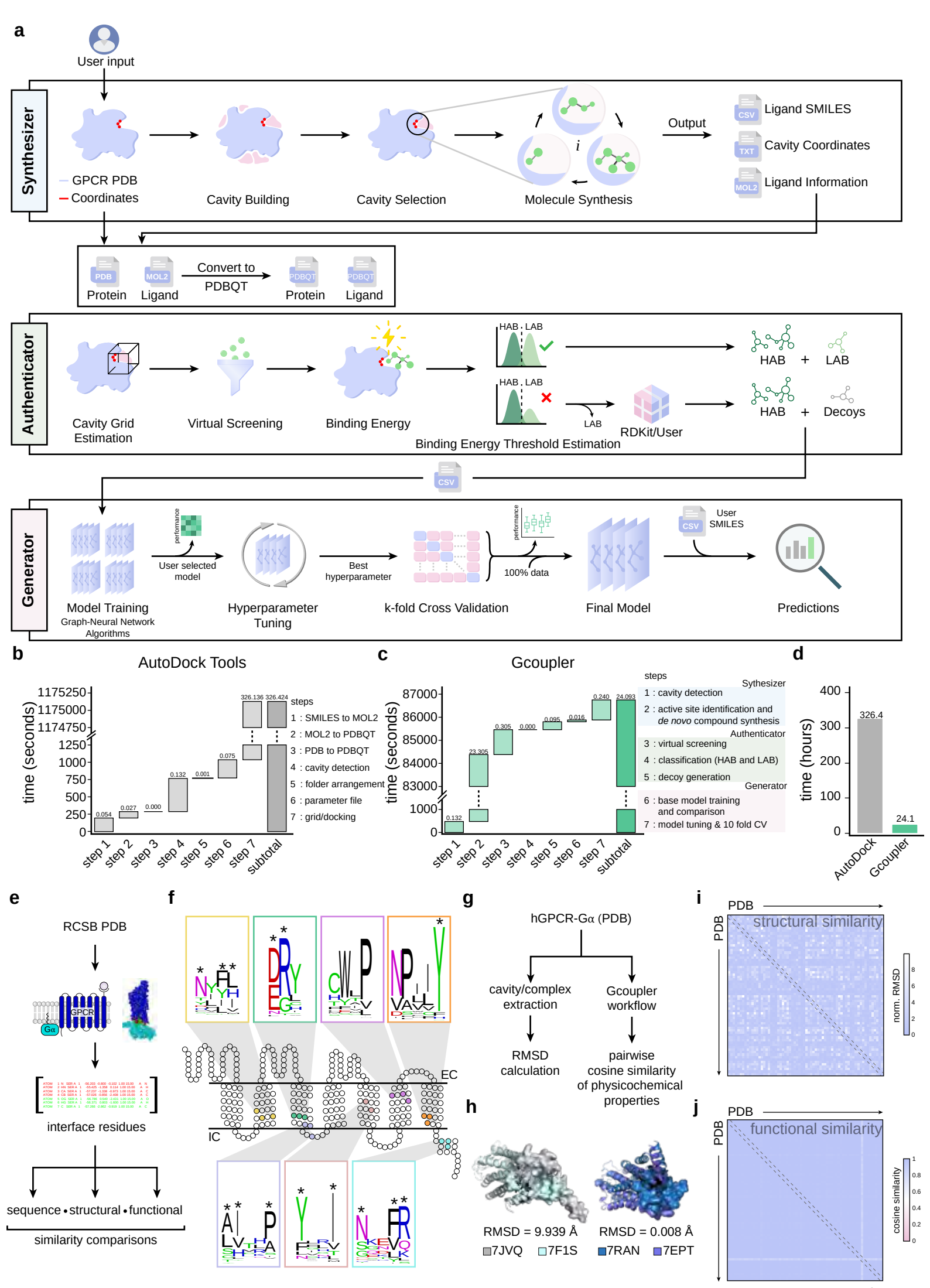
a
User input
Synthesizer
GPCR PDB
Coordinates
Cavity Building
Cavity Selection
Molecule Synthesis
Output
Ligand SMILES
Cavity Coordinates
Ligand Information
Protein
Ligand
Convert to PDBQT
Authenticator
Cavity Grid Estimation
Virtual Screening
Binding Energy
HAB
LAB
RDKit/User
Decoys
Binding Energy Threshold Estimation
Generator
Model Training
Graph-Neural Network Algorithms
User selected model
Hyperparameter Tuning
Best hyperparameter
k-fold Cross Validation
100% data
Final Model
User SMILES
Predictions
b
AutoDock Tools
time (seconds)
steps
1 : SMILES to MOL2
2 : MOL2 to PDBQT
3 : PDB to PDBQT
4 : cavity detection
5 : folder arrangement
6 : parameter file
7 : grid/docking
c
Gcoupler
1 : cavity detection
2 : active site identification and de novo compound synthesis
3 : virtual screening
4 : classification (HAB and LAB)
5 : decoy generation
6 : base model training and comparison
7 : model tuning & 10 fold CV
d
time (hours)
AutoDock
Gcoupler
e
RCSB PDB
interface residues
sequence • structural • functional
similarity comparisons
f
EC
IC
g
hGPCR-Gα (PDB)
cavity/complex extraction
Gcoupler workflow
RMSD calculation
pairwise cosine similarity of physicochemical properties
h
RMSD = 9.939 Å
RMSD = 0.008 Å
7JVQ
7F1S
7RAN
7EPT
i
PDB
structural similarity
norm. RMSD
j
functional similarity
cosine similarity

Figure 1

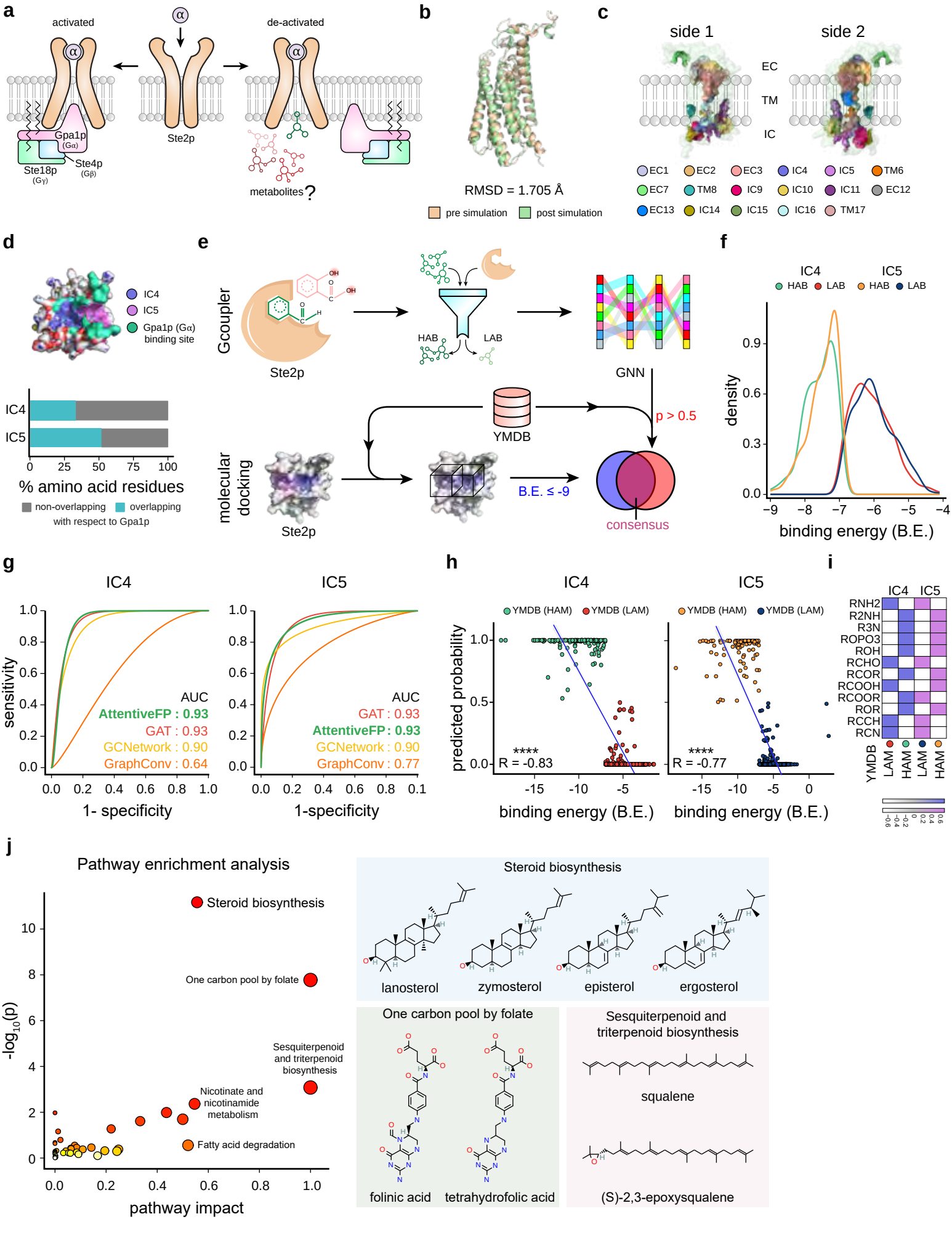
a
activated
de-activated
Gpa1p
(Gα)
Ste2p
Ste18p
(Gγ)
Ste4p
(Gβ)
metabolites?
b
RMSD = 1.705 Å
pre simulation
post simulation
c
side 1
side 2
EC
TM
IC
EC1
EC2
EC3
IC4
IC5
TM6
EC7
TM8
IC9
IC10
IC11
EC12
EC13
IC14
IC15
IC16
TM17
d
IC4
IC5
Gpa1p (Gα) binding site
% amino acid residues
non-overlapping
overlapping
with respect to Gpa1p
e
Gcoupler
Ste2p
HAB
LAB
GNN
YMDB
p > 0.5
molecular docking
B.E. ≤ -9
consensus
f
IC4
IC5
HAB
LAB
density
binding energy (B.E.)
g
IC4
IC5
sensitivity
1- specificity
AUC
AttentiveFP : 0.93
GAT : 0.93
GCNetwork : 0.90
GraphConv : 0.64
GAT : 0.93
AttentiveFP : 0.93
GCNetwork : 0.90
GraphConv : 0.77
h
YMDB (HAM)
YMDB (LAM)
predicted probability
R = -0.83
R = -0.77
binding energy (B.E.)
i
RNH2
R2NH
R3N
ROPO3
ROH
RCHO
RCOR
RCOOH
RCOOR
ROR
RCCH
RCN
YMDB
LAM
HAM
j
Pathway enrichment analysis
Steroid biosynthesis
One carbon pool by folate
Sesquiterpenoid and triterpenoid biosynthesis
Nicotinate and nicotinamide metabolism
Fatty acid degradation
-log10(p)
pathway impact
lanosterol
zymosterol
episterol
ergosterol
folinic acid
tetrahydrofolic acid
squalene
(S)-2,3-epoxysqualene

Figure 2

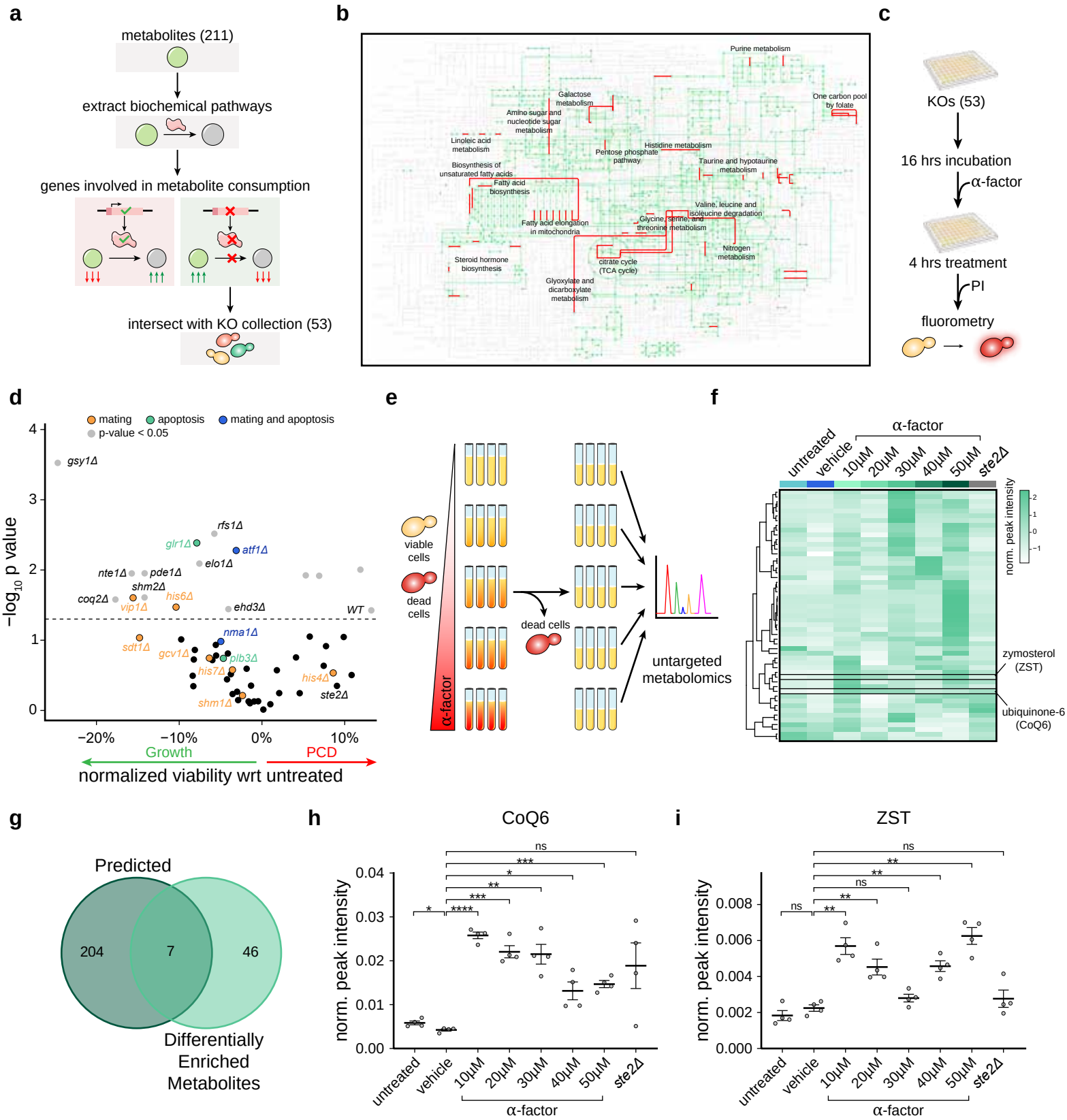
a
metabolites (211)
extract biochemical pathways
genes involved in metabolite consumption
intersect with KO collection (53)
b
Purine metabolism
Galactose metabolism
Amino sugar and nucleotide sugar metabolism
One carbon pool by folate
Linoleic acid metabolism
Histidine metabolism
Pentose phosphate pathway
Taurine and hypotaurine metabolism
Biosynthesis of unsaturated fatty acids
Fatty acid biosynthesis
Valine, leucine and isoleucine degradation
Fatty acid elongation in mitochondria
Glycine, serine, and threonine metabolism
Nitrogen metabolism
Steroid hormone biosynthesis
citrate cycle (TCA cycle)
Glyoxylate and dicarboxylate metabolism
c
KOs (53)
16 hrs incubation
α-factor
4 hrs treatment
PI
fluorometry
d
mating
apoptosis
mating and apoptosis
p-value < 0.05
−log10 p value
gsy1Δ
rfs1Δ
glr1Δ
atf1Δ
elo1Δ
nte1Δ
pde1Δ
shm2Δ
coq2Δ
his6Δ
vip1Δ
ehd3Δ
WT
nma1Δ
sdt1Δ
gcv1Δ
plb3Δ
his7Δ
his4Δ
ste2Δ
shm1Δ
−20%
−10%
0%
10%
Growth
PCD
normalized viability wrt untreated
e
viable cells
dead cells
α-factor
dead cells
untargeted metabolomics
f
α-factor
untreated
vehicle
10μM
20μM
30μM
40μM
50μM
ste2Δ
norm. peak intensity
zymosterol (ZST)
ubiquinone-6 (CoQ6)
g
Predicted
204
7
46
Differentially Enriched Metabolites
h
CoQ6
norm. peak intensity
untreated
vehicle
10μM
20μM
30μM
40μM
50μM
ste2Δ
α-factor
i
ZST
norm. peak intensity
untreated
vehicle
10μM
20μM
30μM
40μM
50μM
ste2Δ
α-factor

Figure 3

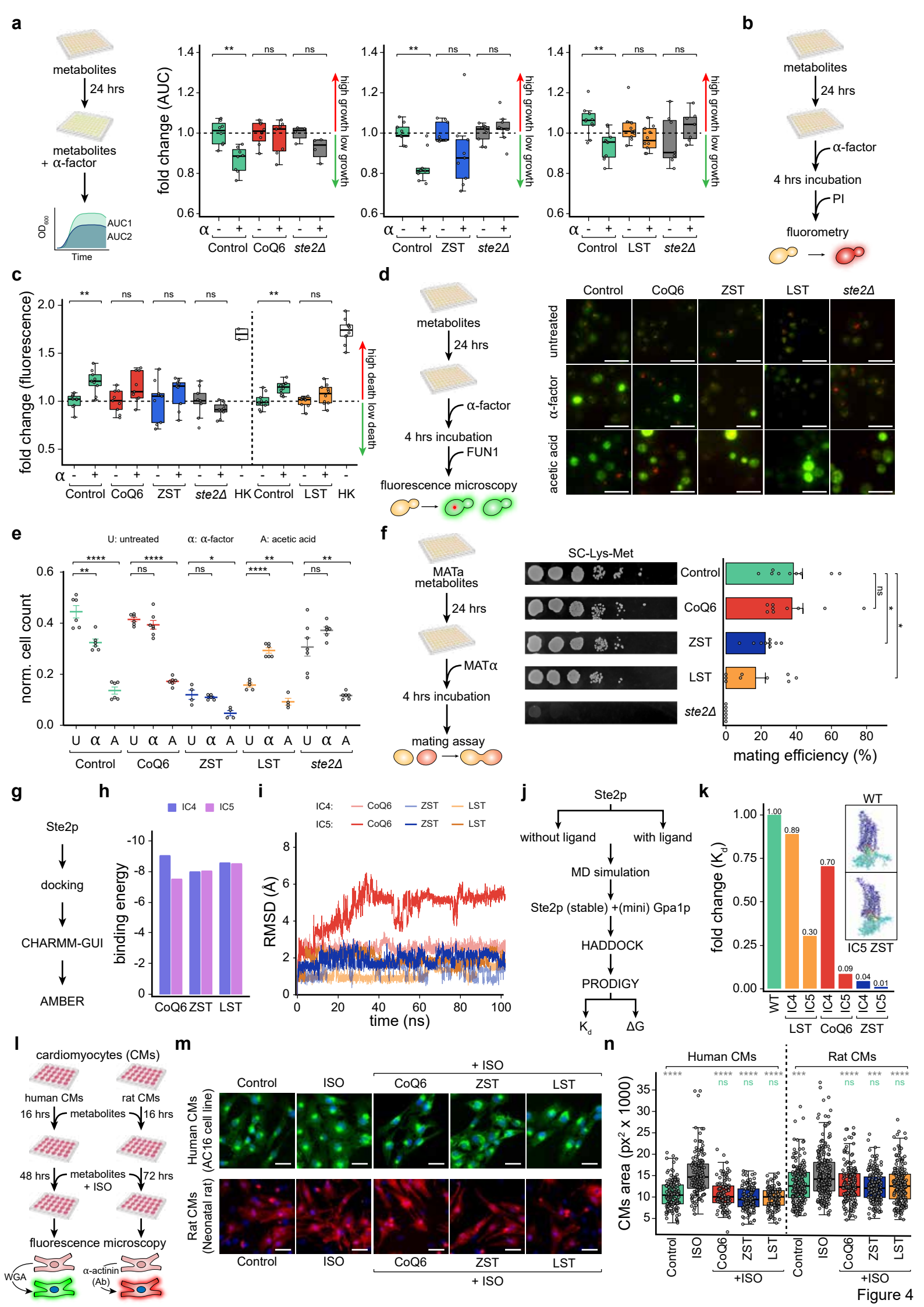
a
metabolites
24 hrs
metabolites
+ α-factor
OD600
AUC1
AUC2
Time
fold change (AUC)
high growth low growth
Control CoQ6 ste2Δ
Control ZST ste2Δ
Control LST ste2Δ
b
metabolites
24 hrs
α-factor
4 hrs incubation
PI
fluorometry
c
fold change (fluorescence)
high death low death
Control CoQ6 ZST ste2Δ HK Control LST HK
d
metabolites
24 hrs
α-factor
4 hrs incubation
FUN1
fluorescence microscopy
Control CoQ6 ZST LST ste2Δ
untreated
α-factor
acetic acid
e
U: untreated
α: α-factor
A: acetic acid
norm. cell count
Control CoQ6 ZST LST ste2Δ
f
MATa
metabolites
24 hrs
MATα
4 hrs incubation
mating assay
SC-Lys-Met
Control
CoQ6
ZST
LST
ste2Δ
mating efficiency (%)
g
Ste2p
docking
CHARMM-GUI
AMBER
h
IC4
IC5
binding energy
CoQ6 ZST LST
i
IC4: CoQ6 ZST LST
IC5: CoQ6 ZST LST
RMSD (Å)
time (ns)
j
Ste2p
without ligand
with ligand
MD simulation
Ste2p (stable) +(mini) Gpa1p
HADDOCK
PRODIGY
Kd
ΔG
k
WT
IC5 ZST
fold change (Kd)
WT IC4 IC5 IC4 IC5 IC4 IC5
LST CoQ6 ZST
l
cardiomyocytes (CMs)
human CMs
rat CMs
16 hrs
metabolites
16 hrs
48 hrs
metabolites
+ ISO
72 hrs
fluorescence microscopy
WGA
α-actinin
(Ab)
m
+ ISO
Control ISO CoQ6 ZST LST
Human CMs
(AC16 cell line)
Rat CMs
(Neonatal rat)
n
Human CMs
Rat CMs
CMs area (px² x 1000)
Control ISO CoQ6 ZST LST
+ISO

Figure 4