Title
Directed evolution of VanR biosensor specificity in yeast

Authors
D'Ambrosio, Vasil
Pramanik, Subrata
Goroncy, Kati
et al.

Publication Date
2020

DOI
10.1016/j.biotno.2020.01.002

Peer reviewed

Directed evolution of VanR biosensor specificity in yeast

Vasil D'Ambrosio[1], Subrata Pramanik[2], Kati Goroncy[3], Tadas Jakociunas[1] , David Schönauer[3], Mehdi D. Davari[2] , Ulrich Schwaneberg[2, 4] Jay D. Keasling[1, 5-8], & Michael K. Jensen[1]*

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

[2] Institute of Biotechnology, RWTH Aachen University, Worringer Weg 3, D-52074 Aachen, Germany

[3] SeSaM-Biotech GmbH, Aachen, Germany

[4] DWI − Leibniz Institute for Interactive Materials, Forckenbeckstraße 50, 52074 Aachen, Germany

[5] Joint BioEnergy Institute, Emeryville, CA, USA

[6] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[7] Department of Chemical and Biomolecular Engineering & Department of Bioengineering, University of California, Berkeley, CA, USA

[8] Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes of Advanced Technologies, Shenzhen, China

* To whom correspondence should be addressed. Michael K. Jensen: Email: mije@biosustain.dtu.dk, Tel: +45 6128 4850

Keywords

biosensor, specificity, directed evolution, vanillic acid, vanillin,

Abstract

Allosterically regulated transcription factors (aTFs) based biosensors from prokaryotes have been widely used to screen large gene libraries, stabilize engineered microbes from evolutionary drifting, and for detection of soil pollutants, among many other applications. However, even though aTF-based biosensors have been established as successful tools for bioengineering and remediation, rational engineering of aTF small molecule-specificity have so far not been demonstrated, highlighting the need for a deeper understanding of the sequence-function relationships that govern aTF allostery. Here, by combining directed evolution of a naïve library of VanR, a vanillic acid transcriptional regulator from *Caulobacter crescentus* in yeast, followed by saturation mutagenesis of selected positions we identify residues required for vanillic acid responsiveness, while at the same

time maintaining responsiveness to vanillin. Selected single-position VanR mutants show both complete repression of transcription in the absence of any ligand, complete loss of vanillic acid responsiveness, while still maintaining high derepression in the presence of vanillin. By computational ligand docking analyses we also discuss the structure-function relationship single mutations can have on aTF specificity, an attribute potentially accounting for the wide-spread arise of aTF members belonging to the GntR superfamily of transcriptional regulators.

1. Introduction

Advances in the field of genetic engineering supports the construction of large combinatorial libraries where we can simultaneously swap homologous genes from different organisms, promoters, and other regulatory sequences at high precision and efficiencies.[1] These advances have been fueled by the advent of CRISPR/Cas9, combined with a drop in DNA synthesis costs.[2] However, as the precision and efficiency by which new genetically-encoded designs can be engineered, the means to prototype genetically encoded library information has become a bottleneck, and the development of screening systems adjusted to user-specific requirements, both in terms of input specificity and dynamic output signals, are thus in high demand.[3]

Allosterically regulated transcription factors (aTFs) are a well-established class of biosensors which has been developed and applied for high-throughput screening and directed evolution of enzymes,[4] bioremediation,[5,6] and for building and characterising population dynamics.[7] While aTFs are almost exclusively found in the prokaryotic kingdom, the diverse environmental cues and metabolic perturbations perceived *in vivo* by aTFs has made them engineering hotspots for the synthetic biology community focusing on the design and implementation of small molecule-inducible logics controls across all kingdoms of life. Indeed, allosterically regulated biosensors have been successfully employed for biotechnological purposes in yeast, bacterial cells,[8–11,12] and in mammals in order to control gene expression for clinical applications.[13,14]

Mechanistically, being transcription factors, aTF functionality is constituted by a DNA-binding domain (DBD) and a effector-binding domain (EBD), with the DBD being conserved among aTF-family members.[15] In prokaryotes, aTFs are abundantly present, where they via sequence-diverse effector-binding domains (EBDs) detect perturbations in specific metabolite abundances, and in response to such perturbations regulate target gene expression via allosterically regulated binding or release to operator sequences through conserved DNA-binding domains (DBD).[16,171617] The most

prominent example of aTF-mediated control of gene expression is exemplified by the transcriptional repressor TetR from *Escherichia coli*, which blocks the transcription of the *tetA* gene, encoding the tetracycline efflux pump, in the absence of intracellular tetracycline [18]. Beyond TetR, and TetR family of aTF repressors, several other families encode aTFs with specificity towards a broad range of metabolites, using diverse modes-of-action to control gene expression.[19]

While several successful applications of aTFs have been exemplified, rational engineering of aTF EBDs for user-defined small molecule-specificity have so far not been demonstrated. This is largely due to lack of mechanistic and sequence-function understanding of aTF allostery itself. Hence, unless a biosensor for the candidate metabolite already exists in nature, the ability to engineer and develop aTFs with high specificity for metabolites of interest constitutes a high gain endeavour of synthetic biology. To this end, several studies have highlighted the power of evolution-guided engineering of biosensor specificity, including key residues involved in aTF response curves and ligand specificity.[20–23] Furthermore, aTF ligand promiscuity is acknowledged in a number of examples.[24–26]

Here we present the development of a biosensor based on the aTF VanR from *Caulobacter crescentus* for the detection of vanillic acid and vanillin in baker's yeast *Saccharomyces cerevisiae*. From an initial successful design of a vanillic acid biosensor, we adopt directed evolution of the VanR EBD to identify mutations involved in VanR responsiveness towards vanillin only. From this, we deep-scan the mutational space using targeted NNK/NNS mutagenesis and domain-docking analyses, in order to search for mutations related to the sensing of vanillic acid and vanillin in VanR, and based on this we identify single-residue changes diminishing the responsiveness of VanR to its native ligand, vanillic acid, while maintaining responsiveness to vanillin.

2. Method and materials

2.1 Strains, chemicals and media.

Saccharomyces cerevisiae CEN.PK113-11C (*MATa MAL2-8C SUC2 ura3-52 his3Δ*) strain was obtained from P. Kötter (Johann Wolfgang Goethe University, Frankfurt, Germany) and was used as the background strain in this study. The complete list of yeast strains used in this study can be found in Supplementary Table 1. The chemicals used in this study were obtained from Sigma-Aldrich, unless otherwise specified. *S. cerevisiae* strains were grown at 30°C, 250rpm in Minimal medium, Synthetic Complete medium and YP medium. Synthetic complete (SC) medium and YP

medium were purchased from Sigma-Aldrich, while mineral medium was prepared according to previous publications.[27] The pH of the Mineral medium was adjusted to 4 for applications involving vanillic acid while a pH of 6 was used for medium supplemented with vanillin.

2.2 Synthetic genes and oligonucleotides.

Oligonucleotides and synthetic genes were commercially synthesized by Integrated DNA Technologies, Inc. Sequences of promoters and synthetic genes used in this study can be found in Supplementary Table 2.

2.3 Plasmids, strains and library construction.

The complete list of plasmids used in this study can be found in supplementary information (Supplementary Table 3). The episomal plasmids used in this study were assembled by USER cloning as previously decribed.[28] The All-in-one backbone plasmid used for marker-free integration of TEF1p_2xVanO:GFP in the XI-3 site was assembled as follows: first, Cas9 was amplified from plasmid (p414-TEF1p-Cas9-CYC1t; Addgene id: 43802) with primers TJOS-9F and TJOS-160, there is RNR2 promoter was amplified from CEN.PK genome with primers MeLS062-F and MeLS063-R. Amplified parts were treated with USER enzyme (NEB) and cloned to USER cloning ready vector (pTAJAK-96)[29,30] to constitute plasmid pRNR2-Cas9-CYC1t. Due to cloning compatibility issues the PvuII site in Cas9 and AsiSI in RNR2 promoter was mutated by first, amplifying the pRNR2-Cas9-CYC1t plasmid with TJOS-104F and TJOS-104R, and re-ligating to silently mutate PvuII site; second amplifying plasmid again with primers TJOS-105F and TJOS-105R, and re-ligating to mutate AsiSI restriction site. Finally, pRNR2-Cas9-CYC1t construct, with PvuII and AsiSI sites mutated, was amplified with TJOS-169F and TJOS-169R, and USER treated/cloned with pTAJAK-72[31] previously amplified with TJOS-170F and TJOS-170R, and resulting plasmid was named pTAJAK-168. The marker in pTAJAK-168 from KanMX to NatMX was replaced as described by Jessop-Fabre and co-workers for construction of pTAJAK-71[32]. The resulting plasmid with NatMX marker carrying USER cloning site and expressing Cas9 from RNR2 promoter was named pTAJAK-177. *Escherichia coli* strain DH5α was used as a host for cloning and plasmid propagation. E. coli strains were grown at 37 °C in Luria-Bertani (LB) medium where 100 μg/mL ampicillin was added.

All the genes used in this study were codon optimised for expression in yeast and, unless otherwise stated, the plasmids were assembled by USER cloning. The USER tails were amplified by PCR. All yeast transformations were carried with the lithium acetate/polyethylene glycol/single

carrier DNA method described by Gietz and coworkers.[33] The transformants were then selected on the appropriate drop-out medium plates or supplemented with the appropriate antibiotic.

2.4 VanR mutagenesis

In order to optimise VanR, three steps of error prone PCR (epPCR) were performed on the predicted effector binding site (position 243-719). The GeneMorph II kit was used for this purpose, according to manufacturer's description (Agilent Technologies).

Homologous tails were then introduced by PCR and the insert was introduced in a *S. cerevisiae* strain carrying a yeast enhanced GFP gene under the control of the TEF1 promoter containing the operator sequence by gap-repair.[34] Similarly, NNK/NNS libraries were prepared by introducing random nucleotides (NNS or NNK) instead of the codon encoding the amino acid to be randomised.[35] Primers carrying the randomized nucleotides were used to amplify the full plasmid, which was then introduced in S. cerevisiae by gap-repair.[34] Single colonies were then selected randomly for screening purposes.

2.5 Fluorescence-activated cell sorting

The VanR mutant library was inoculated in mineral medium without ligand and were incubated at 30 °C for 24 hours. Next, the cells were diluted 1:50 in fresh mineral medium, with and without the inducer, depending on the sorting step, and incubated for 24 hours at 30 °C. The cells were then diluted in PBS and the GFP intensity of individual cells was measured using a BD Biosciences Aria (Becton Dickinson) with a blue laser (488 nm). The yeast strains were then collected depending on the desired phenotype by setting tight gates, as exemplified in Fig. 2A. The gates were drawn on the FITC-A vs FSC-A plot, to avoid any bias related to the cell size. The cells were then rescued in SC-Histidine and grown overnight at 30°C at 250rpm. For each sorting step 10,000 cells were collected in 2 mL SC-Histidine medium.

2.6 Flow cytometry measurements and data analysis

For vanillic acid assays, yeast cells were cultivated overnight at 30°C hours in mineral medium at pH4. The next day, the cells were diluted 1:50 into fresh mineral medium with and without the inducer. The cells were then grown for 16h and the fluorescence was then analysed by flow cytometry using a Fortessa flow cytometer (Becton Dickinson) with a blue laser (488 nm), for validation of single strains. For each strain 10,000 single-cell events were recorded. The fluorescence arithmetic

mean of the total population was calculated and the fold change, determined as the ratio between the induced and non-induced state. Similarly, for vanillin applications, the cells were grown over-night 30°C hours in mineral medium at pH6. The next day, the cells were diluted 1:50 into fresh mineral medium with and without the inducer. The cells were then grown for 16h and the fluorescence was then analysed by flow cytometry. Unless otherwise stated, the data is represented by an average of three biological replicates and the error bars represent the standard deviation between these measurements.

The IC_{50} values were determined by polynomial fitting the dose-response curves. The equations were then solved for the ligand concentration corresponding to 50% of the maximum de-repression.

2.7 Structure modelling and computational analysis

The sequence of VanR (WP_010920250.1) consists of 244 amino acids (AAs), was retrieved from the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov/protein (sequence is provided in Supplementary Table 2). Protein BLAST analysis against Protein Data Bank (PDB) database showed that 3D protein structure of VanR is not available experimentally. Full length sequence of VanR (244 AAs) was used for 3D protein structure modeling using YASARA Structure Version 17.1.28.[36] Quality of model structures were evaluated according to statistical parameters on YASARA protocol[36] to select the best model for molecular docking studies. The structure of VanR variants (F165V, F165R) was generated using the FoldX method[36–38] implemented in the YASARA Structure Version 17.1.28.[37] A FoldX mutation run including rotamer search, exploring alternative conformations (3 independent runs) of substitutions (F165V, F165R) and the surrounding side chains below 6 Å distance from F165 residue was performed during the FoldX energy minimization employing a probability-based rotamer library.[39–41]

Chemical structures of vanillin and vanillic acid were retrieved from NCBI-PubChem database (https://pubchem.ncbi.nlm.nih.gov/). Structures were checked and cleaned up using ChemDraw v16.0. Molecular docking of vanillin and vanillic acid with VanR WT and variants was performed using AutoDock implemented in the YASARA software package.[36] The protein residues were treated using the AMBER ff99.[42] The ligand atoms were treated using GAFF[43,44] with AM1-BCC partial charges[45] employing particle mesh Ewald[46] for long-range electrostatic interactions and a direct force cutoff of 10.5 Å. Twenty five docking runs were performed and the obtained docking

poses were clustered by applying a root- mean- square deviation (RMSD) cutoff of 1.0 Å with the default settings implemented within the YASARA dock_run macro file.[36] Docking poses were analyzed in YASARA Structure Version 17.1.28[36] and graphics were prepared in PyMOL.[47]

3. Results

3.1 Design and characterization of VanR in yeast

While an AraC mutant with responsiveness to vanillin in *E. coli* has recently been published,[23] VanR from *Caulobacter crescentus* was selected for this study, motivated by the previous implementation of VanR as a vanillic acid biosensor in mammalian cells, and the relevance of vanillic acid in the yeast-based biosynthesis of vanillin commercially from a biotechnological point of view.[48–50] Furthermore, VanR is a transcription factor belonging to the GntR family of transcriptional regulators, a large aTF family found in different bacteria, known to regulate several metabolic responses, including amino-acid metabolism and ABC transporter systems, under various environmental conditions.[51,52]

In order to engineer VanR as a vanillic acid biosensor in yeast, nine different biosensor designs were initially tested (Fig. 1A, B). Specifically, for the reporter output promoter, two copies of the operator sequence (VanO; ATTGGATCCAAT) recognized by the DBD of VanR, were inserted into three different yeast promoters (TEF1p, CYC1p, and a truncated pCYC1 (209bp-CYC1p) covering approx. two orders of magnitude in expression levels.[10,53] The two operator copies were separated by the Eco47III restriction site, as previously shown by Gitzinger and co-workers.[13] For all three reporter promoters, the two copies of VanO separated by the Eco47III restriction site were inserted 7 bp downstream the TATA elements (Supplementary Table 2). Similarly, VanR was cloned under the control of two different promoters, namely weak REV1p, and strong TEF1p, also covering approximately two orders of magnitude in expression levels (Fig. 1B).[54] Additionally, the three different reporter promoters were characterized in cells without expression of VanR. Next, the GFP reporter expression for all nine designs were tested on the microtiter plate reader in the absence of any ligand, in order to first assess the ability of VanR to repress the transcription of the reporter gene (Fig. 1B). From this experiment, the biosensor design with both VanR and GFP under the control of pTEF1 resulted in approx. 6-fold repression compared to strain designs without expression of VanR, whereas a maximum of 2-fold repression was observed for the other designs (Fig. 1B).

Having established a functional repressive design, the VanR system expressing TEF1p2xVanO:GFP and TEF1p:VanR was next tested more thoroughly in terms of dynamic output and operational ranges upon feeding vanillic acid to the cultivation medium. For dynamic output range the fluorescence increased up to 3.7-fold when vanillic acid was fed, while an operational range of more than two orders of magnitude, from 16µM up to 4mM, was observed (Fig. 1C). Moreover, when feeding vanillin, the reduced form of vanillic acid and a natural aroma compound, it was observed that VanR also responds to vanillin at concentrations above 2 mM (Fig. 1C).

3.2 Directed evolution of VanR ligand specificity

In order to investigate and identify residues related to the observed ligand promiscuity of VanR between vanillic acid and vanillin, we carried out directed evolution using mutagenesis and toggled selection.[55] More specifically, we initially carried out three rounds of error-prone PCR (epPCR) of the predicted EBD of wild-type VanR, and next pooled the amplicons for construction of a VanR EBD-variant library expressed from a centromeric plasmid through plasmid gap repair (Fig. 2A).[10,34] The library of estimated 30,000 VanR-EBD variants was introduced into a platform TEF1p_2xVanO:GFP reporter strain and screened through a three-step selection method using fluorescence-activated cell sorting (FACS)(Fig. 2A). For the first sorting, the library was cultivated in the absence of ligands to select the functional VanR variants maintaining repression of GFP expression controlled by TEF1p_2xVanO in the apo-form (no ligand bound). This first sorting enabled selection of approx. 10% of the initial library. In the second selection step, cells expressing the VanR library were grown in the presence of 0.8 mM vanillic acid, and FACS was used to remove VanR variants with vanillic acid-mediated depreression of GFP expression. From this selection approx. 2.5% of the VanR EBD-variant sublibrary from the first sorting was selected. Following recovery of the second sub-population, 191 clones (approx 2.5x coverage of sublibrary) were tested in the presence and absence of 4 mM vanillin (Fig. 2A). Based on fold-changes in reporter gene outputs, top-three best performing variants, herein labelled VanR-F9, VanR-D10 and VanR-E3 were isolated, re-transformed into a clean platform reporter strain (i.e. TEF1p_2xVanO:GFP), and validated by flow cytometry (Fig. 2B).

Among the three variants, we discarded the VanR-F9 and VanR-D10 variants from further analysis, focusing on the VanR-E3 variant with a relatively low OFF state and five mutations spanning residues 123-193 central to the VanR EBD (Supplementary Figure 1). This quintuple VanR

EBD-variant (G123W, D131G, F165L, Q191L, Q193L), had a 2.8-fold induction upon feeding 4 mM vanillin, yet without responsiveness to vanillic acid (Fig. 2B, Supplementary Figure 1).

Next, in order to investigate the minimal mutational space enabling ligand-specificity of VanR to only vanillin, we constructed an NNK/NNS library for each of the five mutated single-positions in the VanR-E3 variant (Fig. 3A). In total 84 randomly picked colonies from each single-position library (5 x 84) were first tested in the absence of ligands in order to assess OFF states of the VanR mutants. While most variants showed wild-type-like repression across all five positions, a few variants at positions 123, 165 and 191 showed high GFP expression, indicating that DNA-binding could be abolished for these variants (Supplementary Figure 2). Next, out of the functional repressors, the response of the library variants to vanillic acid and vanillin was assessed. From this screen, different single-position mutants of VanR were observed to impact differently on reporter gene expression (Fig. 3B). More specifically, we observed that mutants at positions 131, 165, and 191 contain mutants which can either increase, decrease or maintain a comparable response to that of the VanR wild-type for both ligands, whereas mutants at position 123 showed lowered, or VanR wild type-like, responses to both inducers. Surprisingly, mutants in position 193 showed an increased response to both the ligands compared to the wild type VanR. Taken together this analysis indicate that mutations in positions 123 and 193 are largely dispensable for the sought-for vanillin-only biosensor phenotype of the parental VanR-E3 variant.

3.3 Single-residue resolution of vanillin-specific VanR mutants

In order to further identify and select for single-position vanillin-specific VanR mutants, we filtered the libraries by selection for mutants that showed no response to vanillic acid, and no less than a 1.7-fold derepression in the presence of vanillin, as determined by the 25th percentile of VanR-E3 replicate samples (Fig. 3B). From deep-scanning the mutational hotspots for specificity derived from directed evolution and toggled selection, we identified seventeen F165 and eight Q191 mutants that were able to detect vanillin but had lost the native VanR wild-type vanillic acid response (Fig. 3B). From sequencing of 25 selected variants, 16 unique mutants were identified, 10 for position F165 and 6 for position Q191 (Supplementary Table 4). These mutants were re-transformed into clean TEF1p_2xVanO:GFP reporter strains and validated (Fig. 3C). Indeed, we observed that none of the mutants responded to vanillic acid (Supplementary Figure 3), while enabling a response to vanillin from 2.5-3.3 fold and 2-2.5 fold for positions F165 and Q191, respectively (Fig. 3C). Notably, the best-performing single-position VanR variants (F165R/V) show vanillin-induced de-repression

of reporter gene expression comparable or even higher than VanR wild-type and the previously selected VanR-E3 variant (Fig. 3C).

3.4 Structure-function relationship for VanR vanillin-specificity

According to previous reports of GntR transcriptional regulators, and the model structure of VanR based on PDB ID 3SXY (resolution 1.65 Å), it is expected that VanR wild-type constitutes dimeric form (Supplementary Table 5).[56,57] Indeed, the model structure has the canonical domain architecture of the GntR family, with an N-terminal winged helix–turn–helix (WH) DNA-binding domain and a C-terminal all-α-helical regulatory ligand-binding domain (Fig. 3D, Supplementary Figure 4).[56] The C-terminal domain contains six α-helices (A1-A6), which is a characteristic feature of VanR family protein while the N-terminal contains the winged-helix DNA-binding domain, which is the hallmark of the GntR family, and is composed of a canonical order of secondary-structure elements: three α-helices (i.e. α1, α2, α3) and two antiparallel β-sheets (i.e. β1, β2)(Fig. 3D).

In order to elucidate the structure-function relationship of vanillin-specific VanR variants F165V and F165R compared to VanR wild-type, molecular docking studies of vanillic acid and vanillin were performed. The docking analyses were analysed to identify the binding modes of the vanillic acid and vanillin as well as crucial interactions and structural properties that are governing the binding affinity. It was observed that vanillic acid and vanillin majorly interact at the dimeric interface F165 of VanR (Fig. 3E). Moreover, vanillic acid exerts higher binding affinity towards F165 of VanR wild-type compared to vanillin (binding energy -5.02 kcal/mol vs -4.67 kcal/mol) (Fig. 3E, Supplementary Fig. 5). Vanillic acid interacts with both F165 residues at the dimer interface of VanR through hydrophobic interactions, whereas vanillin is predicted to interact only with F165 in chain A at the VanR dimeric interface (Supplementary Figure 4). Additionally, vanillic acid is predicted to form a H-bond with the P164 residue (Fig. 3E). In the case of variant F165V, the binding affinity for vanillin (-4.59 kcal/mol) is comparable to VanR wild-type (-4.67 kcal/mol), whereas F165V has decreased binding affinity for vanillic acid (-4.26 kcal/mol) and a different binding orientation, compared to the VanR wild-type (Fig. 3F). Similarly, variant F165R is predicted to have both decreased binding affinity of vanillic acid (-4.28 kcal/mol) as well as different binding orientation compared to VanR wild-type (-5.02 kcal/mol). However, comparable interaction was observed for vanillin with F165R (-4.73 kcal/mol) to VanR wild-type (-4.67 kcal/mol) (Fig. 3G, Supplementary Figure 5). For comparison, similar changes in binding energies was observed to result in 26-fold changes in activity of *Candida parapsilosis* alcohol dehydrogenase 5.[58] Taken together, the molecular

docking studies reveal that the two single-substitution VanR variants F165R and F165V presumably lost responsiveness to vanillic acid due to decreased ligand affinity as well as, in the case of F165R, strengthened interaction with vanillin.

4. Discussion

In this study we have characterized the VanR-based vanillic acid and vanillin biosensor systems in *S. cerevisiae*, a well-established model organism and biotechnology workhorse. By employing directed evolution in combination with toggled selection we were able to evolve the specificity of the VanR wild-type biosensor towards vanillin, a secondary ligand with a significantly lower binding affinity towards the effector binding domain (Fig. 3D-F). Moreover, by deep-scanning the mutation landscape of mutant VanR-E3, a mutational hotspot required for the specificity evolution, namely position F165, was identified. Mutations in this position lead to the identification of variants with an improved response to vanillin compared to both VanR wild-type and the parental VanR-E3 variant, highlighting deep-mutational scanning and meticulous phenotypic characterization (dose-response curves) to enable identification of sequence-function relationships useful for the scientific community to build up a much-needed knowledge base for future predictive engineering of biosensor characteristics. Furthermore, comparing the dynamic output ranges of the best-performing AraC-derived vanillin biosensor reported in *E. coli* (35x, 2 mM vanillin) with the one obtained from this study (3x, 4 mM vanillin),[23] it is evident that it would be relevant to further investigate vanillin biosensor engineering in yeast.

The knowledge generated during this study is considered useful as a steppingstone to further improvement of other structurally similar GntR-type biosensors. One such example could be the fatty acid-responsive transcription factor FadR, which was successfully employed in yeast as a functional biosensor.[24] Most importantly, FadR presents several structural similarities with VanR, and respond promiscuously to fatty acids of different lengths.

In conclusion, in this study we have used directed evolution to identify VanR variants with user-defined changes in biosensor specificity. From this, we have identified a vanillin-specific biosensor in *S. cerevisiae*, which should be relevant for both the biotech industry and synthetic biology community as a new logic gate.

Author contribution

V.D., M.K.J., and J.D.K. conceived the study; V.D. and M.K.J. designed the experiments, V.D. performed the experiments; S.P. and M.D.D. performed 3D structure modeling and molecular docking simulations. T.J. developed the all-in-one Cas9 plasmid, K.G and D.S. provided technical assistance for library construction. V.D. and M.K.J. wrote the manuscript.

Acknowledgements

This work was supported by the Novo Nordisk Foundation and the European Commission Horizon 2020 programme (PAcMEN, No. 722287). Authors would also like to thank colleagues at the Novo Nordisk Foundation Center for Biosustainability for fruitful discussions and comments.

Conflict of interest

J.D.K. has a financial interest in Amyris, Lygos, Demetrix, Constructive Biology, Maple Bio, and Napigen.

References

1. Luo, Y., Enghiad, B. & Zhao, H. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat. Prod. Rep.* **33**, 174–182 (2016).
2. Hughes, R. A. & Ellington, A. D. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).
3. De Paepe, B., Peters, G., Coussement, P., Maertens, J. & De Mey, M. Tailor-made transcriptional biosensors for optimizing microbial cell factories. *J. Ind. Microbiol. Biotechnol.* **44**, 623–645 (2017).
4. Flachbart, L. K., Sokolowsky, S. & Marienhagen, J. Displaced by Deceivers: Prevention of Biosensor Cross-Talk Is Pivotal for Successful Biosensor-Based High-Throughput Screening Campaigns. *ACS Synth. Biol.* **8**, 1847–1857 (2019).
5. Sanseverino, J. *et al.* Use of Saccharomyces cerevisiae BLYES expressing bacterial bioluminescence for rapid, sensitive detection of estrogenic compounds. *Appl. Environ. Microbiol.* **71**, 4455–4460 (2005).
6. Moser, F., Horwitz, A., Chen, J., Lim, W. A. & Voigt, C. A. Genetic sensor for strong methylating

compounds. *ACS Synth. Biol.* **2**, 614–624 (2013).

7. Karig, D. *et al.* Stochastic Turing patterns in a synthetic bacterial population. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6572–6577 (2018).
8. Feng, J. *et al.* A general strategy to construct small molecule biosensors in eukaryotes. *Elife* **4**, (2015).
9. David, F., Nielsen, J. & Siewers, V. Flux Control at the Malonyl-CoA Node through Hierarchical Dynamic Pathway Regulation in Saccharomyces cerevisiae. *ACS Synth. Biol.* **5**, 224–233 (2016).
10. Skjoedt, M. L. *et al.* Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat. Chem. Biol.* **12**, 951–958 (2016).
11. Snoek, T. *et al.* An Orthogonal and pH-Tunable Sensor-Selector for Muconic Acid Biosynthesis in Yeast. *ACS Synth. Biol.* **7**, 995–1003 (2018).
12. Rugbjerg, P., Sarup-Lytzen, K., Nagy, M. & Sommer, M. O. A. Synthetic addiction extends the productive life time of engineered Escherichia coli populations. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2347–2352 (2018).
13. Gitzinger, M. *et al.* The food additive vanillic acid controls transgene expression in mammalian cells and mice. *Nucleic Acids Res.* **40**, e37 (2012).
14. Rössger, K., Charpin-El-Hamri, G. & Fussenegger, M. A closed-loop synthetic gene circuit for the treatment of diet-induced obesity in mice. *Nat. Commun.* **4**, 2825 (2013).
15. AlQuraishi, M., Tang, S. & Xia, X. An affinity-structure database of helix-turn-helix: DNA complexes with a universal coordinate system. *BMC Bioinformatics* **16**, 390 (2015).
16. Zhang, J., Jensen, M. K. & Keasling, J. D. Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol.* **28**, 1–8 (2015).
17. Beckett, D. Regulating transcription regulators via allostery and flexibility. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22035–22036 (2009).
18. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5547–5551 (1992).
19. Ishihama, A., Shimada, T. & Yamazaki, Y. Transcription profile of Escherichia coli: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Res.* **44**, 2058–2074 (2016).

20. Xiong, D. *et al.* Improving key enzyme activity in phenylpropanoid pathway with a designed biosensor. *Metab. Eng.* **40**, 115–123 (2017).
21. Tang, S.-Y., Fazelinia, H. & Cirino, P. C. AraC regulatory protein mutants with altered effector specificity. *J. Am. Chem. Soc.* **130**, 5267–5271 (2008).
22. Taylor, N. D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177–183 (2015).
23. Frei, C. S., Qian, S. & Cirino, P. C. New engineered phenolic biosensors based on the AraC regulatory protein. *Protein Eng. Des. Sel.* **31**, 213–220 (2018).
24. Teo, W. S., Hee, K. S. & Chang, M. W. Bacterial FadR and synthetic promoters function as modular fatty acid sensor- regulators in Saccharomyces cerevisiae. *Eng. Life Sci.* **13**, 456–463 (2013).
25. Siedler, S., Stahlhut, S. G., Malla, S., Maury, J. & Neves, A. R. Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into Escherichia coli. *Metab. Eng.* **21**, 2–8 (2014).
26. Cebolla, A., Sousa, C. & de Lorenzo, V. Effector Specificity Mutants of the Transcriptional Activator NahR of Naphthalene Degrading Pseudomonas Define Protein Sites Involved in Binding of Aromatic Inducers. *J. Biol. Chem.* **272**, 3986–3992 (1997).
27. Jensen, N. B. *et al.* EasyClone: method for iterative chromosomal integration of multiple genes Saccharomyces cerevisiae. *FEMS Yeast Res.* **14**, 238–248 (2014).
28. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34**, e122 (2006).
29. Jakočiūnas, T. *et al.* Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. *Metab. Eng.* **28**, 213–222 (2015).
30. Jakočiūnas, T. *et al.* CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in Saccharomyces cerevisiae. *ACS Synth. Biol.* **4**, 1226–1234 (2015).
31. Ronda, C. *et al.* CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae. *Microb. Cell Fact.* **14**, 97 (2015).
32. Jessop-Fabre, M. M. *et al.* EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes

into Saccharomyces cerevisiae via CRISPR-Cas9. *Biotechnol. J.* **11**, 1110–1117 (2016).

33. Gietz, R. D. & Schiestl, R. H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 38–41 (2007).
34. Eckert-Boulet, N., Pedersen, M. L., Krogh, B. O. & Lisby, M. Optimization of ordered plasmid assembly by gap repair in Saccharomyces cerevisiae. *Yeast* **29**, 323–334 (2012).
35. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).
36. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114–122 (2009).
37. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–8 (2005).
38. Krieger, E. & Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **30**, 2981–2982 (2014).
39. Chinea, G., Padron, G., Hooft, R. W., Sander, C. & Vriend, G. The use of position-specific rotamers in model building by homology. *Proteins* **23**, 415–421 (1995).
40. Dunbrack, R. L., Jr & Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681 (1997).
41. Canutescu, A. A., Shelenkov, A. A. & Dunbrack, R. L., Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **12**, 2001–2014 (2003).
42. Wang, J., Cieplak, P. & Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).
43. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).
44. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
45. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-

BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623–1641 (2002).

46. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
47. De Lano, W. L. & California, U. S. A. PyMOL: An Open-Source Molecular Graphics Tool.
48. Folcher, M., Xie, M., Spinnler, A. & Fussenegger, M. Synthetic mammalian trigger-controlled bipartite transcription factors. *Nucleic Acids Res.* **41**, e134–e134 (2013).
49. Kim, T., Folcher, M., Doaud-El Baba, M. & Fussenegger, M. A synthetic erectile optogenetic stimulator enabling blue-light-inducible penile erection. *Angew. Chem. Int. Ed Engl.* **54**, 5933–5938 (2015).
50. Hansen, E. H. *et al.* De novo biosynthesis of vanillin in fission yeast (Schizosaccharomyces pombe) and baker's yeast (Saccharomyces cerevisiae). *Appl. Environ. Microbiol.* **75**, 2765–2774 (2009).
51. Jain, D. Allosteric control of transcription in GntR family of transcription regulators: A structural overview. *IUBMB Life* **67**, 556–563 (2015).
52. Suvorova, I. A., Korostelev, Y. D. & Gelfand, M. S. GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLoS One* **10**, e0132618 (2015).
53. Reider Apel, A. *et al.* A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **45**, 496–508 (2017).
54. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J. & Dueber, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* **41**, 10668–10678 (2013).
55. Snoek, T. *et al.* Evolution-guided engineering of small-molecule biosensors. *bioRxiv* 601823 (2019) doi:10.1101/601823.
56. Zheng, M. *et al.* Structure of Thermotoga maritima TM0439: implications for the mechanism of bacterial GntR transcription regulators with Zn2+-binding FCD domains. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 356–365 (2009).
57. Lord, D. M. *et al.* McbR/YncC: implications for the mechanism of ligand and DNA binding by a bacterial GntR transcriptional regulator involved in biofilm formation. *Biochemistry* **53**, 7223–7231

(2014).

58. Ensari, Y., Dhoke, G. V., Davari, M. D., Ruff, A. J. & Schwaneberg, U. A Comparative Reengineering Study of cpADH5 through Iterative and Simultaneous Multisite Saturation Mutagenesis. *Chembiochem* (2018) doi:10.1002/cbic.201800159.

Figure legends

Figure 1. VanR characterisation

(A) Schematic overview of the combination of different promoters tested in this study. (B) Repression strength of different promoter combinations. Mean yeGFP fluorescence values are expressed in arbitrary units and were measured on the microtiter plate reader. The values are represented as mean ± s.d. from three (n = 3) biological replicate experiments. (C) Dose-response function for VanR upon vanillic acid and vanillin feeding at concentrations of, 16µM, 64µM, 256µM, 512µM, 1mM, 2mM and 4mM and in the absence of the ligand. Mean fluorescence values represented in arbitrary units were measured on flow cytometer. The values are represented as mean ± s.d. from three (n = 3) biological replicate experiments.

Figure 2. Schematic overview of VanR directed evolution approach and isolated strains

(A) Schematic overview of the directed evolution strategy. VanR-EBD was used as a template for 3 rounds of epPCR. The library was then introduced in *S. cerevisiae* by exploiting the yeast homology recombination machinery. Next, sub-populations were collected by fluorescence-activated cell in a 2-step selection where the gates were drawn to first select for functional repressors and then to select for variants that could not respond to vanillic acid. Next, 191 single colonies were tested for their response to 4mM vanillin. Responses lower than WT-VanR (orange) are highlighted in green while stronger are highlighted in grey. (B) Characterisation of the GFP-Fold induction in response to vanillic acid and vanillin of three selected variants from the 3-step toggled selection. The fluorescence was evaluated in the absence of the ligands and at 800µM for vanillic acid and at concentrations of 250µM, 500µM, 1mM, 2mM and 4mM for vanillin. The values represent the average yeGFP fold

induction of the strains when grown in the presence or absence of vanillic acid and vanillin, respectively. The values are represented as mean ± s.d. from two (n = 2) biological replicate experiments.

Figure 3. NNK/NNS library generation and screening

(A) Schematic representation of the mutations in VanR-E3 and the NNK/NNS library generation strategy. For each mutated residue in the VanR-E3 variant, a wild type VanR* gene was constructed, carrying three random nucleotides at that specific location. The 5 different libraries were then transformed into yeast by gap-repair, next 84 colonies from each library were selected and tested for their response to vanillic acid and vanillin (B) GFP fold induction of the NNK/NNS library in the presence or absence of vanillic acid and vanillin, respectively. Dashed lines represent lower and upper cut-offs. (C) Validation of 16 unique mutations identified from the NNK/NNS library screening. Yeast strains were grown in the presence or absence of 4mM vanillin. The values are represented as mean ± s.d. from three (n = 3) biological replicate experiments. (D) Predicted 3D structure for wild type VanR. (E-G) Ligand docking simulation with binding energy for vanillic acid for wild type VanR, VanR(F165V) and VanR(F165R), respectively.

Supplementary Figure 1. Sequences and functional characterization of epPCR-derived VanR variants. (A) Sequence alignment of the 3 selected variants compared to wild-type VanR. (B) Mean fluorescence intensity and yeGFP fold-change of the 3 selected VanR variants from epPCR compared to wild-type VanR. The values represent the average yeGFP and yeGFP fold-change in response to 0, 0.8, 1.6, 2.4, 3.2, 4.0 mM vanillic acid. The values are represented as mean ± s.d. from two (n = 3) biological replicate experiments. au = arbitrary units. (C) Estimates of IC_{50} values and maximum fold-changes (FC_{max}) based on vanillic acid and vanillin responses of the 3 selected variants compared to wild-type VanR.

Supplementary Figure 2. MFI of the NNK/NNS library in the absence of any ligand

Mean fluorescence intensity of GFP of the NNK/NNS library tested strains when cultivated in the absence of the ligand.

Supplementary Figure 3. GFP fold induction of selected mutants in the presence of vanillic acid

GFP fold induction in response to vanillic acid of 16 unique mutations identified from the NNK/NNS

library screening compared to the VanR_E3 and wild type VanR. Yeast strains were grown in the presence or absence 0.8 mM vanillic acid. The values are represented as mean ± s.d. from three (n = 3) biological replicate experiments.

Supplementary Figure 4. Ligand docking simulation for vanillin

Binding of vanillin at the dimer interface of wild type VanR (Chain A is green and Chain B is orange)

Supplementary Figure 5. Ligand docking simulation for vanillin

Ligand docking with binding energy for vanillin for wild-type VanR, VanR(F165V) and VanR(F165R), respectively.

Figure

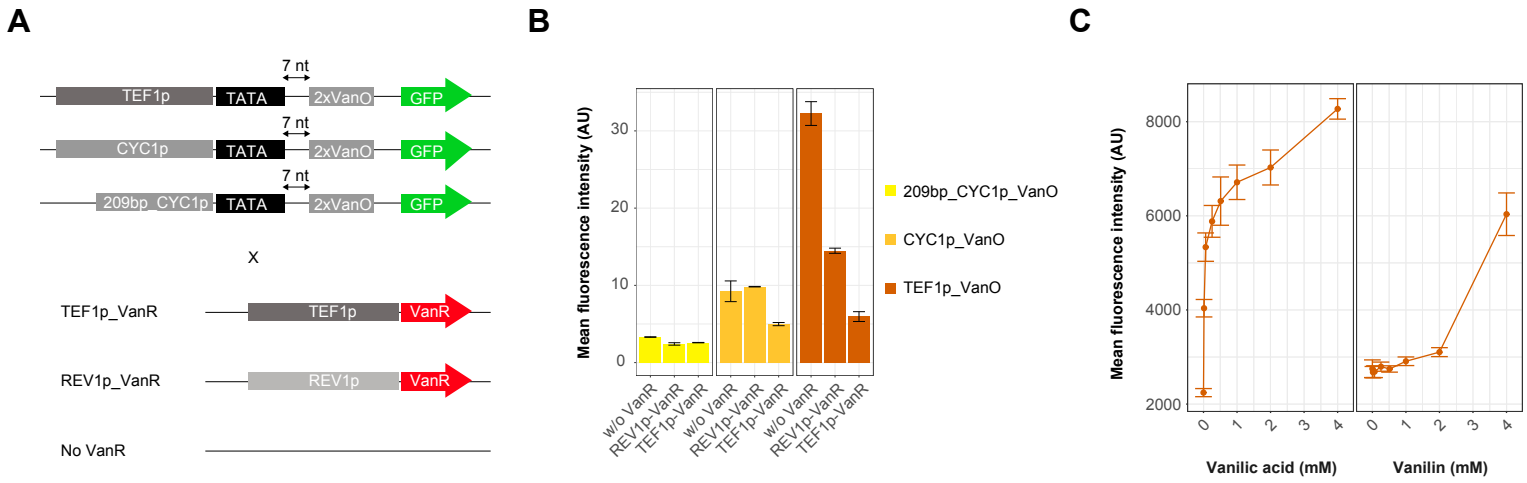

Figure 1.

Figure

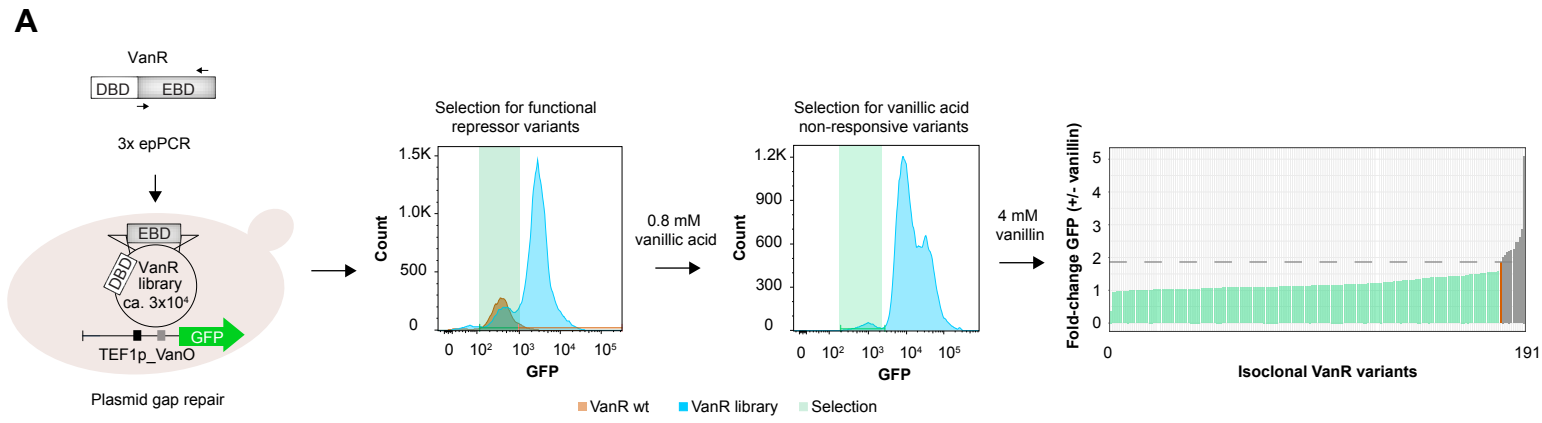

Figure 2.

Figure

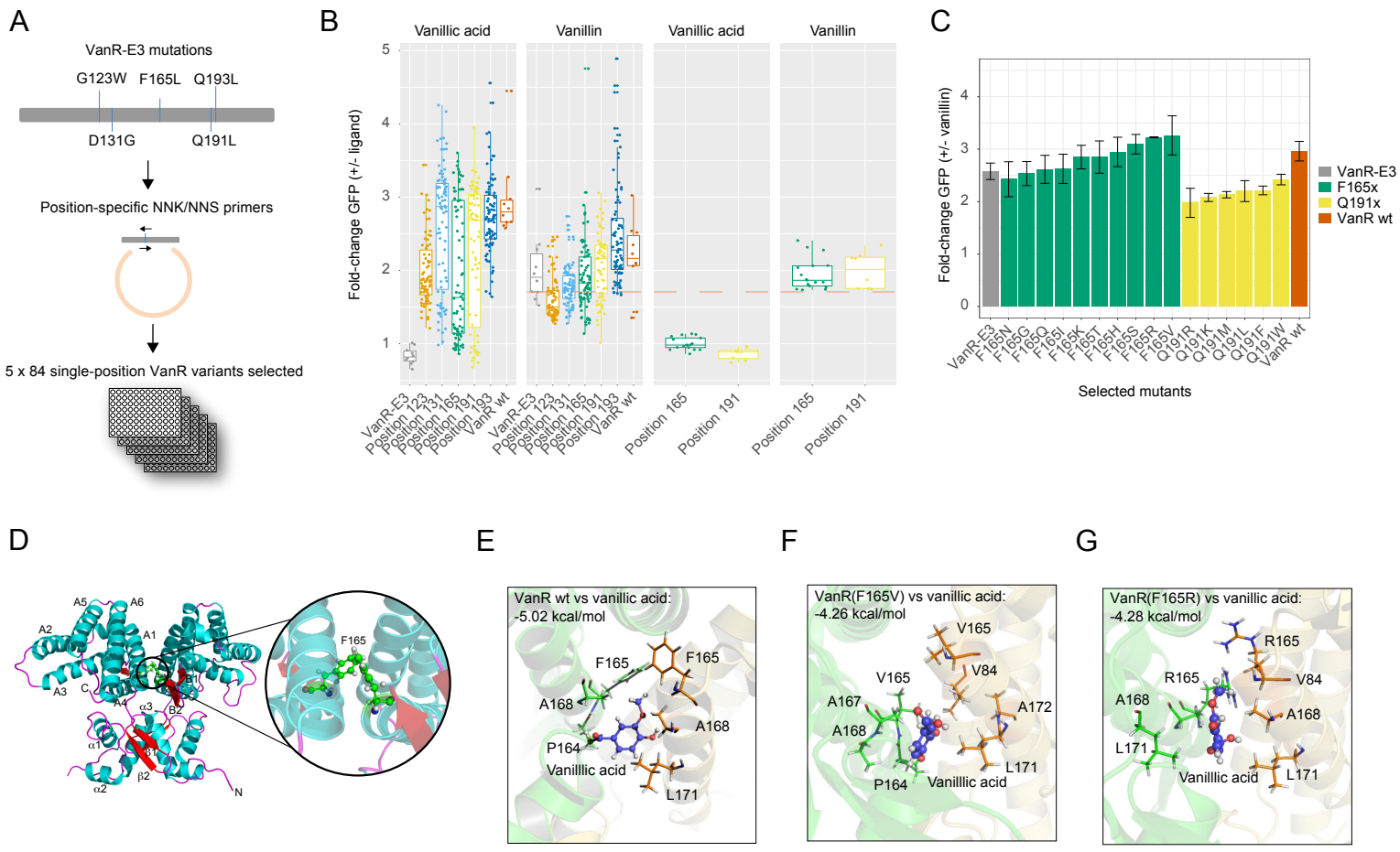

Figure 3.