Decoding Complex Chemical Mixtures with a Physical Model of a Sensor Array

Julia Tsitron[1], **Addison D. Ault[2]**, **James R. Broach[2]**, **Alexandre V. Morozov[1]***

1 Department of Physics & Astronomy and BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, New Jersey, United States of America, **2** Department of Molecular Biology, Princeton University, Princeton, New Jersey, United States of America

Abstract

Combinatorial sensor arrays, such as the olfactory system, can detect a large number of analytes using a relatively small number of receptors. However, the complex pattern of receptor responses to even a single analyte, coupled with the non-linearity of responses to mixtures of analytes, makes quantitative prediction of compound concentrations in a mixture a challenging task. Here we develop a physical model that explicitly takes receptor-ligand interactions into account, and apply it to infer concentrations of highly related sugar nucleotides from the output of four engineered G-protein-coupled receptors. We also derive design principles that enable accurate mixture discrimination with cross-specific sensor arrays. The optimal sensor parameters exhibit relatively weak dependence on component concentrations, making a single designed array useful for analyzing a sizable range of mixtures. The maximum number of mixture components that can be successfully discriminated is twice the number of sensors in the array. Finally, antagonistic receptor responses, well-known to play an important role in natural olfactory systems, prove to be essential for the accurate prediction of component concentrations.

Citation: Tsitron J, Ault AD, Broach JR, Morozov AV (2011) Decoding Complex Chemical Mixtures with a Physical Model of a Sensor Array. PLoS Comput Biol 7(10): e1002224. doi:10.1371/journal.pcbi.1002224

Editor: Christopher V. Rao, University of Illinois at Urbana-Champaign, United States of America

Received April 4, 2011; **Accepted** August 28, 2011; **Published** October 20, 2011

Copyright: © 2011 Tsitron et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This project was supported by NIH grant GM076562 (http://www.nih.gov) and US Army Research Office contract W911NF0810296 (http://www.arl.army.mil) to JRB and by NIH grant HG004708 to AVM, with additional support from a Center for Quantitative Biology/NIH grant P50 GM071508. AVM is a recipient of an Alfred P. Sloan Research Fellowship (http://www.sloan.org). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: morozov@physics.rutgers.edu

These authors contributed equally to this work.

Introduction

Mammalian and insect olfactory systems are capable of recognizing tens of thousands of odors – mostly organic compounds with diverse chemical structures and properties [1–4]. The olfactory tasks commonly faced by such systems include detecting odors, estimating their strength, identifying their source, and recognizing one specific odor in the background of another [5]. The sense of smell exhibits amazing sensitivity and discriminatory power, distinguishing between closely related compounds and detecting vanishingly small odorant concentrations [6]. Olfactory signaling is mediated by a superfamily of several hundred G-protein-coupled receptors (GPCRs) – a significant fraction of the total number of genes in many higher eukaryotes [7–11]. In mammals, GPCRs are located on the surfaces of the cilia projected from olfactory sensory neurons; typically receptors of only one type are expressed in a given neuron [12]. Odor recognition is combinatorial, with one odorant activating multiple receptors and one receptor responding to multiple odorants [12–16]. The resulting complex patterns of receptor activation enable robust identification of many more odors than would have been possible with "lock and key" receptors reacting to only one analyte. Moreover, several studies provide evidence for widespread inhibitory responses in which receptors are antagonized by odorants [14,15,17–20].

The idea of combinatorial recognition has been adapted to artificial arrays in which multiple sensors with partially overlapping selectivities respond to a given analyte [21–24]. While the output of these cross-specific arrays in response to single compounds can generally be interpreted through pattern recognition algorithms [24–27], computational analysis becomes more difficult when the array is presented with a mixture of compounds. Indeed, the non-linear nature of sensor responses to multiple ligands makes it hard to train discriminatory algorithms on a "typical" subset of patterns. The non-linear dependence of sensor output on ligand concentrations is generic in reporter systems and may be compounded by potential binding interference of the two ligands, saturation of the sensor output [28] and, of particular concern, potential antagonistic action of one ligand on another's activity [19]. As a result, responses to complex mixtures have primarily been used to "fingerprint" specific mixtures rather than identify their constituents quantitatively [29–32]. There are relatively few studies which focus on the quantitative analysis of mixtures: for example, Heilig et al. used a single sensor and Fourier transformation techniques to analyze a binary mixture of CO and NO_2 [33], White et al. trained artificial neural networks to identify relative concentrations in binary mixtures [34], and Woodka et al. used a non-negative least squares method to quantify the composition of analyte mixtures with up to five components [35].

Author Summary

Mammalian and insect olfactory systems are combinatorial in nature - instead of activating a single specialized receptor, each analyte invokes a complex pattern of responses across the receptor array. The advantage of such systems lies in their ability to detect a large number of analytes with a relatively small number of receptors. However, the complexity of array responses to mixtures of analytes makes quantitative prediction of component concentrations a challenging task. Here we show that combinatorial output from an array of four engineered G-protein-coupled receptors can be used to predict the concentration of each component in mixtures of highly related sugar nucleotides. We employ a physical model of ligand-receptor interactions and carry out Bayesian analysis of the array output. Furthermore, our *in silico* designs of receptor arrays reveal that antagonistic responses, in which the receptor is bound by the ligand but there is no downstream reporter activity, are necessary for precise recognition of mixture components. This conclusion provides a rationale for the widespread inhibitory responses observed in olfactory systems. Our methodology can be employed with both biological systems and artificial receptor arrays ("electronic noses") designed for various industrial needs.

Here we describe a physical model of receptor-ligand recognition that explicitly relates observed response patterns to component concentrations and receptor properties, making it easier to quantify mixture constituents. We use Bayesian inference to predict absolute concentrations of each ligand in arbitrary mixtures of uridine diphosphate (UDP) sugar nucleotides applied to a combinatorial array of four GPCRs. Furthermore, we develop a universal metric of receptor array performance, and use it to study the fundamental limits imposed on the accuracy of ligand recognition by the physics and biology of receptor-ligand interactions. Finally, we provide design guidelines for constructing cross-specific arrays optimized for mixture recognition, and demonstrate that inhibitory responses are essential for simultaneous detection of all components in a complex mixture.

Results

Biological implementation of the sensor array

Our sensor array is comprised of four engineered receptors (L-3, H-20, K-3 and 2211) with distinct but overlapping specificities for four types of nucleotide sugars: UDP-glucose (UDP-Glc), UDP-galactose (UDP-Gal), UDP-glucosamine (UDP-GlcNAc) and UDP. The receptors were evolved *in vitro* from the human UDP-glucose receptor using directed mutagenesis of the residues involved in ligand binding (see **Materials and Methods**) [36]. Nucleotide sugars and their derivatives are key constituents in polysaccharide synthesis and other cellular processes. Their structural similarity makes them a challenging target for array-based discriminatory analysis. To assess receptor-ligand interactions quantitatively in our sensor array, we functionally expressed the receptors in *S.cerevisiae*. To do so, we replaced the yeast pheromone receptor with one of the sensor GPCRs in strains in which the pheromone response pathway was modified to respond to the heterologous receptor by inducing transcription of the *E. coli* lacZ gene [37]. In this fashion, the extent of GPCR activation following ligand addition could be directly monitored as the level of $\beta-galactosidase$ produced in the cell, which we measured using a fluorescence-based assay (**Materials and Methods**). Applying a mixture of nucleotide sugars to the receptor array yields a complex pattern of responses of the four receptor-bearing strains. The response of each receptor depends on the concentration of all components in the mixture, on the receptor-ligand binding affinities, and on the efficacy with which each ligand activates the receptor. Nonetheless, the contents of arbitrary nucleotide sugar mixtures can be deciphered using array readout as input to a physical model of receptor-ligand interactions.

Physical model of the sensor array

Single-receptor, single-ligand model. We start with the simplest case in which a receptor interacts with a single ligand. We assume that the observed signal in our receptor-bearing reporter strain is proportional to the probability that the receptor is bound by the ligand. This proportionality value, A, which we refer to as the receptor efficacy, can range from 1, for a full agonist, to 0, for a full antagonist. Thus, for a single receptor interacting with a single ligand, the amount of activation of the reporter in the receptor-bearing strain is given by Eq. (1) (**Materials and Methods**). Reporter activation measurements as a function of single ligand concentration are shown in **Fig. S1a**. We use these data to estimate the parameters of Eq. (1) (**Fig. 1a**) and the amount of experimental noise $\tilde{\sigma}$ for each single-receptor, single-ligand combination using Bayesian inference with nested sampling [38] (**Materials and Methods**). The most likely values of the parameters (**Table S1**) are then used for subsequent evaluation of mixtures of compounds. The accuracy of parameter predictions depends on the range of concentrations available for these calibration experiments (**Fig. S2**) and on the amount of experimental noise (**Fig. S3**).

Multiple-receptor, multiple-ligand model. Once all receptor-ligand interaction parameters have been determined through the analysis of single-ligand calibration experiments, we can proceed to interrogating mixtures of ligands with receptor arrays. In considering the response of receptor-bearing strains to ligand mixtures, we note that each ligand contributes to the overall receptor occupancy and that each receptor molecule on the cell surface activates the reporter with an efficacy specified by the ligand to which it is bound, which is often different for different ligands (**Table S1**). Assuming that all ligands bind competitively to the same site on the receptor, we model the response of the receptor-bearing strain to mixtures of compounds by calculating the total intensity as a sum of fractional occupancies of the receptor by each ligand weighted by the corresponding efficacies (Eq. (6)). We treat each of the receptor-bearing strains with an unknown mixture, sequentially diluted to provide a series of samples across a million-fold range of concentrations (**Fig. S1b**). We carry out Bayesian inference for the entire receptor array, predicting the total concentration of all ligands and the concentration ratios of ligand pairs (**Fig. 1b**). From these values we can deduce the absolute concentration of each ligand in the mixture.

Tests of the physical model of mixture recognition. We have tested our approach using a series of assays in which a known combination of ligands was applied to the receptor-bearing strains. As an initial test, we mixed equal proportions of two, three and four ligands in all possible combinations and predicted absolute ligand concentrations. We used a model in which four ligands interacted with four receptors, even if only one, two or three ligands were actually present in the mixture. As can be seen in **Fig. 2** and **Table S2**, our approach is generally quite successful in identifying both zero and non-zero ligand concentrations in the mixtures. For example, with single ligands and binary mixtures the correct chemical or pair of chemicals is predicted to have the highest

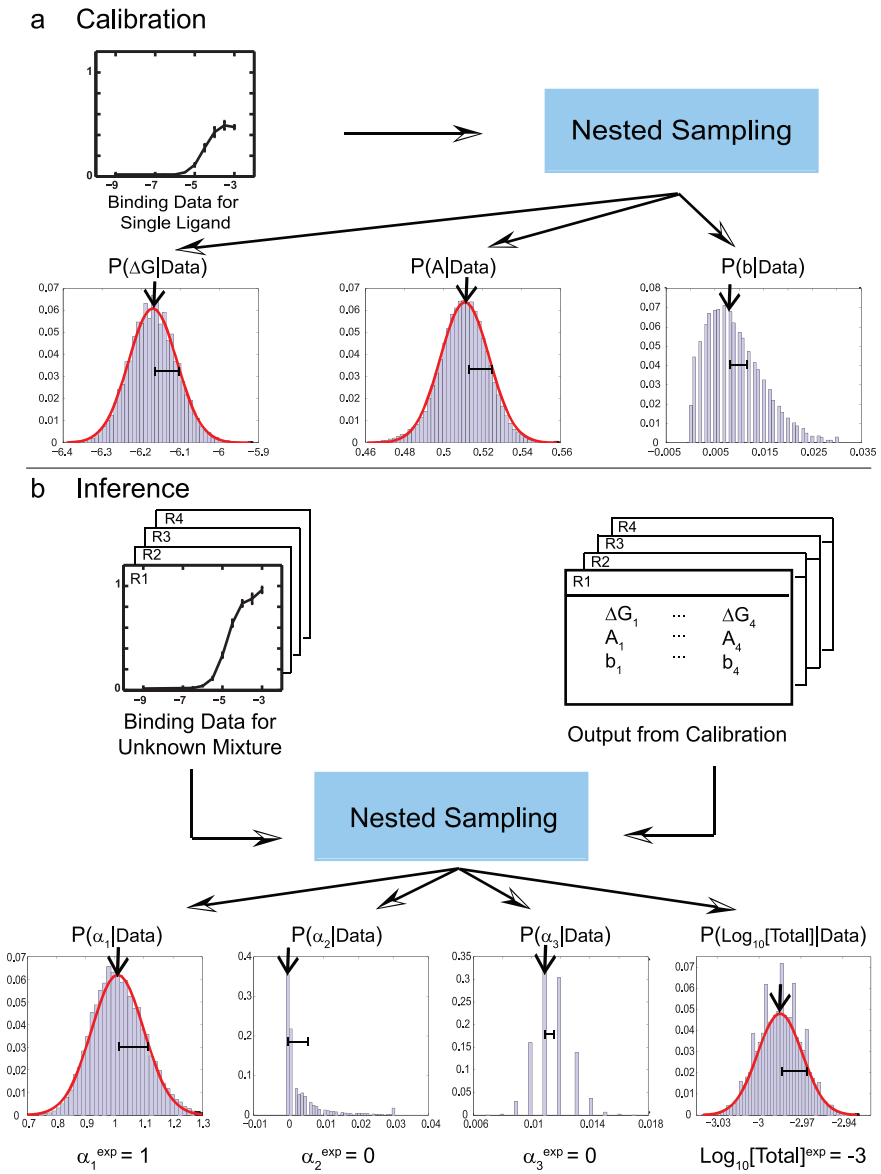

Figure 1. Bayesian algorithm for predicting ligand concentrations in mixtures. (**a**) Calibration of the algorithm: single-ligand, single-receptor binding curves are used to infer binding free energy ΔG, efficacy A and background intensity b for every receptor-ligand combination. Histograms for each predicted parameter are based on an ensemble of 50000 models sampled by Metropolis Monte Carlo [42] starting from the log-likelihood maximum found by nested sampling [38]. Arrows and error bars indicate the most likely value of each parameter and its standard deviation. (**b**) Inference of ligand concentrations in an unknown mixture. Model parameters from (**a**) together with the response curves for all receptors serve as input to the nested sampling algorithm which predicts relative concentrations α_i for each component (with respect to one arbitrarily chosen component, cf. Eq. (7)) and the total concentration $\log_{10} n$ of all ligands in the mixture. Together these predictions yield absolute concentrations for each constituent ligand. Histograms, arrows and error bars have the same meaning as in (**a**), and experimental values are shown below each panel ([Total] = 1 mM at the reference point). For each binding curve, intensity normalized by the maximum intensity on the plate is plotted against $\log_{10} n$ (n is the total ligand concentration in M).
doi:10.1371/journal.pcbi.1002224.g001

concentrations in all 10 cases. However, the inference is consistently less accurate with UDP-containing mixtures, due in part to larger errors in the predicted total concentration. Thus UDP-related efficacies and binding free energies are less optimal than those of other ligands, as will be demonstrated in detail below.

Our second test involved combining UDP-Glc and UDP-Gal in several unequal proportions and applying the resulting mixture to the four-receptor array (**Fig. 3**, **Table S3**). As before, we use a four-ligand model, which should predict zero concentrations for UDP-GlcNAc and UDP. The predicted values of $\alpha_1 = [\mathrm{UDP-Glc}]/[\mathrm{UDP-Gal}]$ show that the ratio of [UDP-Glc] to [UDP-Gal] is successfully ranked in all cases except for the 60/40 and 40/60 mixtures. Apart from the excessive values of $\alpha_2 = [\mathrm{UDP-GlcNAc}]/[\mathrm{UDP-Gal}]$ in the 90/10 and 80/20 cases, which are nonetheless not as large as α_1, concentrations of all ligands absent from the mixture are correctly inferred to be close to zero. We obtain similar results with the alternative definition of α's ($\alpha_1 = [\mathrm{UDP-Gal}]/[\mathrm{UDP-Glc}]$, etc.) (**Table S4**), showing that our approach is not overly sensitive to the arbitrary definition of relative concentrations.

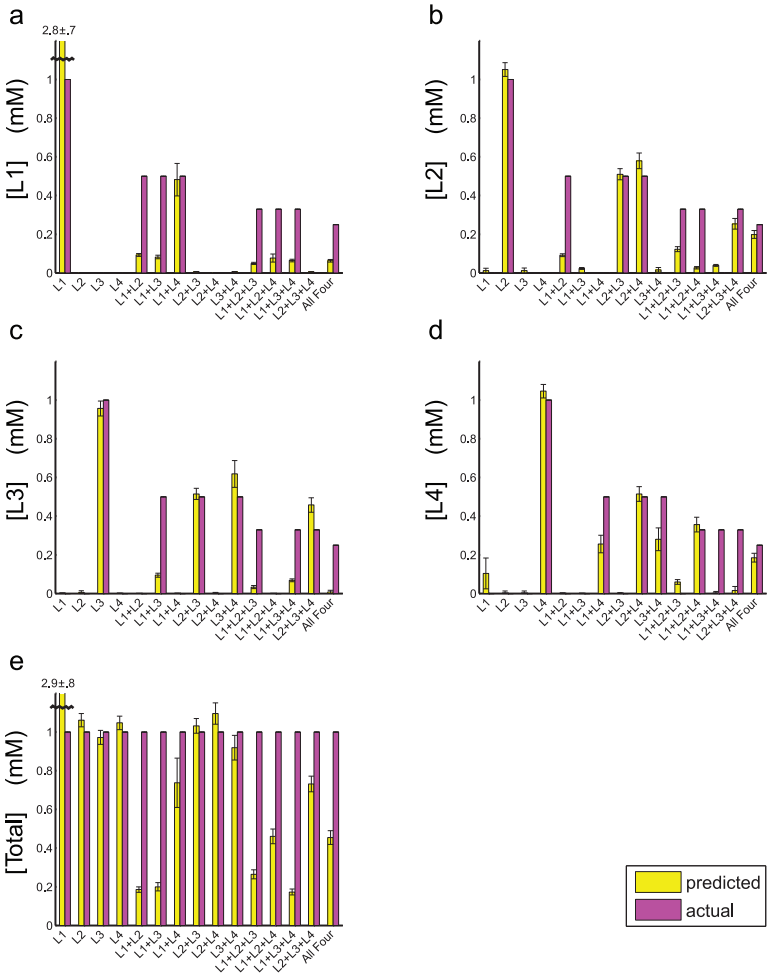

Figure 2. Prediction of ligand concentrations in equal-proportion mixtures. We used nested sampling of a four-receptor, four-ligand model to estimate means and standard deviations for the relative concentrations of all ligands in the mixture and the total ligand concentration at the 1 mM reference point (see **Materials and Methods**). These predictions were converted into absolute concentrations (mM) for each ligand at the 1 mM reference point. L1: UDP, L2: UDP-Gal, L3: UDP-Glc, L4: UDP-GlcNAc.
doi:10.1371/journal.pcbi.1002224.g002

Increasing the number of receptors should improve prediction accuracy by providing additional information about the mixture. To see the extent of these improvements, we have used a variable number of receptors to infer component concentrations in six equal-proportion mixtures of two nucleotide sugars from **Fig. 2** (**Fig. 4**, **Fig. S4**). As expected, the errors rapidly get smaller as the number of receptors is increased, making larger arrays unnecessary. Surprisingly, in several cases adding extra receptors makes the errors somewhat worse before they become better again (see e.g. the R3 and R3/R4 error bars in the UDP+UDP-Gal α_1 panel of **Fig. S4**), indicating that the noise in the new data outweighs the benefit of additional measurements.

As evident from the activation profile of each receptor in response to each ligand (**Fig. S1a**), the receptors differ from each other in fairly subtle ways. In particular, different ligands do not invoke markedly orthogonal profiles of receptor responses. Nonetheless, even with this suboptimal array design, our algorithm provides accurate identification of ligands present in a mixture and a reasonable assessment of the relative amounts of each.

Optimization of sensor array performance

Hessian analysis of sensor arrays. Our Bayesian approach estimates posterior probabilities for the concentration of each component in an arbitrary mixture. With sufficient data, variation of the posterior probability with model parameters is determined by the corresponding log-likelihood (Eq. (8)), which can be visualized as a multidimensional landscape. The global maximum on this landscape corresponds to the model that best describes the data, while the curvature at the maximum shows how sensitive the likelihood is to the change in each parameter. Narrow peaks result in precisely defined parameter values, whereas wide plateaus yield many nearly equivalent predictions and therefore sizable uncertainties in parameter estimates. Expanding the log-likelihood in the vicinity of its maximum yields a Hessian matrix (Eq. (9)), which contains information about standard deviation σ_i of each model parameter γ_i (Eq. (10)) [39]. For example, if the observed receptor response does not depend on γ_i, zero entries appear in the Hessian, leading to the infinite uncertainty σ_i. Making all Hessian matrix elements uniformly larger leads to the smaller σ_i for each predicted parameter γ_i.

Hessian analysis relies on the quadratic expansion in the vicinity of the log-likelihood maximum and hence it is important to check how well it captures the behavior of the more general but computationally intensive nested sampling approach. To create a test case for which the answer is known, we have used Eq. (6) to

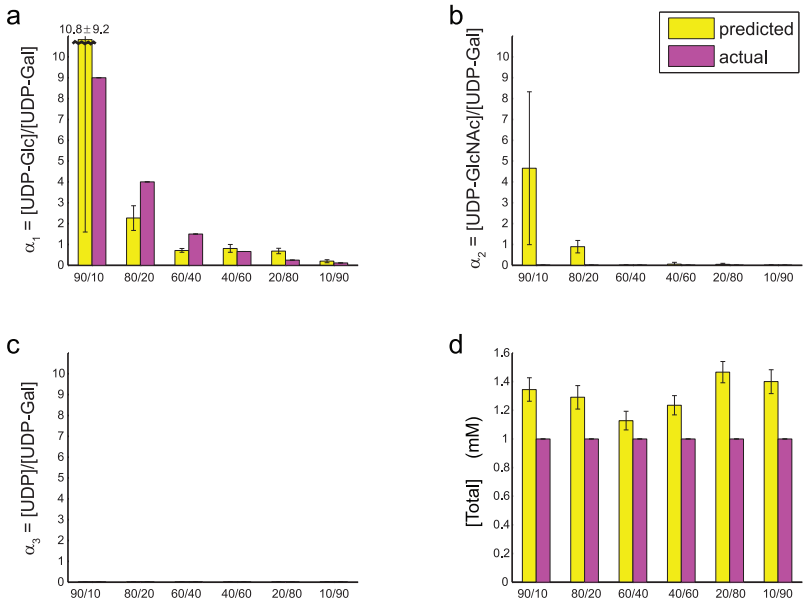

Figure 3. Prediction of ligand concentrations in unequal-proportion binary mixtures of [UDP-Gal] and [UDP-Glc]. We used nested sampling of a four-receptor, four-ligand model to estimate means and standard deviations for the relative concentrations $\alpha_1 = [\mathrm{UDP-Glc}]/[\mathrm{UDP-Gal}]$, $\alpha_2 = [\mathrm{UDP-GlcNAc}]/[\mathrm{UDP-Gal}]$, $\alpha_3 = [\mathrm{UDP}]/[\mathrm{UDP-Gal}]$ and the total concentration (M) [Total] = [UDP-Gal]+[UDP-Glc]+[UDP-GlcNAc]+[UDP] at the 1 mM reference point (see **Materials and Methods**). We found that our predictions were improved if A's and b's were refit to account for "plate bias" (cf. header of **Table S4**): small deviations in the values of A and b (from the standard values shown in **Table S1** and used everywhere else) between measurements 1–3 (Plate 1) and 4–6 (Plate 2).
doi:10.1371/journal.pcbi.1002224.g003

generate synthetic data for 15 equal-proportion mixtures from **Fig. 2** in the low-noise limit ($\tilde{\sigma} = 0.01$ for all receptors, several times smaller than experimental values from **Table S1**). We observe close correspondence between parameter uncertainties inferred from nested sampling vs. Hessian analysis (**Fig. S5**). Moreover, since larger uncertainties make it easier for the average values of predicted parameters to be incorrect, there is also correlation between Hessian errors and the absolute differences between mean predicted and true values (**Fig. S6**). The Hessian-based approach remains useful when experimental data, for which the precise model is unknown and the noise is substantially higher (**Table S1**), is analyzed in the same way: the average over 4 correlation coefficients between Hessian errors and standard deviations from nested sampling (computed for α_1, α_2, α_3 and $\log_{10} n$) is 0.85, and the average over 4 correlation coefficients between Hessian errors and absolute differences between predicted and true values is 0.69. In both real and synthetic cases, the Hessian matrix was computed with correct relative and total concentrations and $\{\Delta G, A, b\}$ values from **Table S1**. We conclude that Hessian errors are a reasonable measure of sensor array performance.

Not all receptors are equally good candidates for inclusion into biosensor arrays – for example, receptors with similar sets of efficacies and binding affinities should be less useful than receptors with more orthogonal binding and activation patterns. Here we make such qualitative insights precise by developing a Hessian approach to biosensor array design. That is, given a certain number of measurements with an array of fixed size (typically, a series in which the total concentration is changed step-by-step within a certain range), we wish to derive the most optimal choice of receptor properties for deciphering the mixture. From the Hessian point of view, the best array will have the smallest errors in predicting component concentrations (Eq. (10)). Because each error is inversely proportional to the determinant of the Hessian, we maximize the determinant instead of minimizing the errors directly. Similarly to the prediction of constituent concentrations, the maximization is carried out by nested sampling [38]. In general, the most optimal receptor parameters and their robustness will depend on the relative concentration of each component in a mixture and on the number of measurements made with the array. For example, an array fine-tuned to detect small admixtures of compound B in the background of compound A may function less well if the concentrations of A and B become approximately equal.

Optimal parameters for a single receptor interacting with a two-ligand mixture. To demonstrate our approach, we first optimize parameters of a single receptor discriminating a mixture of two ligands. By maximizing the determinant of the Hessian, in this case a 2×2 matrix, as a function of two efficacies and two binding energies, we find that the best discrimination is achieved if one ligand acts as an agonist and the other as an antagonist: $A_1 = 1$ and $A_2 = 0$ *or* $A_1 = 0$ and $A_2 = 1$ (for simplicity, background intensities were set to 0 in all sensor array designs). Although in both cases each ligand binds strongly to the receptor, there is a unique set of optimal binding energies ΔG_1 and ΔG_2 for each agonist-antagonist scenario (**Fig. 5a**, **Fig. S7**). The actual values of the binding energies depend on the relative concentration α; for unequal ligand concentrations the two $\{\Delta G_1, \Delta G_2\}$ sets will in general be distinct. This is not surprising since exchanging ligand labels amounts to exchanging relative concentrations of the agonist and the antagonist in the mixture. The height of the peak in both determinant landscapes is the same, indicating that the two alternative solutions lead to equally acceptable array designs as long as the ΔG's are tuned appropriately.

The fine-tuning of binding energies is not necessary if either the total concentration $\log_{10} n$ is known and the task is to minimize the error in predicting the relative concentration α, or vice versa (**Fig. S7** and **S8**). The single-peak landscape structure appears only if

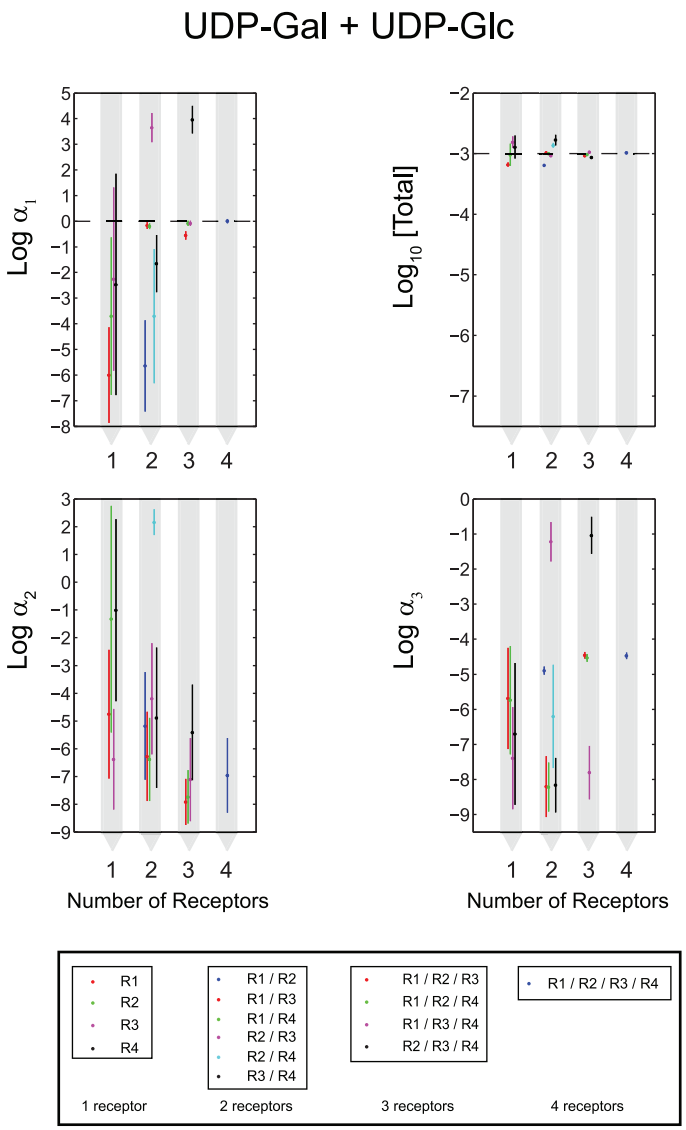

Figure 4. Inference of ligand concentrations is improved with the number of receptors interrogating the mixture. Shown on the log-scale are means and standard deviations for $\alpha_1 = [\mathrm{UDP-Glc}]/[\mathrm{UDP-Gal}]$, $[\mathrm{Total}] = [\text{UDP-Glc}]+[\text{UDP-GlcNAc}]+[\text{UDP-Gal}]+[\mathrm{UDP}]$, $\alpha_2 = [\mathrm{UDP-GlcNAc}]/[\mathrm{UDP-Gal}]$, and $\alpha_3 = [\mathrm{UDP}]/[\mathrm{UDP-Gal}]$. The data are for the 50-50 [UDP-Glc]-[UDP-Gal] binary mixture, leading to $\alpha_1 = 1$, $\alpha_2 = \alpha_3 = 0$, and $\log_{10}[\mathrm{Total}] = -3$ at the reference point. The means and standard deviations were predicted by nested sampling using the four-ligand model and up to four receptors: H-20 (R1), K-3 (R2), L-3 (R3), 2211 (R4).
doi:10.1371/journal.pcbi.1002224.g004

the absolute concentrations of both components need to be predicted together. Strikingly, simultaneous prediction of the total and relative concentrations is impossible with the agonist-agonist receptor response (**Fig. S9**, **Text S1**). The dependence of the optimal binding energies on the value of α is fairly weak (**Fig. S10**). Thus one set of ΔG's optimized for a specific value of α provides a near-optimal solution for a sizable range of ligand concentrations.

Design of multiple-receptor, multiple-ligand arrays. The agonist-antagonist pattern observed in the one-receptor, two-ligand case plays the role of a basic building block when two or more receptors interact with multiple ligands: nested sampling maximization of the Hessian determinant with respect to binding energies ΔG and efficacies A reveals that the array as a whole performs best if each receptor binds one agonist and one antagonist. For example, in the two-receptor, four-ligand case ($N_{rec} = 2$, $N_{lig} = 4$) receptor 1 strongly binds ligands 1 and 3 with $A_1 \simeq 1$ and $A_3 \simeq 0$, whereas receptor 2 strongly binds ligands 2 and 4 with $A_2 \simeq 0$ and $A_4 \simeq 1$ (**Fig. 5b** (***i***)). Each ligand preferentially binds to only one receptor. When another receptor is added to the array, the optimal binding and activation pattern becomes strikingly different: each receptor once again binds both an agonist and an antagonist but ligand 1 now acts as an antagonist to all three receptors (**Fig. 5b** (***ii***)). Each of the other three ligands is an agonist to one of the receptors. In the $N_{rec} = 4$, $N_{lig} = 4$ case each ligand is an agonist for one receptor and an antagonist for another (**Fig. 5b** (***iii***)). Once again, each receptor binds both an agonist and an antagonist. The determinant of the Hessian is dominated by these agonist-antagonist patterns, and is less sensitive to the changes in efficacies and binding energies that do not affect them.

In the light of the observed agonist-antagonist behavior, it is not surprising to see that each receptor can identify concentrations of

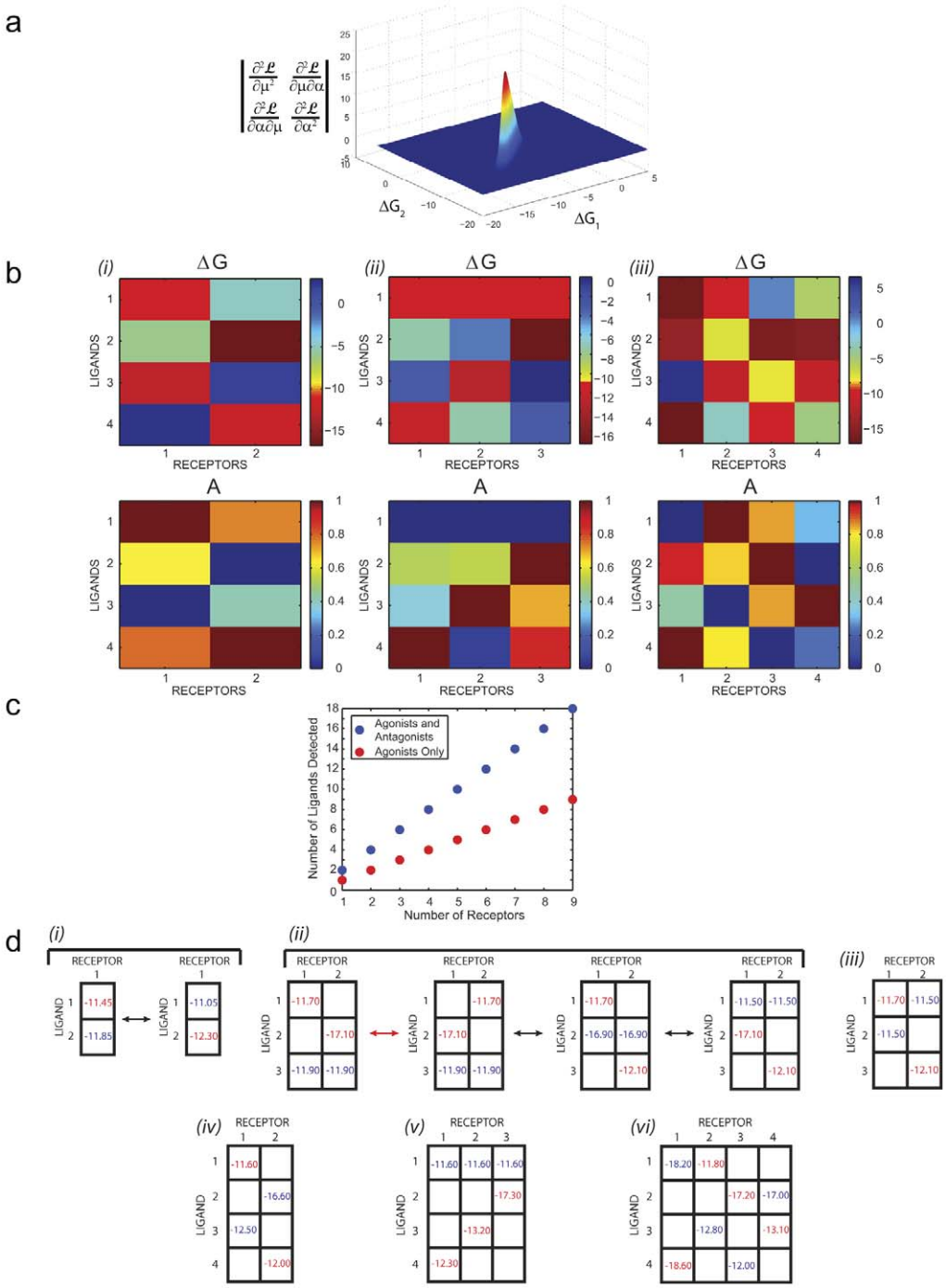

Figure 5. Optimal design of receptor arrays. (**a**) Determinant of the Hessian in the one-receptor, two-ligand case, plotted as a function of binding energies ΔG_1 and ΔG_2. The binding energies at the peak are $\Delta G_1 = -11.45\mathrm{kcal/mol}$ and $\Delta G_2 = -11.85\mathrm{kcal/mol}$. The efficacies are fixed at $A_1 = 1$, $A_2 = 0$; $\alpha = 0.25$. (**b**) Optimal free energies ΔG and efficacies A obtained by maximizing the determinant of the Hessian in the (*i*) two-receptor, four-ligand, (*ii*) three-receptor, four-ligand and (*iii*) four-receptor, four-ligand cases. $\alpha_1 = 10^{-4}, \alpha_2 = 0.1, \alpha_3 = 0.5$. (**c**) The number of successfully discriminated ligands increases linearly with the number of receptors: $N_{lig} = 2N_{rec}$ if the determinant of the Hessian is maximized with respect to all binding energies ΔG and efficacies A (separately for each N_{rec}); $N_{lig} = N_{rec}$ if all ligands are forced to be full agonists with unit efficacies. We call all ligands successfully discriminated if $\sigma_\mu^2 < 2$, $\sigma_{\alpha_i}^2 < 2 (\forall i)$ in a given nested sampling run. Two alternative choices of relative concentrations: $\alpha_i = 1 (\forall i)$ and $\alpha_i = 0.25 (\forall i)$ yielded the same linear dependence on the number of receptors. (**d**) Optimal ΔG values are shown for several cases: (*i*) one-receptor, two-ligand, (*ii*) two-receptor, three-ligand, (*iii*) two-receptor, three-ligand (at a local maximum), (*iv*) two-receptor, four-ligand, (*v*) three-receptor, four-ligand and (*vi*) four-receptor, four-ligand. Values in blue correspond to an efficacy of $A = 0$ (full antagonist), while values in red correspond to $A = 1$ (full agonist). $N_{lig} = 2$: $\alpha = 0.25$; $N_{lig} = 3$: $\alpha_1 = 10^{-4}, \alpha_2 = 0.5$; $N_{lig} = 4$: $\alpha_1 = 10^{-4}, \alpha_2 = 0.1, \alpha_3 = 0.5$. In all cases shown in (**a**)–(**d**), we used 91 datapoints for each receptor (7 replicates with $\log_{10} n^l = \{-3.0, -3.5, \ldots, -8.5, -9.0\}$).
doi:10.1371/journal.pcbi.1002224.g005

at most two ligands (**Fig. 5c**, blue dots). The uncertainty in predicting components of the mixture is minimized if for every receptor one ligand binds strongly as a full agonist and another as a full antagonist. As we have seen, when receptor parameters are less than optimal, the discrimination is still possible but additional receptors may be required: three or four rather than two in the four-ligand case (**Fig. 4**, **Fig. S4**). If we eliminate the agonist-antagonist degree of freedom by setting all efficacies to 1, discriminating N_{lig} requires twice as many receptors (**Fig. 5c**, red dots). In this case each receptor is strongly bound by only one ligand, measuring its concentration independently of the other members of the array. Having access to the full range of receptor responses makes it possible to double the number of ligands in the mixture, but the relationship between N_{rec} and N_{lig} remains linear.

Symmetry properties of optimal sensor arrays. The patterns shown in **Fig. 5b** are not unique – indeed, alternative agonist-antagonist patterns can be generated simply by exchanging receptor labels. Less trivially, a given ligand can be an agonist or an antagonist for different combinations of receptors. In the simplest case of one receptor interacting with two ligands, this symmetry generates two equivalent global maxima discussed above: $A_1=0$, $A_2=1$ and $A_1=1$, $A_2=0$ (**Fig. 5d** (***i***)). In the two-receptor, three-ligand case symmetry arguments combined with extensive sampling yield three global maxima of the Hessian determinant. Each global maximum corresponds to the situation where one of the three ligands acts as an antagonist to both receptors (**Fig. 5d** (***ii***)). The red arrow in **Fig. 5d** (***ii***) indicates a trivial exchange of receptor labels, whereas the black arrows connect three different globally optimal solutions. In addition, there are 9 local maxima with one of the ligands acting either as an agonist to both receptors, or as an agonist to one receptor and an antagonist to the other (e.g. **Fig. 5d** (***iii***); see **Text S1** for a complete enumeration).

In general, $N_{rec}\times N_{lig}$ ΔG's are necessary to characterize all the global and local peaks on the Hessian determinant landscape, with $2N_{rec}$ binding energies describing any given agonist-antagonist pattern. The values of the binding energies depend on the component concentrations in the interrogated mixture. In the $N_{lig}=2N_{rec}$ case all maxima are global and each receptor interacts with two unique ligands. To estimate the benefit of additional receptors, we increased the number of receptors from two to three to four in the four-ligand case (**Fig. 5d** (***iv***), (***v***), (***vi***)). After adding the third receptor the average uncertainty of one total and three relative concentrations, $\langle\sigma_i^2\rangle$, decreased from 0.439 to 0.125. However, only a slight gain was seen when the fourth receptor was added, with $\langle\sigma_i^2\rangle$ becoming 0.101. Thus adding more and more receptors to the array yields increasingly marginal improvements after a certain threshold.

The agonist-antagonist rules described above create readout patterns that are not a simple sum of array responses to single-ligand binding. For one receptor optimized to discriminate two ligands (**Fig. 5d** (***i***)), fluorescent response to the mixture is intermediate between full activation by the agonist and full repression by the antagonist (**Fig. S11a**). This intensity modulation provides enough information for decoding the contents of the mixture. Similarly, in the two-receptor, three-ligand case (**Fig. 5d** (***ii***)) a mixture of all three ligands induces a response with intermediate fluorescense levels (**Fig. S11b**). This pattern is distinct from those induced by single ligands and by binary mixtures with the same relative concentrations as in the three-ligand case.

Performance analysis and improvement of the experimental biosensor array. The design guidelines described above can be used to predict which parameter changes lead to most significant improvements in performance compared to our currently implemented array. Although we do not have direct experimental control over the values of A and ΔG, such insights are useful e.g. for choosing the best combination of several receptors from a larger library. Familiar agonist-antagonist patterns emerge when ΔG's and A's are optimized either separately or together to discriminate an equal-proportion, four-ligand mixture (**Fig. S12**). In particular, if A's are kept fixed, ΔG's for the most distant pair of A's become more favorable for each receptor, creating an agonist-antagonist pair (**Fig. S12c**). Conversely, if ΔG's are fixed, the values of A corresponding to the two lowest ΔG's become more distant from each other (**Fig. S12d**). Not surprisingly, the agonist-antagonist patterns are even more pronounced if both ΔG's and A's are allowed to relax (**Fig. S12e** and **S12f**). Because two and certainly three optimized receptors are sufficient for discriminating four-ligand mixtures (**Fig. 5c**), the fourth receptor, which does not follow the usual pattern as strongly as the other three, appears to be superfluous. Similarly to the cases shown in **Fig. 5d**, **Fig. S12e** and **S12f** represent only one solution from a large family of local and global maxima of the Hessian determinant, which are related by permutations of receptor and ligand indices. Optimizing receptor-ligand parameters leads to a sizable improvement in array performance: with $\tilde{\sigma}=1$ for all receptors, $\sigma_\mu=2.64$, $\sigma_{\alpha_1}=5.46$, $\sigma_{\alpha_2}=16.47$, $\sigma_{\alpha_3}=3.19$ for the original array, whereas $\sigma_\mu=0.69$, $\sigma_{\alpha_1}=0.68$, $\sigma_{\alpha_2}=0.95$, $\sigma_{\alpha_3}=0.68$ for the array in which both A's and ΔG's have been optimized.

For the experimentally implemented four-receptor GPCR array, nested sampling errors are consistently larger when UDP is present in the mixture (**Fig. 2**, **Table S2**). This observation is consistent with Hessian analysis: for example, the average Hessian uncertainties for three UDP-free binary mixtures are $\langle\sigma_\mu\rangle=0.77$, $\langle\sigma_{\alpha_1}\rangle=6.12$, $\langle\sigma_{\alpha_2}\rangle=11.09$, $\langle\sigma_{\alpha_3}\rangle=0.0002$. For three UDP-containing binary mixtures, the average Hessian errors are $\langle\sigma_\mu\rangle=19.09$, $\langle\sigma_{\alpha_1}\rangle=5.57$, $\langle\sigma_{\alpha_2}\rangle=113.44$, $\langle\sigma_{\alpha_3}\rangle=57.72$ (as before, all Hessian errors are computed with correct concentrations and $\tilde{\sigma}=1$ for all receptors). The Hessian determinants are also consistently smaller for UDP-containing binary mixtures. These observations indicate that UDP parameters are further away from the optimal four-receptor array designed to analyze an equal-proportion binary mixture: either A's or ΔG's need to be changed in order to create stronger agonist-antagonist patterns.

Discussion

We have developed a Bayesian algorithm that allows determination of all the constituents in an unknown mixture from the output of a cross-specific sensor array. Our algorithm employs a physical picture of sensor-analyte interactions to model the non-linear relationship between ligand concentrations and the reporter response. After appropriate calibration of each sensor's response to each analyte of interest, the algorithm interprets the integrated output of the entire array and, with a sufficient number of variably tuned sensors, reliably returns the amount of each chemical in a complex mixture.

We also provide quantitative guidelines for designing optimal sets of sensors. Three general principles emerged from our computational and theoretical studies of array design. First, the optimal parameters of the sensors exhibit weak dependence on the relative amounts of compounds in a mixture. Thus a given set of optimal sensors will remain near-optimal through a sizable range of ligand concentrations. Nonetheless, analyzing a mixture where both compounds are present in roughly similar amounts is better accomplished with a set of sensors different from those fine-tuned

to measure a small amount of one compound in the presence of a large excess of the other.

Second, the maximum number of ligands in a mixture whose levels can all be determined simultaneously is simply twice the number of sensors in the array. This linear relationship is different from the exponential relationship between ligands and receptors in olfactory systems [4,12]. The problem addressed by the olfactory system, to recognize a very large number of individual odors with a limited repertoire of receptors, is not the same as that solved by our algorithm, to determine all the constituents in a complex mixture. In fact, even the most skilled human nose can simultaneously detect and distinguish no more than a handful of odorants.

Third, the optimum design of receptors for the array demands that one of the ligands function as a strong agonist of a receptor and a second ligand as a strong antagonist of that receptor. Antagonists sharpen the discriminatory powers of the array by heightening the differences in the receptor response to individual compounds. As a result, a mixture of chemicals produces an array readout which is not a superposition of responses to individual ligands, and whose intensity pattern may be fine-tuned for maximum recognition through receptor-ligand binding energies. Accordingly, odors that function as antagonists to a subset of olfactory receptors could potentially increase the discriminatory power of the olfactory system, and in particular enable it to resolve mixtures that contain those odors. Recent analysis of olfactory receptors suggests that some odorants do possess antagonist activity [14,17–20]. Our theoretical framework provides a rationale for the existence of such antagonists and underscores their role in both olfactory systems and artificial receptor arrays.

Materials and Methods

Targeted mutagenesis and selection of functional receptor mutants

The L-3 mutant was isolated using a procedure similar to that previously employed with the H-20 and K-3 mutants [36]: oligonucleotides with randomized sequences corresponding to the codons to be mutagenized were utilized to generate overlapping PCR products. The L-3 motif corresponds to amino acid residues LLxSA on TM7. Mutant libraries were generated by gap repair using overlapping PCR products and transformed to media selective for recombined plasmids. To select for functional mutants, libraries were replica-plated to selective SC-His media [40] containing one of six ligands: UDP-Gal, UDP-Glc, UDP-galNAc, UDP-GlcNAc, UDP or dTDP-glucose (50 μL of 1 mM solution spread on 30 mL of SC "Leu-His agar medium in 8.5 cm Petri plates). Yeast growth media was supplemented by 1 mM 3AT, a competitive inhibitor of the *HIS3* reporter gene product, which sets the threshold for reporter gene activation. Functional receptor mutants that showed qualitatively disparate responses to the panel of ligands were selected for further analysis. Among these, the H-20 and K-3 mutants, described earlier [36], and the L-3 mutant, described here, were selected to be utilized alongside the 2211 "parent" in a four-receptor array for analysis of mixtures of UDP-Glc, UDP-Gal, UDP-GlcNAc and UDP.

$\beta-$galactosidase assays

Our $\beta-$galactosidase assays were based on microtiter assays described previously [41]. Yeast strains expressing each of the four mutant receptors were diluted to OD_{600} of ~ 0.05 in flasks. Cultures were then grown overnight in 100 mL selective media to an OD_{600} of $0.1-0.2$. Serial dilutions of each ligand or mixture of ligands were prepared in yeast culture medium in 96-well culture blocks. Ligands or mixtures of ligands were transferred in 20 μL aliquots in quadruplicate to deep-well polypropylene 384$-$well plates using a BioMek robotic liquid handler. 180 μL of suspended yeast cells in medium (undiluted from the overnight cultures) were then aliquotted into each well and mixed. The cultures were sealed with foil tape and incubated at 30°C on a plate shaker at 400–500 rpm for 4 hours (H-20, K-3 and L-3 receptors) or overnight (2211 receptor). After incubation, $\beta-$galactosidase substrate [41] (FDG solution; 0.5 mM fluorescein di-beta-D-galactopyranoside, 2.3% Triton X-100, and 0.127 M Pipes, pH 7.2) was mixed with an equal volume of Pierce Y-PER solution (Thermo Scientific) and distributed in 25 μL aliquots to black 384$-$well plates. 50 μL aliquots of the yeast/ligand cultures were then transferred into the black 384$-$well plates and mixed gently but thoroughly by pipetting, taking care to avoid generating bubbles. A single layer of paper towel was placed on top of each plate and the plates were then individually wrapped in aluminum foil and incubated without shaking at 37°C for approximately one hour before reading on an automated fluorescent plate reader (Perkin Elmer EnVision).

Microtiter plate-based assays are often subject to edge- or plate-bias due to uneven heating or discrepancies in timing across a single plate or among plates. While no obvious plate effects were seen, it is very difficult to control for all possible variations in a single experiment. Due to the number of samples and the need to make efficient use of materials, each of the mixture experiments was split across two plates per receptor. In mixtures of equal proportions, samples containing UDP, UDP-Gal and UDP-GlcNAc but lacking UDP-Glc were on Plate 1, while all mixtures containing UDP-Glc were on Plate 2. In the UDP-Gal/UDP-Glc binary mixtures of unequal proportions, samples containing 90%, 80% or 60% UDP-Glc were on Plate 1, while samples containing 40%, 20% or 10% UDP-Glc were on Plate 2.

For each single ligand or combination of ligands, a series of measurements was performed at several values of the total concentration $n^l=\sum_{i=1}^{N_{lig}} n_i^l$ (M): $\log_{10} n^l=\{-3.0,-3.5,\ldots,-6.5,-9.0\}$ for H-20, K-3, L-3 and $\{-5.0,-5.5,\ldots,-9.0\}$ for 2211. The total chemical potential $\mu^l=k_BT\log n^l$ is then given by $\mu^l=\mu-\sum_{i=1}^{l-1}\Delta\mu^i \quad (l=1\ldots N)$, where N is the number of measurements in the series, $\Delta\mu^l$ are known chemical potential differences between two consecutive measurements, and $\mu=-3k_BT\log(10)$ is the chemical potential at the 1 mM reference point. Note that in order to reconstruct the total chemical potential for all points in the series, only μ needs to be predicted. Each series of measurements was replicated four times; fluorescence counts were normalized to 1.0 separately for each plate (**Dataset S1**).

Models of receptor response to ligand binding

For a single receptor interacting with a single ligand, we model the normalized reporter fluorescent intensity as:

$$I^l(\Delta G,A,b)=A\frac{e^{-\beta(\Delta G-\mu^l)}}{1+e^{-\beta(\Delta G-\mu^l)}}+b, \qquad (1)$$

where A is the receptor efficacy, b is the background intensity (a small amount of background fluorescence observed in the absence of ligand binding), ΔG is the free energy of receptor-ligand binding, $\beta=1/k_BT$ (k_B is the Boltzmann constant, and T is the temperature), and μ^l is the chemical potential.

We compute the log-likelihood of the data by assuming that fluorescence measurements are Gaussian-distributed around values from Eq. (1):

$$L = \log P(\{\tilde{I}\}|\Delta G,A,b,\tilde{\sigma}) = -\frac{1}{2\tilde{\sigma}^2}\sum_{l=1}^{N}\left[I^l(\Delta G,A,b)-\tilde{I}^l\right]^2 - \frac{N}{2}\log(2\pi\tilde{\sigma}^2), \quad (2)$$

where $\tilde{I}^l (l=1\ldots N)$ are measured intensities and $\tilde{\sigma}$ is the noise parameter. The log-likelihood is used to estimate the posterior probability of all model parameters according to the Bayes' formula [38]:

$$P(\Delta G,A,b,\tilde{\sigma}|\{\tilde{I}\}) = \frac{P(\{\tilde{I}\}|\Delta G,A,b,\tilde{\sigma})P(\Delta G)P(A)P(b)P(\tilde{\sigma})}{P(\{\tilde{I}\})}, \quad (3)$$

where on the right-hand side the likelihood from Eq. (2) is multiplied by the product of priors for each model parameter and divided by evidence. $\{\tilde{I}\}$ combines data from all experimental replicates. We use uniform priors (invariant with respect to translations, $x \rightarrow x+a$):

$$P(x) = \begin{cases} 1/(x_{max}-x_{min}) & \text{if } x\in[x_{min},x_{max}], \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

for ΔG, A and b, and Jeffrey's priors (invariant with respect to rescaling, $x \rightarrow ax$) for $\tilde{\sigma}$:

$$P(x) = \begin{cases} 1/(x\log(x_{max}/x_{min})) & \text{if } x\in[x_{min},x_{max}], \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

We have used $(\Delta G_{min},\Delta G_{max}) = (-20.0,5.0)$kcal/mol, $(A_{min},A_{max}) = (0.0,1.0)$, $(b_{min},b_{max}) = (0.0,1.0)$, $(\tilde{\sigma}_{min},\tilde{\sigma}_{max}) = (0.001,100.0)$ in our calculations.

The reporter response to a mixture of ligands is given by

$$I_k^l = \sum_{m=1}^{N_{lig}} A_m^k p_m^{k,l} + b^k, \quad (6)$$

where $p_m^{k,l} = e^{-\beta(\Delta G_m^k - \mu_m^l)}/Z^{k,l}$ is the probability that receptor k is bound by ligand m and $Z^{k,l} = 1+\sum_{i=1}^{N_{lig}} e^{-\beta(\Delta G_i^k - \mu_i^l)}$ is the partition function. ΔG_m^k is the binding free energy between receptor k and ligand m ($k=1\ldots N_{rec}$, $m=1\ldots N_{lig}$), A_m^k is the efficacy, and b^k is the background intensity. The background intensity for receptor k is the average from all calibration experiments involving that receptor. $\mu_m^l = k_B T \log n_m^l$ is the chemical potential of ligand m, which can be expressed through the total chemical potential μ^l and the relative concentrations $\alpha_m = n_{m+1}^l/n_1^l$ ($\forall l$, $m=1\ldots N_{lig}-1$):

$$\mu_1^l = \mu^l + k_B T \log\frac{1}{S}, \quad (7)$$

$$\mu_m^l = \mu^l + k_B T \log\frac{\alpha_{m-1}}{S}, m=2,3,\ldots$$

where $S = 1+\sum_{i=1}^{N_{lig}-1}\alpha_i$. Note that an arbitrary choice of the ligand in the denominator leads to several equivalent representations of the relative concentrations.

The log-likelihood of the observed pattern of fluorescence intensities from multiple receptors interrogated by a mixture of ligands is given by

$$L = \log P(\{\tilde{I}\}|\{\alpha\},\mu,\{\tilde{\sigma}\}) = -\sum_{k=1}^{N_{rec}}\left\{\frac{1}{2\tilde{\sigma}_k^2}\sum_{l=1}^{N_k}[I_k^l(\{\alpha\},\mu)-\tilde{I}_k^l]^2 + \frac{N_k}{2}\log(2\pi\tilde{\sigma}_k^2)\right\}. \quad (8)$$

Here $I_k^l(\{\alpha\},\mu)$ is defined in Eq. (6) (in the interests of brevity, we suppress its dependence on $\{A,b,\Delta G\}$ for each receptor-ligand combination). $\tilde{I}_k^l$ denotes fluorescence measured for receptor k at the total chemical potential μ_l, N_k is the total number of measurements, and $\tilde{\sigma}_k$ is the noise parameter. Similarly to Eq. (3), the log-likelihood is used to estimate the posterior probability $P(\{\alpha\},\mu,\{\tilde{\sigma}\}|\{\tilde{I}\})$. We employ a uniform prior for μ with $(\mu_{min},\mu_{max}) = (-10.0,-2.0)$ and a Jeffrey's prior for α's with $(\alpha_{min},\alpha_{max}) = (0.0001,100.0)$.

We estimate all posterior probabilities by nested sampling [38] – a Bayesian Monte-Carlo (MC) technique that yields an ensemble of models from which the average value of each parameter and its standard deviation are computed. Unlike other methods such as MC sampling of the product of likelihood and priors, nested sampling allows us to keep track of the evidence, yielding absolute values of the posterior probability.

Hessian analysis

The Hessian matrix in the low-noise limit can be written as (**Text S1**):

$$\frac{\partial^2 L}{\partial\gamma_i\partial\gamma_j} = -\sum_{k=1}^{N_{rec}}\frac{1}{\tilde{\sigma}_k^2}\sum_{l=1}^{N_k}\frac{\partial I_k^l}{\partial\gamma_i}\frac{\partial I_k^l}{\partial\gamma_j}, \quad (9)$$

where $\{\gamma\} = (\{\alpha\},\mu)$, L is the log-likelihood, N_k is the total number of measurements for receptor k, and $\tilde{\sigma}_k$ is the noise parameter. Uncertainties σ_i for each predicted parameter γ_i are given by the diagonal elements of the inverse Hessian matrix:

$$\sigma_i^2 = -\left\|\frac{\partial^2 L}{\partial\gamma_i\partial\gamma_j}\right\|_{ii}^{-1}. \quad (10)$$

Software used in this study, called RANSA (Receptor Array Nested Sampling Algorithm), is available at http://olfaction.rutgers.edu.

Supporting Information

Figure S1 Overview of the GPCR-based biosensor array. (**a**) Each receptor-ligand combination is tested for functional activation, yielding 16 binding curves. For each curve, intensity normalized by the maximum intensity on the plate is plotted against $\log_{10}n$ (n is the ligand concentration in M). The error bars on each curve are from four biological replicates. The single-receptor, single-ligand binding curves are used to calibrate the physical model by inferring ΔG, A and b separately for each receptor-ligand combination. (**b**) An unknown mixture of four ligands is applied to each of the four receptors. The resulting fluorescent response curves together with the $\{\Delta G,A,b\}$ predictions are used as input to the Bayesian algorithm designed to predict absolute concentrations of each ligand in the mixture. R1: H-20, R2: K-3, R3: L-3, R4: 2211. (EPS)

Figure S2 Prediction accuracy increases with the number of measurements at different total concentrations. Synthetic data was generated using Eq. (1) in the main text, with $A=0.8$, $b=0.2$, and $\Delta G=-6.5$ kcal/mol. To account for experimental error, Gaussian noise with $\tilde{\sigma}=0.02$ was added to the intensity from Eq. (1). The maximum total concentration of the ligand was gradually increased as shown in the nine panels on top, yielding more and more complete binding curves: $\log_{10}[\text{max}]=\{-4.75, -4.5, -4.25, -4.0, -3.75, -3.5, -3.0, -2.75, -2.5\}$. $\log_{10}[\text{min}]$ was -9.0 in all cases, and 4 replicates with 9 datapoints per curve were created for each concentration range. In each panel I^l is plotted as a function of μ^l in the absence of noise. For each concentration range, 1000 nested sampling runs were carried out to predict ΔG, A and b. The standard deviation $\sigma_{\Delta G}$ from each run was averaged and plotted in the bottom panel as a function of the total range of ligand concentrations $k_B T \ln([\text{max}]/[\text{min}])$. Each dot in the bottom panel is color-coded to correspond to a particular binding curve on top.
(EPS)

Figure S3 Prediction accuracy decreases with the amount of noise in the data. Synthetic data was generated using Eq. (1) in the main text, with $A=0.8$, $b=0.2$, and $\Delta G=-6.5$ kcal/mol. In analogy with the experiments, we used the concentration range $\log_{10} n^l=\{-3.0, -3.5, \ldots, -6.5, -9.0\}$ and created 4 replicates, yielding 36 datapoints. To model the increase in experimental error, Gaussian noise with $\tilde{\sigma}$ ranging from 0.01 to 0.17 was added to the intensity from Eq. (1). For each value of $\tilde{\sigma}$, 1000 nested sampling runs were carried out to predict ΔG, A and b. The standard deviation $\sigma_{\Delta G}$ from each run was averaged and plotted as a function of $\tilde{\sigma}$.
(EPS)

Figure S4 Inference of ligand concentrations is improved with the number of receptors interrogating the mixture. Shown on the log-scale are means and standard deviations for three relative concentrations (α_1, α_2, α_3) and the total concentration, predicted by nested sampling using the four-ligand model and up to four receptors: H-20 (R1), K-3 (R2), L-3 (R3), 2211 (R4). Each experiment has a 50-50 binary mixture of two ligands indicated on top of each panel, leading to $\alpha_1=1$, $\alpha_2=\alpha_3=0$, and $\log_{10}[\text{Total}]=-3$ at the reference point ([Total] = [UDP-Glc]+[UDP-GlcNAc]+[UDP-Gal]+[UDP]). **UDP-Gal+UDP-GlcNAc mixture:** $\alpha_1=[\text{UDP}-\text{GlcNAc}]/[\text{UDP}-\text{Gal}]$, $\alpha_2=[\text{UDP}-\text{Glc}]/[\text{UDP}-\text{Gal}]$, $\alpha_3=[\text{UDP}]/[\text{UDP}-\text{Gal}]$. **UDP-Glc+UDP-GlcNAc mixture:** $\alpha_1=[\text{UDP}-\text{GlcNAc}]/[\text{UDP}-\text{Glc}]$, $\alpha_2=[\text{UDP}-\text{Gal}]/[\text{UDP}-\text{Glc}]$, $\alpha_3=[\text{UDP}]/[\text{UDP}-\text{Glc}]$. **UDP+UDP-Glc mixture:** $\alpha_1=[\text{UDP}]/[\text{UDP}-\text{Glc}]$, $\alpha_2=[\text{UDP}-\text{Gal}]/[\text{UDP}-\text{Glc}]$, $\alpha_3=[\text{UDP}-\text{GlcNAc}]/[\text{UDP}-\text{Glc}]$. **UDP+UDP-GlcNAc mixture:** $\alpha_1=[\text{UDP}]/[\text{UDP}-\text{GlcNAc}]$, $\alpha_2=[\text{UDP}-\text{Glc}]/[\text{UDP}-\text{GlcNAc}]$, $\alpha_3=[\text{UDP}-\text{Gal}]/[\text{UDP}-\text{GlcNAc}]$. **UDP+UDP-Gal mixture:** $\alpha_1=[\text{UDP}]/[\text{UDP}-\text{Gal}]$, $\alpha_2=[\text{UDP}-\text{Glc}]/[\text{UDP}-\text{Gal}]$, $\alpha_3=[\text{UDP}-\text{GlcNAc}]/[\text{UDP}-\text{Gal}]$.
(EPS)

Figure S5 Hessian uncertainties vs. standard deviations from nested sampling. Synthetic data was generated as four replicates for each of the 15 equal-proportion mixtures from **Fig. 2**, using parameters from **Table S1** and $\tilde{\sigma}=0.01$ for all receptors (Eq. (6) in the main text). For each receptor, concentration ranges were taken from the corresponding experiment (**Materials and Methods**). For each parameter, a Hessian error computed using Eq. (10) in the main text (x-axis) is compared with the standard deviation from a nested sampling run (y-axis). Nested sampling simultaneously infers relative concentrations α_1, α_2, α_3, the total concentration and $\tilde{\sigma}$'s given receptor-ligand parameters from **Table S1**.
(EPS)

Figure S6 Hessian uncertainties vs. errors in parameter predictions. Synthetic data was generated as four replicates for each of the 15 equal-proportion mixtures from **Fig. 2**, using parameters from **Table S1** and $\tilde{\sigma}=0.01$ for all receptors (Eq. (6) in the main text). For each receptor, concentration ranges were taken from the corresponding experiment (**Materials and Methods**). For each parameter, a Hessian error computed using Eq. (10) in the main text (x-axis) is compared with the absolute magnitude of the difference between the mean value predicted by nested sampling and the correct value (y-axis). Nested sampling simultaneously infers relative concentrations α_1, α_2, α_3, the total concentration and $\tilde{\sigma}$'s given receptor-ligand parameters from **Table S1**.
(EPS)

Figure S7 Matrix elements and the determinant of the Hessian in the antagonist-agonist case, plotted as a function of binding energies ΔG_1 and ΔG_2 in the one-receptor, two-ligand system. The efficacies are fixed at $A_1=0$, $A_2=1$; $\alpha=0.25$. We used 7 replicates with $\log_{10} n^l=\{-3.0, -3.5, \ldots, -8.5, -9.0\}$. The values of the binding energies at the peak of the determinant landscape are $\Delta G_1=-11.05\text{kcal/mol}$ and $\Delta G_2=-12.30\text{kcal/mol}$.
(EPS)

Figure S8 Matrix elements of the Hessian in the agonist-antagonist case, plotted as a function of binding energies ΔG_1 and ΔG_2 in the one-receptor, two-ligand system. The efficacies are fixed at $A_1=1$, $A_2=0$; $\alpha=0.25$. We used 7 replicates with $\log_{10} n^l=\{-3.0, -3.5, \ldots, -8.5, -9.0\}$.
(EPS)

Figure S9 Matrix elements and the determinant of the Hessian in the agonist-agonist case, plotted as a function of binding energies ΔG_1 and ΔG_2 in the one-receptor, two-ligand system. The efficacies are fixed at $A_1=1$, $A_2=1$; $\alpha=1$. We used 7 replicates with $\log_{10} n^l=\{-3.0, -3.5, \ldots, -8.5, -9.0\}$.
(EPS)

Figure S10 Changes in the Hessian determinant with the concentration of the second ligand. Each determinant is plotted as a function of binding energies ΔG_1 and ΔG_2 for a given value of $\alpha=n_2/n_1$. The efficacies are fixed at $A_1=1$, $A_2=0$. We used 7 replicates with $\log_{10} n^l=\{-3.0, -3.5, \ldots, -8.5, -9.0\}$. Shown in each panel are the optimal ΔG_1 and ΔG_2 corresponding to the maximum value of the determinant.
(EPS)

Figure S11 Schematic diagram of receptor activation by single ligands and ligand mixtures with optimized binding affinities and efficacies. (**a**) Two cases of agonist-antagonist single-receptor arrays designed to discriminate a mixture of two ligands (**Fig. 5d** (***i***)). (**b**) Array of two receptors designed to discriminate a mixture of three ligands (**Fig. 5d** (***ii***)). In both cases we show relative intensities (normalized to 1.0) corresponding to $\log_{10} n=-3.0$. $N_{lig}=2$: $\alpha=0.25$; $N_{lig}=3$: $\alpha_1=10^{-4}$,$\alpha_2=0.5$ (leading to $\alpha=10^{-4}$, 0.5 and 5×10^3 for the binary combinations of ligands 1–2, 2–3 and 1–3, respectively). 4.7(0.99) and -1.8(0.47) are ΔG's and A's (in parentheses) for receptor-ligand interactions outside of the dominant agonist-antagonist pattern.
(EPS)

Figure S12 Improving performance of the experimentally implemented sensor array. Free energies ΔG (**a**) and efficacies A (**b**) in the experimentally implemented sensor array with parameters from **Table S1**. Free energies ΔG (**c**) and efficacies A (**d**) from two sensor arrays in which the determinant of the Hessian was maximized only with respect to ΔG's and A's,

respectively. Free energies ΔG (**e**) and efficacies A (**f**) in the optimal sensor array in which the determinant of the Hessian was maximized with respect to both ΔG's and A's. The determinant was computed using four replicates of an equal-proportion mixture of four ligands and $\tilde{\sigma}=1$ for all receptors. For each receptor, concentrations were taken from the corresponding experiment (**Materials and Methods**). In panels **a**–**d**, the order of ligands is L1: UDP, L2: UDP-Gal, L3: UDP-Glc, L4: UDP-GlcNAc. The order of receptors is R1: H-20, R2: K-3, R3: L-3, R4: 2211. Note that ligand and receptor identities are lost in panels **e**,**f** since all parameters have been optimized.
(EPS)

Dataset S1 Normalized fluorescence intensity measurements from experiments with equal-proportion (Fig. 2) and unequal-proportion (Fig. 3) mixtures.
(XLS)

Table S1 Parameters of receptor-ligand interactions predicted from one-receptor, one-ligand binding curves (UDP-Gal, UDP-Glc, UDP-GlcNAc) and one-receptor, two-ligand binding curves (UDP). ΔG is the receptor-ligand binding free energy (kcal/mol), A is the receptor efficacy, b is the background intensity, and $\tilde{\sigma}$ is the noise parameter which quantifies the discrepancy between the model and the observed binding curves. Due to antagonistic activity of UDP, 50/50 UDP+UDP-Glc binary mixture was used with K-3, L-3, 2211 and 50/50 UDP+UDP-Gal binary mixture was used with H-20 to predict UDP parameters (compound concentrations were set to their exact values for these calibration predictions). In each case, the mixture was chosen on the basis of the smallest standard deviation of ΔG.
(PDF)

Table S2 Prediction of ligand concentrations in equal-proportion mixtures (data for Fig. 2).
(PDF)

Table S3 Prediction of ligand concentrations in unequal-proportion binary mixtures of [UDP-Gal] and [UDP-Glc] (data for Fig. 3).
(PDF)

Table S4 Prediction of ligand concentrations in unequal-proportion binary mixtures of [UDP-Gal] and [UDP-Glc] using an alternative definition of relative concentrations. We used nested sampling of a four-receptor, four-ligand model to infer relative concentrations $\alpha_1=[\mathrm{UDP-Gal}]/[\mathrm{UDP-Glc}]$, $\alpha_2=[\mathrm{UDP-GlcNAc}]/[\mathrm{UDP-Glc}]$, $\alpha_3=[\mathrm{UDP}]/[\mathrm{UDP-Glc}]$, as well as the total concentration [Total] = [UDP-Gal]+[UDP-Glc]+[UDP-GlcNAc]+[UDP] at the $1\,\mathrm{mM}$ reference point. A's and b's were refit to account for "plate bias": small systematic deviations in the values of A and b (from the standard values shown in **Table S1** and used everywhere else) between different plates. For Plate 1 (measurements 1–3), $A_{\mathrm{UDP-Gal}}$, $A_{\mathrm{UDP-Glc}}$ and b were set to $\{0.94,0.79,0.64,0.50\}$, $\{0.91,0.81,0.90,0.81\}$ and $\{0.02,0.15,0.02,0.12\}$ for H-20, K-3, L-3 and 2211, respectively. For Plate 2 (measurements 4–6), the corresponding values were $\{0.96,0.93,0.94,0.94\}$, $\{0.85,0.79,0.86,0.95\}$ and $\{0.02,0.13,0.03,0.11\}$. $A_{\mathrm{UDP-GlcNAc}}$ and A_{UDP} were taken from **Table S1**.
(PDF)

Text S1 Supplementary Materials and Methods.
(PDF)

Acknowledgments

JT wishes to thank Allan Haldane for technical advice.

Author Contributions

Conceived and designed the experiments: ADA JRB. Performed the experiments: ADA. Analyzed the data: JT AVM. Wrote the paper: ADA AVM.

References

1. Buck L (1996) Information coding in the vertebrate olfactory system. Annu Rev Neurosci 19: 517–544.
2. Firestein S (2001) How the olfactory system makes sense of scents. Nature 413: 211–218.
3. Breer H (2003) Olfactory receptors: molecular basis for recognition and discrimination of odors. Anal Bioanal Chem 377: 427–433.
4. Touhara K, Vosshall L (2009) Sensing odorants and pheromones with chemosensory receptors. Annu Rev Physiol 71: 307–332.
5. Hopfield J (1999) Odor space and olfactory processing: Collective algorithms and neural implementation. Proc Natl Acad Sci U S A 96: 12506–12511.
6. Su CY, Menuz K, Carlson J (2009) Olfactory perception: Receptors, cells, and circuits. Cell 139: 45–59.
7. Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell 65: 175–187.
8. Zhang X, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. Nature Neurosci 5: 124–133.
9. Malnic B, Godfrey P, Buck L (2004) The human olfactory receptor gene family. Proc Natl Acad Sci U S A 101: 2584–2589.
10. Godfrey P, Malnic B, Buck L (2004) The mouse olfactory receptor gene family. Proc Natl Acad Sci U S A 101: 2156–2161.
11. Hasin-Brumshtein Y, Lancet D, Olender T (2009) Human olfaction: from genomic variation to phenotypic diversity. Trends Genet 25: 178–184.
12. Malnic B, Hirono J, Sato T, Buck L (1999) Combinatorial receptor codes for odors. Cell 96: 713–723.
13. Saito H, Kubota M, Roberts R, Chi Q, Matsunami H (2004) Rtp family members induce functional expression of mammalian odorant receptors. Cell 119: 679–691.
14. Shirokova E, Schmiedeberg K, Bedner P, Niessen H, Willecke K, et al. (2005) Identification of specific ligands for orphan olfactory receptors. J Biol Chem 280: 11807–11815.
15. Hallem E, Carlson J (2006) Coding of odors by a receptor repertoire. Cell 125: 143–160.
16. Saito H, Chi Q, Zhuang H, Matsunami H, Mainland J (2009) Odor coding by a mammalian receptor repertoire. Sci Signal 2: ra9.
17. Spehr M, Gisselmann G, Poplawski A, Riffell J, Wetzel C, et al. (2003) Identification of a testicular odorant receptor mediating human sperm chemotaxis. Science 299: 2054–2058.
18. Araneda R, Peterlin Z, Zhang X, Chesler A, Firestein S (2004) A pharmacological profile of the aldehyde receptor repertoire in rat olfactory epithelium. J Physiol 555(Pt 3): 743–756.
19. Oka Y, Omura M, Kataoka H, Touhara K (2004) Olfactory receptor antagonism between odorants. EMBO J 23: 120–126.
20. Sanz G, Schlegel C, Pernollet JC, Briand L (2005) Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. Chem Senses 30: 69–80.
21. Albert K, Lewis N, Schauer C, Sotzing G, Stitzel S, et al. (2000) Cross-reactive chemical sensor arrays. Chem Rev 100: 2595–2626.
22. Rakow N, Suslick K (2000) A colorimetric sensor array for odour visualization. Nature 406: 710–713.
23. Wright A, Anslyn E (2006) Differential receptor arrays and assays for solution-based molecular recognition. Chem Soc Rev 35: 14–28.
24. Shabbir S, Joyce L, da Cruz G, Lynch V, Sorey S, et al. (2009) Pattern-based recognition for the rapid determination of identity, concentration, and enantiomeric excess of subtly different threo diols. J Am Chem Soc 131: 13125–13131.
25. Jurs P, Bakken G, McClelland H (2000) Computational methods for the analysis of chemical sensor array data from volatile analytes. Chem Rev 100: 2649–2678.
26. Gao T, Woodka MD, Brunschwig BS, Lewis NS (2006) Chemiresistors for array-based vapor sensing using composites of carbon black with low volatility organicmolecules. ChemMater 18: 5193–5202.
27. Kitamura M, Shabbir S, Anslyn E (2009) Guidelines for pattern recognition using differential receptors and indicator displacement arrays. J Org Chem 74: 4479–4489.
28. Feuerstein T, Limberger N (1999) Mathematical analysis of the control of neurotransmitter release by presynaptic receptors as a supplement to experimental data. Arch Pharmacol 359: 345–359.
29. Zhang C, Bailey DP, Suslick KS (2006) Colorimetric sensor arrays for the analysis of beers: A feasibility study. J Agr Food Chem 54: 4925–4931.

30. Zhang C, Suslick KS (2007) Colorimetric sensor array for soft drink analysis. J Agr Food Chem 55: 237–242.
31. Dutta R, Kashwan KR, Bhuyan M, Hines EL, Gardner JW (2003) Electronic nose based tea quality standardization. Neural Netw 16: 847–853.
32. Pardo M, Sberveglieri G (2002) Coffee analysis with an electronic nose. IEEE Trans Inst Measur 51: 1334–1339.
33. Heilig A, Barsan N, Weimar U, SchweizerBerberich M, Gopel W, et al. (1997) Gas identification by modulating temperatures of SnO2-based thick film sensors. Sens Actuators B 43: 45–51.
34. White J, Kauer JS, Dickinson TA, Walt DR (1996) Rapid analyte recognition in a device based on optical sensors and the olfactory system. Anal Chem 68: 2191–2202.
35. Woodka MD, Brunschwig BS, Lewis NS (2007) Use of spatiotemporal response information from sorption-based sensor arrays to identify and quantify the composition of analyte mixtures. Langmuir 23: 13232–13241.
36. Ault A, Broach J (2006) Creation of GPCR-based chemical sensors by directed evolution in yeast. Protein Eng Des Sel 19: 1–8.
37. Klein C, Paul J, Sauve K, Schmidt M, Arcangeli L, et al. (1998) Identification of surrogate agonists for the human fprl-1 receptor by autocrine selection in yeast. Nature Biotechnol 16: 1334–1337.
38. Sivia D, Skilling J (2006) Data analysis: A Bayesian tutorial. Oxford: Oxford University Press.
39. Gutenkunst R, Waterfall J, Casey F, Brown K, Myers C, et al. (2007) Universally sloppy parameter sensitivities in systems biology models. PLoS Comp Biol 3: 1871–1878.
40. Kaiser C, Michaelis S, Mitchell A (1994) Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual. Cold Spring HarborNY: Cold Spring Harbor Laboratory Press.
41. Chambers J, Macdonald L, Sarau H, Ames R, Freeman K, et al. (2000) A G protein-coupled receptor for UDP-glucose. J Biol Chem 275: 10767–10771.
42. Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, E T (1953) Equation of state by fast computing machines. J Chem Phys 21: 1087–1092.