

# Chemogenomic Analysis of G-Protein Coupled Receptors and Their Ligands Deciphers Locks and Keys Governing Diverse Aspects of Signalling

Jörg D. Wichard<sup>1,3</sup>, Antonius ter Laak<sup>3</sup>, Gerd Krause<sup>1</sup>, Nikolaus Heinrich<sup>3</sup>, Ronald Kühne<sup>1\*</sup>, Gunnar Kleinau<sup>1,2</sup>

**1** Leibniz-Institut für Molekulare Pharmakologie, Berlin, Germany, **2** Institute of Experimental Pediatric Endocrinology, Charité Universitätsmedizin Berlin, Berlin, Germany, **3** Bayer-Schering Pharma, Berlin, Germany

## Abstract

Understanding the molecular mechanism of signalling in the important super-family of G-protein-coupled receptors (GPCRs) is causally related to questions of how and where these receptors can be activated or inhibited. In this context, it is of great interest to unravel the common molecular features of GPCRs as well as those related to an active or inactive state or to subtype specific G-protein coupling. In our underlying chemogenomics study, we analyse for the first time the statistical link between the properties of G-protein-coupled receptors and GPCR ligands. The technique of mutual information (MI) is able to reveal statistical inter-dependence between variations in amino acid residues on the one hand and variations in ligand molecular descriptors on the other. Although this MI analysis uses novel information that differs from the results of known site-directed mutagenesis studies or published GPCR crystal structures, the method is capable of identifying the well-known common ligand binding region of GPCRs between the upper part of the seven transmembrane helices and the second extracellular loop. The analysis shows amino acid positions that are sensitive to either stimulating (agonistic) or inhibitory (antagonistic) ligand effects or both. It appears that amino acid positions for antagonistic and agonistic effects are both concentrated around the extracellular region, but selective agonistic effects are cumulated between transmembrane helices (TMHs) 2, 3, and ECL2, while selective residues for antagonistic effects are located at the top of helices 5 and 6. Above all, the MI analysis provides detailed indications about amino acids located in the transmembrane region of these receptors that determine G-protein signalling pathway preferences.

**Citation:** Wichard JD, ter Laak A, Krause G, Heinrich N, Kühne R, et al. (2011) Chemogenomic Analysis of G-Protein Coupled Receptors and Their Ligands Deciphers Locks and Keys Governing Diverse Aspects of Signalling. PLoS ONE 6(2): e16811. doi:10.1371/journal.pone.0016811

**Editor:** Leo Lee, University of Hong Kong, Hong Kong

**Received:** September 2, 2010; **Accepted:** January 12, 2011; **Published:** February 4, 2011

**Copyright:** © 2011 Wichard 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:** JW, AtL and NH are employees of Bayer-Schering Pharma, Berlin, Germany, who had a role in study design, data collection and analysis, decision to publish, and preparation of the manuscript.

**Competing Interests:** JW, AtL and NH are employees of Bayer-Schering Pharma, Berlin, Germany. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials, as detailed online in the guide for authors <http://www.plosone.org/static/policies.action#sharing>, and there are no relevant patents, products in development or marketed products to declare.

\* E-mail: [kuehne@fmp-berlin.de](mailto:kuehne@fmp-berlin.de)

## Introduction

G-protein-coupled receptors (GPCRs) constitute a large super-family of transmembrane receptors which convey extracellular signals into the intracellular region to effect sensory perception, chemotaxis, neurotransmission, cell communication and several other physiological events. The importance of GPCRs arises from their role as signal transmitters and regulators. In humans around 850 GPCRs are known [1] and several diseases are caused by GPCR malfunction [2–4]. They can be activated by a wide variety of endogenous stimuli such as amino acids, peptides, ions and (pher-) hormones [5]. GPCRs are subdivided into several families [6], whereby the largest family is the rhodopsin-like family A. Therefore, understanding these complex proteins and related signaling systems is of enormous importance, not least for drug discovery [7–11]. This is reflected by the fact that GPCRs are the largest target group for therapeutics [12] including up to 40% of currently marketed drugs [13].

Different structural parts of GPCRs are responsible for specific intra- as well as intermolecular functions during a sequential signal

transduction process consisting of: i. receiving a stimulus, ii. transmission of the stimulus by inducing conformational changes of the receptor iii. intracellular presentation of determinants enabling activation of signal transducers such as G-proteins [14]. Most of the endogenous and synthetic ligands of family A GPCRs are thought to bind within the transmembrane domain close to the second extracellular loop 2 (ECL2) [11]. Based on a huge amount of experimental data a “global toggle switching” mechanism is assumed to take place during ligand induced activation, whereby a vertical see-saw movement of transmembrane helix (TMH) 6 occurs around a pivot [15,16]. In consequence activation is characterized by a spatial re-arrangement of the TMHs and to the greatest extent between TMHs 5, 6, and 7 [17,18]. This structural re-arrangement is supported by amino acids acting as “micro-switches” [19,20]. In addition, contacts between ECL2 and the extracellular extensions of the helices have been proposed to participate as regulars during activation [21–23]. Different GPCR conformations are related to different signalling activity states [4,19,24] and several family A GPCR crystal structures were solved either in the inactive [18,25–27] or the active conformation [28,29].

On the intracellular side GPCRs interact with heterotrimeric guanine nucleotide-binding proteins (G-proteins), which play a crucial role in signal transduction towards second messenger cascades. G-protein subtypes are distinguished by their specific alpha-subunits. The main members are termed  $G\alpha_s$ ,  $G\alpha_q$  and  $G\alpha_i$ , whereby the induced effect on secondary messengers is considered (e.g. s-stimulation, or i-inhibition). GPCR mediated G-protein activation is characterized by structural shifts inside and between the G-protein subunits, followed by exchange of GDP for GTP in the  $\alpha$ -subunit and separation of the  $G\alpha$  from the  $G\beta\gamma$ -subunits. This opens up interfaces on the G-protein subunits to potential contact partners such as phospho-diesterase [30]. A detailed understanding of which structural features of GPCRs are related to selectivity of G-protein coupling and which factors or determinants are responsible for promiscuity in G-protein coupling of GPCRs is not yet available. The most prominent hypotheses addressing this topic are: 1) different conformational states of a receptor are selective for a certain G-protein subtype, since extracellular mutations and diverse ligands can cause different G-protein-subtype preferences for a single receptor [31–34]; 2) distinctive selective interaction patterns in terms of particular intracellular residues exist, which are responsible for G-protein subtype specific interactions [35–38]; 3) the G-protein preference is determined by the set of cell-specific G-protein subtype(s) [39]. None of these possibilities can be assigned as the main cause for G-protein preference for all GPCRs. Therefore, it is most likely that a combination of different factors define the G-protein portfolio of a certain GPCR.

The link between the variation in receptor properties and signal transduction specificities was studied using different theoretical methods. Earlier studies involving information theory measures for GPCR analysis focused on the receptor sequences and the interrelationship of the amino acid positions. Oliveira et al. used entropy-variability plots to detect functionally conserved residues [40]. Ye et al. proposed a two-entropy analysis to determine the functional positions in the transmembrane regions of GPCRs [41]. Fatakia et al. combined mutual information (MI) and concepts from graph theory to reveal correlated positions in the major GPCR classes [42].

The underlying study describes for the first time a chemogenomics approach on the activation mechanism and G-protein coupling of GPCRs. Chemogenomics is originally defined as a method that discovers active and/or selective ligands for biologically related targets in a systematic manner [43] and an extremely broad and significant example of chemogenomics is reported by Keiser et al. who quantitatively grouped and related hundreds of drug targets based on the chemical similarity of 65,000 ligands [44]. For GPCRs and other target families, Jabob and Vert proposed a more integrated chemogenomics method, as they apply their *Support Vector Machine* (SVM) machine learning algorithm to the joint chemical and the biological space, which is an advantage as it makes targets with few known ligands benefit from the data points of similar targets [45]. Recently, Van der Horst et al. presented a classification of GPCRs that is purely based on their ligands, complementing the classical sequence-based phylogenetic classifications of these receptors [46]. Such substructure-based and ligand-based phylogenetic classifications of GPCRs may help to unravel potential cross-reactivities of GPCR ligands.

The aim of this study is to identify features that link the sequence variation of family A GPCRs and their G-protein preference with the corresponding structural variation of agonistic and antagonistic ligands applying the concept of mutual information (MI). The novelty of our approach lies in the

application of the MI concept to chemogenomics of GPCRs by including both the target sequence and ligand properties. In contrast to previous studies, we investigated the MI between GPCR residue variations in specific positions and the molecular properties of ligands. Furthermore, we also analyzed such receptor/ligand property correlations according to types of preferred G-protein activation. Strikingly, our analyses provide detailed indications about the effects of specific ligand properties and the determination of G-protein preferences in the transmembrane domain of GPCRs.

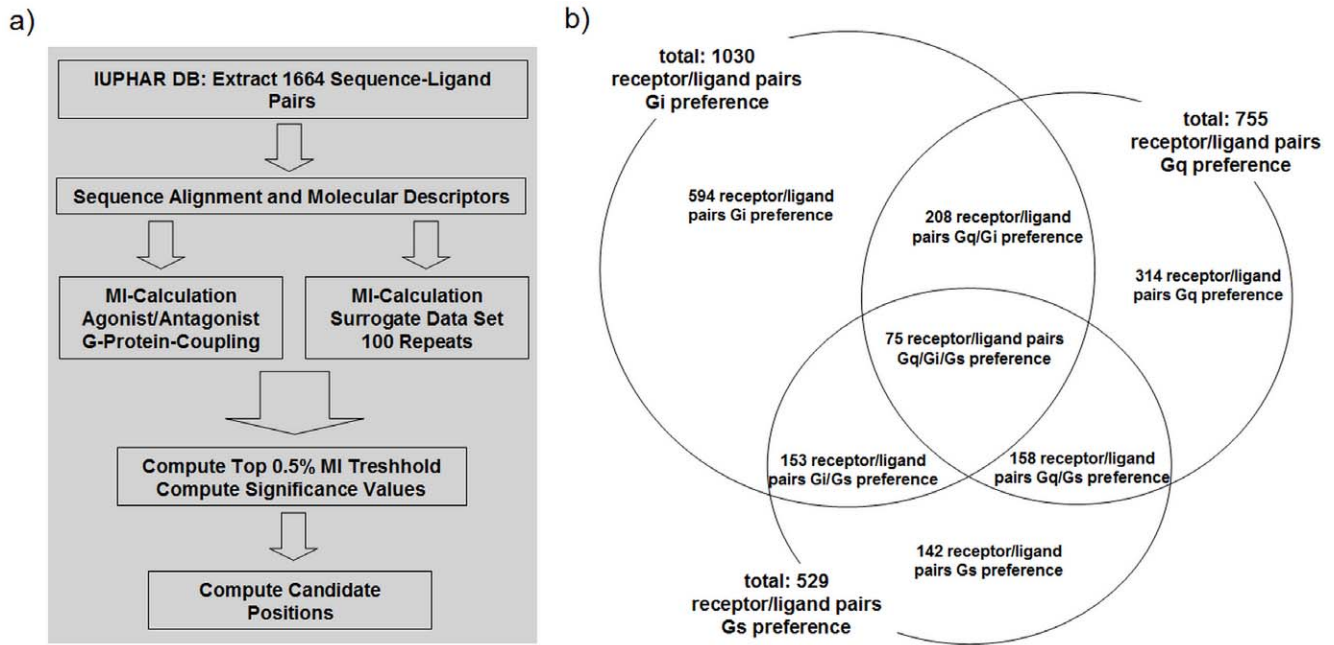
## Results

### Database, GPCR alignment, ligand description and mutual information

We gathered 100 family A GPCRs with known ligands and information regarding their preferred G-protein subtype from the standardized IUPHAR [47] database (figure 1a). These GPCRs cover over 30 receptor sub-families. Several of the extracted GPCRs couple only to a particular G-protein subtype, while most of them are able to undergo dual G-protein coupling to Gq/Gs, Gq/Gi or Gs/Gi pairs (figure 1b). Few receptors are known to activate all three G-protein subtypes. We assigned agonistic and antagonistic ligands to each receptor, whereby the individual receptor preference for G-protein coupling was also noted. More than 1660 receptor-ligand pairs containing 767 full agonists, 184 partial agonists and 713 antagonists were collected (figure 1b). In order to calculate the mutual information it is necessary to use appropriate descriptions of the sequence space and the ligand space. The representation of the sequence space is based on the multiple sequence alignment of the 100 analysed GPCRs using a profile hidden Markov model. The sequence alignment is provided under: <http://fimp-berlin.info/research/structuralbiology/researchgroups/drug-design/downloads.html>. To represent the ligand space we calculated a set of discrete and countable molecular descriptors (see section *Material and Methods*). The goal of mutual information consists in finding correlations between the variation of residues at a certain sequence position and the variation of ligand properties. Several data sets were built in order to calculate the MI for sequence-agonist, sequence-antagonists, sequence-Gi, sequence-Gs and sequence-Gq correlations. The data are shown tabulated in Table S1.

### Hot spot positions for antagonistic and agonistic effects on GPCRs

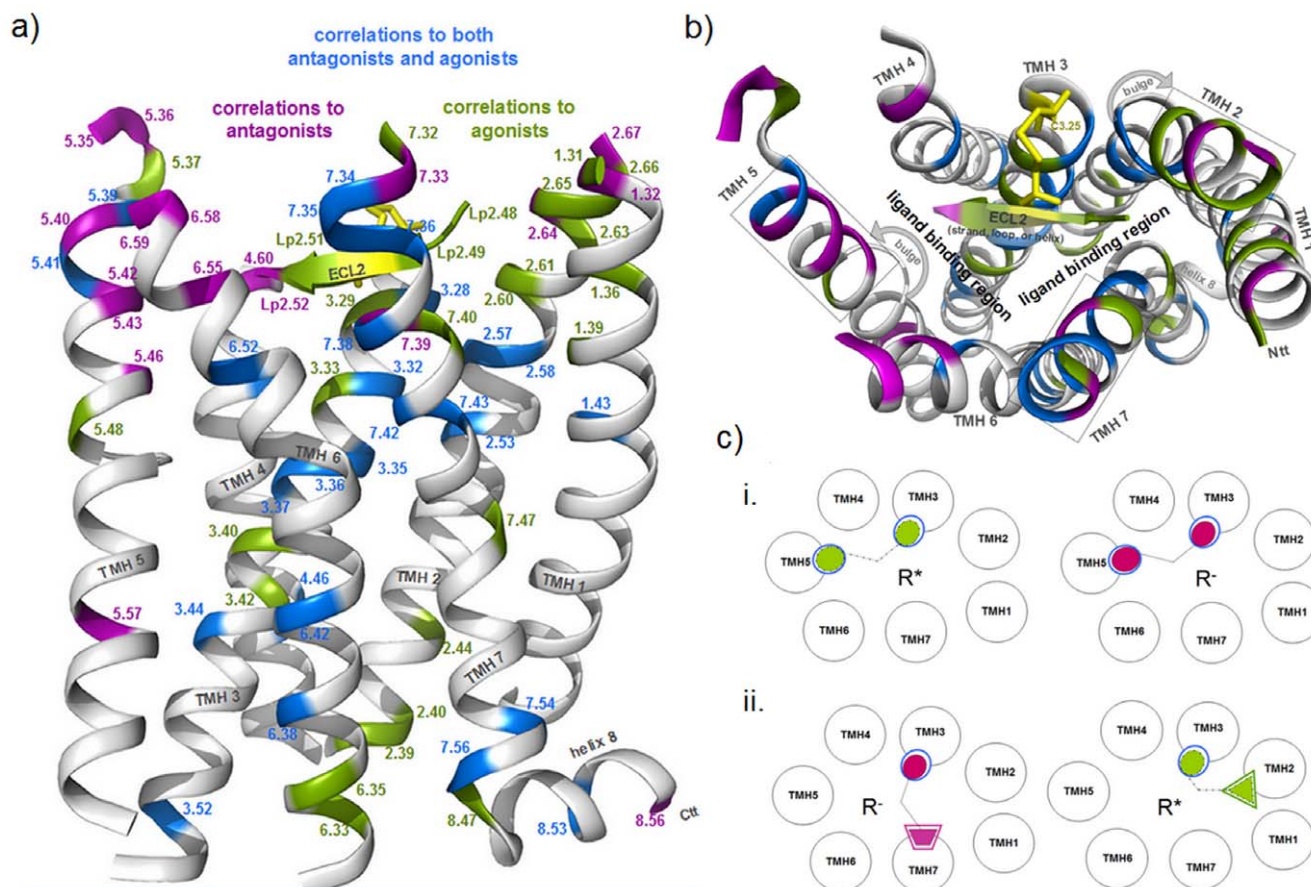
A relatively high value for the mutual information between receptor residue variations and ligand properties with respect to antagonistic or agonistic ligand induced effects, lead to the identification of common as well as selective GPCR positions. The MI-identified hot spot positions involved in ligand induced stabilization of the GPCR inactive state and ligand induced receptor activation are shown in figure 2. We have mapped this information onto the three-dimensional structure of rhodopsin for spatial assignment of sensitive positions (figure 3a). First of all, our analysis shows that the common family A GPCR ligand-binding region is highly sensitive to correlated residue-ligand properties (figure 3a,b). It is of special note that the side chains identified by our study are mostly oriented inwards towards the transmembrane helical bundle (figure 3a). Selective antagonistic positions are clustered around the extracellular region, especially between TMH5 and TMH6. Positions extracted for selective agonistic effects in the helices are mainly located between TMH1, TMH2 and TMH3 (e.g. 1.31, 1.36; 2.40, 2.44; 3.33) compared to selective antagonistic sensitive positions located mainly between TMH5 and TMH6 (e.g. 5.36, 5.40, 5.57).



**Figure 1. Extracted information from the IUPHAR database used for analyses.** a) This workflow presents the methods we used to reveal the candidates with the highest mutual information between receptor positions and ligand features. b) We gathered the GPCR information from the *IUPHAR database* and extracted the 2D chemical structures from the *PubChem Structure Database*. In total 1664 receptor-ligand pairs were collected for analysis, whereby 100 different family A GPCRs from 30 sub-classes with known ligands were considered. Only full agonists, partial agonists and antagonists were explored. For the receptors their Gs, Gi and Gq coupling preferences were assigned based on the annotations in the *IUPHAR database*. Several receptors are known as to be promiscuous by their capability to activate two or three G-protein subtypes. doi:10.1371/journal.pone.0016811.g001

| Pos           | Ago | Antago | Pos    | Ago | Antago | Pos    | Ago | Antago | Pos      | Ago | Antago | Pos    | Ago | Antago | Pos    | Ago | Antago | Pos           | Ago | Antago |               |    |   |
|---------------|-----|--------|--------|-----|--------|--------|-----|--------|----------|-----|--------|--------|-----|--------|--------|-----|--------|---------------|-----|--------|---------------|----|---|
| 1.31          | 2   | -      | 2.67   | 3   | 7      | 3.25   | -   | -      | Lp2.52   | -   | 1      | 5.35   | 6   | 11     | 6.59   | -   | 4      | 7.32          | 4   | -      | 8.56          | -  | 1 |
| 1.32          | -   | 1      | 2.66   | 1   | -      | 3.26   | -   | -      | Lp2.51   | 6   | -      | 5.36   | -   | 2      | 6.58   | 3   | 11     | 7.33          | -   | 4      | 8.55          | -  | - |
| 1.33          | -   | -      | 2.65   | 4   | 1      | 3.27   | -   | -      | Lp2.50 C | -   | -      | 5.37   | 2   | -      | 6.57   | -   | -      | 7.34          | 2   | 1      | 8.54          | -  | - |
| 1.34          | -   | -      | 2.64   | -   | 3      | 3.28   | 2   | 1      | Lp2.49   | 7   | -      | 5.38   | -   | -      | 6.56   | -   | -      | 7.35          | 1   | 3      | 8.53          | 2  | 4 |
| 1.35          | -   | -      | 2.63   | 3   | -      | 3.29   | 6   | 2      | Lp2.48   | 1   | -      | 5.39   | 3   | 4      | 6.55   | 1   | 4      | 7.36          | 5   | 6      | 8.52          | -  | - |
| 1.36          | 3   | -      | 2.62   | -   | -      | 3.30   | -   | -      | 4.63     | -   | -      | 5.40   | -   | 3      | 6.54   | -   | -      | 7.37          | -   | -      | 8.51          | -  | - |
| 1.37          | -   | -      | 2.61   | 1   | -      | 3.31   | -   | -      | 4.62     | -   | -      | 5.41   | 5   | 5      | 6.53   | -   | -      | 7.38          | 3   | 4      | 8.50 F        | -  | - |
| 1.38          | -   | -      | 2.60   | -   | -      | 3.32   | 2   | 4      | 4.61     | -   | -      | 5.42   | 5   | 12     | 6.52   | 6   | 6      | 7.39          | -   | 3      | 8.49          | -  | - |
| 1.39          | 8   | 3      | 2.59   | -   | -      | 3.33   | 2   | -      | 4.60     | -   | 11     | 5.43   | 2   | 6      | 6.51   | -   | -      | 7.40          | 6   | 2      | 8.48          | -  | - |
| 1.40          | -   | -      | 2.58   | 1   | 1      | 3.34   | -   | -      | 4.59     | -   | -      | 5.44   | -   | -      | 6.50 P | -   | -      | 7.41          | -   | -      | 8.47          | 10 | 2 |
| 1.41          | -   | -      | 2.57   | 4   | 6      | 3.35   | 2   | 3      | 4.58     | -   | -      | 5.45   | -   | -      | 6.49   | -   | -      | 7.42          | 5   | 5      | n descriptors | -  | - |
| 1.42          | -   | -      | 2.56   | -   | -      | 3.36   | 16  | 12     | 4.57     | 3   | 3      | 5.46   | 4   | 9      | 6.48   | -   | -      | 7.43          | 2   | 2      | -             | -  | - |
| 1.43          | 2   | 2      | 2.55   | -   | -      | 3.37   | 12  | 16     | 4.56     | -   | -      | 5.47   | -   | -      | 6.47   | -   | -      | 7.44          | -   | -      | -             | -  | - |
| 1.44          | -   | -      | 2.54   | -   | -      | 3.38   | -   | -      | 4.55     | -   | -      | 5.48   | 1   | -      | 6.46   | -   | -      | 7.45          | -   | -      | -             | -  | - |
| 1.45          | -   | -      | 2.53   | 7   | 5      | 3.39   | -   | -      | 4.54     | -   | -      | 5.49   | -   | -      | 6.45   | -   | -      | 7.46          | -   | -      | -             | -  | - |
| 1.46          | -   | -      | 2.52   | -   | -      | 3.40   | 1   | -      | 4.53     | -   | -      | 5.50 P | -   | -      | 6.44   | -   | -      | 7.47          | 8   | -      | -             | -  | - |
| 1.47          | -   | -      | 2.51   | -   | -      | 3.41   | -   | -      | 4.52     | -   | -      | 5.51   | -   | -      | 6.43   | -   | -      | 7.48          | -   | -      | -             | -  | - |
| 1.48          | -   | -      | 2.50 D | -   | -      | 3.42   | 9   | 1      | 4.51     | -   | -      | 5.52   | -   | -      | 6.42   | 2   | 2      | 7.49 N        | -   | -      | -             | -  | - |
| 1.49          | -   | -      | 2.49   | -   | -      | 3.43   | -   | -      | 4.50 W   | -   | -      | 5.53   | -   | -      | 6.41   | -   | -      | 7.50 P        | -   | -      | -             | -  | - |
| 1.50 N        | -   | -      | 2.48   | -   | -      | 3.44   | 2   | 4      | 4.49     | -   | -      | 5.54   | -   | -      | 6.40   | -   | -      | 7.51          | -   | -      | -             | -  | - |
| 1.51          | -   | -      | 2.47   | -   | -      | 3.45   | -   | -      | 4.48     | -   | -      | 5.55   | -   | -      | 6.39   | -   | -      | 7.52          | -   | -      | -             | -  | - |
| 1.52          | -   | -      | 2.46   | -   | -      | 3.46   | -   | -      | 4.47     | -   | -      | 5.56   | -   | -      | 6.38   | 2   | 5      | 7.53 Y        | -   | -      | -             | -  | - |
| 1.53          | -   | -      | 2.45   | -   | -      | 3.47   | -   | -      | 4.46     | 2   | 1      | 5.57   | -   | 2      | 6.37   | -   | -      | 7.54          | 7   | 8      | -             | -  | - |
| 1.54          | -   | -      | 2.44   | 2   | -      | 3.48   | -   | -      | 4.45     | -   | -      | 5.58   | -   | -      | 6.36   | -   | -      | 7.55          | -   | -      | -             | -  | - |
| 1.55          | -   | -      | 2.43   | -   | -      | 3.49 D | -   | -      | 4.44     | -   | -      | 5.59   | -   | -      | 6.35   | 1   | -      | 7.56          | 4   | 4      | -             | -  | - |
| 1.56          | -   | -      | 2.42   | -   | -      | 3.50 R | -   | -      | 4.43     | -   | -      | 5.60   | -   | -      | 6.34   | -   | -      | n descriptors | -   | -      | -             | -  | - |
| 1.57          | -   | -      | 2.41   | -   | -      | 3.51 Y | -   | -      | 4.42     | -   | -      | 5.61   | -   | -      | 6.33   | 2   | -      | -             | -   | -      | -             | -  | - |
| 1.58          | -   | -      | 2.40   | 2   | -      | 3.52   | 1   | 2      | 4.41     | -   | -      | 5.62   | -   | -      | 6.32   | -   | -      | -             | -   | -      | -             | -  | - |
| n descriptors | -   | -      | 2.39   | 5   | 2      | 3.53   | -   | -      | 4.40     | -   | -      | 5.63   | -   | -      | 6.31   | -   | -      | -             | -   | -      | -             | -  | - |
| n descriptors | -   | -      | 2.38   | -   | -      | 3.54   | -   | -      | 4.39     | -   | -      | 5.64   | -   | -      | -      | -   | -      | -             | -   | -      | -             | -  | - |
| n descriptors | -   | -      | 3.55   | -   | -      | -      | -   | -      | 4.38     | -   | -      | 5.65   | -   | -      | -      | -   | -      | -             | -   | -      | -             | -  | - |

**Figure 2. Identified positions for antagonistic and agonistic effects.** This table summarizes the identified hot-spots of correlated GPCR residue positions and specific ligand properties by separating antagonistic and agonistic effects on signaling. Details can be found in Table S1. GPCR positions are given using the Ballesteros and Weinstein numbering [92]. For the second extracellular loop the highly conserved cysteine in ECL2, which in most family A GPCRs is linked to the cysteine in TMH3 (position 3.25), is numbered as Lp2.50. "Lp" indicates that this is a loop and number 2 that it is the second loop. For visualisation different colours are used according to the assigned functional effects: green - selective for agonists; lilac - selective for antagonist; blue - chemical descriptors of both agonists and antagonists are correlated and the number of descriptors is comparable. Highly conserved amino acids or motifs in each helix are provided in one letter code. n - number of descriptors that are correlated significantly. doi:10.1371/journal.pone.0016811.g002



**Figure 3. Identified positions for antagonistic and agonistic effects mapped to the rhodopsin crystal structure.** a) GPCR positions correlated with chemical properties of ligands with either antagonistic or agonistic effects are mapped onto the inactive conformation of rhodopsin (pdb code 1U19). A colour code as in figure 2 enables the translation of functional information to the three-dimensional structure. The side-view with rendered backbone (white) highlights spatial localization and clustering of identified positions. b) The extracellular top view shows preferred ligand binding regions of family A GPCRs between the helices and the ECL2 close to the extracellular side. At the extracellular end especially of TMH2, 5 and 7 (boxes) several here identified amino acid side-chains are pointing towards the membrane. There involvement in ligand gating mechanisms or structural differences between the GPCRs can be speculated. c) Scheme of hypothetical scenarios for specific ligand-receptor interaction: i) both agonistic (green circular surface) or antagonistic effects (magenta circular surface) can be triggered at one receptor position (blue circle) dependent to the interrelated physico-chemical ligand properties. We assume that agonistic as well as antagonistic effects might be multiplied and simultaneously triggered at different parts of the receptor. ii) This also includes combinations between selective (triangle, trapeze) and non-selective contact points.

doi:10.1371/journal.pone.0016811.g003

Amino acids in the ECL2 (two residues before, and two behind the highly conserved cysteine which forms a disulphide bridge with cysteine 3.25 on TMH3) contribute significantly to ligand action, mostly with agonistic effects.

The amino-acid side chains at several positions point towards the membrane. These are located on the extracellular ends of TMHs 1, 2, 5 and 7. Our analysis also reveals amino acid positions which are most likely not involved in interactions with ligands, especially those that are located close to the intracellular side.

#### Specific ligand properties can be assigned to antagonistic and/or agonistic effects

The number of correlated ligand descriptors is in the range of 1-16 (figure 4). High numbers of correlated descriptors might be an indication for positions of priority in terms of ligand effect specificity. Several descriptors are found for both types of ligands (agonists and antagonists) but few are observable only for a particular type of ligand-effect. Interestingly, despite occurrence of shared descriptors to an equal amount, several of the descriptors

assigned either to agonists or to antagonists do appear with different frequencies (high numbers versus low numbers). The fact that similar descriptors occur with different frequencies for the two effects could be an indication that these descriptors are related to ligand properties causing the agonistic or antagonistic response.

#### Determination of G-protein preferences encoded in the transmembrane region of family A GPCRs

To investigate details of the general relationship between GPCR amino acid properties, ligands and the G-protein coupling preference we linked known G-protein (signaling pathway) preference to each particular receptor-ligand pair. We found three principal categories (figure 5):

1. Positions on all transmembrane helices are correlated to ligand properties and promiscuous G-preferences for  $G_i$ ,  $G_q$  and  $G_s$ .
2. Selective amino acid positions can be found for  $G_s$ ,  $G_q$  and  $G_i$ . Interestingly, this also includes the ECL2 and helix 8. The

| Frequency | Antagonist   | Num | Num | Agonist  | Frequency |
|-----------|--|-----|-----|--|-----------|
| 18        | ES-Count: dssC   |     |     | Number of H-Donors                                     | 14        |
| 15        | Number of H-Donors                                     |     |     | O in phenol, enol, carboxyl OH                         | 13        |
| 15        | Number of H-Acceptors                                  |     |     | ES-Count: ssCH2  | 12        |
| 13        | Number of Rings  |     |     | C in CH2R2   | 12        |
| 12        | Molecular Surface Area                                 |     |     | O in alcohol   | 11        |
| 12        | C in CH2R2   |     |     | H attached to heteroatom                               | 11        |
| 11        | Number of Aromatic Bonds                               |     |     | Number of Aromatic Rings                               | 10        |
| 10        | H attached to heteroatom                               |     |     | Molecular Polar Surface Area                           | 10        |
| 10        | H attached to C0sp3 with one X attached to next carbon |     |     | C in CHR2X   | 10        |
| 10        | ES-Count: aasC   |     |     | Number of Atoms  | 9         |
| 9         | H attached to C0sp3 with no X attached next to C       |     |     | Molecular Weight                                       | 9         |
| 9         | ES-Count: sOH  |     |     | ES-Count: dsCH   | 9         |
| 8         | Number of Ring Bonds                                   |     |     | C in R-CR-R  | 9         |
| 8         | Number of Atoms  |     |     | Number of Rotatable Bonds                              | 7         |
| 8         | N in Al3N  |     |     | Number of Rings  | 7         |
| 8         | H attached to C1sp3 or C0sp2                           |     |     | Number of Bonds  | 7         |
| 7         | Number of Bonds  |     |     | H attached to C1sp3 or C0sp2                           | 7         |
| 7         | ES-Count: dO   |     |     | ES-Count: sssCH  | 7         |
| 6         | Molecular Weight                                       |     |     | H attached to C0sp3 with no X attached next to carbon  | 6         |
| 4         | O in =O  |     |     | ES-Count: aasC   | 6         |
| 4         | ES-Count: sssCH  |     |     | C in CHR3  | 6         |
| 4         | ES-Count: ssNH   |     |     | C in =CHR  | 6         |
| 3         | O in phenol, enol, carboxyl OH                         |     |     | Number of Ring Bonds                                   | 5         |
| 3         | Number of Bridge Bonds                                 |     |     | Number of Hydrogens                                    | 4         |
| 3         | ES-Count: sssN   |     |     | N in R-N-R or R-N-X                                    | 3         |
| 2         | Number of Hydrogens                                    |     |     | H attached to C0sp3 with one X attached to next carbon | 3         |
| 2         | C in CH2RX   |     |     | ES-Count: aaN  | 3         |
| 1         | Number of Rotatable Bonds                              |     |     | Number of H-Acceptors                                  | 2         |
| 1         | Number of Aromatic Rings                               |     |     | Molecular Surface Area                                 | 2         |
| 1         | Molecular Polar Surface Area                           |     |     | Number of Aromatic Bonds                               | 1         |
| 1         | H attached to alpha-Cd                                 |     |     | N in Ar2NH, Ar3N, Ar2N-Al or R...N...R                 | 1         |
| 1         | ES-Count: aaCH   |     |     | H attached to alpha-Cd                                 | 1         |
| 1         | C in R-C(=X)-X or X=C=X                                |     |     | ES-Count: ssO  | 1         |
| 1         | C in R-CH-R  |     |     | ES-Count: dO   | 1         |
|           |  |     |     | ES-Count: aaaC   | 1         |
|           |  |     |     | C in X-CH...X  | 1         |
|           |  |     |     | C in CHR2  | 1         |

**Figure 4. Frequency of descriptors for antagonists and agonists.** The most significant ligand descriptors are sorted according to their frequency of occurrence in the results of our approach. Most of them are found for both types of ligands (coloured), agonistic and antagonistic, but few of them (white) are specifically observed only for one of each ligand-type. Interestingly, despite occurrence of descriptors for both ligand types to an equal amount (gray), few shared descriptors are found in high or low amount for agonists or antagonists (cyan, yellow), respectively. doi:10.1371/journal.pone.0016811.g004

numbers of selective positions for the G-protein subtypes considered are: Gi-7, Gq-14, Gs-18.

3. Correlations of amino acid positions to pairs of Gi/Gq, Gs/Gq or Gi/Gs can be identified.

In cases 1. and 3. the number of correlated descriptor variables can vary to a high degree (e.g. 21 vs. 4) potentially indicating the priority of one out of two signaling pathways. This is the case for instance at position 3.36 with Gi n = 21, Gq n = 8, Gs n = 3, or position 8.53 with Gi n = 21, Gq n = 4.

## Discussion

Expansion of rhodopsin-like GPCRs started ~500 million years ago [48]. Apart from a few exceptions such as the Glycoprotein hormone receptors (GPHRs) and Leucine-rich repeat containing G-protein coupled receptors (LGRs 4-8) [49], family A GPCRs are characterized by ligand binding close to or inside the transmembrane helices and the extracellular loops [11,50,51]. This is in contrast to other GPCR families and it has been proposed that this circumstance may be the basis for their evolutionary success, reflected by the highest number of members compared to the more structurally complex receptor proteins that have long ligand-binding N termini [9]. The advantages of such a ligand binding region might be the direct link to transmembrane signal transduction components (helices) and the general stability of the seven transmembrane helix structure which provides a scaffold for the evolution of new ligand binding partners. However, also for

family B GPCRs few reports pointing to a potential allosteric ligand binding site between the transmembrane helices [52,53].

An excellent analysis of this well-known classical transmembrane ligand binding region, supported by various mutagenesis studies, is given by Surgand and others for the entire collection of non-olfactive GPCRs [11]. These authors report a clear relationship between known ligand chemotypes (e.g. amines, carboxylic acids, phosphates, peptides, eicosanoids and lipids) and the cognate transmembrane cavities defined by 30 critical amino acid positions in the transmembrane helical region. Receptors for bulky ligands (e.g., phospholipids, prostanoids) appear to have a transmembrane cavity significantly larger than that for smaller compounds, and receptors for charged ligands (cationic amines, phosphates, mono and di-carboxylic acids) always present, among the 30 critical residues, one or more conserved amino acid exhibiting the opposite charge.

With exceptions such as rhodopsin which has a permanently bound inverse agonist, retinal, ligand binding is the initial event for signal transduction through the membrane via conformational changes in the helix arrangement [16,19,20,27,54]. Finally, heterotrimeric G-proteins are intracellularly activated which induces second messenger cascades [14,30,55]. Alternative signal transducers in addition to G-proteins are also known [56]. However, signal transduction across the membrane through GPCRs is governed by a complex set of interplaying components. A survey of their interrelationships is important for understanding the entire system. Several open but fundamental questions remain:



reason for the outward pointing side chains might be caused by structural differences between the GPCRs, because precisely these helices are characterized by a very diverse set of helix features [75] such as bulges and proline induced or supported kinks. These bulges and kinks have a strong effect on the helical twist and in consequence on the orientation of the side chains. In consequence, we assume that these regions and amino acids might be involved in ligand-gating mechanisms as shown in a previous study of Hildebrand and co-workers on opsin [76]. In conclusion, the mapping of our results for 100 different family A GPCRs onto one structural template, the rhodopsin structure, might not be generally representative for these particular regions. Equally, certain structural features in the extracellular regions of the transmembrane helices might be different for some of the family A GPCRs, despite the high structural overlap observable in the transmembrane region of the available GPCR crystal structures and our results may not be completely representative. Furthermore, our method has also identified residues outside the proposed general binding pocket region as correlated to specific ligand properties (figure 3a). They are mainly located close to or inside the intracellular region, which suggests their involvement in G-protein recognition and activation. Of special note are amino acids in helix 8. Their importance for GPCR signal transduction and G-protein coupling was pointed out in several studies on different GPCRs [37,77–79]. In conclusion, the identified correlations between agonistic or antagonistic effects and positions outside the ligand-binding region are likely related to their importance for selective G-protein recognition or interaction with effectors.

Raymond [32] and Wess [39] have hypothesized from previous data that certain receptor amino acids inside the transmembrane region might be involved in the regulation of G-protein subtype preference. In 2000 it was shown experimentally for the  $\beta$ 2-adrenergic receptor [33] that diverse ligands are indeed able to induce different signaling pathways at one receptor. By using FRET measurements on the  $\alpha$ 2-adrenergic receptor a correlation between diverse ligands and specific receptor conformations was recently shown [80]. Here we have found links between ligand effects and certain GPCR amino acid positions to be correlated with the activation of G-protein subtypes. Apart from the observation of amino acid positions at all transmembrane helices with correlations to effects on several G-protein subtypes (Gi, Gq, and Gs), we also revealed residues selective for either Gs, Gq or Gi (figures 5-6). We conclude that - comparable to the suggested scenario for agonistic or antagonistic effects - the ligand-specific effect on G-protein preference could be determined by the particular contact point(s) or the combination of receptor-ligand contacts [81,82] (figure 3c). It might be that particular structural features on the intracellular side of the activated receptors are involved in regulation of G-protein subtype preference [24,32,37,39,83]. These structural features, including the intracellular surface-shape, are regulated by specific ligand-receptor contacts directing helical re-arrangements during activation. This regulation is constituted by induction or blocking of receptor movements. Subsequently, it would be feasible that Gs, Gi and Gq adjust slightly differently to the receptor conformation to allow optimal complementary intermolecular side-chain interactions in each case. The assumption and possibility of different active GPCR conformations, even if slight, is supported by several published data [24,33,34,80,84].

### Significance and limitations of the analyses

All findings, results and conclusions depend strongly on the quality of the database. The standardized IUPHAR database can be assumed to be one of the most complete collections of GPCR

ligands. The IUPHAR database also contains GPCRs with promiscuous G-protein coupling capability (figure 2) but where the effect of dedicated ligands is not always evidenced by experimental studies on all types of G-protein coupling on that particular receptor. Therefore, it cannot be excluded that some receptor/ligand pairs assigned to a specific signaling pathway are potentially not linked to that pathway *in vivo*. As an example: for the LHCGR it was shown recently, that small allosteric ligands that bind to and activate this receptor do not induce both signaling pathways observed for the endogenous hormone-ligand [85].

The designated G-protein coupling (Gs, Gq, Gi) of the GPCRs are an indirect conclusion from measurements of signaling pathway intermediates like cAMP or calcium. The direct GPCR/G-protein interaction is hardly ever shown or proven. This is of relevance, because it has been discussed that the different subunits of a G-protein subtype may not be restricted to activating only one signaling pathway [30]. For example, for the lutropin receptor (LHCGR) it has been suggested that the beta-gamma subunit of Gi activates phospholipase C [86]. Therefore, exceptions might exist from the assumed direct one to one link between a signaling pathway and a specific G-protein subtype (subunit).

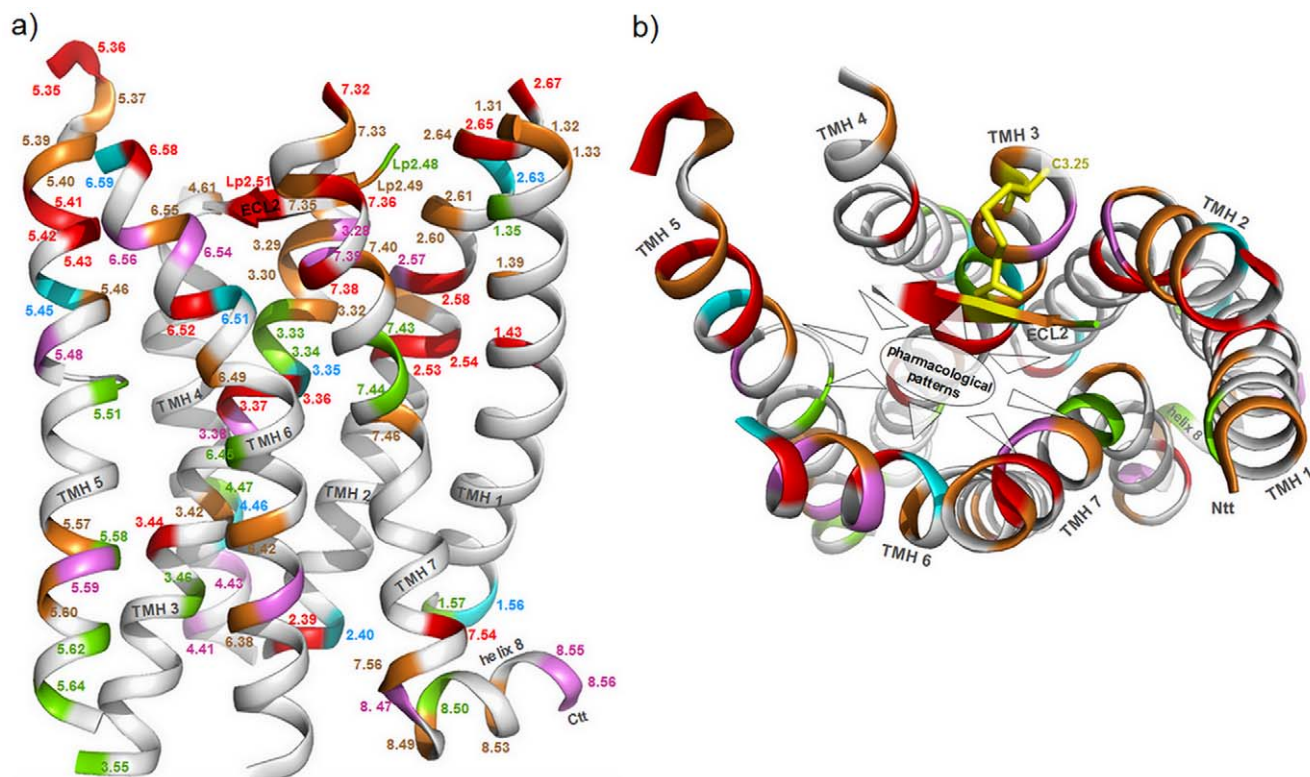
Highly conserved residues produce low values of mutual information and are therefore not included in the outcome of our analysis. This does of course not exclude their potential involvement in ligand-interaction or their participation in signal transduction.

In summary, we conducted this analysis with a maximum of data validation, but specific deviations from the general findings and conclusions are to be assumed given the statistical nature of our method and the restricted data set. A growing amount of reliable data will lead to deeper and more detailed insights.

Taken together, our results help with respect to set of limitations to understand GPCR signaling based on a distinct combination of interplaying parameters constituted by the chemical-physical properties of ligands and biophysical features of transmembrane GPCR residues. Our results can be condensed into three main hypotheses concerning regulation of signaling in family A GPCRs: 1. Residue positions in distinct receptor regions are specific for antagonistic or agonistic effects induced by ligands. 2. Important common ligand descriptors can be found for both types of ligands, but few are unique to either antagonists or agonists and might discriminate between the induced effects. 3. Preferences for activation of different G-protein subtypes are causally linked to specific residues in the transmembrane region of family A GPCRs. Our study, therefore, provides comprehensive information for understanding GPCR signaling and reveals new implications for the evolutionary derived intrinsic capacities of this protein superfamily. We conducted this analysis with a maximum of data validation, but specific deviations from the general findings and conclusions are to be assumed given the statistical nature of our method and the restricted data set. A growing amount of reliable data will lead to deeper and more detailed insights.

### Materials and Methods

Our analysis follows the workflow that we show in figure 1a. After extraction of sequence-ligand pairs from the database, we align the sequences and calculate various molecular descriptors for the related ligands. For this data set we then calculate the mutual information between each alignment position and each molecular descriptor. For the significance test we repeat this calculation for 100 surrogate data sets. The top 0.5% of the mutual information values are selected as hot spot positions.



**Figure 6. Receptor positions correlated with G-protein coupling preferences mapped to the rhodopsin crystal structure.** a) The amino acid positions with highly correlated mutual information to specific ligand properties (antagonists and agonists) are mapped to the structure of rhodopsin (b, top-view), but in combination with the G-protein preference of a particular GPCR subtype indicated by specific colours according to figure 5.

doi:10.1371/journal.pone.0016811.g006

## The Database

Annotated compound libraries have emerged as a strong information basis for computational drug design and several vendors provide annotated libraries for different purposes [87]. There are several open-source databases available such as the *MOAD database* collected by Hu et al. [88], the *GLIDA database* by Okono et al. [89] or the *IUPHAR database* by Harmar et al. [47]. We gathered GPCR information from the *IUPHAR database* [47] and extracted the 2D chemical structures from the *PubChem Structure Database*. In total we collected 1664 receptor-ligand pairs covering 100 family A GPCRs (figure 1b). This collection represents the maximum amount of family A GPCRs with annotated ligands that was available from open source at the time of our study. Only full and partial agonists/antagonists were taken into account. Inverse agonists and allosteric modulators are not included in this study.

Additionally the known preferences for Gs, Gi and Gq activation which are given by the measured signalling pathway were assigned. Several receptors are known to be promiscuous for their ability to activate two or three G-protein subtypes (figure 1b).

## The GPCR sequence alignment

The Rhodopsin-like family A GPCRs share highly conserved residues in the seven helices [75]. Therefore the multiple sequence alignment of the 100 analysed GPCRs (available under: <http://www.fmp-berlin.de/1129.html>) is straightforward and was achieved for the transmembrane domain with a Profile Hidden Markov Model (HMM) as included in the HMMER software [90] and the PF00001 profile for the rhodopsin family taken from the

*Pfam database* [91]. In addition, we adjusted the outcome of the HMM by hand in order to align the second extracellular loop (ECL2) around the cysteine residue involved in the highly conserved disulfide bridge between ECL2 and TMH3. Manual modifications were made to avoid gaps in the transmembrane helices. The extracted TMHs and conserved residues were numbered using the Ballesteros and Weinstein [92] numbering scheme to describe the common amino acid positions.

We also modified this numbering scheme by an extension to the extracellular (ECL) loop 2, whereby the highly conserved cysteine in ECL2, which in most family A GPCRs interacts with the cysteine in TMH3 (position 3.25), is numbered as Lp2.50. The “Lp” marks this as a loop and the “2” as the second loop. Estimation of these residues into analyses is supported by several previous studies on ligand interactions in GPCRs in the ECL2 [22].

## Molecular descriptors of analysed ligands

According to Todeschini and Consonni [93], a molecular descriptor is the ‘final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment’.

In order to characterize the chemical space and the properties of the GPCR ligands we preferred discrete and countable descriptors such as element and molecular property counts, Ghose-Crippen AlogP [94] and electro-topological state counts [95,96] which are easy to calculate using the two-dimensional structural formula. The only two continuous descriptors used were



the molecular surface area and the molecular polar surface area. All molecular descriptors were calculated with the Pipeline Pilot software (<http://www.scitegic.com>).

### Mutual information

Mutual information is a basic concept in information theory and provides a general measure of interdependence between random variables.

If we consider a discrete random variable  $X$  with  $n$  possible values taken from the alphabet  $A_x = \{x_1, x_2, \dots, x_n\}$  and the associated probability distribution  $P(X) = \{p(x_1), p(x_2), \dots, p(x_n)\}$  then the entropy  $H(X)$  is defined as

$$H(X) = - \sum_{i=1}^n p(x_i) \log(p(x_i)).$$

The joined entropy  $H(X, Y)$  of two random variables  $X$  and  $Y$  with the alphabet  $A_y = \{y_1, y_2, \dots, y_m\}$  and the joined probability distribution

$$P(X, Y) = \{p(x_1, y_1), p(x_1, y_2), \dots, p(x_n, y_m)\}$$

is defined as

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log(p(x_i, y_j)),$$

where  $p(x_i, y_j)$  denotes the probability of the joined occurrence of  $x_i$  and  $y_j$ . The mutual information  $I(X, Y)$  between two random variables  $X$  and  $Y$  is defined as

$$I(X, Y) = H(X) + H(Y) - H(X, Y).$$

The mutual information is zero if  $X$  and  $Y$  are statistically independent. Since it makes no assumption about the type of relationship between  $X$  and  $Y$ , mutual information is sometimes considered to be an extension of the linear correlation coefficient [97].

The histogram approach is the most common way to estimate the mutual information of a finite sample set. It calculates the relative frequencies in the histogram bins as an estimate of the probability distribution of the random variables.

In general the total number and the width of the bins is crucial and has effects on the outcome of the estimation process [98]. Let us consider  $K$  simultaneous measurements of the two random variables  $X$  and  $Y$  with the alphabets  $A_x = \{x_1, x_2, \dots, x_n\}$  and  $A_y = \{y_1, y_2, \dots, y_m\}$  and let  $k_{ij}$  be the total number of measurements with  $X = x_i$  and  $Y = y_j$ . Then the probabilities  $p(x_i, y_j)$  are approximated by the corresponding relative frequencies of the pairwise occurrence  $p(x_i, y_j) = \frac{k_{ij}}{K}$  and accordingly for the single probabilities  $p(x_i) = \frac{k_i}{K}$  and  $p(y_j) = \frac{k_j}{K}$ .

The mutual information  $I(X, Y)$  of the two random variables  $X$  and  $Y$  turns into

$$I(X, Y) = \log(K) + \frac{1}{K} \sum_{ij} k_{ij} \log \left( \frac{k_{ij}}{k_i k_j} \right).$$

It is known that the estimation of information theoretical functions such as entropy and mutual information may be affected by systematic errors. Steuer et al. pointed out [98] that the systematic error could be fairly approximated by

$$\langle I^{observed} \rangle \approx I(X, Y)^{true} + \Delta I(X, Y) \quad (1)$$

with

$$\Delta I(X, Y) = \frac{M_{xy} - M_x - M_y - 1}{2K},$$

where  $M_x$ ,  $M_y$  and  $M_{xy}$  are the number of histogram bins with nonzero probability and  $K$  is the total number of data samples.

### Estimating the Mutual Information between Sequence Positions and Ligand Features

The random variable  $X$  is given by the distribution of the residues in a specific alignment position whereby the 20 natural amino acids define the canonical alphabet. The random variable  $Y$  is given by the distribution of a specific molecular descriptor and the alphabet is based on the discrete (countable) values. We selected the particular subsets for agonists, antagonists and the G-Protein coupling from the entire data set of 1664 ligand-receptor pairs and calculated the mutual information between each alignment position and each molecular descriptor. Only the top 0.5% of the mutual information values were selected and the results are reported in Table S1.

### Significance-Test

We formulate the null hypothesis that  $X$  and  $Y$  are independent. To test the null hypothesis we generate 100 different surrogate datasets  $\{X^S, Y^S\}$  by creating random permutations of the original dataset  $\{X, Y\}$ . We calculate the corrected mutual information according to Equ. 1 for each surrogate realization of the null hypothesis and estimate the mean and the standard deviation  $\sigma^{sur}$ .

The test statistic  $S$  is then given by

$$S = \frac{I(X, Y)^{data} - \langle I(X, Y)^{sur} \rangle}{\sigma^{sur}}.$$

Under the assumption that the surrogate data follows a normal distribution we performed a one sided t-test at an alpha = 0.005 significance level which rejects the null hypothesis for  $S > 2.63$ . The values for the test statistic are reported in Table S1.

We performed a *Lilliefors* test in order to check the assumption that the surrogate data sets follow a normal distribution. We test the default null hypothesis that the data set comes from a distribution in the normal family, against the alternative that it does not come from a normal distribution. The test rejects the null hypothesis at the 5% significance level in less than 10% of all cases and justifies the assumption that the surrogate data sets are normally distributed.

### Supporting Information

**Table S1**  
(PDF)

## Acknowledgments

The authors would like to thank the members of the Computational Chemistry Department of Bayer-Schering Pharma for helpful discussions. We thank Victoria Higman for critical reading of the manuscript and her constructive suggestions.

## References

- Bjarnadóttir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, et al. (2006) Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* 88: 263–273.
- Dorsam RT, Gutkind JS (2007) G-protein-coupled receptors and cancer. *Nat Rev Cancer* 7: 79–94.
- Schöneberg T, Schulz A, Biebertmann H, Hermsdorf T, Rompler H, et al. (2004) Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacol Ther* 104: 173–206.
- Seifert R, Wenzel-Seifert K (2002) Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors. *Naunyn Schmiedeberg Arch Pharmacol* 366: 381–416.
- Kristiansen K (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol Ther* 103: 21–80.
- Fredriksson R, Lagerstrom MC, Lundin LG, Schiöth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogous groups, and fingerprints. *Mol Pharmacol* 63: 1256–1272.
- Tyndall JD, Sandilya R (2005) GPCR agonists and antagonists in the clinic. *Med Chem* 1: 405–421.
- Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, et al. (2004) G protein-coupled receptors: in silico drug discovery in 3D. *Proc Natl Acad Sci U S A* 101: 11304–11309.
- Lagerstrom MC, Schiöth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* 7: 339–357.
- Schlyer S, Horuk R (2006) I want a new drug: G-protein-coupled receptors in drug development. *Drug Discov Today* 11: 481–493.
- Surgand JS, Rodrigo J, Kellenberger E, Rognan D (2006) A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* 62: 509–538.
- Jacoby E, Bouhelal R, Gerspacher M, Seuwen K (2006) The 7 TM G-protein-coupled receptor target family. *ChemMedChem* 1: 761–782.
- Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1: 727–730.
- Oldham WM, Hamm HE (2008) Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* 9: 60–71.
- Smit MJ, Vischer HF, Bakker RA, Jongejan A, Timmerman H, et al. (2007) Pharmacogenomic and structural analysis of constitutive G-protein-coupled receptor activity. *Annu Rev Pharmacol Toxicol* 47: 53–87.
- Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM, Elling CE (2006) Molecular mechanism of 7TM receptor activation—a global toggle switch model. *Annu Rev Pharmacol Toxicol* 46: 481–519.
- Scheerer P, Heck M, Goede A, Park JH, Choe HW, et al. (2009) Structural and kinetic modeling of an activating helix switch in the rhodopsin-transducin interface. *Proc Natl Acad Sci U S A* 106: 10660–10665.
- Schertler GF (2008) Signal transduction: the rhodopsin story continued. *Nature* 453: 292–293.
- Ahuja S, Smith SO (2009) Multiple switches in G protein-coupled receptor activation. *Trends Pharmacol Sci* 30: 494–502.
- Hofmann KP, Scheerer P, Hildebrand PW, Choe HW, Park JH, et al. (2009) A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem Sci* 34: 540–552.
- Kleinau G, Claus M, Jaeschke H, Mueller S, Neumann S, et al. (2007) Contacts between extracellular loop two and transmembrane helix six determine basal activity of the thyroid-stimulating hormone receptor. *J Biol Chem* 282: 518–525.
- Massotte D, Kieffer BL (2005) The second extracellular loop: a damper for G protein-coupled receptors? *Nat Struct Mol Biol* 12: 287–288.
- Conner M, Hawtin SR, Simms J, Wootten D, Lawson Z, et al. (2007) Systematic analysis of the entire second extracellular loop of the V(1a) vasopressin receptor: key residues, conserved throughout a G-protein-coupled receptor family, identified. *J Biol Chem* 282: 17405–17412.
- Kobilka BK, Deupi X (2007) Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol Sci* 28: 397–406.
- Hanson MA, Stevens RC (2009) Discovery of new GPCR biology: one receptor structure at a time. *Structure* 17: 8–14.
- Kobilka B, Schertler GF (2008) New G-protein-coupled receptor crystal structures: insights and limitations. *Trends Pharmacol Sci* 29: 79–83.
- Rosenbaum DM, Rasmussen SG, Kobilka BK (2009) The structure and function of G-protein-coupled receptors. *Nature* 459: 356–363.
- Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 454: 183–187.

## Author Contributions

Conceived and designed the experiments: JDW RK AL G. Kleinau. Performed the experiments: JDW RK G. Kleinau. Analyzed the data: JDW AL G. Krause NH RK G. Kleinau. Contributed reagents/materials/analysis tools: JDW RK NH G. Kleinau. Wrote the paper: JDW AL G. Krause NH RK G. Kleinau.

- Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, et al. (2008) Crystal structure of opsin in its G-protein-interacting conformation. *Nature* 455: 497–502.
- Smrcka AV (2008) G protein betagamma subunits: central mediators of G protein-coupled receptor signaling. *Cell Mol Life Sci* 65: 2191–2214.
- Evans PD, Robb S, Cheek TR, Reale V, Hannan FL, et al. (1995) Agonist-specific coupling of G-protein-coupled receptors to second-messenger systems. *Prog Brain Res* 106: 259–268.
- Raymond JR (1995) Multiple mechanisms of receptor-G protein signaling specificity. *Am J Physiol* 269: F141–F158.
- Wenzel-Seifert K, Seifert R (2000) Molecular analysis of beta(2)-adrenoceptor coupling to G(s)-, G(i)-, and G(q)-proteins. *Mol Pharmacol* 58: 954–966.
- Perez DM, Karnik SS (2005) Multiple signaling states of G-protein-coupled receptors. *Pharmacol Rev* 57: 147–161.
- Horn F, van der Wenden EM, Oliveira L, Ijzerman AP, Vriend G (2000) Receptors coupling to G proteins: is there a signal behind the sequence? *Proteins* 41: 448–459.
- Moller S, Vilo J, Croning MD (2001) Prediction of the coupling specificity of G protein-coupled receptors to their G proteins. *Bioinformatics* 17(Suppl 1): S174–S181.
- Kleinau G, Jaeschke H, Worth CL, Mueller S, Gonzalez J, et al. (2010) Principles and determinants of G-protein coupling by the rhodopsin-like thyrotropin receptor. *PLoS One* 5: e9745.
- Hu J, Wang Y, Zhang X, Lloyd JR, Li JH, et al. (2010) Structural basis of G-protein-coupled receptor-G protein interactions. *Nat Chem Biol* 6: 541–548.
- Wess J (1998) Molecular basis of receptor/G-protein-coupling selectivity. *Pharmacol Ther* 80: 231–264.
- Oliveira L, Paiva AC, Vriend G (2002) Correlated mutation analyses on very large sequence families. *Chembiochem* 3: 1010–1017.
- Ye K, Lameijer EW, Beukers MW, Ijzerman AP (2006) A two-entropies analysis to identify functional positions in the transmembrane region of class A G protein-coupled receptors. *Proteins* 63: 1018–1030.
- Fatakia SN, Costanzi S, Chow CC (2009) Computing highly correlated positions using mutual information and graph theory for G protein-coupled receptors. *PLoS One* 4: e4681.
- Kubinyi H, Mueller G, Mannhold R, Folkers G (2004) *Chemogenomics in Drug Discovery: A Medicinal Chemistry Perspective. Methods and Principles in Medicinal Chemistry*. New York: Wiley-VCH.
- Keiser MJ, Roth BL, Armbuster BN, Ernsberger P, Irwin JJ, et al. (2007) Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 25: 197–206.
- Jacob L, Vert JP (2008) Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics* 24: 2149–2156.
- van der HE, Peironcelly JE, Ijzerman AP, Beukers MW, Lane JR, et al. (2010) A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization. *BMC Bioinformatics* 11: 316.
- Harmar AJ, Hills RA, Rosser EM, Jones M, Buneman OP, et al. (2009) IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic Acids Res* 37: D680–D685.
- Schöneberg T, Hofreiter M, Schulz A, Rompler H (2007) Learning from the past: evolution of GPCR functions. *Trends Pharmacol Sci* 28: 117–121.
- Kleinau G, Krause G (2009) Thyrotropin and homologous glycoprotein hormone receptors: structural and functional aspects of extracellular signaling mechanisms. *Endocr Rev* 30: 133–151.
- Bywater RP (2005) Location and nature of the residues important for ligand recognition in G-protein coupled receptors. *J Mol Recognit* 18: 60–72.
- Klabunde T, Hessler G (2002) Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* 3: 928–944.
- Bhattacharya S, Subramanian G, Hall S, Lin J, Laoui A, Vaidehi N (2010) Allosteric antagonist binding sites in class B GPCRs: corticotropin receptor 1. *J Comput Aided Mol Des* 24: 659–674.
- Cascieri MA, Koch GE, Ber E, Sadowski SJ, Louzides D, et al. (1999) Characterization of a novel, non-peptidyl antagonist of the human glucagon receptor. *J Biol Chem* 274: 8694–8697.
- Wess J, Han SJ, Kim SK, Jacobson KA, Li JH (2008) Conformational changes involved in G-protein-coupled-receptor activation. *Trends Pharmacol Sci* 29: 616–625.
- Van Eps N, Oldham WM, Hamm HE, Hubbell WL (2006) Structural and dynamical changes in an alpha-subunit of a heterotrimeric G protein along the activation pathway. *Proc Natl Acad Sci U S A* 103: 16194–16199.
- Sun Y, McGarrigle D, Huang XY (2007) When a G protein-coupled receptor does not couple to a G protein. *Mol Biosyst* 3: 849–854.

57. Gloriam DE, Foord SM, Blaney FE, Garland SL (2009) Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. *J Med Chem* 52: 4429–4442.
58. Bokoch MP, Zou Y, Rasmussen SG, Liu CW, Nygaard R, et al. (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* 463: 108–112.
59. Unal H, Jagannathan R, Bhat MB, Karnik SS (2010) Ligand-specific conformation of extracellular loop-2 in the angiotensin II type 1 receptor. *J Biol Chem* 285: 16341–16350.
60. Urizar E, Claeysen S, Deupi X, Govaerts C, Costagliola S, et al. (2005) An activation switch in the rhodopsin family of G protein-coupled receptors: the thyrotropin receptor. *J Biol Chem* 280: 17135–17141.
61. Ye S, Zaitseva E, Caltabiano G, Schertler GF, Sakmar TP, et al. (2010) Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* 464: 1386–1389.
62. Chugunov AO, Simms J, Poyner DR, Dehouck Y, Rooman M, et al. (2010) Evidence that interaction between conserved residues in transmembrane helices 2, 3, and 7 are crucial for human VPAC1 receptor activation. *Mol Pharmacol* 78: 394–401.
63. Angel TE, Gupta S, Jastrzebska B, Palczewski K, Chance MR (2009) Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc Natl Acad Sci U S A* 106: 14367–14372.
64. Pardo L, Deupi X, Dolker N, Lopez-Rodriguez ML, Campillo M (2007) The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *ChemBiochem* 8: 19–24.
65. Tarnow P, Rediger A, Brumm H, Ambrugger P, Rettenbacher E, et al. (2008) A heterozygous mutation in the third transmembrane domain causes a dominant-negative effect on signalling capability of the MC4R. *Obes Facts* 1: 155–162.
66. Pogozheva ID, Chai BX, Lomize AL, Fong TM, Weinberg DH, et al. (2005) Interactions of human melanocortin 4 receptor with nonpeptide and peptide agonists. *Biochemistry* 44: 11329–11341.
67. Hogan K, Peluso S, Gould S, Parsons I, Ryan D, et al. (2006) Mapping the binding site of melanocortin 4 receptor agonists: a hydrophobic pocket formed by I3.28(125), I3.32(129), and I7.42(291) is critical for receptor activation. *J Med Chem* 49: 911–922.
68. Yang YK, Fong TM, Dickinson CJ, Mao C, Li JY, et al. (2000) Molecular determinants of ligand binding to the human melanocortin-4 receptor. *Biochemistry* 39: 14900–14911.
69. Nickolls SA, Cismowski MI, Wang X, Wolff M, Conlon PJ, et al. (2003) Molecular determinants of melanocortin 4 receptor ligand binding and MC4/MC3 receptor selectivity. *J Pharmacol Exp Ther* 304: 1217–1227.
70. Jaakola VP, Lane JR, Lin JY, Katritch V, Ijzerman AP, et al. (2010) Ligand binding and subtype selectivity of the human A(2A) adenosine receptor: identification and characterization of essential amino acid residues. *J Biol Chem* 285: 13032–13044.
71. Jiang Q, Van Rhee AM, Kim J, Yehle S, Wess J, et al. (1996) Hydrophilic side chains in the third and seventh transmembrane helical domains of human A2A adenosine receptors are required for ligand recognition. *Mol Pharmacol* 50: 512–521.
72. Runyon SP, Mosier PD, Roth BL, Glennon RA, Westkaemper RB (2008) Potential modes of interaction of 9-aminomethyl-9,10-dihydroanthracene (AMDA) derivatives with the 5-HT2A receptor: a ligand structure-affinity relationship, receptor mutagenesis and receptor modeling investigation. *J Med Chem* 51: 6808–6828.
73. Westkaemper RB, Glennon RA (1994) Molecular modeling of the interaction of LSD and other hallucinogens with 5-HT2 receptors. *NIDA Res Monogr* 146: 263–283.
74. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, et al. (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318: 1258–1265.
75. Worth CL, Kleinau G, Krause G (2009) Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PLoS One* 4: e7011.
76. Hildebrand PW, Scheerer P, Park JH, Choe HW, Piechnick R, et al. (2009) A ligand channel through the G protein coupled receptor opsin. *PLoS One* 4: e4382.
77. Huynh J, Thomas WG, Aguilar MI, Pattenden LK (2009) Role of helix 8 in G protein-coupled receptors based on structure-function studies on the type 1 angiotensin receptor. *Mol Cell Endocrinol* 302: 118–127.
78. Lehmann N, Alexiev U, Fahmy K (2007) Linkage between the intramembrane H-bond network around aspartic acid 83 and the cytosolic environment of helix 8 in photoactivated rhodopsin. *J Mol Biol* 366: 1129–1141.
79. Verzijl D, Pardo L, van Dijk M, Grijthuijsen YK, Jongejan A, et al. (2006) Helix 8 of the viral chemokine receptor ORF74 directs chemokine binding. *J Biol Chem* 281: 35327–35335.
80. Zurn A, Zabel U, Vilardaga JP, Schindelin H, Lohse MJ, et al. (2009) Fluorescence resonance energy transfer analysis of alpha 2a-adrenergic receptor activation reveals distinct agonist-specific conformational changes. *Mol Pharmacol* 75: 534–541.
81. Huang J, Hamasaki T, Ozoe F, Ozoe Y (2008) Single amino acid of an octopamine receptor as a molecular switch for distinct G protein couplings. *Biochem Biophys Res Commun* 371: 610–614.
82. Robb S, Check TR, Hannan FL, Hall LM, Midgley JM, et al. (1994) Agonist-specific coupling of a cloned *Drosophila* octopamine/tyramine receptor to multiple second messenger systems. *EMBO J* 13: 1325–1330.
83. Wong SK (2003) G protein selectivity is regulated by multiple intracellular regions of GPCRs. *Neurosignals* 12: 1–12.
84. Vauquelin G, Van Liefde I (2005) G protein-coupled receptors: a count of 1001 conformations. *Fundam Clin Pharmacol* 19: 45–56.
85. van Koppen CJ, Zaman GJ, Timmers CM, Kelder J, Mosselman S, et al. (2008) A signaling-selective, nanomolar potent allosteric low molecular weight agonist for the human luteinizing hormone receptor. *Naunyn-Schmiedeberg Arch Pharmacol* 378: 503–514.
86. Kuhn B, Gudermann T (1999) The luteinizing hormone receptor activates phospholipase C via preferential coupling to Gi2. *Biochemistry* 38: 12490–12498.
87. Balakin K, Tkachenko A, Kiselyova SE, Savchuk N (2005) Focused chemistry from annotated libraries. *Drug Discovery Today: Technologies* 3: 397–403.
88. Hu L, Benson ML, Smith RD, Lerner MG, Carlson HA (2005) Binding MOAD (Mother Of All Databases). *Proteins* 60: 333–340.
89. Okuno Y, Tamon A, Yabuuchi H, Nijima S, Minowa Y, et al. (2008) GLIDA: GPCR—ligand database for chemical genomics drug discovery—database and tools update. *Nucleic Acids Res* 36: D907–D912.
90. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14: 755–763.
91. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, et al. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34: D247–D251.
92. Ballesteros JA, Weinstein H (1995) Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relationships in G-Protein Coupled Receptors. *Methods Neurosci* 25: 366–428.
93. Todeschini R, Consonni V (2000) Handbook of Molecular Descriptors. Weinheim, New York: Wiley-VCH.
94. Ghose A, Viswanadhan V, Wendoloski J (1998) Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *Journal of Physical Chemistry A* 102: 3762–3772.
95. Hall LH, Kier LB (2000) The E-state as the basis for molecular structure space definition and structure similarity. *J Chem Inf Comput Sci* 40: 784–791.
96. Lowell L, Hall H, Mohny B, Kier L (1991) The electrotopological state: Structure information at the atomic level for molecular graphs. *J Chem Inf and Comput Sci* 31: 76–82.
97. Li W (1990) Mutual information functions versus correlation functions. *Journal of Statistical Physics* 60: 823–837.
98. Steuer R, Kurths J, Daub CO, Weise J, Selbig J (2002) The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* 18 Suppl 2: S231–S240.