Published in final edited form as:

Neuroscience. 2019 August 01; 412: 48–59. doi:10.1016/j.neuroscience.2019.05.013.

Coevolution of residues provides evidence of a functional heterodimer of 5-HT_{2A}R and 5-HT_{2C}R involving both intracellular and extracellular domains

Bernard Fongang^{1,2,3,4,*}, Kathryn A. Cunningham⁵, Maga Rowicka^{1,6,7}, Andrzej Kudlicki^{1,6,7,*}

¹Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX 77555, USA

²Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, UTHSCSA, San Antonio, TX 78229, USA

³Department of Biochemistry and Structural Biology, UTHSCSA, San Antonio, TX 78229, USA

⁴Department of Epidemiology and Biostatistics, UTHSCSA, San Antonio, TX 78229, USA

⁵Center for Addiction Research and Department of Pharmacology and Toxicology, University of Texas Medical Branch, Galveston, TX 77555, USA

⁶Institute for Translational Sciences, University of Texas Medical Branch, Galveston, TX 77555, USA

⁷Sealy Center for Molecular Medicine, University of Texas Medical Branch, Galveston, TX 77555, USA

Abstract

Serotonin is a neurotransmitter that plays a role in regulating activities such as sleep, appetite, mood and substance abuse disorders; serotonin receptors 5-HT_{2A}R and 5-HT_{2C}R are active within pathways associated with substance abuse. It has been suggested that 5-HT_{2A}R and 5-HT_{2C}R may form a dimer that affects behavioral processes. Here we study the coevolution of residues in 5-HT_{2A}R and 5-HT_{2C}R to identify potential interactions between residues in both proteins. Coevolution studies can detect protein interactions, and since the thus uncovered interactions are subject to evolutionary pressure, they are likely functional. We assessed the significance of the 5-HT_{2A}R/5-HT_{2C}R interactions using randomized phylogenetic trees and found the coevolution significant (p-value = 0.01). We also discuss how co-expression of the receptors suggests the predicted interaction is functional. Finally, we analyze how several single nucleotide polymorphisms for the 5-HT_{2A}R and 5-HT_{2C}R genes affect their interaction. Our findings are the

*To whom correspondence should be addressed. Tel: +1 409 772 2253; Fax: +1 409 772 0725; askudlic@utmb.edu.

Authors contributions:

AK and BF designed research; BF and AK wrote computer code; BF prepared illustrations; BF calculated coevolution maps and performed statistical analysis; AK analyzed expression profiles across tissues; BF, MR and AK interpreted results; KC integrated results with experimental data; BF, MR, AK and KC wrote the paper.

Supplementary files:

The supplementary files of this publication can be accessed through the following link <http://dx.doi.org/10.17632/5hkxztmvpk.2>

first to characterize the binding interface of 5-HT_{2A}R/5-HT_{2C}R and indicate a correlation between this interface and location of SNPs in both proteins.

Keywords

Serotonin receptors; Evolutionary coupling; 5-HT_{2A} receptor; 5-HT_{2C} receptor; Direct Coupling Analysis; Gaussian convolution; evolutionary trees

INTRODUCTION

Serotonin (5-hydroxytryptamine, 5-HT) is a monoamine neurotransmitter with broad regulatory roles within the central and peripheral nervous systems with complex actions transduced by 5-HT binding to one ligand-gated cation channel and 13 G protein-coupled receptors (GPCRs). Of the 5-HT GPCRs, the 5-HT₂ receptor family includes the 5-HT_{2A} receptor (5-HT_{2A}R), 5-HT_{2B}R and 5-HT_{2C}R on the basis of their structural and operational characteristics (Hannon J and Hoyer D, 2008). The brain is particularly enriched in expression of the 5-HT_{2A}R and 5-HT_{2C}R that govern normal behavioral (e.g., appetite, mood, sleep) and physiological function (e.g., circadian rhythm, endocrine secretion). Furthermore, dysfunctional 5-HT_{2A}R and 5-HT_{2C}R systems are implicated in a broad range of neuropsychiatric disorders (e.g., addiction, depression, schizophrenia) and obesity (Hannon J and Hoyer D, 2008; Howell LL and Cunningham KA, 2015). These receptors couple to G $\alpha_{q/11}$ to activate the enzyme phospholipase C β which generates intracellular second messengers inositol-1,4,5-trisphosphate and diacylglycerol, leading to increased calcium release from intracellular stores (Hannon J and Hoyer D, 2008). The homomeric form of the 5-HT_{2A}R (Iglesias A et al., 2017; Iglesias A et al., 2017; Iglesias A et al., 2016) and 5-HT_{2C}R (Herrick-Davis K et al., 2012; Herrick-Davis K et al., 2015) is proposed to be main signaling unit, while both are reported to form functional heterodimeric complexes with other GPCRs (Albizu L et al., 2006; Albizu L et al., 2010) (Schellekens H et al., 2013; Schellekens H et al., 2013). Employing co-immunoprecipitation (Co-IP) assays, we recently demonstrated that the two receptors form a protein complex in rodent brain, the expression of which tracked with phenotypic behaviors (Anastasio NC et al., 2015) while bioluminescence resonance energy transfer (BRET) assays detected a very close (within 10 nM) interaction in transfected cells (Felsing DE et al., 2018; Moutkine I et al., 2017). The 5-HT_{2A}R:5-HT_{2C}R protein complex triggers intracellular signaling in live cultured cells that is uniquely distinct from each protomer alone (Moutkine I, et al., 2017). Further (Felsing DE, et al., 2018) studies are required to predict biological features of relevance such as residue contacts within the 5-HT_{2A}R:5-HT_{2C}R protein structures and protein functional sites.

Several experimental techniques are useful in the study of protein dimerization, including Co-IP, BRET, tandem affinity purification with mass spectrometry, yeast two-hybrid screens, protein-fragment complementation assays, NMR spectroscopy and split luciferase complementation assays (Keene JD et al., 2006; Rigaut G et al., 1999; Rochette S et al., 2015; Sekar RB and Periasamy A, 2003; Uetz P et al., 2000). Although useful in the detection of dimerization interfaces, these methods cannot distinguish between direct and indirect interactions, are frequently performed under non-physiological conditions, and remain time

consuming and costly. Moreover, each approach probes a different aspect of the interaction as protein dimerization is believed to be a dynamic and reversible process (Calebiro D et al., 2013; Scarselli M et al., 2016), and suffers from the potential for assay-specific experimental errors; also, experimental approaches may miss transient interactions crucial to biological processes, limiting their utility in providing definitive answers as to whether an interaction is functional *in vivo* (Phizicky EM and Fields S, 1995). An alternative approach for identifying functionally important residues is to use the coevolution of amino acids. These methods are generally classified in two groups: the first group aim to predict structural contacts and protein-protein interactions through evolutionary couplings of residues in these proteins along the phylogenetic tree (Bitbol AF, 2018; Bitbol AF et al., 2016; dos Santos RN et al., 2015; Ekeberg M et al., 2014; Gueudre T et al., 2016; Hopf TA et al., 2012; Hopf TA et al., 2014; Marks DS et al., 2011; Marks DS et al., 2012; Morcos F et al., 2011; Sulkowska JI et al., 2012; Weigt M et al., 2009) and the second group is focused on studying the collective correlations of amino acids (Halabi N et al., 2009; Lockless SW and Ranganathan R, 1999; Suel GM et al., 2003; Tesileanu T et al., 2015). The concept underlying evolutionary coupling is that, to preserve function, a mutation in one interacting residue is likely to be compensated by a complementary mutation in the other (de Juan D et al., 2013). The key advantage of this approach is that interactions between residues are detected not only based on their physical proximity, but upon evolutionary pressure, and therefore are more likely to be functionally relevant for biology. For almost two decades, co-evolution of residues in protein sequences has been valuable for inferring residue-residue interactions (contacts) within small, bacterial proteins with large numbers of homologs. Several statistical approaches to studying co-evolution have been developed (Ekeberg M, et al., 2014; Hopf TA, et al., 2014; Marks DS, et al., 2012; Morcos F, et al., 2011; Sulkowska JI, et al., 2012), including direct coupling analysis (DCA) (Morcos F, et al., 2011). With the rapid increase in the number of sequenced animal genomes, these methods have become feasible for assessments of protein interactions in metazoans, including humans, and for larger proteins and even protein complexes (Abriata LA et al., 2016; Champeimont R et al., 2016; dos Santos RN, et al., 2015; Feinauer C et al., 2016; Hopf TA, et al., 2012; Hopf TA, et al., 2014; Kim DE et al., 2014; Lua RC et al., 2016; Naveed H et al., 2012; Neuwald AF, 2016; Ovchinnikov S et al., 2014; Tesileanu T, et al., 2015; Yu JC et al., 2016).

It is now well accepted that most G Protein-coupled receptors (GPCRs) achieve their function through dimerization or oligomerization. Although the whole concept of dimerization is of high interest in pharmacology, the determination of native contacts or binding interfaces is still very challenging experimentally and computationally. Nonetheless, several GPCRs heterodimer interfaces have been determined through various experimental methods. For example, it has been shown that the heterodimerization of serotonin 5-HT_{2A}R and dopamine D₂ receptor involves the 3rd intracellular loop of D₂ and the c-terminal domain of 5-HT_{2A}R (Lukasiewicz S et al., 2010). Consistent with this result, 2 amino-acid residues in the 3rd intracellular loop of D₂ and c-terminal domain of D₁ were also identified as native contacts for the D₁/D₂ heterodimer (O'Dowd BF et al., 2012). The 5-HT_{2A}R/mGlu₂ heterodimer has been also extensively studied and 3 residues have been proposed as stabilizing the interaction (Moreno JL et al., 2012). A broad review of GPCRs receptor oligomerization and their most likely interfaces is provided by Gerda E. Breitwieser and

several others (Breitwieser GE, 2004;Guo W et al., 2005;Lee SP et al., 2003). Although very challenging, it appears substantial efforts are made to identify binding interfaces of GPCRs heterodimers including serotonin receptors. The burden of these experimental efforts may be reduced if we could use computational predictions to prioritize the most likely binding interfaces of large and multi-domain proteins including GPCRs.

Here, we adapted DCA to studying interactions between the 5-HT_{2A}R and 5-HT_{2C}R, which are particularly challenging due to their size, α -helical transmembrane domain structure, and limited structural resolution for these proteins. Next, we utilized simulated phylogenetic trees to demonstrate that the identified interactions are indeed statistically significant. We interpret the results in the context of the structure of the 5-HT_{2A}R and 5-HT_{2C}R proteins, as well as the known polymorphisms in the genes coding for these proteins. Finally, we present additional evidence supporting the 5-HT_{2A}R and 5-HT_{2C}R interaction hypothesis, based on temporal co-expression of the receptor genes in the primate circadian cycle. We will also discuss how our method can be adapted and extended to other challenging classes of protein interactions, such as interactions between other G-protein coupled receptors that are targets of 30% of FDA approved medications (Filmore D, 2004;Overington JP et al., 2006).

MATERIALS AND METHODS

Multiple Sequence Alignment and Evolutionary Scores

Evolutionary coupling analysis depends on a large collection of genes orthologous to each of the genes whose interaction is under investigation; the orthologs of both genes have to come from the same set of species. We obtained 241 and 235 orthologous sequences of 5-HT_{2A}R and 5-HT_{2C}R respectively by querying *GENBANK* (Benson D et al., 1993) and running genome-wide *tblastn* (Altschul SF et al., 1990) on genomes not present in *GENBANK*. The orthologous sequences were joined by species and aligned using *Clustal-Ω* (Sievers F et al., 2011) with manually introduced corrections (based on domain annotation) to generate the concatenated Multiple Sequence Alignments (cMSA) of 5-HT_{2A}R/5-HT_{2C}R in 231 vertebrates. For species with several isoforms, we considered only the isoform with highest sequence similarity to the human protein, which was chosen as reference during the DCA analysis step.

We analyzed evolutionary coupling using Direct Coupling Analysis (DCA) algorithm (Morcos F, et al., 2011). DCA was chosen among Evolutionary Coupling (EC) methods because of its better ability to model the whole data set rather than individual pairs of residues, helping to distinguish between direct functional residue interactions and correlations resulting from indirect interactions that are given a lower weight in DCA. The primary advantage of DCA lies in this ability to assign higher scores to direct correlations than to indirect ones, which is achieved through computing direct statistical couplings for each pair of positions and an appropriate sampling correction. The pseudo-likelihood maximization Direct-Coupling Analysis (*plmDCA*) variant of the algorithm (Ekeberg M, et al., 2014) that we utilized has a lower computational cost and higher precision than alternative approaches. Options for *plmDCA* were set as follows: Optimization method (conjugate gradient, cg); Sequence of interest (human, use first member as sequence of interest); and other parameters as default.

DCA post-processing: convolution of Evolutionary Coupling scores (ECs): 2D Gaussian filter applied to DCA maps

DCA-based methods have been shown to be reliable in predicting native contacts and have long been used to reconstruct the 3D structure of proteins. However, the ECs scores, as provided by these methods, are noisy and contain a large number of False Positives (FP), especially for large proteins, such as 5-HT_{2A}R and 5-HT_{2C}R. Such FPs, due to random correlations between evolutionary changes, appear as isolated peaks in the DCA maps. To account for the contribution of neighboring residues in the formation of interfaces, we propose here to use a Gaussian convolution with a kernel constructed on the structural information of both proteins. Convolution or deconvolution methods have been successfully used in the past to extract the signal in various processes (Fongang B and Kudlicki A, 2013; Fongang B and Kudlicki A, 2016; Homouz D et al., 2015; Kudlicki A et al., 1996; Rowicka M et al., 2007) in which the background noise or experimental artifacts overshadow the true signal.

Our global approach is outlined in Figure 1: For two studied proteins, a cMSA is constructed from orthologous sequences and ECs are computed using the *plmDCA* algorithm as previously described. For each pair of residues i, j , a 2D Gaussian convolution, centered on (i, j) with different variances is applied resulting in convolved ECs that depend on the local structure of peptides. ECs provided by the DCA algorithm constitute a M by N matrix where M and N are the lengths of the two interacting proteins. Each point of the matrix, $P_{i,j}$, represents the Evolutionary Score of residues i and j of the two proteins. The convolved DCA signal is defined as:

$$Q_{i,j} = \sum_{\alpha=i-l}^{i+l} \sum_{\beta=j-l}^{j+l} P_{\alpha,\beta} K_{\alpha-i, \beta-j}^{\gamma} \quad (1)$$

where $2l+1$ is the number of residues in the neighborhood and $K_{i,j}^{\gamma}$ is a kernel function adjusted to the secondary structure. γ denotes the dependence to the secondary structure of the protein ($\gamma=1$ if the same type of structure is present in both proteins, $\gamma=2$ otherwise). The biological interpretation of l is that it represents the average length of the stretch of protein sequence involved in the interaction.

In the present study, the kernel $K_{i,j}^{\gamma}$ is a Gaussian function

$$K_{i,j}^{\gamma} = \exp\left(-\left\{a(\alpha-i)^2 + b(\beta-j)^2\right\}\right) \quad (2)$$

with a and b the normalized variances (optimized values of a and b for $\gamma = 1$ or 2 are given in Table 1). The general idea behind the local convolution of ECs signal is depicted in Figure 1: Residues with the same secondary structure (alpha helix, beta barrel or coil) but less significant overall ECs have greater chance to be in contact than isolated residues with stronger ECs. The final convolved ECs depends on the variances a and b as well as l , the total number of amino-acids (AA) at the interface. Optimized values of the parameters a and b , used for the present study are listed in Table 1. The parameters were optimized using another complex, the INTS9-INTS11 complex, for which we had crystal structure (Bernard

Fongang, Yingjie Zhu, Eric Wagner, Andrzej Kudlicki, Maga Rowicka, Co-evolutionary analysis accurately predicts details of interactions between the Integrator subunits, *in preparation*). We have estimated the parameter $l = 21$ AA for the 5-HT_{2A}R- 5HT_{2C}R heterodimer (see Results and Discussion section). The Gaussian convolution depends on the accurate prediction of the structural information of individual protein. Here, we predicted the secondary structure of 5-HT_{2A}R and 5-HT_{2C}R using stand-alone versions of PSIPRED (McGuffin LJ et al., 2000), with manual correction for consistency. This bioinformatics tool has been shown to accurately predict protein structure information based on sequences and we have verified that predicted secondary structures, in the case of 5-HT_{2A}R and 5-HT_{2C}R agree with known properties (GPCRs domain architecture). The secondary structures were checked for consistency and corrected where required using our own program written in R. Due to the relatively small number of orthologous sequences of many proteins of interest, leading to noisy contact predictions, several methods have been developed to improve the quality of the predictions. Liu et al (Liu Y et al., 2018) used convolutional neural network to learn structural interaction motifs from thousands of proteins with experimentally solved structures which combine to evolutionary coupling scores significantly improve contact prediction. Other methods include the deep mutational scanning introduced by Schmiedel et al (Schmiedel J and Lehner B, 2018) which statistically quantify interactions between mutations to identify tertiary structure contacts within a protein.

Single Nucleotide Polymorphisms (SNPs)

Human SNPs localized in the coding sequence of 5-HT_{2A}R and 5-HT_{2C}R were retrieved from the NCBI SNPs database (dbSNP) (Sherry ST et al., 2001) and analyzed in terms of coevolution and phenotypic annotation (see Results).

Coexpression during baboon circadian rhythm.

To assess coexpression of the receptor genes in primate circadian cycle, we have downloaded transcriptome data from the diurnal transcriptome atlas of Baboon (Papio Anubis) across all major neural and peripheral tissues (64 tissues over 12 time points), available at GEO under accession number GSE98965 (Mure LS et al., 2018). Correlation between the genes coding for 5-HT_{2A}R and 5-HT_{2C}R has been computed over time. To assess the significance of co-expression the correlations were compared with correlations between 10,000 randomly chosen pairs of named genes, as well as between the serotonin receptors and all other named genes in the dataset.

RESULTS AND DISCUSSION

Coevolution of residues in the putative 5-HT_{2A}R/5-HT_{2C}R heterodimer.

Coevolution between residues of heterodimeric proteins can be computed using a wide range of methods from Mutual Information (MI) to the Direct Coupling Analysis (DCA).

Here, we chose the DCA algorithm (Morcos F, et al., 2011) among Evolutionary Coupling (EC) methods because of its superior ability to model the whole data set rather than individual pairs of residues, helping to distinguish between direct functional residue interactions and correlations resulting from indirect interactions. Sequences of 5-HT_{2A}R and

5-HT_{2C}R from 231 (Table 2) vertebrates were obtained from *GENBANK* (Benson D, et al., 1993) and aligned using *Clustal-Q* (Sievers F, et al., 2011), resulting in a concatenated Multiple Sequence Alignments (cMSA) of the 5-HT_{2A}R and 5-HT_{2C}R containing only orthologous sequences (one pair of proteins per species).

We compared the cMSA with structural domain data and secondary structure prediction for the receptor proteins computed using stand-alone versions of PSIPRED (McGuffin LJ, et al., 2000) (Methods). Both in 5-HT_{2A}R and 5-HT_{2C}R, the transmembrane domains are highly conserved compared to the extra- and intra-cellular domains.

The DCA map of 5-HT_{2A}R and 5-HT_{2C}R interactions.

Evolutionary pressure between two proteins can be visualized as a N residues × M residue map (a DCA map), where each point represents the strength of the coupling between a pair of residues. The DCA map of the predicted residue interactions in the 5-HT_{2A}R – 5-HT_{2C}R heterodimer was generated with the top 5% of the 471×458 interactions (471 and 458 are the lengths of 5-HT_{2A}R and 5-HT_{2C}R, respectively) (Fig. 1). For DCA to be successful, a certain level of variation in residues in a given position is needed, completely conserved residues will not give any coevolution signal. In our case, such variability is sufficient in intra- and extracellular domains of 5-HT_{2A}R and 5-HT_{2C}R, but their transmembrane domains are too well conserved (60% of transmembrane residues have a 100% conservation). Due to this high level of sequence conservation of the Transmembrane domains (TM), they do not produce strong signal, unlike IC and EC. This result should not be interpreted to mean that there are no evolutionary conserved interactions involving TM domains, but that, due to high conservation of TM domains, their functional interactions are more challenging, or inaccessible, to studying using DCA methods.

Postprocessing the 5-HT_{2A}R/5-HT_{2C}R DCA map: Filtering out the false positives from the 5-HT_{2A}R/5-HT_{2C}R 2A-2C DCA map by Local convolution

Traditional DCA maps are interpreted to mean strongest signal between residues, but since a strong signal in an island of background noise is not necessarily an indication of dimerization, additional post-processing is needed to improve the inference of interacting peptides. Moreover, even though DCA is designed to assign higher scores to direct contacts, this goal is not always perfectly met resulting in statistical noise hindering its interpretation as observed in Figure 2. This statistical noise is even higher when predicting native contacts between large proteins for which generally, only a small number of orthologous sequences is available (N/L < 0.7, see Table 2). The resulting false positive signals, responsible for the peaks connecting intracellular and extracellular domains (Figure 2), are difficult to interpret and must be filtered out.

Because DCA based methods are known to generate large amount of isolated False Positive (FP) contacts (dos Santos RN, et al., 2015; Ekeberg M, et al., 2014; Morcos F, et al., 2011; Sulkowska JI, et al., 2012), we have developed a method (see Methods section), based on local convolution of Evolutionary Coupling Scores (ECs) with a Gaussian kernel specific to the secondary structures of the protein, to reduce the FP and thus prioritize the most likely interacting peptides in protein complexes. The resulting filtered map of 5-HT_{2A}R/5-HT_{2C}R,

for several values of the normalized variances and average length of interacting peptides, is shown in Figure 3. It appears that local convolution of ECs helps identify most likely interacting peptides (hot spots regions in Figure 3, panel d).

Protein domains involved in the predicted 5-HT_{2A}R/5-HT_{2C}R interaction.

Because one of the main applications of the present study is to guide experiments determining the interacting peptides of oligomers, it is important to estimate the frequencies of domain-domain interactions which are the focus of experimental protocols such as point mutations. Both 5-HT_{2A}R and 5-HT_{2C}R consist of four Intra-Cellular domains (labelled IC1 to IC4), four Extra-Cellular domains (labelled EC1 to EC4), and seven Transmembrane domains (TM1 to TM7). For each individual domain in one receptor, we evaluated how frequently it interacts with other domains in adjacent protein. Only the top 5% interactions were considered. EC1 is the domain on 5-HT_{2A}R with more significant interactions whereas IC3 on 5-HT_{2C}R is the domain with more interactions. (Figure 4 and Figure 5).

Randomized phylogeny to assess the statistical significance of the predicted heterodimerization contacts

To compute the statistical significance of the heterodimerization, we used the randomized phylogeny which is defined as follows: in the joint MSA, we randomized the assignment of 5-HT_{2C}R sequences to species while keeping the 5-HT_{2A}R sequences unchanged and computed the ECs as previously described. For each randomized ECs, the number of predicted contacts higher than a threshold (minimum ECs score of the top 5% of the original ECs) was recorded. Of the 100-randomized phylogeny ECs performed, only one predicted a number of contacts higher than the actual ECs suggesting the predicted interacting peptides between the two receptors are not random and resulted from evolutionary pressure (Figure 6.c). The basic idea behind this method is that any randomization of orthologous sequences of one protein will disrupt the phylogeny and destroy evolutionary pressure between the two proteins (see Figure 6.b). The statistical significance of the heterodimerization is therefore computed as the total number of randomized ECs with more significant scores than the original, 1 EC, divided by the total number of randomized phylogenies, 100, leading to a statistical significant score of 0.01. This score can rigorously be considered as p-value considering our null hypothesis that there is no significant interaction between the two proteins and signals are due to noise.

Polymorphisms associated with inferred contacts

It has been hypothesized that the 5-HT_{2A}R/5-HT_{2C}R interaction plays a role in substance use disorders. There is significant variation in individual propensity to addiction disorders, attributed to genetic variants, including polymorphisms in receptor proteins (Bond C et al., 1998; Comings DE et al., 1991; Kreek MJ et al., 2005; Levran O et al., 2014; Levran O et al., 2015; Liu QR et al., 2006; Oslin DW et al., 2003; Szeto CYK et al., 2001; van den Wildenberg E et al., 2007; Wang DX et al., 2001). Also, systematic difference exists between male and female subjects (Brady KT and Randall CL, 1999; Perkins KA and Scott J, 2008). It is possible that some of these differences might be explained by genetic variants affecting the interaction between 5-HT_{2A}R and 5-HT_{2C}R, specifically by polymorphisms in receptor coding sequences that are translated to residues involved in the interaction. Also, the

statistical differences observed between the sexes may be caused by different propagation of genotypic variants to phenotype in males vs females in the case of the X-linked 5-HT_{2C}R gene.

To identify genetic polymorphisms of 5-HT_{2A}R and 5-HT_{2C}R that may affect the interaction, we queried the *NCBI* database of genetic variation (Sherry ST, et al., 2001) and extracted common variants coinciding with the coding sequences of HTR2A and of HTR2C. Next, we selected the SNPs that are located on our predicted interacting residues. We considered the affected residues when convolved ECs was greater than 0.01 (corresponding to the top 0.1% of the map, figure 7). Table 3 highlights the SNPs frequently associated with diseases and the full list of 5-HT_{2A}R and 5-HT_{2C}R SNPs is provided in table 4 and 5.

Several SNPs that our study predicts to be involved in the dimerization of 5-HT_{2A}R and 5-HT_{2C}R have been implicated in neuropsychiatric diseases and aggressive behavior (impulsivity, hostility) (Banlaki Z et al., 2015; Oades RD et al., 2008). On the HTR2C gene, rs6318 also known as Cys23Ser or 68G>C, is implicated in eating and anxiety disorders, obesity, depression, bipolar disorder, schizophrenia, and cocaine use disorder. On the HTR2A gene, rs6314, also known as His452Ty or C1354>T, is associated with depression and Cocaine use disorder.

An interesting observation is that polymorphisms in the genes coding for the HTR2A and HTR2C proteins are more prevalent in the intra- and extracellular domains (75 SNPs in 5-HT_{2C}R, on average 26 SNPs/100aa; 93 SNPs in 5-HT_{2A}R, on average 30 SNPs/100aa) than in the transmembrane domains (31 SNPs in 5-HT_{2C}R, on average 19 SNPs/100AA; 25 SNPs in 5-HT_{2A}R, on average 16 SNPs/100aa). Therefore, polymorphisms of clinical relevance affecting the function (including dimerization) of these proteins and potentially related to propensity toward substance use disorders, are more likely to exist in the IC and EC domains, even if interactions also exist in the TM regions of the two proteins. Intriguingly, this pattern is similar to the coincidence between interspecific and intraspecific variation previously observed in mRNA processing regions (Castle JC, 2011), and in disease-associated genes (Miller MP and Kumar S, 2001).

Coexpression of HTR2A and HTR2C in baboon circadian cycle

Analysis of gene expression data is a valuable tool for functional analysis of genes and proteins. Specifically, products of co-expressed genes are often involved in direct protein-protein interaction. Recently, a large timecourse study of baboon circadian circle has been published that covers circadian transcriptome in 64 different tissues, including 22 different brain regions, sampled every 2 h over 24 h (Mure LS, et al., 2018).

We computed the coexpression between the circadian profiles of the HTR2A and HTR2C genes as the Pearson Correlation coefficient between the two transcripts (Kudlicki A et al., 2007). We find that for most brain regions the correlation is positive. To assess the significance of the positive correlation, we calculated the fraction of named genes that have correlation with HTR2A lower than HTR2R and the fraction of named genes that have correlation with HTR2C lower than HTR2A. The results are summarized in Table 4. While this result does not prove that 5-HT_{2A}R and 5-HT_{2C}R are involved in a direct interaction, it

indicates in which tissues such functional contact may be taking place. One notable finding is that the correlation is especially high and statistically significant in the prefrontal cortex, a brain region associated with forming addiction (Goldstein RZ and Volkow ND, 2011). Other regions with high correlation of HTR2A and HTR2C are the putamen, habenula, Suprachiasmatic nuclei, and the cerebellum, all part of the central nervous system (CNS). Their coexpression in CNS is not surprising as 5-HT_{2A}R and 5-HT_{2C}R are involved in many other brain activities. The result is consistent with the hypothesis that the 5-HT_{2A}R/5-HT_{2C}R interaction indeed takes place and that it may play a role in substance abuse disorders (Anastasio NC et al., 2013; Anastasio NC et al., 2014; Anastasio NC et al., 2014; McAllister CEA, N.C.; Sonja J. Stutz, Rachel M. Hartley, Latham H. Fink, Bernard Fongang, Andrzej Kudlicki, Scott R. Gilbertson, YingChuChen, Harshini Neelankantan, Cheryl S. Watson, F. Gerard Moeller, Kathryn A. Cunningham 2016).

Whether 5-HT_{2A}R and 5-HT_{2C}R form a functional heterodimer has been a subject of debate over the last decade, while recent evidence strongly points towards the heterodimer forming at least in some conditions. The two receptors are particularly important as they are involved in substances abuse and may be used to design drugs targeting substance addiction. Several experimental evidences point towards the formation of a 5HT_{2A}R/5-HT_{2C}R dimer in the prefrontal cortex of rats and other model animals, but where and how this heterodimerization is formed remains elusive. Using computational methods and the co-evolution principle of functional residues of proteins, we concluded that the 5-HT_{2A}R and 5-HT_{2C}R form a functional heterodimer with interacting peptides either on the N-terminal domain, 3rd intracellular loop or on the C-terminal domain of the receptors. Although significant interactions are highlighted using the local convolution method, signal also remains that points to several interactions involving physically distant domains. A possible explanation is the propagation of spatially correlated noise which is persistent even after the local convolution step.

Traditional DCA methods work for large number of sequences ($N/L > 0.7$), but here we show that if one is only interested in predicting the peptides that are most likely to be involved in the interaction, a local convolution of ECs signals with the kernel adjusted to the secondary structure element is an efficient post-processing method. Moreover, the post-processing opens the door to the investigation of the heterodimerization in larger multi-domain proteins. We assessed the statistical significance of the predicted 5-HT_{2A}R/5-HT_{2C}R dimerization contacts using the randomized phylogeny and concluded that the p-value (probability that the observed interaction is due to noise) = 0.01. As it has been reported that the Single Nucleotide Polymorphisms (SNPs) play a significant role in differential propensity to substance abuse disorders, we determined the SNPs and found that a significant number of SNPs are located at the binding interfaces. Finally, an analysis of coexpression in the recently published primate transcription atlas suggests the prefrontal cortex as one possible tissue where the interaction between 5-HT_{2A}R/5-HT_{2C}R may be functional.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This study was conducted with the support by NIH grants P50 DA033935, R01 GM112131, the Institute for Translational Sciences, supported in part by a Clinical and Translational Science Award (UL1TR000071) from the National Center for Advancing Translational Sciences, and the Center for Addiction Research at the University of Texas Medical Branch. The authors are very grateful to Noelle C. Anastasio, Scott R. Gilbertson, F. Gerard Moeller for insightful discussions.

REFERENCES

- Abriata LA, Bovigny C, Dal Peraro M (2016), Detection and sequence/structure mapping of biophysical constraints to protein variation in saturated mutational libraries and protein sequence alignments with a dedicated server(vol 1, 242, 2016). *Bmc Bioinformatics* 17.
- Albizu L, Balestre MN, Breton C, Pin JP, Manning M, Mouillac B, Barberis C, Durroux T (2006), Probing the existence of G protein-coupled receptor dimers by positive and negative ligand-dependent cooperative binding. *Mol Pharmacol* 70:1783–1791. [PubMed: 16926282]
- Albizu L, Moreno JL, Gonzalez-Maesos J, Sealfon SC (2010), Heteromerization of G protein-coupled receptors: relevance to neurological disorders and neurotherapeutics. *CNS Neurol Disord Drug Targets* 9:636–650. [PubMed: 20632964]
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990), Basic local alignment search tool. *J Mol Biol* 215:403–410. [PubMed: 2231712]
- Anastasio NC, Gilbertson SR, Bubar MJ, Agarkov A, Stutz SJ, Jeng YJ, Bremer NM, Smith TD, et al. (2013), Peptide inhibitors disrupt the serotonin 5-HT_{2C} receptor interaction with phosphatase and tensin homolog to allosterically modulate cellular signaling and behavior. *Journal of Neuroscience* 33:1615–1630. [PubMed: 23345234]
- Anastasio NC, Liu S, Maili L, Swinford SE, Lane SD, Fox RG, Hamon SC, Nielsen DA, et al. (2014), Variation within the serotonin (5-HT) 5-HT_{2C} receptor system aligns with vulnerability to cocaine cue reactivity. *Translational Psychiatry* 4.
- Anastasio NC, Stutz SJ, Fink LH, Swinford-Jackson SE, Sears RM, DiLeone RJ, Rice KC, Moeller FG, et al. (2015), Serotonin (5-HT) 5-HT_{2A} Receptor (5-HT_{2AR}):5-HT_{2CR} Imbalance in Medial Prefrontal Cortex Associates with Motor Impulsivity. *ACS Chem Neurosci* 6:1248–1258. [PubMed: 26120876]
- Anastasio NC, Stutz SJ, Fox RG, Sears RM, Emeson RB, Di Leone RJ, O'Neil RT, Fink LH, et al. (2014), Functional Status of the Serotonin 5-HT_{2C} Receptor (5-HT_{2CR}) Drives Interlocked Phenotypes that Precipitate Relapse-Like Behaviors in Cocaine Dependence. *Neuropsychopharmacology* 39:360–372.
- Banlaki Z, Elek Z, Nanasi T, Szekely A, Nemoda Z, Sasvari-Szekely M, Ronai Z (2015), Polymorphism in the serotonin receptor 2a (HTR2A) gene as possible predisposal factor for aggressive traits. *PLoS One* 10:e0117792. [PubMed: 25658328]
- Benson D, Lipman DJ, Ostell J (1993), GenBank. *Nucleic Acids Res* 21:2963–2965. [PubMed: 8332518]
- Bitbol AF (2018), Inferring interaction partners from protein sequences using mutual information. *PLoS Comput Biol* 14:e1006401. [PubMed: 30422978]
- Bitbol AF, Dwyer RS, Colwell LJ, Wingreen NS (2016), Inferring interaction partners from protein sequences. *Proc Natl Acad Sci U S A* 113:12180–12185. [PubMed: 27663738]
- Bond C, LaForge KS, Tian MT, Melia D, Zhang SW, Borg L, Gong JH, Schluger J, et al. (1998), Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: Possible implications for opiate addiction. *Proceedings of the National Academy of Sciences of the United States of America* 95:9608–9613. [PubMed: 9689128]
- Brady KT, Randall CL (1999), Gender differences in substance use disorders. *Psychiatr Clin North Am* 22:241–252. [PubMed: 10385931]
- Breitwieser GE (2004), G protein-coupled receptor oligomerization: implications for G protein activation and cell signaling. *Circ Res* 94:17–27. [PubMed: 14715532]
- Calebiro D, Rieken F, Wagner J, Sungkaworn T, Zabel U, Borzi A, Cocucci E, Zurn A, et al. (2013), Single-molecule analysis of fluorescently labeled G-protein-coupled receptors reveals complexes

with distinct dynamics and organization. *Proc Natl Acad Sci U S A* 110:743–748. [PubMed: 23267088]

- Castle JC (2011), SNPs occur in regions with less genomic sequence conservation. *PLoS One* 6:e20660. [PubMed: 21674007]
- Champeimont R, Laine E, Hu SW, Penin F, Carbone A (2016), Coevolution analysis of Hepatitis C virus genome to identify the structural and functional dependency network of viral proteins. *Scientific Reports* 6.
- Comings DE, Comings BG, Muhleman D, Dietz G, Shahbahrami B, Tast D, Knell E, Kocsis P, et al. (1991), The Dopamine-D2 Receptor Locus as a Modifying Gene in Neuropsychiatric Disorders. *Jama-J Am Med Assoc* 266:1793–1800.
- de Juan D, Pazos F, Valencia A (2013), Emerging methods in protein co-evolution. *Nature Reviews Genetics* 14:249–261.
- dos Santos RN, Morcos F, Jana B, Andricopulo AD, Onuchic JN (2015), Dimeric interactions and complex formation using direct coevolutionary couplings. *Scientific Reports* 5.
- Ekeberg M, Hartonen T, Aurell E (2014), Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences. *Journal of Computational Physics* 276:341–356.
- Feinauer C, Szurmant H, Weigt M, Pagnani A (2016), Inter-Protein Sequence Co-Evolution Predicts Known Physical Interactions in Bacterial Ribosomes and the Trp Operon. *Plos One* 11.
- Felsing DE, Anastasio NC, Miszkil JM, Gilbertson SR, Allen JA, Cunningham KA (2018), Biophysical validation of serotonin 5-HT_{2A} and 5-HT_{2C} receptor interaction. *PLoS One* 13:e0203137. [PubMed: 30157263]
- Filmore D (2004), It's a GPCR world. *Modern Drug Discovery*:24–28.
- Fongang B, Kudlicki A (2013), The precise timeline of transcriptional regulation reveals causation in mouse somitogenesis network. *BMC Dev Biol* 13:42. [PubMed: 24304493]
- Fongang B, Kudlicki A (2016), Comparison between Timelines of Transcriptional Regulation in Mammals, Birds, and Teleost Fish Somitogenesis. *PLoS One* 11:e0155802. [PubMed: 27192554]
- Goldstein RZ, Volkow ND (2011), Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci* 12:652–669. [PubMed: 22011681]
- Guede T, Baldassi C, Zamparo M, Weigt M, Pagnani A (2016), Simultaneous identification of specifically interacting paralogs and interprotein contacts by direct coupling analysis. *Proceedings of the National Academy of Sciences of the United States of America* 113:12186–12191. [PubMed: 27729520]
- Guo W, Shi L, Filizola M, Weinstein H, Javitch JA (2005), Crosstalk in G protein-coupled receptors: changes at the transmembrane homodimer interface determine activation. *Proc Natl Acad Sci U S A* 102:17495–17500. [PubMed: 16301531]
- Halabi N, Rivoire O, Leibler S, Ranganathan R (2009), Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138:774–786. [PubMed: 19703402]
- Hannon J, Hoyer D (2008), Molecular biology of 5-HT receptors. *Behav Brain Res* 195:198–213. [PubMed: 18571247]
- Herrick-Davis K, Grinde E, Lindsley T, Cowan A, Mazurkiewicz JE (2012), Oligomer size of the serotonin 5-hydroxytryptamine 2C (5-HT_{2C}) receptor revealed by fluorescence correlation spectroscopy with photon counting histogram analysis: evidence for homodimers without monomers or tetramers. *J Biol Chem* 287:23604–23614. [PubMed: 22593582]
- Herrick-Davis K, Grinde E, Lindsley T, Teitler M, Mancia F, Cowan A, Mazurkiewicz JE (2015), Native serotonin 5-HT_{2C} receptors are expressed as homodimers on the apical surface of choroid plexus epithelial cells. *Mol Pharmacol* 87:660–673. [PubMed: 25609374]
- Homouz D, Chen G, Kudlicki AS (2015), Correcting positional correlations in Affymetrix (R) Genome Chips. *Sci Rep-Uk* 5.
- Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS (2012), Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. *Cell* 149:1607–1621. [PubMed: 22579045]
- Hopf TA, Scharfe CPI, Rodrigues JPGLM, Green AG, Kohlbacher O, Sander C, Bonvin AMJJ, Marks DS (2014), Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* 3.

- Howell LL, Cunningham KA (2015), Serotonin 5-HT₂ receptor interactions with dopamine function: implications for therapeutics in cocaine use disorder. *Pharmacol Rev* 67:176–197. [PubMed: 25505168]
- Iglesias A, Cimadevila M, Cadavid MI, Loza MI, Brea J (2017), Serotonin-2A homodimers are needed for signalling via both phospholipase A₂ and phospholipase C in transfected CHO cells. *Eur J Pharmacol* 800:63–69. [PubMed: 28216047]
- Iglesias A, Cimadevila M, la Fuente RA, Marti-Solano M, Cadavid MI, Castro M, Selent J, Loza MI, et al. (2017), Serotonin 2A receptor disulfide bridge integrity is crucial for ligand binding to different signalling states but not for its homodimerization. *Eur J Pharmacol* 815:138–146. [PubMed: 28899696]
- Iglesias A, Lage S, Cadavid MI, Loza MI, Brea J (2016), Development of a Multiplex Assay for Studying Functional Selectivity of Human Serotonin 5-HT_{2A} Receptors and Identification of Active Compounds by High-Throughput Screening. *J Biomol Screen* 21:816–823. [PubMed: 27095818]
- Keene JD, Komisarow JM, Friedersdorf MB (2006), RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* 1:302–307. [PubMed: 17406249]
- Kim DE, DiMaio F, Wang RYR, Song YF, Baker D (2014), One contact for every twelve residues allows robust and accurate topology-level protein structure modeling. *Proteins-Structure Function and Bioinformatics* 82:208–218.
- Kreek MJ, Bart G, Lilly C, Laforge KS, Nielsen DA (2005), Pharmacogenetics and human molecular genetics of opiate and cocaine addictions and their treatments. *Pharmacol Rev* 57:1–26. [PubMed: 15734726]
- Kudlicki A, Plewa T, Rozyczka M (1996), CPPA - A new hydrodynamical code for cosmological large-scale structure simulations. *Acta Astronomica* 46:297–309.
- Kudlicki A, Rowicka M, Otwinowski Z (2007), SCEPTRANS: an online tool for analyzing periodic transcription in yeast. *Bioinformatics* 23:1559–1561. [PubMed: 17400726]
- Lee SP, O'Dowd BF, George SR (2003), Homo- and hetero-oligomerization of G protein-coupled receptors. *Life Sci* 74:173–180. [PubMed: 14607244]
- Levrán O, Peles E, Randesi M, Correa da Rosa J, Ott J, Rotrosen J, Adelson M, Kreek MJ (2014), Dopaminergic pathway polymorphisms and heroin addiction: further support for association of CSNK1E variants. *Pharmacogenomics* 15:2001–2009. [PubMed: 25521358]
- Levrán O, Peles E, Randesi M, Correa da Rosa J, Ott J, Rotrosen J, Adelson M, Kreek MJ (2015), Susceptibility loci for heroin and cocaine addiction in the serotonergic and adrenergic pathways in populations of different ancestry. *Pharmacogenomics* 16:1329–1342. [PubMed: 26227246]
- Liu QR, Drgon T, Johnson C, Walther D, Hess J, Uhl GR (2006), Addiction molecular genetics: 639,401 SNP whole genome association identifies many “cell adhesion” genes. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics* 141b:918–925.
- Liu Y, Palmedo P, Ye Q, Berger B, Peng J (2018), Enhancing Evolutionary Couplings with Deep Convolutional Neural Networks. *Cell Syst* 6:65–74 e63. [PubMed: 29275173]
- Lockless SW, Ranganathan R (1999), Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286:295–299. [PubMed: 10514373]
- Lua RC, Wilson SJ, Konecki DM, Wilkins AD, Venner E, Morgan DH, Lichtarge O (2016), UET: a database of evolutionarily-predicted functional determinants of protein sequences that cluster as functional sites in protein structures. *Nucleic Acids Research* 44:D308–D312. [PubMed: 26590254]
- Lukasiewicz S, Polit A, Kedracka-Krok S, Wedzony K, Mackowiak M, Dziedzicka-Wasylewska M (2010), Hetero-dimerization of serotonin 5-HT_{2A} and dopamine D₂ receptors. *Biochim Biophys Acta* 1803:1347–1358. [PubMed: 20831885]
- Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C (2011), Protein 3D structure computed from evolutionary sequence variation. *PLoS One* 6:e28766. [PubMed: 22163331]
- Marks DS, Hopf TA, Sander C (2012), Protein structure prediction from sequence variation. *Nature Biotechnology* 30:1072–1080.

- McAllister CEA NC; Stutz Sonja J., Hartley Rachel M., Fink Latham H., Fongang Bernard, Kudlicki Andrzej, Gilbertson Scott R., Chen YingChu, Neelankantan Harshini, Watson Cheryl S., Moeller F. Gerard, Cunningham Kathryn A. (2016), Investigating a putative Protein:Protein interaction between the serotonin (5HT) 5-HT_{2A} Receptor (5-HTR_{2AR}) and 5-HTR_{2CR}. *International Society for Serotonin Research*:11.
- McGuffin LJ, Bryson K, Jones DT (2000), The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405. [PubMed: 10869041]
- Miller MP, Kumar S (2001), Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet* 10:2319–2328. [PubMed: 11689479]
- Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, et al. (2011), Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences of the United States of America* 108:E1293–E1301. [PubMed: 22106262]
- Moreno JL, Muguruza C, Umali A, Mortillo S, Holloway T, Pilar-Cuellar F, Mocci G, Seto J, et al. (2012), Identification of three residues essential for 5-hydroxytryptamine 2A-metabotropic glutamate 2 (5-HT_{2A}.mGlu₂) receptor heteromerization and its psychoactive behavioral function. *J Biol Chem* 287:44301–44319. [PubMed: 23129762]
- Moutkine I, Quentin E, Guiard BP, Maroteaux L, Doly S (2017), Heterodimers of serotonin receptor subtypes 2 are driven by 5-HT_{2C} protomers. *J Biol Chem* 292:6352–6368. [PubMed: 28258217]
- Mure LS, Le HD, Benegiamo G, Chang MW, Rios L, Jillani N, Ngotho M, Kariuki T, et al. (2018), Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science* 359.
- Naveed H, Xu Y, Jackups R, Liang J (2012), Predicting Three-Dimensional Structures of Transmembrane Domains of beta-Barrel Membrane Proteins. *Journal of the American Chemical Society* 134:1775–1781. [PubMed: 22148174]
- Neuwalder AF (2016), Gleaning structural and functional information from correlations in protein multiple sequence alignments. *Current Opinion in Structural Biology* 38:1–8. [PubMed: 27179293]
- O'Dowd BF, Ji X, Nguyen T, George SR (2012), Two amino acids in each of D1 and D2 dopamine receptor cytoplasmic regions are involved in D1-D2 heteromer formation. *Biochem Biophys Res Commun* 417:23–28. [PubMed: 22100647]
- Oades RD, Lasky-Su J, Christiansen H, Faraone SV, Sonuga-Barke EJ, Banaschewski T, Chen W, Anney RJ, et al. (2008), The influence of serotonin- and other genes on impulsive behavioral aggression and cognitive impulsivity in children with attention-deficit/hyperactivity disorder (ADHD): Findings from a family-based association test (FBAT) analysis. *Behav Brain Funct* 4:48. [PubMed: 18937842]
- Oslin DW, Berrettini W, Kranzler HR, Pettinati H, Gelernter J, Volpicelli JR, O'Brien CP (2003), A functional polymorphism of the mu-opioid receptor gene is associated with naltrexone response in alcohol-dependent patients. *Neuropsychopharmacol* 28:1546–1552.
- Ovchinnikov S, Kamisetty H, Baker D (2014), Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* 3.
- Overington JP, Al-Lazikani B, Hopkins AL (2006), How many drug targets are there? *Nat Rev Drug Discov* 5:993–996. [PubMed: 17139284]
- Perkins KA, Scott J (2008), Sex differences in long-term smoking cessation rates due to nicotine patch. *Nicotine Tob Res* 10:1245–1250. [PubMed: 18629735]
- Phizicky EM, Fields S (1995), Protein-protein interactions: methods for detection and analysis. *Microbiol Rev* 59:94–123. [PubMed: 7708014]
- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999), A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17:1030–1032. [PubMed: 10504710]
- Rochette S, Diss G, Filteau M, Leducq JB, Dube AK, Landry CR (2015), Genome-wide protein-protein interaction screening by protein-fragment complementation assay (PCA) in living cells. *J Vis Exp*.

- Rowicka M, Kudlicki A, Tu BP, Otwinowski Z (2007), High-resolution timing of cell cycle-regulated gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 104:16892–16897. [PubMed: 17827275]
- Scarselli M, Annibale P, McCormick PJ, Kolachalam S, Aringhieri S, Radenovic A, Corsini GU, Maggio R (2016), Revealing G-protein-coupled receptor oligomerization at the single-molecule level through a nanoscopic lens: methods, dynamics and biological function. *FEBS J* 283:1197–1217. [PubMed: 26509747]
- Schellekens H, Dinan TG, Cryan JF (2013), Taking two to tango: a role for ghrelin receptor heterodimerization in stress and reward. *Front Neurosci* 7:148. [PubMed: 24009547]
- Schellekens H, van Oeffelen WE, Dinan TG, Cryan JF (2013), Promiscuous dimerization of the growth hormone secretagogue receptor (GHS-R1a) attenuates ghrelin-mediated signaling. *J Biol Chem* 288:181–191. [PubMed: 23161547]
- Schmiedel J, Lehner B (2018), Determining protein structures using genetics.
- Sekar RB, Periasamy A (2003), Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol* 160:629–633. [PubMed: 12615908]
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001), dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* 29:308–311. [PubMed: 11125122]
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, et al. (2011), Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7.
- Suel GM, Lockless SW, Wall MA, Ranganathan R (2003), Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 10:59–69. [PubMed: 12483203]
- Sulkowska JJ, Morcos F, Weigt M, Hwa T, Onuchic JN (2012), Genomics-aided structure prediction. *Proceedings of the National Academy of Sciences of the United States of America* 109:10340–10345. [PubMed: 22691493]
- Szeto CYK, Tang NLS, Lee DTS, Stadlin A (2001), Association between mu opioid receptor gene polymorphisms and Chinese heroin addicts. *Neuroreport* 12:1103–1106. [PubMed: 11338173]
- Tesileanu T, Colwell LJ, Leibler S (2015), Protein Sectors: Statistical Coupling Analysis versus Conservation. *Plos Computational Biology* 11.
- Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, et al. (2000), A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623–627. [PubMed: 10688190]
- van den Wildenberg E, Wiers RW, Dessers J, Janssen RGJH, Lambrichs EH, Smeets HJM, van Breukelen GJP (2007), A functional polymorphism of the mu-opioid receptor gene (OPRM1) influences cue-induced craving for alcohol in male heavy drinkers. *Alcohol Clin Exp Res* 31:1–10. [PubMed: 17207095]
- Wang DX, Quillan JM, Winans K, Lucas JL, Sadee W (2001), Single nucleotide polymorphisms in the human mu opioid receptor gene alter basal G protein coupling and calmodulin binding. *Journal of Biological Chemistry* 276:34624–34630. [PubMed: 11457836]
- Weigt M, White RA, Szurmant H, Hoch JA, Hwa T (2009), Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A* 106:67–72. [PubMed: 19116270]
- Yu JC, Vavrusa M, Andreani J, Rey J, Tuffery P, Guerois R (2016), InterEvDock: a docking server to predict the structure of protein-protein interactions using evolutionary information. *Nucleic Acids Research* 44:W542–W549. [PubMed: 27131368]

Figure 1: Convolved DCA signal.

Neighborhood contribution to DCA Evolutionary score between residues. The original DCA signal (a, green circle) is convolved by a kernel centered around residues at position (50,50) on a window of 21 amino-acid. The final convolved signal (b, green circle) has been scaled to include the contribution of the neighboring residues.

Figure 2: The raw, non-optimized DCA map of 5-HT_{2A}R – 5-HT_{2C}R.

The figure shows top 5% of the interaction between the 2 receptors. Receptor domains are color coded (TM= Transmembrane (blue), IC = Intra-Cellular (orange), EC = Extra-Cellular (green)). Due to high conservation of the TM domain, most signal are located at EC and IC domains.

Figure 3: Convolved ECs of 5-HT_{2A}R – 5-HT_{2C}R highlights most likely interacting peptides. The local convolution method was applied to raw ECs (a) using optimized parameters as described in the Methods section with l (total number of interacting aa) = 17 (panel b), 19 (panel c), and 21 (panel d). The variances of the Gaussian function are $a = b = 0.01$ for $\gamma=1$ and $1 = b = 0.001$ for $\gamma=2$.

Figure 4: Domain interaction of 5-HT_{2A}R – 5-HT_{2C}R.

We evaluate how frequently domains interact with each other (EC: Extra-cellular, IC: Intra-Cellular, TM: Transmembrane domain). A) frequency of 5-HT_{2A}R domains predicted to interact with 5-HT_{2C}R; B) frequency of 5-HT_{2C}R domains predicted to interact with 5-HT_{2A}R; C) inter-domain interactions between 5-HT_{2A}R and 5-HT_{2C}R; D-L) subdomain interactions between 5-HT_{2A}R and 5-HT_{2C}R: D = EC/EC, E = EC/IC, F = IC/TM, G = IC/EC, H = IC/IC, I = IC/TM, J = TM/EC, K = TM/IC, L = TM/TM.

Figure 5: Domain interaction of 5-HT_{2A}R – 5-HT_{2C}R.

A graph representing the interactions between the specific domains. Arc widths corresponds to numbers of inter-protein residue pairs of the top 5% of the convolved DCA signal. (EC: Extra-cellular, IC: Intra-Cellular, TM: Transmembrane domain).

Figure 6: Randomized phylogeny to assess the statistical significance of the heterodimerization. We randomized species assignments in the MSA of the 5-HT_{2C}R while keeping the 5-HT_{2A}R. We estimated the number of residues with signal larger than the smallest among of the 5% top original signals and found that only one randomized ECs has more significant ECs leading to a p-value = 0.01 (see text for description)

Figure 7: The density of SNPs is correlated with the convolved EC scores.

Non-synonymous SNPs are shown as red lines for 5-HT_{2A}R and green for 5-HT_{2C}R. Their densities are correlated with the convolved ECs and known SNPs involved in depression and cocaine use disorder are located at the binding interfaces.

Table 1:
Optimized values of the Gaussian Convolution:

Values were optimized using serotonin proteins and GPCRs. a and b are the optimized variances of the Gaussian Kernel, l is the average length of the interacting peptides and $\gamma = 1$ if the residues belong to the same secondary structure, if not $\gamma = 2$.

	a	b	l (AA)
$\gamma = 1$	0.01 – 0.05	0.01 – 0.05	17 – 24
$\gamma = 2$	0.001 – 0.008	0.001 – 0.008	17 – 24

Table 2:**Receptor properties.**

The significant length of the complex (Sig. length; total length of the cMSA / # of sequences) was computed to be 0.25, lower than the 0.7 previously suggested for DCA analysis in single proteins (Hopf TA, et al., 2014). Because the number of sequences is too small for traditional DCA analysis, we have introduced a Gaussian convolution of local ECs scores and secondary structure to filter noise and spurious contacts, thus highlighting the most likely interacting peptides.

	Protein length	# individual sequences	cMSA	Sig. length	Sequence coverage
5-HT _{2C} R	458 AA	244	231	0.25	55 – 90%
5-HT _{2A} R	471 AA	262			

Table 3:
Significant SNPs at interacting peptides of 5-HT_{2A}R – 5-HT_{2C}R:

full list of SNPs located on the interacting peptides is provided in supplementary data file SDF3.

SNPs ID	SNP in	5-HT _{2A} R position	5-HT _{2A} R AA	5-HT _{2A} R Domain	5-HT _{2C} R position	5-HT _{2C} R AA	5-HT _{2C} R Domain	Convolved ECs
rs6314	5-HT _{2A} R	452	H	IC4	278	A	IC3	0.049
rs769223465	5-HT _{2A} R	467	K	IC4	286	A	IC3	0.030
rs773466572	5-HT _{2A} R	78	A	TM1	278	A	IC3	0.025
rs749275325	5-HT _{2A} R	463	G	IC4	280	P	IC3	0.043
rs781643370	5-HT _{2A} R	214	G	TM4	279	N	IC3	0.023
rs75634836	5-HT _{2A} R	241	V	TM5	428	S	IC4	0.022
rs6318	5-HT _{2C} R	25	T	EC1	23	C	EC1	0.025
rs143566182	5-HT _{2C} R	453	S	IC4	42	D	EC1	0.034
rs781896215	5-HT _{2C} R	457	S	IC4	59	S	TM1	0.027
rs144618891	5-HT _{2C} R	454	E	IC4	188	P	TM4	0.022
rs144618891	5-HT _{2C} R	457	S	IC4	188	P	TM4	0.022
rs368641796	5-HT _{2C} R	457	S	IC4	214	F	TM5	0.022
rs782627259	5-HT _{2C} R	460	N	IC4	323	M	TM6	0.023

Table 4:
Gene expressions of HTR2A and HTR2C are correlated in baboon circadian cycle. PCC:

Pearson Correlation Coefficient; **FRPG – Random**: Fraction of Random Pairs of named Genes with lower PCC in the same dataset (based on 10000 random pairs); **FNG - HTR2A**: Fraction of Named Genes that have lower PCC with HTR2A in the same dataset; **FNG - HTR2C**: fraction of named genes that have lower PCC with HTR2C in the same dataset.

TISSUE	PCC	FRPG - Random	FNG - HTR2A	FNG - HTR2C	TISSUE NAME
VIC	-0.4740	0.0871	0.0390	0.2852	Visual cortex
PUT	-0.2670	0.1822	0.1420	0.2591	Putamen
HAB	-0.1570	0.2773	0.3048	0.3273	Habenula
SCN	-0.1040	0.2993	0.2258	0.4452	Suprachiasmatic nuclei
CER	-0.0020	0.3133	0.5560	0.3693	Cerebellum
LH	0.0020	0.6056	0.5847	0.6835	Lateral hypothalamus
PON	0.0270	0.6426	0.5594	0.7179	Pons
VMH	0.0460	0.6386	0.5765	0.5239	Ventromedial hypothalamus
ARC	0.0590	0.6076	0.5665	0.5481	Arcuate nucleus
DMH	0.0940	0.6757	0.5685	0.6707	Dorsomedial hypothalamus
MGP	0.2150	0.7568	0.7730	0.8283	Medial globus pallidus
LGP	0.2470	0.7928	0.7237	0.8563	Lateral globus pallidus
PVN	0.2880	0.7648	0.8034	0.6770	Paraventricular nuclei
SUN	0.3040	0.7998	0.9069	0.8041	Substantia nigra
PRA	0.3050	0.8038	0.8143	0.7510	Preoptic area
HIP	0.3570	0.8228	0.7940	0.8293	Hippocampus
AMY	0.4920	0.7778	0.9025	0.9463	Amygdala
ONH	0.5860	0.8939	0.8970	0.9507	Optic nerve head
SON s	0.6250	0.9489	0.9138	0.9200	Supraoptic nucleu
OLB	0.6660	0.9449	0.8654	0.9791	Olfactory bulb
PRC	0.6880	0.9509	0.9851	0.9747	Prefrontal cortex
MMB	0.7490	0.9640	0.9771	0.9590	Mammilliary body
THA	0.7780	0.9770	0.9742	0.9363	Thalamus