FEBS Letters 589 (2015) 548–552

journal homepage: www.FEBSLetters.org

Characteristic molecular vibrations of adenosine receptor ligands

Hyun Keun Chee [a], Jin-San Yang [b], Je-Gun Joung [c], Byoung-Tak Zhang [b], S. June Oh [d,*]

[a] Department of Thoracic and Cardiovascular Surgery, Konkuk University School of Medicine, Seoul 143-729, Republic of Korea
[b] School of Computer Science and Engineering, Seoul National University, Seoul 151-742, Republic of Korea
[c] Translational Bioinformatics Lab (TBL), Samsung Genome Institute (SGI), Samsung Medical Center, Seoul 135-710, Republic of Korea
[d] Department of Pharmacology, Inje University College of Medicine, Busan 614-735, Republic of Korea

ARTICLE INFO

Article history:
Received 17 September 2014
Revised 27 November 2014
Accepted 15 January 2015
Available online 23 January 2015

Edited by Alfonso Valencia

Keywords:
Adenosine receptor
Molecular vibration-activity relationship
Information gain
t-Test
Support vector machine
G protein-coupled receptor

ABSTRACT

Although the regulation of membrane receptor activation is known to be crucial for molecular signal transduction, the molecular mechanism underlying receptor activation is not fully elucidated. Here we study the physicochemical nature of membrane receptor behavior by investigating the characteristic molecular vibrations of receptor ligands using computational chemistry and informatics methods. By using information gain, *t*-tests, and support vector machines, we have identified highly informative features of adenosine receptor (AdoR) ligand and corresponding functional amino acid residues such as Asn (6.55) of AdoR that has informative significance and is indispensable for ligand recognition of AdoRs. These findings may provide new perspectives and insights into the fundamental mechanism of class A G protein-coupled receptor activation.
© 2015 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

Molecular recognition of membrane receptors in biological systems plays a crucial role in intercellular and intracellular transduction of signals. G protein-coupled receptors (GPCRs), also known as seven-transmembrane-segment receptors (7TMRs), are integral membrane proteins that are connected by 3 extracellular and 3 intracellular loops of variable length, and they transmit ligand information by interacting with trimeric GTP-binding proteins or β-arrestins to modulate intracellular pathways. GPCRs constitute the largest family of proteins encoded in the human genome [1,2] and play pivotal roles in the transmission of extracellular signals into cells. This family of proteins is known to react with a broad range of ligands such as hormone molecules [3], volatile organic compounds [4], tastants [5,6], and even photons [7]. They are also major targets of modern drugs and are associated with more than one-third of pharmaceuticals [8].

Among the GPCR classes, rhodopsin-like class A GPCRs have the simplest polypeptide ends and the greatest number of reported three-dimensional (3-D) structures [9]. Firing of a signal and the accompanying information transfer is elicited by an agonist that activates its cognate receptor; however, the molecular mechanism underlying receptor activation is not simple. Lefkowitz and his colleagues have reported that agonist-biased activation of class A GPCRs is related to β-arrestins [10–14].

The fundamental mechanism of olfaction mediated by odorant receptors, members of the class A GPCR family, is controversial [15–17]. Various attempts have been made to describe the molecular mechanism of ligand-receptor recognition during olfaction, such as the classical binding theory and vibration theory [18–20] of electron transfer [21–23]. The former theory, in which ligand specificity is explained by its molecular shape, has been developed into the pharmacophore concept and is generally accepted by researchers. However, this theory is not sufficient to account for the diversity of ligands and complexity of GPCR agonism. In recent decades, various models and experiments have been used to explain the activation of the olfactory receptor, a class A GPCR, by means of a molecular vibrationally assisted electron tunneling mechanism [17,21–24].

Borea et al. reported the thermodynamic discrimination in AdoRs [25,26] and neuronal nicotinic receptor [27] as a method of studying ligand-receptor interactions. According to these papers, agonistic binding was both enthalpy- and entropy-driven, while antagonistic binding was entirely entropy-driven. Pivonka made a report that the spectral trends of infrared (IR) and/or Raman analyses of human estrogen receptor β (ER-β) ligands mirror the trends in binding strength values obtained from biological assays [28]. Takane et al. showed the existence of a structure-odor relationship by a ligand-based approach using EigenVAlue (EVA) descriptor and

* Corresponding author. Fax: +82 51 797 8569.
E-mail address: o@biophilos.org (S.J. Oh).

http://dx.doi.org/10.1016/j.febslet.2015.01.024
0014-5793/© 2015 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

hierarchical clustering [29]. Of the various molecular descriptors, EVA is a unique approach based on IR range molecular vibrational frequencies [30]. More recently, computational approaches were used to search for molecular vibration-activity relationships in the agonism of histamine and adenosine receptors, and the authors suggested that the molecular vibrational frequency pattern may serve as a possible molecular descriptor for the classification of agonist and antagonist class A GPCR ligands [31,32]. Thus, in this study, we focus on the possible characteristics that contribute to the activation of class A GPCRs rather than their conventional biochemical properties.

Although class A GPCRs have relatively low sequence similarity, their endogenous agonists are highly conserved. Four AdoR subtypes have been identified: AdoRA1, AdoRA2A, AdoRA2B, and AdoRA3 [33]. While each AdoR subtype interacts with and is activated/inactivated by specific ligands, biogenic nucleoside adenosine commonly activates all AdoRs. Adenosine interacts with AdoRs involved in various diseases including cardiac ischemia, arrhythmia, neurodegeneration, diabetes, glaucoma, and inflammation [34]. It also plays an important role in managing asthma and arthritis, and in finding applications for the treatment of pain, cancer and other disorders [35]. We thus designed and utilized a computational approach to investigate characteristic molecular vibrations of AdoR agonists and non-agonists (antagonists and inverse agonists).

2. Materials and methods

To facilitate their classification, AdoR ligands were grouped into two categories: agonists and non-agonists. Non-agonists included antagonists and inverse agonists that block and decrease agonist-mediated receptor activation, respectively.

2.1. Dataset

A 64-ligand dataset consisting of 30 AdoR agonists and 34 non-agonists was used (Table S1). Three-dimensional structure data format (SDF) files for the AdoR ligands were downloaded from the PubChem Compound Database at the National Center for Biotechnology Information and subjected to geometry optimization, molecular vibrational pattern analysis, and further study.

2.2. Molecular vibration calculation and data formulation

First, geometry optimization was carried out since the calculation of molecular vibrational frequencies requires a given 3-D structure of a given molecule. The theoretical 3-D conformer SDF of each molecule was modeled as a single low-energy conformation by using the Becke and Lee, Yang, Parr correlation (BLYP) density functional theory (DFT) and standard split-valence basis set 6-31G(d,p). The results of geometry optimization were then subjected to vibrational frequency calculations. All calculations of geometry optimization and normal modes of molecular vibration were performed using the GAMESS program package [36,37].

To compare the molecular vibration patterns of AdoR ligands while maintaining their characteristics of molecular vibration, the corralled intensity of molecular vibrational frequency (CIMVF) of each ligand was generated as a vector of 800 elements as previously reported [32]. To restrict the number of non-discrete vibrational frequency spaces and retain the properties of molecular vibration, we discretized the molecular vibration dataset of each AdoR ligand as follows.

Let (x_i, a_i) represent the ith pair of vibrational frequency and amplitude (intensity) among n observed pairs. Transform x_i by $y_i = |\frac{x_i}{c}|$ where $|x|$ is the largest integer not greater than x and c is the corral size. Denote M distinct (ascending) integer values of as $\{z_1, \ldots, z_M\}$ and represent I_m for the set of indices corresponding to the discretized z_m (where $I_m = \{i | y_i = z_m, i \in \{1, .., n\}\}$) for $m = 1, \ldots, M$. If b_m is the sum of amplitudes with indices (a_i's) corresponding to I_m for $m = 1, \cdots, M$, then $(z_1,\ b_1),\ (z_2,\ b_2), \cdots, (z_M,\ b_M)$ become our new discretized data pairs in the range of $0 \leqslant z_m \leqslant 4000/c$.

Finally, the CIMVF of a ligand is represented as a one-dimensional vector containing 800 elements of vibrational intensity for the vibrational frequency range from 0 to 4000 cm^{-1} by setting the corral size c to 5 cm^{-1}. It should be noted that the CIMVF did not correspond to the IR or Raman spectrum of the relevant ligand. During feature selection, the corrals of molecular vibration were regarded as features of each ligand.

2.3. Feature selection by information gain

The dimension of a dataset is the number of variables or features that are measured with each observation. One of the challenges with high-dimensional datasets is that not all of the features are important or informative for understanding the underlying mechanism of a particular phenomenon. Feature selection is a method for reducing meaningless and less informative features. The overall procedure of feature selection involves scoring each potential feature according to a particular feature selection metric. Scoring involves separately counting the occurrences of a feature in positive- and negative-class training examples, and then computing a function of these [38]. The information gain (IG) yielded from a dataset is given by the relative entropy (also known as Kullback–Leibler divergence [39]) between the prior and posterior probabilities [40]. IG measures the amount of information about the class prediction in bits, if the only information available is the presence of a feature and the corresponding class distribution [41].

The IG is

$$\mathrm{IG}(S_x, x_i) = H(S_x) - \sum_{v=Values(x_i)} \frac{|S_{x_i=v}|}{|S_x|} \cdot H(S_{x_i=v})$$

where H is the entropy function, S_x is the set of training examples, x_i is the vector of the ith variable in the set, and $|S_{x_i=v}| / |S_x|$ is the fraction of examples of the ith variable having value v.

Thus, we applied IG-based feature selection to identify the corrals of molecular vibrational frequency that were the most informative among the 800 elements for binary classification of AdoR ligands as agonists or non-agonists. An IG of zero implied that the corresponding feature was no better than that of random sampling. We trained and tested the procedure by applying leave-one-out cross-validation to each ligand. The calculation of IG was performed using the Weka machine learning package [42].

2.4. Parametric and non-parametric analyses of informative features

Because each group of agonists and non-agonists has a tendency to show intensities in specific frequency ranges, we wanted to identify a set of meaningful frequency ranges where the mean responses of the two groups were significantly different. After analyzing the intensities over the range of 800 features, a subset of 18 features were selected for testing the equality of the mean intensities of the two groups, where α was equal to 0.01 in two-sample t-tests.

We also selected meaningful features by using the linear support vector machine (SVM) to compare the t-test results. For each feature, the agonist and non-agonist groups were classified by 10-fold cross validation using the SVM classifier. Here, data sets with particular features were randomly divided in two sets; 90% of samples were assigned into a training set and 10% of the samples were

Download English Version:

https://daneshyari.com/en/article/10870009

Download Persian Version:

https://daneshyari.com/article/10870009

Daneshyari.com