Characteristic molecular vibrations of adenosine receptor ligands

Hyun Keun Chee\textsuperscript{a}, Jin-San Yang\textsuperscript{b}, Je-Gun Joung\textsuperscript{c}, Byoung-Tak Zhang\textsuperscript{b}, S. June Oh\textsuperscript{d,\textup{e}}

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

\textsuperscript{e}School of Computer Science and Engineering, Seoul National University, Seoul 151-742, Republic of Korea

\begin{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.
\end{abstract}

\section{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 \(\beta\)-arrestins to modulate intracellular pathways. GPCRs constitute the largest family of proteins encoded in the human genome \cite{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 \cite{3}, volatile organic compounds \cite{4}, tastants \cite{5,6}, and even photons \cite{7}. They are also major targets of modern drugs and are associated with more than one-third of pharmaceuticals \cite{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 \cite{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 \(\beta\)-arrestins \cite{10–14}.

The fundamental mechanism of olfaction mediated by odorant receptors, members of the class A GPCR family, is controversial \cite{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 \cite{18–20} of electron transfer \cite{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 \cite{17,21–24}.

Borea et al. reported the thermodynamic discrimination in AdoRs \cite{25,26} and neuronal nicotinic receptor \cite{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 \(\beta\) (ER-\(\beta\)) ligands mirror the trends in binding strength values obtained from biological assays \cite{28}. Takane et al. showed the existence of a structure-odor relationship by a ligand-based approach using EigenValue (EVA) descriptor and

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 (agonists 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 corrall ed 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 \((x_i, \alpha_i)\) represent the ith pair of vibrational frequency and amplitude (intensity) among n observed pairs. Transform \(x_i\) by \(y_{i1} = \lfloor \frac{x_i}{c} \rfloor\) 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 \(z_m\) \((z_m = \{i | y_{i1} = z_m, i \in \{1, \ldots, n\}\})\) for \(m = 1, \ldots, M\). If \(b_m\) is the sum of amplitudes with indices (\(\alpha_i\)) corresponding to \(I_m\) for \(m = 1, \ldots, M\), then \((z_1, b_1), (z_2, b_2), \ldots, (z_M, b_M)\) become our new discretized data pairs in the range of \(0 \leq z_m \leq 4000/c\).

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

\[
IG(S, x_i) = H(S_x) - \sum_{v=\text{Values}(x_i)} \frac{|S_{x_i=v}|}{|S_x|} H(S_{x_i=v})
\]

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 \(x\) 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 into two sets; 90% of samples were assigned into a training set and 10% of the samples were
assigned to a validation set. Linear kernel was used to fit the training set, and the performance of the trained classifier was then evaluated based on the validation set. The mean classification rate for each part was used to determine the classification rate of the corresponding frequency. After repeating this process for each feature, we selected those showing the best performances with regard to classification rate, where the size of the best feature set was set equal to that of the features selected using the t-test \((n = 18)\). The analyses were performed using MATLAB.

3. Results and discussion

3.1. Comparison of agonist and non-agonist molecular vibrational patterns

To examine the global patterns in molecular vibrational frequency, the mean vibrational intensities of AdoR agonists and non-agonists according to their molecular vibrational frequency were plotted (Fig. 1). The mean intensity value of the agonist group differed significantly from that of the non-agonist group across the entire vibrational frequency range \((P\text{-value} = 3.42 \times 10^{-6}\text{, two-sample t-test})\). According to the overall intensity plot, the two groups showed differences within a certain frequency range. We selected interesting sets of frequencies to further analyze the different behaviors of the two groups by using IG score, t-test, and SVM.

3.2. IG ranking of informative features

Among the 800 corrals, 382 had vibrational intensities of 0. The remaining 418 corrals were regarded as features containing molecular vibrational information to be further analyzed. Each feature within the feature space was ranked depending on its importance for AdoR agonist and non-agonist classification. After the IG score was calculated for each AdoR ligand, only 28 features out of 418 showed IG scores larger than 0 (Table 1). The majority of features within IG scores greater than the mean value were located between wavenumbers 800 and 1800. This frequency range contained the most discriminative features for testing the equality of mean intensities between the agonist and non-agonist groups. Thus, it was possible to focus our analysis on this frequency subset.

3.3. Comparison of selected features by using statistical analysis

To evaluate the selected features by using IG ranking, we used statistical analyses (two-sample t-test and SVM) to compare the classification potential of these features to other feature lists prepared from the same datasets by using statistical analyses. For additional validation of the t-test analysis, the linear SVM classifier, a non-parametric technique, was applied to the same datasets.

As shown in Table 1 and Fig. 2, six features were commonly selected from the three analyses within the wave number range of 800–1800; IG ranking, t-test \((P < 0.01)\), and SVM \((n = 18)\). The apparent differences in mean intensity and corresponding jumps in firing rates were predominantly found around wave numbers 1000 and 1600.

3.4. Considerations from molecular aspects

The numbers of features from each ligand counts from 57 (theophylline) to 330 (UK-432097), and most of the features holding relatively high intensities grouped near wave number 1600. For example, the features around 1600 were apparent when the mean intensity threshold was set to 9 (Fig. 3). The wave numbers around

**Table 1**

<table>
<thead>
<tr>
<th>Feature No.</th>
<th>Wavenumber</th>
<th>IG score</th>
<th>t-Test ((P &lt; 0.01))</th>
<th>SVM ((n = 18))</th>
</tr>
</thead>
<tbody>
<tr>
<td>238</td>
<td>1190–1195</td>
<td>0.267597015</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>1055–1060</td>
<td>0.2279074627</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>990–995</td>
<td>0.227546287</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>625–630</td>
<td>0.201458209</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>203</td>
<td>1015–1020</td>
<td>0.1974880957</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>319</td>
<td>1595–1600</td>
<td>0.1816791045</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>1000–1005</td>
<td>0.1806432836</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>615–620</td>
<td>0.1765880957</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>208</td>
<td>1040–1045</td>
<td>0.1765880957</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>1200–1205</td>
<td>0.1603910448</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>228</td>
<td>1140–1145</td>
<td>0.1600820896</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>136</td>
<td>680–685</td>
<td>0.1579985075</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>214</td>
<td>1070–1075</td>
<td>0.1566671642</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Mean IG</td>
<td></td>
<td>0.1543408692</td>
<td>●</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 1](https://example.com/fig1.png)  
**Fig. 1.** The mean intensities of the two corralled intensities of molecular vibrational frequency (CIMVs) of adenosine receptor (AdoR) agonists (blue) and non-agonists (red) according to the wave number of the molecular vibration.

![Fig. 2](https://example.com/fig2.png)  
**Fig. 2.** Highly informative features for the classification of AdoR ligands as agonists (blue and green) and non-agonists (red and orange). (A) Wave numbers of features for which the confidence level \((P\text{-value})\) was below 0.01 in the t-test and selected from support vector machines. The blue and green marks represent the informative features of agonists and the red and orange marks represent those of non-agonists. (B) Frequency distribution of the molecular vibration (wave number).
1600 in this study were derived from the scissoring modes of vibration in the primary or secondary amines commonly found in all AdoR agonists. The molecular interaction of this moiety with Asn253 (6.55; this number was adopted from the Ballesteros–Weinstein nomenclature [43]) of AdoRA2A was found to be critical in previous studies in which the crystal structures of human AdoR2A bound to its ligands were determined [44–47]. In addition, mutational analyses showed that mutations in this residue disrupt the binding of both agonist and antagonist ligands. For example, the K values for the Asn253 (6.55) AdoRA2A mutants (N253A, N253S and N253Q [48]; N253D [49]), Asn250 (6.55) AdoRA3 mutant (N250A [50]), and Asn254 (6.55) AdoRA1 mutant (N254A [51]) could not be determined, as they lost their ability to bind to either the agonist or antagonist with high affinity. These results indicated that the interaction between the amine moiety of the ligand and the Asn (6.55) residue of AdoR is indispensable for ligand binding and controls of AdoR activity. The molecular parts of most of the selected features around wave number 1000 were reported to participate in the interaction with the ribose moiety of AdoR agonists [45,46].

The functions of proteins are governed by their dynamic characteristics; membrane receptors also elicit their functions via structural dynamics. Many biological processes are controlled by alterations in rates and relative populations rather than by a simple binary “on-off” switch [52]. Recently, Markov state models of simplified GPCR dynamics were stitched into a single statistical model to describe β2-adrenergic receptor (β2AR) activation [53]. As with the existence of receptor-specific agonists, antagonists, and inverse agonists, the existence of biased agonists implies that the binding affinity and activation capability of a ligand should not be regarded as the same action. Although molecular mechanisms of ligand binding to GPCRs were reported from a variety of viewpoints, the fundamental mechanism of GPCR activation is not yet clear. The results of the current study showed that by using IG, t-test, and SVM methods as cross-check validation, it was possible to select highly informative features of molecular vibration for ligand classification of AdoR agonism.

To our knowledge, this is the first reported informatics and statistical approach to elucidating the relationship between molecular vibrations of ligands and their corresponding AdoRs. Though the calculated vibrational spectrum in this study does not simulate actual IR or Raman spectra, its patterns were consistent for the molecular properties of each ligand; thus, we could not disregard these findings. Selecting subsets of highly informative features would be beneficial and useful for developing and presenting new approaches for ligand design and contribute to ligand and drug discovery. When much more ligand data is gathered, characterization of AdoR ligand subtypes corresponding to their cognate receptors by their molecular vibrational patterns will become solid (data not shown), then targeted modification of binding sites in ligands could be possible. Our findings are also expected to give inspirational contribution to the elucidation of the receptor activation mechanism via material transfer, such as electron or proton, and/or biased agonism. Future work is required to extend molecular vibration to the dynamics of membrane receptor activation.

Acknowledgements

The authors would like to thank Prof. C. H. Choi for his invaluable discussion and comments. The authors also wish to thank Ms. Christina Baek who assisted in the proof-reading of the manuscript. This work was supported by the 2013 Inje University research grant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2015.01.024.

References
