



Bridging Chemical and Biological Realms: an Ion Channel/Ligand Case Study

Ismail Ijjaali, Elodie Dubus, François Petitet, Aureus-Pharma, 174 Quai de Jemmapes, 75010 Paris, ismail.ijjaali@aureus-pharma.com, www.aureus-pharma.com.

INTRODUCTION

The intention of the present analysis is to provide a general overview on the ion channels active chemical space, and to evaluate its diversity. Due to their crucial physiological role and their involvement in many genetic diseases, ion channels have been targeted pharmacologically and many drugs have been developed. They include local and general anesthetics, muscle relaxants, antihypertensives, and oral hypoglycemics^[1,2]. Thus the understanding of the structural, functional and pharmacological aspects of ion channels is of great importance. AurSCOPE Ion Channel knowledge database active chemical space was defined and chemical diversity assessed using Jarvis-Patrick fingerprints-based clustering. Biologically relevant space relative to different ion channel families was then delimited using Principal Components Analysis (PCA).

METHODS

Dataset extraction

All the analyzed chemical structures were extracted from Aureus' AurSCOPE Ion Channels knowledgebase. We used AurQUEST, Aureus's web-based query system to extract a dataset of the most active non-peptide ligands tested in binding or electrophysiology protocols on all ion channel families (corresponding to more than 254 wild targets entries in the database at the time of the query, December 05). Biological activity thresholds expressed by K_i , IC_{50} , and EC_{50} were set for retrieving affinities. This resulted in 11 519 molecules. Since our computational analysis is based on 2D approach, duplicates with different stereochemistry as well as molecules with different counter ions were eliminated (1 548 molecules). In addition, molecules with a molecular weight more than 700 were disregarded (74 molecules) leading to a final set of 9 897 unique molecules.

Computational approaches

To encode the chemical space and perform similarity searches, both 2D chemical fingerprints and molecular descriptors were used. Clustering analyses were performed using ChemAxon Chemical Fingerprints (CF)^[3]. For the clustering process, we employed a version of Jarvis-Patrick algorithm based on CF fingerprints. Clustering parameters (nearest neighbors list and top nearest neighbors in common) were adjusted.

Principal components analysis was used as another way to visualize the chemical diversity. PCA latent variables were computed from a set of 2D-molecular descriptors using the MOE cheminformatic suite^[4]. Then, these PCAs served to project different clusters into the chemical space taking into account the corresponding biological activity data.

RESULTS & DISCUSSION

Due to the structural diversity of the dataset, a first step was to optimize different parameters relating to the generation of CF fingerprints. These included fingerprint length, maximum pattern length, and number of bits to be set for patterns^[3]. A larger fingerprints (2 048 bits) and longer paths were more appropriate to enhance the descriptive power of CFs. for the subsequent clustering operation.

As output of this operation, number of singletons and clusters were determined at different similarity thresholds. At an 0.85 similarity threshold, clustering identified 1 663 singletons and 938 clusters with the most populated cluster having 334 molecules. The larger cluster is mainly formed of triazolo-ptalazine derivatives. The second largest cluster contains 286 molecules which include pyrolidin series, and cluster 3 contains pyrido-benzimidazole derivatives. Further major classes of structures are found among the top 15 clusters. Clusters centroids are presented below (Figure 1a). Figure 1b shows examples of singletons present in AurSCOPE Ion Channel.

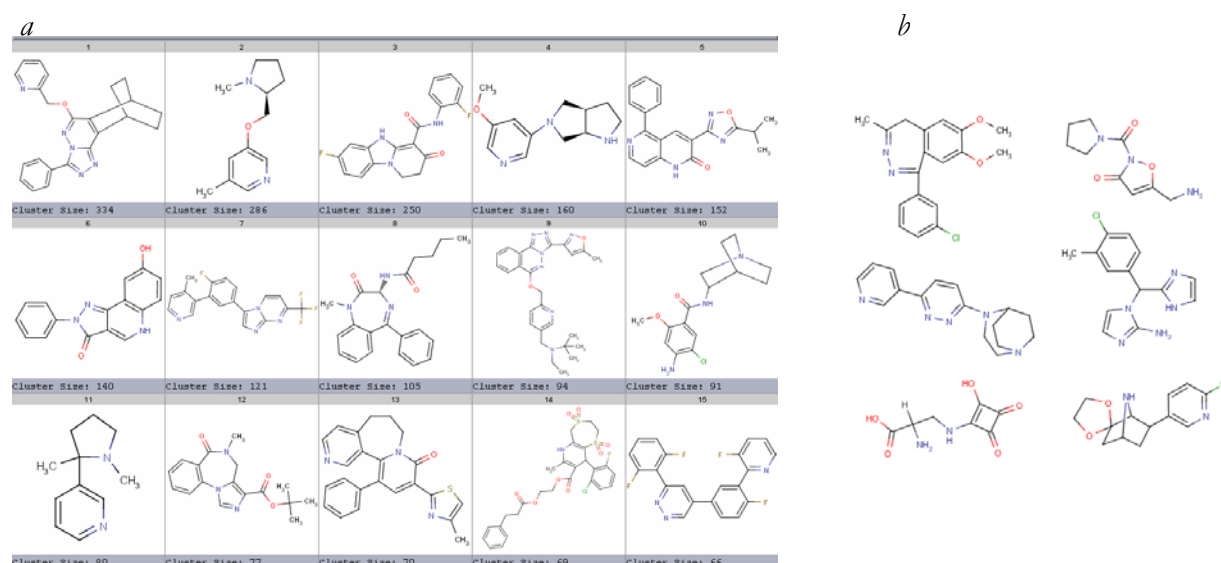


Figure 1. Centroids of the top 15 clusters and singletons among active dataset.

Targets associated with molecules corresponding to the top 12 clusters (representing 1890 compounds and 19% of the whole "active" ion channel set) were identified and each individual cluster was projected on the global ion channel set associated chemical space (Figure 2). Not surprisingly, the majority of the clusters among the most populated are linked to GABA_A receptors, the most represented target in the dataset.

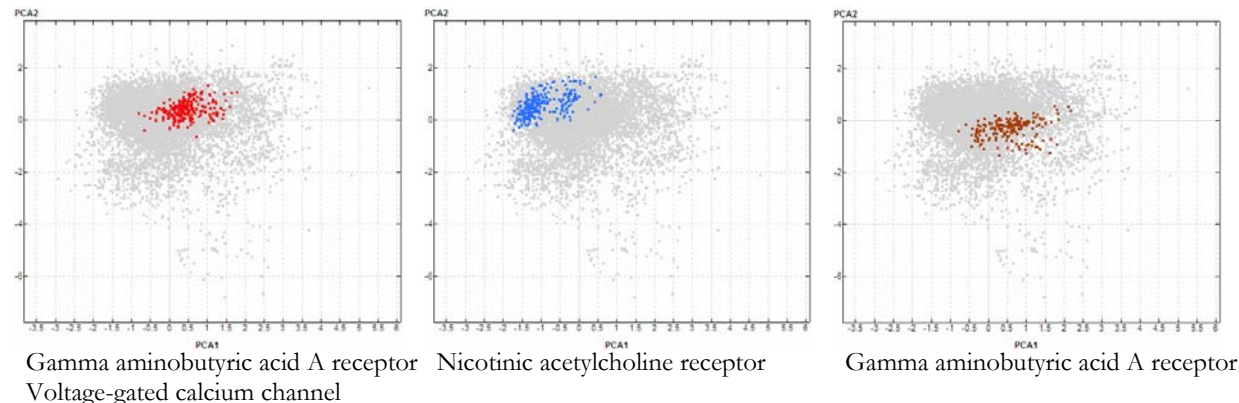


Figure 2. Projection of the three first populated clusters.

From a biological point of view, γ -amino butyric acid (GABA), nicotinic acetylcholine (nAChR), N-methyl D-aspartate (NMDA), serotonin 5-HT₃ receptors, calcium and potassium channels are the most represented targets, with more than 8835 active molecules corresponding to 89 % of the whole set. For a given target family molecules exhibiting an affinity or an activity determined in binding or electrophysiology protocols (< 300 nM) were projected into a 3D representation of the chemical space according to the three main components of the PCA analysis (Figure 3).

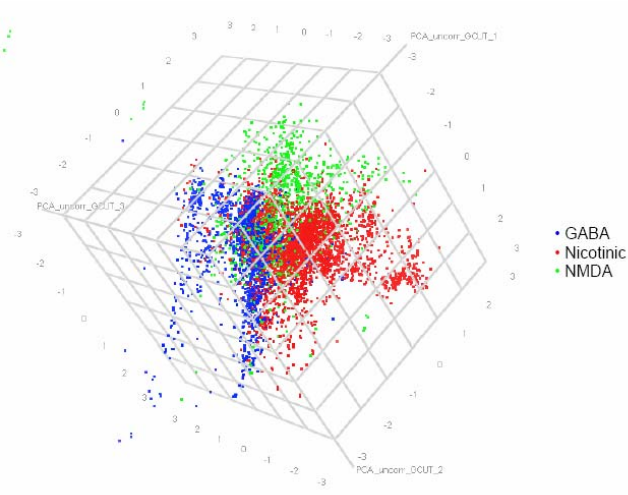


Figure 3. 3D representation along the first three principal components of datasets active on GABA, nicotinic, and NMDA channels.

A large domain is delimited by both GABA and nAChR ligands and clear overlap between different families was shown such as 5-HT₃ and nAChR (not shown for clarity) or NMDA and GABA. This analysis shows that current medicinal chemistry knowledge is mainly represented by these targets as molecules targeting other families are scarce. Each of these families is associated with a specific 3D space as shown for potassium and P2X channels (Figure 4). These spaces are more or less dispersed and resulted in a visual indication of the chemical diversity associated to corresponding target.

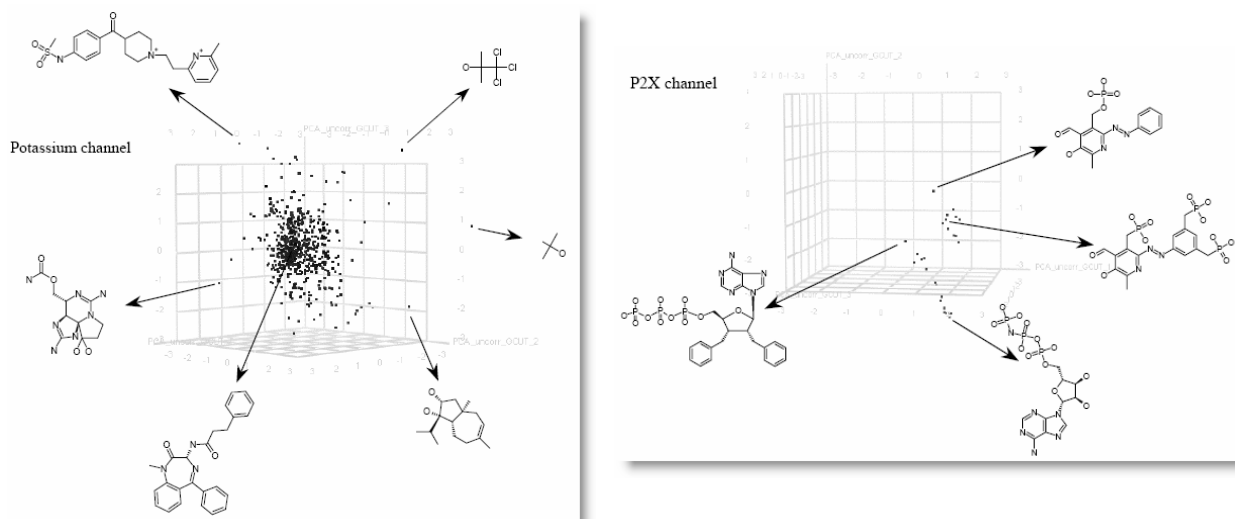


Figure 4. PCA-projection of active molecules on potassium and P2X channels.

Finally, the nAChR ligands were considered more closely and the corresponding dataset (1 654 molecules) was divided and projected according to a supplementary biological filter: subunits composition (Figure 5).

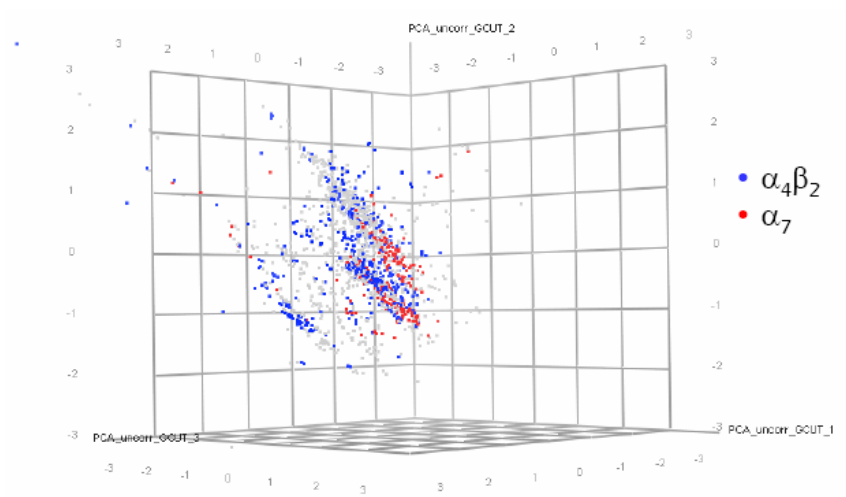


Figure 5. Projection of molecules active on nicotinic acetylcholine receptor with a subdivision corresponding to $\alpha_4\beta_2$ (blue) and α_7 (red) subunits.

The two spaces delimited by $\alpha_4\beta_2$ and α_7 ligands (533 and 208 compounds, respectively) are distinct but overlapping. Consequently, some areas may be associated with selective $\alpha_4\beta_2$ or α_7 compounds and others with non selective compounds able to recognize both receptor sub-types.

In conclusion, a precise analysis of the chemical space associated with a given target and assessing more closely the chemical structures located in regions associated to specific activity/selectivity criteria can aid drug discovery projects during the lead generation and optimization phases.

REFERENCES

- [1] Pardo L.A. et al. *J. Membr. Biol.* **2005**, *205*: 115-124.
- [2] Li S. et al. *Drug Design Technology* **2005**, *10*: 129-137.
- [3] JChem, version 3.5.2, ChemAxon, Budapest, Hungary, www.chemaxon.com
- [4] MOE, version 2005.06, Chemical Computing Group Inc., Montreal, Canada, www.chemcomp.com