Guilherme Lopes Gabriel

In Silico Perspective on Psychedelic Studies:
Characterization of the interaction between small molecules and 5-HT2AR

UNIVERSIDADE D
COIMBRA
1 2 9 0

1 2 9 0

UNIVERSIDADE D
COIMBRA

Guilherme Lopes Gabriel

In Silico Perspective on Psychedelic Studies:
Characterization of the interaction between small molecules and 5-HT2AR

VOLUME 1

Dissertação no âmbito do Mestrado em Biologia Celular e Molecular com especialização em Neurobiologia orientada pela Professora Doutora Irina Moreira e pelo Professor Doutor Miguel Castelo-Branco e apresentada ao Departamento de Ciências de Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Junho de 2022

Faculdade de Ciências e Tecnologia da Universidade de Coimbra

In Silico perspective on Psychedelic Studies

Characterization of the interaction between small molecules and5-HT2AR

Guilherme Lopes Gabriel

Dissertação no âmbito do Mestrado em Biologia Celular e Molecular com Especialização em Neurobiologia orientada pela Professora Doutora Irina Moreira e pelo Professor Doutor MiguelCastelo-Branco e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências eTecnologia da Universidade de Coimbra

Junho de 2022

UNIVERSIDADE D COIMBRA

Agradecimentos

Em primeiro lugar quero agradecer à minha Mãe, por tudo. Pela vida que me proporcionou, pelo exemplo, pelo esforço, mas sobretudo pelo tempo e espaço que faz de mim a pessoa que sou hoje. Tem sido um prazer descobrir a vida contigo e que muitos mais desafios venham na nossa direção. Esta é para ti, para sempre, o meu grande Amor.

Quero agradecer ao meu tio, à minha irmã e à minha avó, incansáveis no seu apoio e um porto seguro onde nunca faltou luz para me guiar. Obrigado pelas partilhas e ensinamentos, mas sobretudo por me ensinarem o significado de Casa.

Talvez o agradecimento mais complicado de fazer, pela sua extensão, vai para a Rápo-Táxo. Teria de desenvolver um segundo Volume deste documento para vos agradecer proporcionalmente ao que significam para mim. Pelas histórias, lições, paciência, mas sobretudo por me abrirem caminho para uma louca série de acontecimentos. A TODOS vocês, presentes, antigos e amigos com quem me orgulho de partilhar a vida, o maior obrigado por estenderem o meu significado de Família.

Preciso ainda de agradecer aos "amigos dos montes" com quem tive o gosto de começar a explorar o mundo e que me motivaram a novos desafios nesta vida. A vocês, a quem a distância nunca afetou, um enorme obrigado pelas voltas ao estacionamento a que chamamos "Terrinha".

Um grande obrigado aos amigos com quem descobri a bioquímica, partilhei bancada e que me continuam a motivar o interesse pelo "Mundo do Lehninger".

Quero ainda agradecer à Professora Irina Moreira e ao Professor Miguel Castelo-Branco por terem acreditado em mim e dado espaço ao desenvolvimento da minha ideia.

Sem esquecer, obrigado aos professores que serviram de exemplo para mim e me motivaram a chegar a esta entrega, pela sua irreverência e gosto do saber.

Obrigado, gente! Pela companhia que faz desta experiência uma Vida cheia.

Resumo

Psicadélicos, anteriormente vistos como um problema central na "Guerra às Drogas", são atualmente muito referenciados entre as comunidades científica e médica. A mudança de perspetiva inerente à classificação, ao uso e à regulamentação destes compostos requer dados recentes e mais quantitativos, por forma a tratar estas drogas com uma fundação baseada no conhecimento científico. Com isto, ao longo dos últimos anos, tem-se vindo a implementar novas técnicas, desde estudos farmacodinâmicos a estudos de imagiologia exploratórios, na investigação dos mecanismos e efeitos destes compostos. Avançando o nível de sensibilidade, métodos computacionais revolucionaram tanto o entendimento de fármacos existentes, como o desenvolvimento de novos sintéticos. Com isto, ao aplicar metodologia *in silico* ao estudo de enteogéneos, uma nova e importante perspetiva é adicionada a esta discussão.

Nesta tese é salientado o possível uso de psicadélicos como novos vetores terapêuticos no tratamento de doenças neurológicas, ou na mitigação dos seus sintomas. Primeiro, ao rever a química destes compostos é notória a sua semelhança com a molécula de serotonina (5-HT). Isto leva à revisão dos recetores serotonérgicos (5-HTR), considerados como o principal tradutor na resposta psicadélica. Para além disso, desequilíbrios no sistema humano de serotonina serão relacionados com problemas neuropsiquiátricos, por forma a obter um sentido prático no possível efeito destas drogas, ou seus análogos, na terapia supramencionada. Paralelamente, alguns métodos computacionais, como Modelação por Homologia ou "Docking" Molecular, serão revistas, mostrando a sua importância no auxílio do desenvolvimento de novos fármacos.

Por último são apresentadas as fundações para a criação de um modelo baseado em "Machine-Learning", com o objetivo de prever o estado de ativação do recetor 5-HT2A, sustentado pela previsão de valores de RMSD. Esta previsão é baseada em características físicas que podem auxiliar na caracterização físico-química dos bolsos hidrofóbicos em estudo. Este novo modelo poderá ajudar no desenvolvimento de novos compostos baseados na estrutura psicadélica, com menores custos e maior segurança, bem como um melhor entendimento do agonismo apresentado por estas substâncias na ligação com o recetor 5-HT2A.

Keywords: Psicadélicos, recetores de serotonina (5-HTR), Doenças Serotonérgicas, Bioquímica Computacional

Abstract

Psychedelics, once seen as a drug enforcement issue, are now receiving a lot of attention by the scientific and medical community. The mind shift regarding the classification, use and regulamentation of these compounds requires more quantitative and recent data, to deal with these drugs with a knowledge-based foundation. With this, in the last years, novel research has used a variety of techniques, ranging from in vitro pharmacodynamic assays to exploratory imaging studies in clinical trials, to study the mechanism and action of these intricate compounds. Further advancing the level of sensitivity, computational methods have revolutionized both the understanding of existing pharmaceuticals as the design and development of new ones. With this, applying *in silico* methodology to the study of entheogens would add a different and important perspective to this discussion.

In this thesis it is highlighted the hypothesized use of psychedelics as therapeutic vectors for the treatment of some neurological disorders, or the mitigation of their symptoms. First, by reviewing the basic chemistry of these compounds it is possible to notice their resemblance with serotonin. This leads to the review of serotonin family receptors (5-HTR), thought to be the main "gates" in the psychedelic response. Next, malfunctions in the serotonergic human system will be associated with neurological disorders, to get a practical sense on how these drugs, or their analogues, might build the next line of therapeutics. Parallelly, some computational methods, such as Homology Modelling and Molecular Docking, will be reviewed, showing how they might aid in the drug development pipeline.

Lastly it presents the foundations of a novel Machine Learning-based model, used to predict the activeness state of 5-HT2A receptor, based on the predicted RMSD value. The prediction is based on physical features that can help on physical characterization of the studied binding pockets. This new model would assist in the development of novel compounds based on the psychedelic structure, with diminished costs and highest safety, and a better understanding on the agonism of these substances with 5-HT2A receptor.

Keywords: Psychedelics, 5-HT receptors (5-HTR), Serotonergic Disorders, Computational Biochemistry

List of Abbreviations

5-CT: 5-CarboxamidoTryptamine
5-HT: 5-HydroxyTryptamine (Serotonin)
5-HTXR (X = 1,2,3,4,5,6,7): 5-hydroxytryptamine receptors (sub-class)
5-MeO-AMT: 5-methoxy-alpha-methyltryptamine
5-MeO-DALT: N,N-diallyl-5-MethOxytryptamine
5-MeO-DiPT: 5-methoxy-diisopropyltryptamine
AA: Arachidonic Acid
AADC: Aromatic Amino Acid Decarboxylase
Acc: Accuracy
Ach: Acetylcholine
AD: Alzheimer Disease
ADHD: Attention-Deficit/Hyperactivity Disorder
Akt: Protein kinase B (PKB)
ANN: Artificial Neural Networks
APP: amyloid precursor protein
ASFP: Artificial Intelligence based Scoring Function Platform
BDNF: Brain-Derived Neurotrophic Factor
BP: Binding Pocket
Ca2+: Calcium bication
CADD: Computer-Aided Drug Design
cAMP: cyclic Adenosine MonoPhosphate
CCp: Pearson's Correlation Coefficient
CCs: Spearman's Correlation Coefficient
Cdc42: Cell division control protein 42 homolog
CHARMM-22: Chemistry at Harvard Macromolecular Mechanics 22
CNS: Central Nervous System
DAG: Diacylglycerol
DG: Dentate Gyrus
ΔGsolv: Solvation/Desolvation effect
DMT: N,N-dimethyltryptamine
DNN: Deep Neural Networks
DOB: Dimethoxybromoamphetamine
DOI: 2,5-Dimethoxy-4-iodoamphetamine
DOM: 2,5-Dimethoxy-4-methylamphetamine

DOPE: Discrete Optimized Protein Energy
DRN: Dorsal Raphe Nucleus
EC50: Half Maximal Effective Concentration
ECL2: Extracellular Loop 2
Eelec: Electrostatic interaction
Epac: Exchange protein activated by cAMP
ERK: Extracellular signal-regulated Kinases
Evdw: van der Waals interaction
FGF: Fibroblast Growth Factor
FGFR2: Fibroblast Growth Factor Receptor 2
FP: False Positives
GABA: Gamma-Aminobutyric Acid
gCa2+: conductance of the voltage gated Ca2+ channel
GDNF: Glial cell Derived Neurotrophic Factor
Gi: inhibitory G protein
GI: gastrointestinal tract
GIRK: G-protein-gated inwardly rectifying gK+
gK+: conductance of the voltage-gated K+ channel
GMP: Guanosine MonoPhosphate
Go: inhibitory G protein (GNAI3)
GPCR: G Protein-Coupled Receptors
Gq/G11: G protein from Gq-family
Gs: stimulatory G protein
Hb: Lateral Habenula
HTS: High-Throughput Screening process
IC50: Half-maximal Inhibitory Concentration
INMT: Indole-N-Methyltransferase
k: Coehn's Kappa
Kd: Dissociation Constant
Ki: Inhibitor Constant
knn: k-Nearest Neighbors
LSD: Lysergic Acid Diethylamide
LSD-25: Lysergic Acid Diethylamide-25
MAE: Mean Absolute Error
MAO: Monoamine Oxidase
MAOIs: Monoamine Oxidase Inhibitors
MAPKs: Mitogen Activated Protein Kinases
MDD: Major Depressive Disorder
MDMA: 3,4-MethyleneDioxyMethAmphetamine

mGlu2/3: metabotropic Glutamate receptor 2/3
ML: Machine Learning
ML-SF: Machine Learning-based Scoring Function
MM: Molecular Mechanics
MMPs: Matrix Metalloproteinases
MOE: Molecular Operator Environment
mRNA: messenger Ribonucleic Acid
MSE: Mean Square Error
NIH3T3: fibroblast cell line isolated from mouse NIH/Swiss embryo,
NMDA: N-Methyl-D-Aspartate
NMR: Nuclear Magnetic Resonance
NOS: Nitric Oxide Synthase
NT: NeuroTransmitter
OCD: Obsessive-Compulsive Disorder
PAG: PeriAqueductal Gray
PDB: Protein Data Bank
PFC: PreFrontal Cortex
PI3K: PhosphoInositide 3-Kinases
PLA2: Phospholipase A2
PLC: Phospholipase C
QM/MM: Quantum Mechanical/Molecular Mechanics
R^2: Coefficient of Determination
Rab4: small G protein from the Ras superfamily
Rac1: Ras-related C3 botulinum toxin substrate 1
RBF: Radial Basis Function
RF: Random Forest
RMSD: Root-Mean-Square Deviation
RNA: Ribonucleic Acid
RRF: Regularized Random Forest
RSK2: Serine/threonine-protein kinase
RVM: Rostral Ventromedial Medulla
S188: Serine residue index 188
S421: Serine residue index 421
SAD: Seasonal Affective Disorder
SAM: S-Adenosyl-Methionine
SAR: Structure-Activity Relationship
SBC: Subtractive Clustering and Fuzzy c-Means Rules
SCN: hypothalamic SupraChiasmatic Nuclei
Sen: Sensitivity

SER: Serine

SERT: Serotonin Transporters

SF: Scoring Functions

SNRIs: Serotonin-Norepinephrine Reuptake Inhibitors

Spf: Specificity

SSRIs: Selective Serotonin Reuptake Inhibitors

SVR: Support Vector Regression

TAAR: Trace-Amine-Associated Receptors

TBG: Tabernanthalog

TCAs: Tricyclic Antidepressants

TMA-2: TriMethoxyAmphetamine-2

TP: True Positive

Tph1: Tryptophan Hydroxylase 1

Tph2: Tryptophan Hydroxylase 2

TYR: Tyrosine

VMAT2: Vesicular Monoamine Transporter 2

VTA: Ventral Tegmental Area

Index

Figure Index

1. Introduction

1.a. A long, lasting trip

Psychoactive drugs have gained a crescent interest inside the scientific community (Castelhano et al, 2021). Their recognition is due to their potential therapeutic effect. This is particularly true for tryptamine psychedelics, which are structurally very similar to serotonin, a major neuromodulator, associated with mood, memory, cognition, perception, and consciousness (Bryan L, Roth 2011). During the last three decades there was an opinion shift about psychedelics (Nichols 2016). These were first referenced as psychotomimetic, a term coined with mental states like psychosis (Hoffer and Osmond 1967). However, this term was later abandoned when it was realized that the states that these compounds produced were not that similar to psychotic episodes, then becoming known as hallucinogens, due to their capacity in producing hallucinations (Nichols 2016). Though this term is still very little descriptive, it continues to be preferred by the scientific community. This asks for a more precise definition of these compounds, due to the variety of compounds that this classification includes and the diversity in the effects that they produce. Accordingly, it has been tried to better classify them according to cell signaling pathways that they activate, and by the selective action of specific receptors. One example is the effect of classic serotonergic hallucinogens (psychedelics), whose main effect was firstly reported as agonists (or partial agonists) of 5-hydroxytryptamine (5-HT) 2A receptors (Nichols 2016), although there is now evidence that this is not the only type of receptor being activated by these molecules.

For a long time, the notion that psychedelics may have a therapeutic effect had little acceptance by the scientific circle, a fact that arose due to the lack of legal support or the appreciation of these drugs as being illegal with no useful properties. The dispute over psychedelics has its days started since the 60s with the American counterculture against the Vietnam War, which helped to criminalize these substances due to political and social pressures. However, it is often forgotten that "between the 1950 and the mid-1960s there were more than a thousand clinical papers discussing 40,000 patients, several dozen books, and six international conferences on psychedelic drug therapy. It aroused the interest of many psychiatrists who were in no sense cultural

rebels or especially radical in their attitudes" (Vattano 1981). This highlights the nonsensical way in which these drugs were seen in society and how it slowed down the progress of innovative research that may have helped to advance both psychiatry and neuroscience.

The research of psychedelics opened some doors that lead to major discoveries, helping to tackle the brain's functional pharmacology. A main example was the discovery of Lysergic acid diethylamide (LSD) by Albert Hofmann in 1943. Due to its structural resemblance with serotonin, in 1953 when this last was discovered in the mammalian brain (Twarog and Page 1953), LSD helped to better understand how serotonin works on the brain, as inferred by their tryptamine moiety (Figure 1).

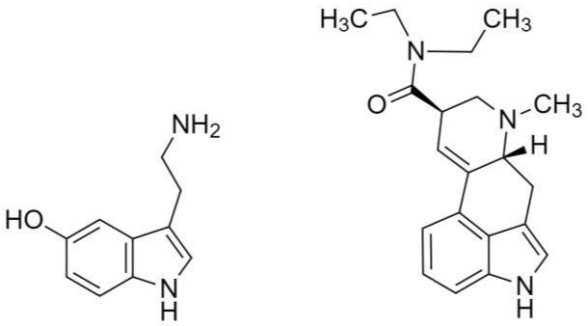


Figure 1. Serotonin (left) and LSD (right).

This discovery led to a serotonergic hypothesis of mental disturbances, when Woolley and Shaw in 1954 stated that the effects of LSD occurred due to its interference on the normal function of the serotonergic system (Nichols 2016). In other words, the wholefield of serotonin neuroscience was accelerated by the discovery of LSD. Some even claimed that these classes of compounds contributed to neuroscientific advances in the same way that the microscope contributed to cell biology (Grof 2008).

Psychedelics are a class of compounds that need to be discussed from a multidisciplinary perspective, from anthropology to ethnopharmacology, psychiatry and psychology. They stand as one of the most intricate and older substances used by mankind. There is registry of numerous religious and spiritual practices that evoke the use of substances capable to induce mind altering experiences, such as the *Soma* in Rigveda (Wasson and Gordon Wasson 1971) and the use of Psilocybin mushrooms by the Aztec shamans, known as *teonanacatl* or "god's flesh" (A, H, Smith, Ott, and Bigwood 1979), (Schultes, Hofmann, and Rätsch 2001). Moreover, in the Native American practices, the Peyote *(Lophophora*

williamsii) cactus has been a sacrament in the Native American Church for almost 5700 years (Schultes, Hofmann, and Rätsch 2001; Bruhn et al, 2002), With this, in the classification of classic psychedelics we include LSD, mescaline, psilocybin and N,N-dimethyltryptamine (DMT) (Figure 2).

One main characteristic of these compounds, highlighted by Daniel X, Freedman, is their ability to reveal a state of 'portentousness'- meaning the capacity of the mind to see more and experience more that it can tell (D, E, Smith and Rose 1968; Freedman 1968). These compounds had such a tremendous and notorious effect on religious practices, that Ruck et al, in 1979 proposed the term *entheogen*, instead of hallucinogen or psychedelic, that had more negative connotations (*entheos: "God (theos) within" + genesthe: "to generate") (Ruck et al, 1979)*. This novel term refers essentially to a substance capable of generating God or the divine within someone. It is because of these profound psyche alterations that it becomes mandatory to understand the biochemistry inherent to these changes and how they can benefit or not brain function. Moreover, they can help to answer questions that remain unanswered for ages, such as the neural basis for consciousness, and its neurobiological correlates.

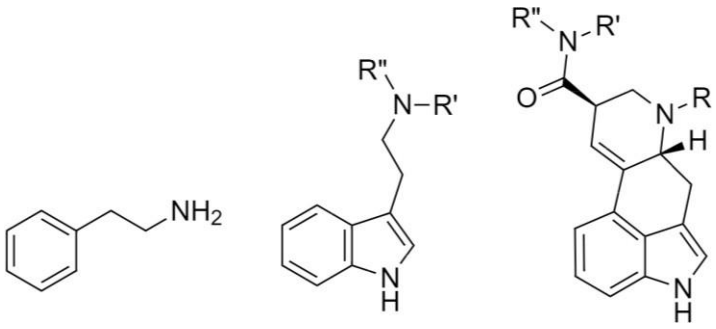


Figure 2. Psychedelics (general structures): Phenethylamines (left), tryptamines (center) andergolines (right).

1.b. The trip building blocks

We can classify the classic serotonergic hallucinogens according to the relation between their structure and activity (Structure-Activity Relationship, SAR) as tryptamines, ergolines and phenethylamines (Nichols 2018). Serotonin 5-hydroxytryptamine 2A receptors (5-HT2AR) were seen as the main target for the effect of psychedelic drugs, with a key role in the regulation of cortical function and cognition. There are other types of molecules, better called psychotomimetic, that also activate the serotonin 5-HT2A receptors, among others, such as 3,4-methylenedioxymethamphetamine (MDMA), ketamine analogues (that are NMDA receptor antagonists) or cannabinoids, but they will not be addressed in this thesis.

1.b.1. Tryptamines

Tryptamines present themselves as the most similar chemotype to serotonin (5-HT, 5-Hydroxytryptamine). They are very similar to ergolines, the last ones considered as rigidified tryptamines. Although they are precursors for a variety of molecules, tryptamines can't be modified to retain their activity. Shulgin dedicated a major part of his career to study these compounds, but only a small number of N,N-substituent variations were analyzed in humans (Shulgin and Shulgin 1997). Some of these occur naturally like dimethyltryptamine (DMT), psilocybin and psilocin (present in magic mushrooms). Serotonin and melatonin are tryptamines that occur naturally on animals, including Humans. All the compounds mentioned derive from tryptamine, that is a monoamine alkaloid, related to the amino acid tryptophan. Generally, tryptamines consist in an indole ring structure, comprising a fused double ring of a pyrrole and a benzene ring, adding to a two-carbon side chain. The addition to both the chemical moieties leads to the formation of a multitude of both natural and synthetic compounds (Figure 3.).

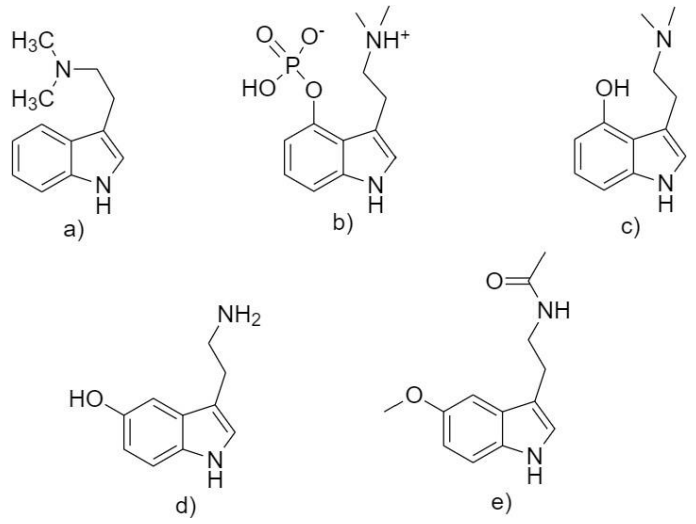


Figure 3. Structure of Tryptamines, a) DMT, b) Psilocybin, c) Psilocin, d) serotonin and e)melatonin.

Focusing on DMT, it is known that it is first metabolized by the monoamine oxidase (MAO), making it not active when taken orally (Dargan and Wood 2021; Sitaram et al, 1987). As such, traditional cultural rituals, e.g. Ayahuasca usage in Amazonian indigenous tribes, needs to contain both the DMT as the psychoactive component and MAO inhibitors, such as harmaline, also an indole alkaloid. It's the presence of these inhibitors that turns DMT orally active, inactivating the first step on its metabolism. It can also be administered via insufflation, inhalation and intramuscular (IM) or intravenous (IV) injection. Mechanistically, DMT is degraded by MAO via oxidative deamination of the side chain. Novel investigation states that there are other non-MAO metabolic pathways, including N-oxidation and N-methylation. The last step consists of the conversion to 3-indoleacetic acid (Riba et al. 2012) (Figure 4.). Another example of a tryptamine is psilocybin. This compound, present in some species of mushrooms, is metabolized to psilocin, via dephosphorylation, a molecule with hallucinogenic properties in animals. This follows a different metabolic pathway, when compared to DMT, being degraded via hepatic glucuronidation (Kamata, Katagi, and Tsuchihashi 2010). This one differs from other 5-substituted simple tryptamines, such as 5-methoxy-alpha-methyltryptamine (5-MeO-AMT), 5-methoxy-diisopropyltryptamine (5-MeO-DiPT) and N,N-diallyl-5-methoxytryptamine (5-MeO-DALT) that are processed through 6-hydroxylation, O-methylation or N-dealkylation by hepatic cytochrome P450 enzymes. The final metabolites are conjugated with glucuronide or sulphide, preparing them for

waste (Shulgin and Shulgin 1997). Concerning their mechanism of action, the predominant clinical effect produced by these compounds are hallucinations, mediated mainly by agonism at 5-HT1A and 5-HT2A receptors. However, they also interact with other receptors in the central nervous system, including vesicular monoamine transporter 2 (VMAT2), sigma-1 receptor, trace-amine-associated receptors (TAAR) and serotonin transporters (SERT) (Ray 2010).

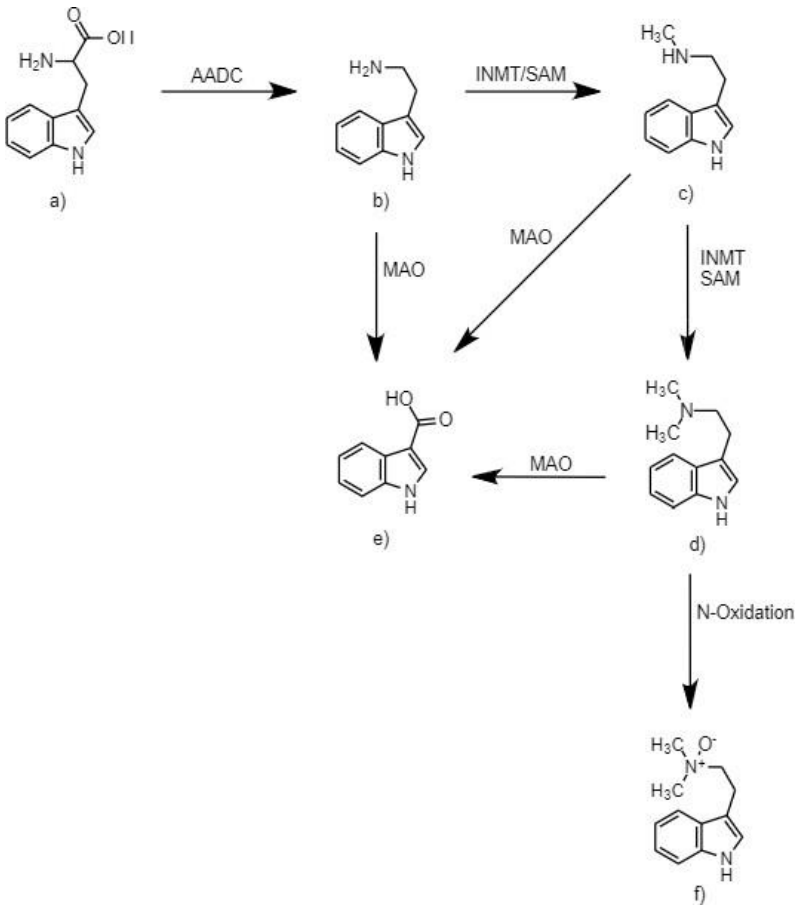


Figure 4. Biosynthesis, and metabolism of DMT, Tryptophan (a) is first converted to tryptamine (b) by the aromatic amino acid decarboxylase (AADC), Tryptamine is then dimethylated to yield N-methyltryptamine (c) and then DMT (d) by the indole-N-methyltranferase.

1.b.2. Ergolines

They are tetracyclic molecules derived from ergot alkaloids (ergot fungus, genus *Claviceps*) (Hylin and Watson 1965). Like tryptamines, they are 5-HT2A agonists and the most important one is lysergic acid diethylamide (LSD) or LSD-25, Though LSD is a very potent hallucinogen, its affinity with the 5-HT2A receptor is not high, when compared to simpler molecules such as 2,5-dimethoxy-4-iodoamphetamine (DOI) (Nichols 2018). They are similar to tryptamines in their indole system; however, they have a tetracyclic ring, giving them a more rigid structure (Dargan and Wood 2021). Another example of ergoline is the psychoactive alkaloid Ergoline found in seeds of the morning glory family (*Convolvulaceae*). It has a similar structure to LSD (lysergamide). Because of their similarity to tryptamine's molecular structure, they show similar pharmacodynamics.

1.b.3. Phenethylamines

Phenethylamines are the most explored class of psychedelics, due to their facile synthesis (Nichols 2018). This is a very broad class of compounds, including mescaline and MDMA, as psychedelics, but also other compounds such as derivatives from amphetamines, e,g, benzo difuran derivatives (Figure 5). Some substituted amphetamines were developed by Alexander Shulgin, being the most potent ones with its substituent in positions 2,4,5 of the ring, like TMA-2 (Trimethoxyamphetamine-2) (Shulgin and Shulgin 1991). This classification is used to describe any structure derived from an aromatic group adjoined to a terminal amine by an ethyl group.

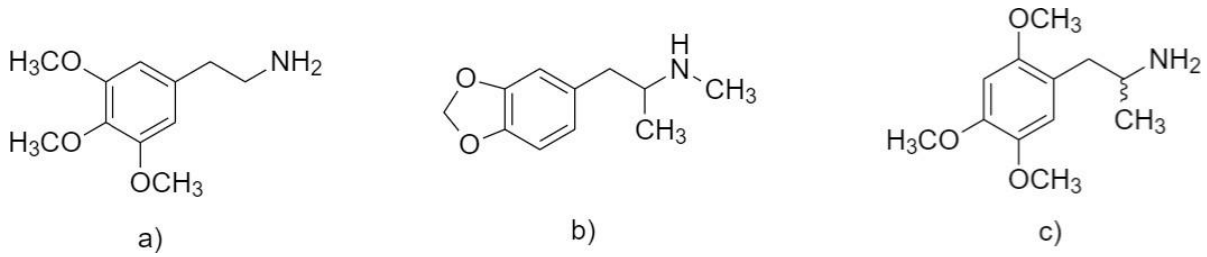


Figure 5. Structure of some phenethylamines: a) mescaline, b) MDMA and c) TMA-2.

1.c. The Entrance Door: 5-HT receptors

Neurotransmitters (NT), such as serotonin, evolved to better regulate the ion channels necessary to maintain a stable membrane potential (*Handbook of the Behavioral Neurobiology of Serotonin* 2020; Muller and Jacobs 2009) and to regulate neural communication. In an evolutionary perspective, basic serotonin receptor subtypes evolved very early in geological periods and records point to the early development of serotonin receptors in the development of an organism (early ontogeny). This gave time for the development of a diverse set of genes, leading to a variable class of receptors (*Handbook of the Behavioral Neurobiology of Serotonin* 2020; Muller and Jacobs 2009). In 2006, Moroz et al, reported at least 20 separate neuronal transcripts of 5-HT receptors (Moroz et al, 2006). The specificity of serotonin is very diverse, being able to bind to at least 16 specific receptors, increasing the variety of signaling cascades that it can activate. This makes serotonin a powerful regulator of ion channels, c-AMP levels, and kinase activity in neurons. The presented data obliges the thinking of why the necessity of such a diverse set of translators is needed for serotonin. This might be due to the lack of tryptophan available in animals, resulting in a low net concentration of serotonin (*Handbook of the Behavioral Neurobiology of Serotonin* 2020). The quantity of expressed receptors in each cell will define and assure that the chemical signal starts a cell response. If the ligand necessary to produce this effect is present in a low concentration, this leads to an increase in the number of receptors needed to restore the cell communication. In addition, a fine-tuned transport system to carry both tryptophan and serotonin also evolved in animals, enabling the necessary transport of these molecules through the blood to their specific targets, such as the brain (Bachmann 2002). In sum, loss of tryptophan has promoted a highly branched, diffuse neural network and a huge variety of specific receptors to maximize serotonin's actions.

There are seven classes of 5-HT receptors, 5-HT1 to 5-HT7, coded by 17 different genes. These are part of a major family of membrane receptors, activated by a range of chemical ligands, which integrate critical cell responses (Azam et al. 2020), called G protein-coupled receptors (GPCR) with the exception of 5-HT3R, that is an ionic channel. GPCRs are one of the oldest molecular devices concerned with signal transduction, presented even before plants, fungi and animals evolved. GPCRs were divided into families based on their structure due to their considerable diversity (i.e. amino acid sequence: rhodopsin, adhesion, secretin, glutamate and frizzled) (Shahbazi et al, 2020). An extracellular N-terminus, seven transmembrane helices with intracellular and extracellular loops, and an intracellular C-terminus are the three parts of GPCRs.

Moreover, GPCRs play important roles in regulating mood, appetite, pain, vision, immune responses, cognition, and synaptic transmission. The GPCR superfamily does not share any sequence homology. Nonetheless, these receptors share enough short sequences and even individual amino acid residues to allow for a common three-dimensional structure and signal transduction mechanism. Some of these structural motifs (Lagerström and Schiöth 2008), or sequence similarities, are dedicated to sustaining receptor inflection and others mediate the agonist-dependent activity.

The genetic code concerned with the expression of 5-HT receptors has suffered some edits to generate a more diverse group of molecules, necessary for a fine control of the molecular pathways and consequently their cellular functions. There were two periods on the classification of 5-HT receptors. The first one being the pharmacological period, where the synthesis of different ligands helped in the classification of the different receptors. The second one was a molecular biology period, where the cloning of the different receptors helped integrate the classification chart of this family of receptors (Kroeze and Roth, 2002). In a physiological and evolutionary perspective, these receptors have different cell expression and control differently, via different signaling events, the cell physiological response. Each receptor subclass has a different signaling repertoire.

5-HT contracting effect, reported by Gaddum and colleagues, blocked by dibenzyline (became the D receptor, finally characterized as a postsynaptic 5-HT2 receptor) and other part by morphine (giving the name to the M receptor, finally characterized as a presynaptic 5-HT3 receptor) (Hannon and Hoyer, 2008.; Gaddum and Picarelli 1997). Laterwith the studies from Bradley and co-workers (Bradley et al, 1986), they were split into three 5-HT receptor classes: 5-HT1, 5-HT2 and 5-HT3 (being this last one the correspondent tothe M receptor). Other tools started being used such as the measurement of the productionof second messengers, following 5-HT stimulation, aiding on the study of structure- response relationship. This enabled the discovery of the 5-HT4 receptor, able to stimulate cAMP (cyclic adenosine monophosphate) production in colliculus neurons (Bockaert et al, 2006).

The receptor 5-HT1A presents substantial homology similarity with the rhodopsin receptor, being included in its family (Nowak et al, 2006). From a genetic perspective the 5-HT1A receptor is estimated to have evolved 750 million to 1 billion years ago, before muscarinic, dopaminergic, and adrenergic receptor systems (Peroutka and Howell 1994),

indicating that the receptor existed before the evolution of the most primitive animal form, sponges, that evolved some 600 million years ago.

Binding studies with a variety of radioligands and second messenger studies were used to define the subtypes of 5-HT1 receptors, 5-HT1A, in neurons, is negatively coupled to adenylyl-cyclase (De Vivo and Maayani 1986), 5-HT1B, a presynaptic receptor, is negatively coupled to adenylyl cyclase in substantia nigra (Bouhelal, Smounya, and Bockaert 1988), 5-HT1C has a very transient classification and is found at high density in the choroid plexus. They are coupled to inositol phosphate production and Ca2+ signaling (like 5-HT2 receptors), making it join the 5-HT2 family, 5-HT1D has a more complicated history, being separated concerning its genes (there's an alpha and a beta) (Hannon and Hoyer, 2008). The 5-HT1E was found in the human frontal cortex, but missing in rodents' brains, and no specific drugs targeting this receptor are available. On the other hand, 5- HT2 receptors were first divided in 5-HT2A and 5-HT2C. The first one corresponds to the D receptor of Gaddum as well as the ones found by Peroutka and Snyder. Ketanserin became the reference ligand for 5-HT2A. Both the receptors are coupled to phospholipase C. During the "pharmacological period" the study of 5-HT2A receptor was very subliminal, giving it its name 5-HT2F ("present in the fundus"). This was reported as capable of inducing contraction but lacking trustful tools for evaluation (Hannon and Hoyer, 2008).

The second research period, the cloning one, started with the cloning of the 5-HT1A receptor (Fargin et al, 1988). The possibility of cloning newly discovered genes brought the possibility to characterize new receptors, missing in the pharmacological research, such as 5-HT1F, 5-HT5A, 5-HT6 and 5-HT7, 5-HT1F, cloned based on 5-HT1B/1D receptors, is found in many brain areas, and has an elevated affinity for triptans and LSD. Concerning the signaling pathway, it is negatively coupled to adenylyl-cyclase (Adham et al, 1993). The cloned 5-HT5a and 5-HT5b still need to have their signal transduction and physiological function defined, to become recognized as functional receptors (*Handbook of the Behavioral Neurobiology of Serotonin* 2020). The 5-HT6 is an essential receptor in the brain, positively coupled to adenylyl cyclase in neurons from the striatum and reported to be sensitive to antipsychotic drugs. A last 5-HT receptor was reported to be positively coupled to adenylyl cyclase, 5-HT7 (Ruat et al. 1993).

Lastly, the only serotonergic receptor that is a ligand-gated ion channel, 5-HT3, enabled a second type of transmission, being an ionotropic receptor, it manages a rapid (milliseconds) response to the binding of ligands, opposed to the slow response given by

the other metabotropic 5-HT receptors. The 5-HT3 receptors form pentameric structures, co-assembling with each other (5-HT3B and 5-HT3A), to form a functional receptor (Boess, Beroukhim, and Martin 1995). The role of 5-HT3C-E receptors is yet to be elucidated, but it is known to modulate the function of 5-HT3A (Jensen et al, 2008).

The diverse function of 5-HT appears because of the concerted actions of several receptor subtypes, splicing and editing of the same and by the coupling of the multiple subtypes of this family of receptors (Millan et al. 2008). It is important to provide a comprehensive overview of what is known about the signaling mechanisms comported by the different families of receptors (Figure 6.).

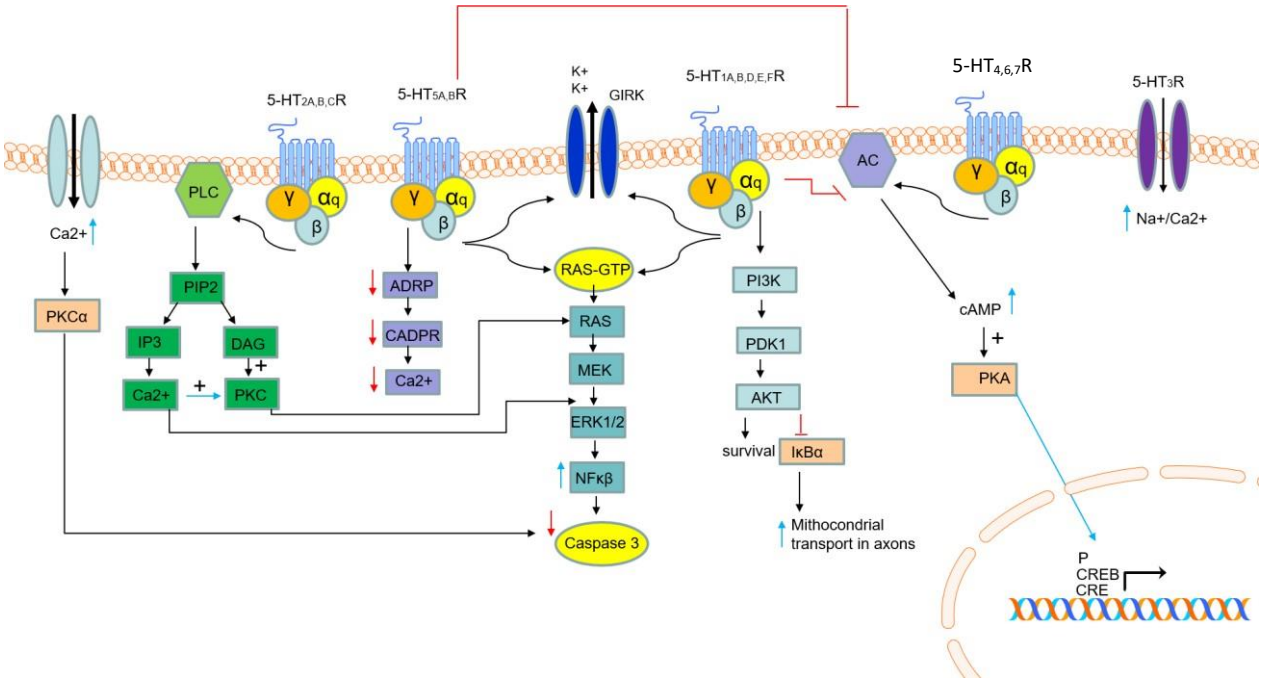


Figure 6. Review of proposed signaling pathways activated by 5-HT family receptors (based in (Pourhamzeh et al, 2021)).

The 5-HT1A receptor inhibits adenylyl-cyclase by the coupling with Gi/Go, but it is also reported to open G-protein-gated inwardly rectifying gK+ (GIRK) and inhibit gCa2+ (conductance of the voltage-gated Ca2+ channel) (Muller and Jacobs 2009) subtype of receptors and presents different distributions in the human brain, some being auto-receptors, while others are postsynaptic heterologous neurons, adjusting the drug action. One differential aspect is the much greater ability to desensitize the auto-receptor type. More research will be needed to clarify this difference, but it is thought that lies on the coupling of different G-proteins (auto receptors activating Gi3 and the postsynaptic

receptors, Go) (Cour et al, 2006). Regarding the signaling of 5-HT2A, summing its mechanical aspects, 5-HT2A receptors turn on PLC (phospholipase C), PLA2 (phospholipase A2) and the ERK pathway (downstream of PLC) in neurons (Bockaert et al, 2006). Moreover, the activation of this receptor is coined with the activation of small G proteins such as RhoA, Rab4 and Rac1 (Dai et al, 2008). An interesting aspect was reported by Yuen et al, who showed that 5- HT2A receptors block the action mediated by 5-HT1A receptors, when modulating NMDA currents is pyramidal neurons via an arrestin/Src/ERK pathway, highlighting the biased signaling in these family of receptors (Yuen and Yan 2007). Moreover, the action of these subtype of receptors is associated with the activation of matrix metalloproteinases (MMPs) via Src pathway, the release of FGF (fibroblast growth factor) receptor 2 ligands, the activation of FGFR2 receptor, the activation of downstream ERK and transcription of GDNF (Tsuchioka et al, 2008). Belonging to the same subfamily as 5-HT2A, the receptor 5-HT2B also activates Gq/G11 proteins, inducing PLC action. They are also reported to stimulate the Ca2+ mobilization in astrocytes in the rat brain. Moreover, their stimulation may lead to increase in cyclic GMP (guanosine monophosphate) through the dual activation of constitutive and inducible NOS (Nitric oxide synthase). Lastly, the activation of Gq/11 by 5-HT2B receptors can activate PI3K/Akt and ERK1/2 signaling cascades (Bockaert et al, 2006). The last in the subtype of 5-HT2 receptors, the 5-HT2C subfamily also stimulates the activity of PLC by Gq/11 in many brain regions, such as the choroid plexus (Conn, Janowsky, and Sanders-Bush 1987). There's a variety of signals initiated by these receptors, because of the RNA editing, originating different isoforms (McGrew, Chang, and Sanders-Bush 2002).

The best characterized signaling pathway activated through 5-HT4Rs is Gs/cAMP/PKA. The activation of PKA leads to the modulation of ionic currents, through long-lasting inhibition of K+ currents, enhancing neuronal excitability and a decrease in spike accommodation (Ansanay et al, 1995). Still a lot of research is needed with 5-HT5 receptors both in vivo and in vitro (Hannon and Hoyer, 2008). The 5-HT6 receptors also stimulate positively to adenylyl-cyclase in neurons (Sebben et al, 1994). On the other hand, the 5-HT7 receptors also positively couple to adenylyl cyclase, showing high affinity for 5-CT (5-carboxamidotryptamine) (Shenker et al. 1987). This subfamily also activates ERK1/2 pathway, via Epac (exchange protein activated by cAMP) or PKA, reported in hippocampal neurons (S, L, Lin et al, 2003). Moreover, it was also reported that these receptors can potentiate neurite length by Galfa/RhoA/Cdc42 pathway.

The 5-HT receptors present different characteristics, but generally they are heteroreceptors and expressed postsynaptically in non-serotonergic neurons. Some of them are auto receptors located in the presynaptic soma (like 5-HT1ARs) or in axon terminal (such as 5-HT1B and 5-HT1D receptors) (S, L, Lin et al, 2003; Pourhamzeh et al, 2021) of serotonergic neurons, controlling the release of serotonin, in a negative feedback mechanism, with the aid of serotonin transporters, adjusting the neuronal firing rate. Each neuron can express a variety of 5-HTRs, creating a complex crosstalk system, which tunes in a very sensible fashion brain function (Sahu et al, 2018). Some studies relate the possible regulative role of 5-HT3Rs in the production of acetylcholine (Ach) and its possible use in disorders such as Alzheimer's Disease (Iidaka et al, 2005). Some recent research is also trying to relate the 5-HT7Rs with dopamine, GABA, and glutamate transmission (Blattner et al, 2019).

As it was previously said, 5-HT receptors, being from the GPCR class, can communicate their chemical signal through other pathways different from the ones coupled to G-proteins. One classic non-G signaling pathway uses beta-arrestin and is coined with ERK pathway, in a long-lasting activation (>5 minutes up to several hours). It contrasts with the short-lasting activation of ERK by GPCRs, in a G-protein-dependent way (Preto et al, 2020). The internalization of GPCR-beta-arrestin normally occurs in the cytosol and is normally coupled to the Src-Raf-MEK-ERK module. Other differences in the recruitment of signaling pathways are reported when comparing different subfamilies of 5-HT receptors. For instance, 5-HT2C mediate their action in a process independent of G proteins (Gq, Gi/o), requiring physical interaction of calmodulin with the C-terminal domain of the receptor, recruiting beta-arrestin1 and 2 (Labasque et al, 2008). These last findings came to highlight the diversity of signals mediated by the activation of 5-HT receptors, both including the participation of beta-arrestin and other molecular partners, culminating in the activation of ERK pathway by these family of receptors. Moreover, it is important to highlight the constitutive activity of 5-HT receptors. In some cases, such for 5-HT2C receptors, this type of activity is dependent on the mRNA processing and editing steps, deciding the fate and functional role of these receptors (Chanrion et al, 2008). A physiological example of the constitutive effect might relate 5-HT2A receptors with impairment of associative learning by inverse agonists (Berg et al, 2008).

1.d. The psychedelic role of the 5-HT2A receptor

As stated above, the 5-HT2A receptor is the major psychedelic effector in the nervous system. This receptor is coupled with Gq/11 protein, which is then linked to the phosphoinositide hydrolysis signaling cascade (Nichols and Nichols 2008). Both tryptamines and phenethylamines, such as 2,5-Dimethoxy-4-methylamphetamine (DOM), 2,5-Dimethoxy-4-iodoamphetamine (DOI) or Dimethoxybromoamphetamine (DOB), classes of hallucinogens bind to 5-HT2AR with high affinity (Pierce and Peroutka 1989). A frequent administration of psychedelics can lead to a very rapid development of tolerance known as tachyphylaxis; a phenomenon believed to result from 5-HT2A receptor downregulation (Nichols 2016). Contrary to the majority of GPCRs, 5-HT2A suffers downregulation in response either to agonist or antagonist treatment (J, A, Gray and Roth 2001). Moreover, studies revealed the importance of two non-conserved residues in the 5-HT2A receptor, S421 in the C terminus and S188 in the intracellular loop2, in the desensitization induced by agonism on these receptors (John A, Gray, Compton-Toth, and Roth 2003). Reports show that desensitization of 5-HT2A receptor signaling is not due to reduced ability of Galfaq/11 proteins to stimulate PLC but rather to changes in 5-HT2A receptors or their coupling to G proteins. Relating this desensitization to post-translational modifications in the receptor (e,g, phosphorylation) and Galfaq or Galfa11 proteins, altering the receptor interface. There is some evidence that the shaking behavior, characterized by small tremors or involuntary movements, caused by some of these compounds might be related to metabotropic glutamate mGlu2/3-sensitive glutamate release downstream of frontocortical 5-HT2A activation (Nichols 2016). Similarities found between schizophrenia and model-psychosis induced by hallucinogenic drugs lead to the hypothesis that dopamine receptors might also modulate the cell response to hallucinogens (Nichols 2016). In 2015, Buchborn et al, stated that the difference in adaptation of different receptors, when administered LSD or DOI, pointed out that tolerance to serotonergic hallucinogens might come at two levels (Buchborn et al. 2015). As such, if a psychedelic, such as LSD, fails in downregulating 5-HT2AR, glutamate receptors might adapt instead and thus prevent cortical overstimulation.

The data brought by Moreno et al. (2013) supports the hypothesis that a constant blockage of the signaling lead by mGlu2 receptor, downregulates the binding to 5-HT2A receptors in the somatosensory cortex, measured in mice (Moreno et al, 2013). Consequently, it influences both cells signaling and ultimately behavior. In addition, since mGlu2 are presynaptic receptors, their blockage leads to an excessive release of

glutamate, which might result in a feedback mechanism of downregulating 5-HT2A gene expression on the pyramidal apical dendrites.

During the past 20 years there was a shift in the way we think about GPCRs. Now it is known that they can couple to more than one signaling pathway, with different molecules involved. The canonical pathway for 5-HT2A is coupled to Galfaq, which activates PLC. This differential activation is ligand dependent (Zhou and Bohn 2014). The final assembly between the receptor and the ligand is very much dependent on the rigidity of the ligand. This means that the flexible ethylamine chain of serotonin allows for an adaptive conformation of the complex, where the receptor can adapt to the small ligand. This adaptation is steric, electronic, and conformational in both parts. In the case of LSD, due to its rigid structure, the adaptation to a final complex is very different, resulting in a much different final assembly. Due to this, different ligands can lead to completely different cell responses, due to the diverse complexes formed between the receptor and the small ligand. Because of this it is important to get a deeper sense when we talk about 5-HT2A agonism, meaning that it is necessary to understand which signaling pathways are being activated when giving different sets of ligands and how these differences influence the psychedelic response.

As it was stated previously, the most well understood signaling pathway activated by 5-HT2A is the one coupled to Gαq, that induces the stimulation of PI-specific PLC (B, L, Roth et al, 1984). This enzyme breaks phosphatidylinositol membrane lipids, generating inositol-1,4,5-trisphosphate, and diacylglycerol (DAG) (Williams 1999). The inositol phosphates lead to release of Ca2+ from intracellular stores and diacylglycerol remains bound to the membrane and activates protein kinase C (PKC). Although, some studies such as the one from Rabin et al, came to highlight the lack of correlation between the activation of this signaling pathway and the discriminative stimulus effects of hallucinogens. This indicates the possible effects of additional transition states of the receptor-ligand complex and their contribution to the agonist efficacy (Rabin et al, 2002).

When 5-HT2A is activated, it can also stimulate phospholipase A2 (PLA2), which preferentially hydrolyzes arachidonic acid (AA)-containing phospholipids, producing free AA and lysophospholipid. This pathway is independent of PLC-mediated signaling. The PLA2 signaling pathway is more complex than the PI turnover cascade, apparently involving multiple G proteins and the extracellular signal-regulated kinase (ERK) 1/2 and p38 mitogen activated protein kinases (MAPKs), at least in NIH3T3 cells (Kurrasch-Orbaugh, Parrish, et al, 2003). Different ligands exhibit different affinities (EC50, half

maximal effective concentration) with 5-HT2A, reflected also in different activation ratios of both AA and PI pathways (Kurrasch-Orbaugh, Watts, et al, 2003).

Some of the kinases participating in 5-HT2A phosphorylation are known, such as RSK2 (one of the p90 ribosomal S6 kinases, RSKs, member of the ERK/MAPK cascade) (Sheffler et al, 2006). RSK was associated with a "tonic brake" on the production of a second messenger, serving as a regulator to the GPCR function. In addition, the group also reported an important interaction between RSK2 and the loop3 (ICL3) of 5-HT2AR. Other kinases were identified that also affect 5-HT2A phosphorylation, such as the PKC phosphorylation of serine residue S291.

Beta-arrestins are scaffolding proteins that also mediate the GPCR signaling, being determinant to the effects of specific ligands, Schmid et al. did research to evaluate if the recruitment of beta-arrestin was necessary for the in vivo behavior as an effect to some psychedelics (Schmid, Raehal, and Bohn 2008). They concluded that the effect of some of these molecules, e.g., DOI, are beta-arrestin2 independent, contrary to the effect of 5HTP. They performed other experiments using MEFs, concluding that different stimuli (DOI and 5-HT), by different molecules, stimulate the production of ERK1/2, by different pathways (PLC-dependent and beta-arrestin, respectively) (Figure 7).

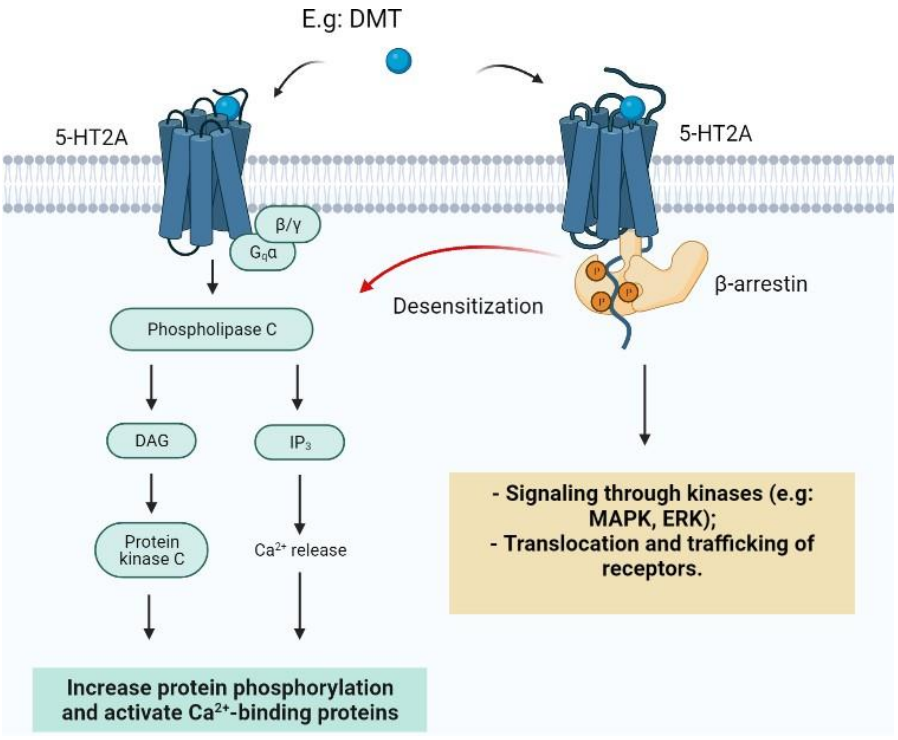


Figure 7. Example of biased signaling present in 5-HT2A receptor signaling.

As confirmed from the reports presented, a lot of research is still needed to get a better sense on how 5-HT2A receptors are activated by the chemical diverse class of psychedelics. Ranging from structural studies of how the binding of a specific ligand prefers a determined conformation of the receptor to systemic studies on how this choice leads to the activation of specific signaling pathways. This is a major step to design novel therapeutics, based on entheogens, if trying to develop a more specific and problem-oriented pharmaceutical. The objective is to dissect the pharmacophores present in these compounds and try to establish connections with specific signaling pathways. This might pave the way for effective modulation of brain disorders.

1.e. 5-HT system and Neurological Disorders

1.e.1. Anatomical distribution of 5-HT neurons

It is important to know the anatomical distribution of serotonergic neurons to better understand how serotonergic neurotransmission influences brain function and how its dysfunction might lead to neurological disorders. We know that cell bodies that contain serotonin are organized in clusters located in the midline of the brainstem (Brady and Siegel 2012). These groups of cell bodies have been previously identified as the raphe nuclei. Later, Dahlstrom and Fuxe identified and characterized nine groups, based on their structural characteristics and organization, from B1 to B9, with the most part belonging to the raphe nuclei area (Brady and Siegel 2012; Törk 1990). Controversially, most neurons in the raphe nuclei are non-serotonergic. Serotonergic neurons having their cell bodies outside this area have projections entering the raphe nuclei.

The serotonergic ascending pathway, from the midbrain raphe nuclei to the forebrain, is divided in two main axes, the dorsal periventricular path, and the ventral tegmental radiations (Brady and Siegel 2012). In the caudal hypothalamus, both pathways join with the medial forebrain bundle. This region is also enervated with dopaminergic and noradrenergic axons (Molliver 1987). The raphe nuclei can be divided in two sub-regions, the dorsal raphe nuclei innervate the ventral hippocampus, amygdala, and striatum, whereas the median raphe nuclei projects to the dorsal hippocampus, septum, and hypothalamus. They form dissimilar yet partially overlapping patterns of innervation. On the other hand, both areas send overlapping neuronal projections to the neocortex,

exhibiting a topographical organization, affecting different cortical neurons. Moreover, they present different morphology (type M and type D) and consequently they are affected differently by compounds such as amphetamine derivatives, like MDMA (Brady and Siegel 2012). Yet, the raphe nuclei receive projections from other clusters in the brainstem like substantia nigra and ventral tegmental area (dopaminergic neurons), superior vestibular nucleus (containing acetylcholine), locus coeruleus (with norepinephrine) and nucleus prepositus hypoglossi and nucleus of the solitary tract (epinephrine) (Hensler 2006).

1.e.2. The Serotonergic Dance

Some brain disorders are not only dependent on the low concentrations of serotonin but also from its precursors. Serotonin is made from the amino acid tryptophan in two following reactions. The first being the addition of a hydroxyl group to the indole ring of tryptophan by the enzyme tryptophan hydroxylase. The second one is the removal of the carboxylic group from the end of the lateral chain, by the aromatic amino acid decarboxylase. Moreover, the tryptophan used for the synthesis of serotonin is only obtained by diet (Figure 8.). If the concentration of this is lowered, and not enough precursor material is reaching the brain, the lowered levels of 5-HT in the brain might lead to changes in behavior. The mechanism of action of novel psychotherapeutic drugs can be studied using the same strategy, by lowering the levels of 5-HT in the brain. Theoretically, raising the intake of dietary tryptophan leads to an increase in the brain's 5-HT concentration. It is interesting to point out that there are two isoforms of the tryptophan hydroxylase, one being expressed in peripheral systems (Tph1) and other in the brain (Tph2). This gives the opportunity to have a more selective way of action when designing novel drugs. Moreover, it is reported that the production of serotonin is dependent on the frequency of the electrical stimulus in serotonergic soma. This might be an explanation for the plasticity of neurons to adapt after short or long-term activation of function (Boadle-Biber 1993).

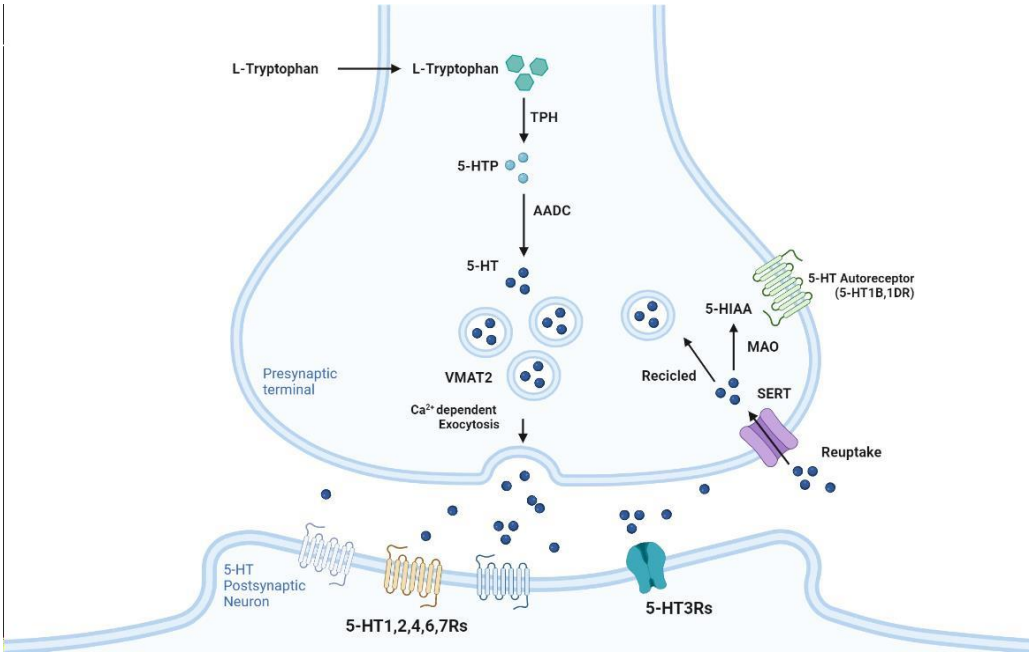


Figure 8, Synthesis, and metabolism of serotonin (5-HT) (base in (Pourhamzeh et al, 2021)).

The serotonergic system has a very diverse action, ranging from the central nervous system (CNS) to the gastrointestinal tract (GI) (Pourhamzeh et al, 2021). It modulates a variety of behaviors, such as mood, cognition, anxiety, learning, memory, reward processing and sleep. With this, any deficit in the serotonergic concentration or metabolism in the brain can lead to a broad range of pathological conditions, such as depression, schizophrenia, mood disorders and autism. Recent review also points to the role of serotonin in sexuality, respiratory stability, circadian rhythms, and embryonic development (Abela et al, 2020), (Paulus and Mintz 2016). As it was stated previously, there is a different concentration of serotonin between the peripheral system (approximately 95%), and the central nervous system (5%) (Pourhamzeh et al, 2021). It is important to notice that the blood-brain barrier is not permeable to 5-HT, making these two systems independent from each other (Sahu et al, 2018), 5-HT has a role in neuronal development, by stimulating the synapse formation and connectivity, to build a complex network. Yet, it regulates cell adhesion molecules, influencing neuronal plasticity in developing brains and adult brains (Dalva, McClelland, and Kayser 2007), mediated by interactions with brain-derived neurotrophic factor (BDNF). Moreover, serotonin is capable of modulating other neurotransmitters and hormones, such as dopamine, epinephrine, gamma-aminobutyric acid (GABA), cortisol, prolactin, acetylcholine (ACh), oxytocin, substance P and vasopressin.

Concerning the transport system of free serotonin, this is carried by serotonin reuptake transporters, or SERTs, in an active transportation and recycling process (Figure 8). This controls the duration and extent of 5-HT activation. Variations on the expression of SERTs, such as edits during the processing of their correspondent mRNA, were coined with depression, anxiety disorder, suicidality, and autism (White, Walline, and Barker 2005). These disorders are also related with low levels of 5-HT in the synaptic cleft, due to a decrease in the production of the NT or by an increase in the functioning of SERTs. Because of this, SSRIs, or selective serotonin reuptake inhibitors, were developed to block the 5-HT reuptake by SERTs, having a very high specificity to these systems. Some examples include fluoxetine, citalopram, paroxetine, fluvoxamine, sertraline, and escitalopram. These drugs are globally used in the treatment of psychiatric disorders like major depressive disorder (MDD), obsessive-compulsive disorder (OCD), bulimia nervosa, anxiety disorders and some non-psychiatric disorders, such as migraines and pain syndromes (Lorman 2018). Nonetheless about one-third of clinical patients do not respond to SSRIs (Zugliani et al, 2019), calling for the design and development of better drugs for the treatment of these disorders. Moreover, several side effects have been reported because of the excessive use of these pharmaceuticals, mainly in the CNS by activation of 5-HT1A autoreceptors and 5-HT2C heteroreceptors (Burghardt et al, 2007).

One deregulation on this system is serotonin syndrome, which surges because of the increase in activity in both peripheral and central 5-HTRs, being a response to high levels of serotonin. The raise in 5-HT content is normally caused by interaction between drugs. These include monoamine oxidase inhibitors (MAOIs), SSRIs, SNRIs (serotonin-norepinephrine reuptake inhibitors), TCAs (tricyclic antidepressants), 5-HT releasers, precursors, and agonists of 5-HTRs, and some opiates (Baldo and Rose 2020). Some of the symptoms of 5-HT syndrome involve autonomic hyperactivity, changes in the mental status (like disorientation, anxiety, restlessness) and neuromuscular abnormalities (Baldo and Rose 2020; Simon and Keenaghan 2022).

1.e.3. The Serotonin Manifestation

As stated previously, serotonin has a key role in the functioning of the healthy brain. To better relate neuropsychiatric disorders with serotonergic imbalances, it is important to highlight how serotonin modulates behavior and other neurological processes.

1.f.3.a. Serotonin and Memory

Firstly, serotonin modulates learning and memory processing, in memory consolidation processes (Bostancıklıoğlu 2020) and in the formation of both associative and non-associative memories. Moreover, 5-HT is also associated with the formation of short- and long-term, verbal episodic, spatial working memories and long- term stimulation of 5-HT has been coined with memory impairment, by destroying axon terminals (He et al, 2020; Volle et al, 2018), (Hritcu, Clicinschi, and Nabeshima 2007),(van Goethem et al, 2015). A lot of research is still needed to properly correlate memory and serotonin, however it is thought that it might be related to various subtypes of 5-HTR (Cowen and Sherwood 2013). Some of the subtypes of receptors were already associated with some impairments, such in emotional memory that occurs when the neuronal activity decreases, due to the activation of 5-HT1ARs, located postsynaptically. On the other hand, the use of 5-HT1A, 5-HT1B and 5-HT3 receptor antagonists or 5- HT2A, 5-HT2C and 5-HT4 receptor agonists have been related to prevention of memory impairments and facilitators in learning (Stiedl et al, 2015). This last sub-group might normalize NMDA receptor function, and consequently improve cognitive abilities. As an example, pimavanserin is described as a 5-HT2ARs antagonist, able to alleviate Parkinson's disease psychosis (Sahli and Tarazi 2018). In a recent study, 5-HT4 receptors were associated with the non-amyloidogenic pathway of amyloid precursor protein (APP), giving an opportunity for action in Alzheimer's disease (Mdawar, Ghossoub, and Khoury 2020). Serotonin was also reported to be able to bind to intermediate aggregates of alpha-synuclein, having a potential role in alpha-syn pathology, such as in the development of Parkinson's disease (Falsone et al, 2011).

On the other hand, a decrease in the content of 5-HT4 receptors have been observed in cortical regions, mainly in the hippocampus, of patients with Alzheimer disease. At the same time, this subtype of receptors is linked to adult neurogenesis. Their agonists can increase proliferation of new cells in the dentate gyrus (DG) (Mendez-David et al, 2014). Recently, it was proposed that 5-HT4 and 5-HT1A can promote neuronal maturation, which could contribute to eventual remission of AD, activating neurogenesis in a BDNF dependent manner. BDNF interacts with the MEK-ERK pathway, implicated in an enhanced activity of alpha-secretase, with inhibition of gamma-secretase, reducing the production of toxic amyloid-beta (Liu et al, 2019).

1.f.3.b. Serotonin and Motivation

Serotonin, in combination with dopamine, is well acknowledged as an important player in the function of reward systems, accommodating motivational and reinforcement behaviors (Cohen, Amoroso, and Uchida 2015). The flood of 5-HT in the limbic-corticostriatal circuit is coined with the value of natural rewards. A pivotal area is the raphe nuclei, which takes part in the activation of self-stimulation, associated with the induction of the medial brain forebrain bundle or the VTA (ventral tegmental area) (Pollak Dorocic et al, 2014). Yet about these systems, inputs, mediating excitability, or inhibition effects, originate from the prefrontal cortex (PFC) and the lateral habenula (Hb) to serotonergic dorsal raphe nucleus (DRN) and dopaminergic VTA neurons (Geddes et al, 2016). Finally, RN (raphe nucleus) also receives GABA inputs from the lateral Hb, through the rostromedial tegmental nucleus. This aids in the hypothesis that reinforcement learning, occurring during habit formation, is a consequence of serotonergic stimulation (Iigaya et al, 2018).

1.f.3.c. Serotonin and Pain

Serotonin has also been reported to modulate pain and nociceptive inputs, in the dorsal horn of the spinal cord, where these signals are then sent to the thalamus, reaching the cortex. The accepted gate control theory of pain states that the interneurons from the dorsal horn release GABA and glycine to diminish the nociceptive signal. At the same time, the periaqueductal gray (PAG) receives information from higher brain centers, mediating an analgesic effect (Provenzi et al, 2020). On the other hand, the rostral ventromedial medulla (RVM), that contain serotonergic neurons from Raphe Magnus, can modulate nociceptive inputs, through the descending inhibitory system (Heinricher et al, 2009). Generally, mal function in serotonergic systems might lead to analgesia or hyperalgesia, though the respective intrinsic mechanisms are still difficult to differentiate (Tao et al, 2019). A wide variety of receptors are associated with analgesic effects, such as 5-HT1A, 5-HT1B, 5-HT1D, 5-HT4, 5-HT5A and 5-HT7. On the contrary, the combined action of 5-HT2B and 5-HT3 receptors promote nociception. Also, the subfamily of 5-HT1 receptors have been coined with inhibition by GABAergic and glutamatergic transmission in the rat midbrain PAG (Cortes-Altamirano et al, 2018).

1.f.3.d. Serotonin and Sleep

Lastly, the circadian system, which regulates sleep-wake cycles, is a biological clock, controlling behavior, physiology, and mood. It involves the action of the hypothalamic suprachiasmatic nuclei (SCN), and it has been reported that there is a crosstalk between the circadian and serotonergic systems. Both systems influence a similar set of mood disorders, such as seasonal affective disorder (SAD), depression, bipolar disorder, and autism (Ciarleglio, Resuehr, and McMahon 2011). 5-HT affects the circadian system in the non-photic phase of the circadian cycle, acting in opposition to the activation in SCN by light. In the other way, the content in serotonin is under influence of the circadian rhythm, where the synthesis of this NT depends on the release within SCN and other limbic projections (Duet and Fonken 2019). Manipulation of the circadian cycle might come as an effective treatment for mood disorders, when interacting with 5-HT receptors, by the design of novel pharmaceuticals.

1.e.4. Serotonin Imbalance and neuropsychiatric disturbances

After reviewing some of the serotonin roles in the healthy brain, it is mandatory to review the imbalances in its metabolism, distribution or signaling that can lead to neuropsychiatric disorders.

1.e.4.a. Serotonin and Anxiety

During the last years anxiety disorders gained a principal role in neuroscience research, relating them to 5-HT disturbances (Ohmura et al, 2020). Based on the developmental role of serotonin, dysregulation of its transmission in critical stages on the development of an organism can have long-lasting effects and/or alterations, leading to anxiety in adulthood (Teissier, Soiza-Reilly, and Gaspar 2017). 5-HT1A receptors have been used as a target for the treatment of anxiety, since there is data pointing to a smaller number of 5-HT1A receptors in the forebrain of patients with panic disorder (Nash et al, 2008) and in the amygdala of patients with social anxiety disorder (SAD) (Lanzenberger et al, 2007). Nowadays, novel 5-HT1ARs agonists, such as buspirone, are being prescribed as effective anxiolytics, though not substituting benzodiazepines (Yamashita et al, 2018). Other receptors like 5-HT2, 5-HT3 and 5-HT7 receptors are also possible targets for the treatment of anxiety disorders (Griebel and Holmes 2013). 5-HT6 receptors have been

linked to a therapeutic role in depression, because of its high affinity for antidepressant drugs, like amitriptyline and amoxapine, and some antipsychotics such as olanzapine and clozapine, This difference in their role depends on the brain region where the receptor is being expressed (Żmudzka et al, 2018). The content in SERTs has also been linked to GAD (generalized anxiety disorder). Blocking the action of these transporters has been one major strategy in the chronic treatment of these disorders. Both SSRIs and SNRIs are considered standard treatments for anxiety disorders, like GAD, PD, SAD and PTSD (post-traumatic stress disorder) (Bandelow, Michaelis, and Wedekind 2017). The underlying mechanism of these therapies is linked to a decrease in the amygdala response, after negative stimulus. Because of the long time needed for these pharmaceuticals to influence humans, more research is needed in the design of more effective and fast-acting medication.

1.e.4.b. Serotonin and Depression

Another major neuropsychiatric disorder that has puzzled neuroscientists is depression. The neurobiological implications of this imbalance in the brain are still not completely understood and consequently the design for therapeutic drugs is an extremely demanding work. Though a serotonergic hypothesis has been thought for depression, concerning the hypofunction of 5-HT neurons. Alteration in both SERTs and 5-HTRs and enhanced activity of presynaptic receptors are linked with major depressive disorder (MDD) (S,-H, Lin, Lee, and Yang 2014). The blockade of several subtypes of serotonin receptors have been coined with antidepressant effects: 5-HT2ARs (Aznar and Hervig 2016), 5-HT2CR (McCorvy et al, 2011) and 5-HT3Rs (Gupta, Prabhakar, and Radhakrishnan 2016). On the other hand, 5-HT2BRs (Hamati, El Mansari, and Blier 2020) and 5-HT4Rs are associated with depressive behaviors when stimulated. Recent research points to a detrimental interaction between serotonin and noradrenaline, where the lack of efficacy of SSRIs in resistant depressive cases might lie in this crosstalk communication (Dremencov, El Mansari, and Blier 2007). The blockade of 5-HT3Rs in the hypothalamic-pituitary-adrenal (HPA) axis has been a target for antidepressant effects (Gupta, Prabhakar, and Radhakrishnan 2016). Research relating 5-HT6Rs and 5-HT7Rs is being conducted, where their blockage might offer a faster therapeutic approach (Mnie-Filali et al, 2011). The availability and distribution of SERTs across the brain may also be a signature of depression (Hsieh et al, 2014). Nowadays, the main pharmaceuticals for treatment of depression include TCAs, SSRIs and SNRIs, that act by increasing the level

of monoamines, such as 5-HT and noradrenaline, enhancing the activity of postsynaptic serotonergic and noradrenergic receptors. Of these groups of compounds, the most prescribed are MAOIs and SSRIs, such as amitriptyline and imipramine. This last group of drugs is the most effective one in treating depression, nowadays, associated with possible long-term treatment (Naber and Bullinger 2018).

1.e.4.c. Serotonin and OCD

Another neuropsychiatric disorder that has been central in recent pharmacological advances is OCD or obsessive-compulsive disorder. It is characterized by intrusive thoughts (obsessions) and repetitive behaviors (compulsions). Recent advances coined this disorder with the cortico-striato-thalamocortical pathway (Lissemore et al, 2018). Several genetic studies have been conducted, highlighting the connection between gene expression of serotonergic, dopaminergic, and glutamatergic systems with the pathophysiology of this disease. Moreover, abnormal levels in expression of 5-HT2ARs or SERTs were reported in OCD patients, as well as alterations in the brain volumes (Nazeer et al, 2020). During the last years, some drugs were developed, such as granisetron and ondansetron, that act as antagonists of 5-HT3 receptors, described as a treatment for OCD patients (Askari et al, 2012). SSRIs are again one of the major lines of treatment for this disease since most OCD patients respond to their action. Recent advances coin the summed action between dopamine antagonists and SSRIs as a more sophisticated form of treatment.

1.e.4.d. Serotonin and ADHD

ADHD or Attention-Deficit/Hyperactivity Disorder is one of the most diagnosed neurological disorders in children, described by hyperactivity, lack of attention and impulsivity. It was postulated that most patients exhibiting ADHD, also tend to display other disorders such as oppositional defiant disorder, conduct disorder, depression, anxiety disorders and learning disabilities (Bélanger et al, 2018). The phenotype of ADHD disorder has been related with a chronic reduction in the bioavailability of serotonin, reported by the low concentration of 5-HT in the blood of patients with this condition. 5-HT is linked to behaviors such (Bolaños et al, 2008) as impulsivity, inhibition, and attention, by coupling its action with the dopaminergic system (Hou et al,

2018). In addition, the orbitofrontal cortex was reported to have its content in 5-HT altered in ADHD patients, resulting in emotional imbalances, inhibition, and reversal learning (Curatolo, D'Agati, and Moavero 2010). Besides, the participation of 5-HT receptors, like 5-HT1B, 5-HT2A and 5-HT2C was also reported (Hou et al, 2018). On the other hand, there was no difference noted in the expression of SERTs between healthy and ADHD patients. However, it was noted a reduced affinity of SERT in individuals with ADHD (Oades 2007). Due to their interaction with 5-HT, SSRIs and TCAs are the main line for drug treatment of ADHD, such as fluoxetine, methylphenidate (MPH) or buspirone, that was reported to normalize ADHD-like behavior in preadolescent rats (Oades 2007; Bolaños et al, 2008).

1.e.4.e. Serotonin and Autism

Other neuropsychiatric diseases related to the serotonergic system are disorders in the autism spectrum (ASDs). They are characterized by difficulties in communication, social interaction and repetitive or obsessive behaviors. Contrary to the other disorders described in here, ASDs are characterized by hyperserotonemia, or an increased level of 5-HT in the blood in one third of autistic male adults. This contrasts to the low levels of the same neurotransmitter in the brain of children with ASDs (Marler et al, 2016). Though it is not completely understood how the serotonergic system influences the ASD neurobiology, it was reported that abnormalities in 5-HT neurons in the brainstem might lead to synaptic and network modifications in projection areas controlling social behavior, like the frontal cortex (Takumi et al, 2020). Studies point to the involvement of SERTs and some 5-HTRs, in the pathophysiology of ASDs, like 5-HT1A and 5-HT2A, that were reported to have a reduced binding potential in the thalamus, in the posterior cingulate cortex and in the fusiform gyrus (Oblak, Gibbs, and Blatt 2013). Moreover, SERTs were also described with lower affinity in ASDs phenotypes, postulating that any alteration in the normal activity of these transporters, increases the odds to develop ASDs-like neurobiology (X, Chen et al, 2015). Early exposure of infants to SSRIs is coined with an increased incidence of ASDs, due to the development of larger areas in the amygdala and insula regions, influencing the control of anxiety, mood states and social behaviors (Andalib et al, 2017). Because of this, more studies are needed to better understand the neurobiology of ASD to find suited pharmaceuticals for their treatment.

1.e.4.f. Serotonin and Schizophrenia

One of the most related neuropsychiatric disorders with the administration of psychedelics is schizophrenia-like psychosis. This disorder is characterized by periods of chronic psychosis, and it distinguishes two types of symptoms. Positive symptoms relate to active imbalances, such as hallucinations, while negative symptoms describe lack of normal capacities, such as inappropriate emotional responses. Adding to these, schizophrenia includes impairments in cognition, attention, memory, and executive functions. Neurodevelopmental abnormalities of the brain and dysregulation of NTs, such as dopamine, in several pathways have been linked with this disorder. Recent imaging techniques support the idea that high 5-HT activity can disrupt brain areas such as the cerebral cortex, anterior cingulate cortex, and dorsolateral frontal lobe. Moreover, altered 5-HTR, such as 5-HT1A and 5-HT2A, and SERT expression are found in schizophrenic patients. However, the results about the expression of 5-HT1AR are dubious (Andalib et al, 2017; Selvaraj et al, 2014). On the other hand, the story for 5-HT2A is a little bit more clear, and it has been reported that the binding potential of these receptors in the frontal cortex of schizophrenic patients is a lot smaller when compared to healthy brains (Aznar and Hervig 2016). However, Selvaraj et al, (2014) also reports a downregulation of 5-HT2AR in schizophrenia (Andalib et al, 2017; Selvaraj et al, 2014). Antagonists of these receptors, such as olanzapine and risperidone, can act in the nigrostriatal pathway, enhancing the release of dopamine in the striatal area, inhibiting the serotonin effect (Stępnicki, Kondej, and Kaczor 2018). Stahl et al. (2018) also related the hyperactivity of 5-HT2ARs with the release of glutamate in VTA and with activity in the mesolimbic pathway, resulting in a load of dopamine in the ventral striatum (Stahl 2018). This points to a better therapeutic response when using multi-target antagonists of 5-HT2ARs and D2 receptors, with a bigger affinity for the first class of receptors (Stępnicki, Kondej, and Kaczor 2018). It was also reported a possible therapeutic window with 5-HT3 and 5-HT6 receptors, in alleviating the cognitive symptoms associated with this disorder, 5-HT5A and 5-HT7 receptors were also mentioned in the disruption of cognitive impairments and negative symptoms (Nikiforuk et al, 2016). The expression of SERT has not been linked to schizophrenia symptoms. Additional work is mandatory to better relate the neurobiology of schizophrenia with the serotonergic system to facilitate the process of design of more efficient anti-psychotic drugs.

1.e.4.g. Serotonin and Addiction

Lastly, the serotonergic system is of major importance in the mediation of addictive behavior. Addiction is a major downfall of nowadays society, given rise to several neuropsychiatric disorders both during adulthood and in younger ages, due to genetic implications that might influence child and adolescent development. The abuse of drugs such as cocaine, amphetamines (AMPH), methamphetamine (METH), MDMA, morphine and alcohol increase the activity of extracellular serotonin (Trigo et al, 2007). This raise in serotonin levels, after a continuous intake of these drugs, was reported in several brain areas, like subcortical structures, such as Nucleus Accumbens, dorsal striatum, ventral pallidum, hippocampus, thalamus, VTA, amygdala and in neocortical regions, including frontal, prefrontal, temporal, occipital, entorhinal and perirhinal cortices. Yet the use of these drugs can lead to peaks of 5-HT, that in the long-term lead to a reduction of serotonergic activity due to an enhanced stimulation of inhibitory auto receptors (Kirby, Zeeb, and Winstanley 2011). Serotonergic receptors have been coined with possible therapeutic roles, such as the action of 5-HT1AR's agonists, like buspirone, in the reduction of cocaine intake (Collins and France 2018). The blockade of other receptors such as 5-HT1B, 5-HT2A, 5-HT2C and 5-HT6 were reported to control drug-seeking behavior (Dhonnchadha et al, 2009). With this said, drugs that would interact with serotonergic receptors might lead to the reshape of addiction therapy.

2. Motivation

It is now evident how essential it is to explore the structural changes in the serotonergic receptors when bound to different ligands that activate a variety of signaling pathways, ending in a diverse group of effects. *In vitro* and *in vivo* assessments are being developed on the characteristics of the binding between these psychotropics and receptors from the serotonin family, including the kinetics of bindings, signaling pathways that may be activated and the integration of this activity with certain behavioral responses, such as anti-addictive effects in both humans and animals. However, the structural perspective of this story remains poorly elusive. The release of the first crystallographic structure between LSD and the receptor 5-HT2B in 2017 (Wacker et al, 2017) led to a growing interest in the way these ligands modulate the structure of the receptors and how these different conformations lead to different cellular responses. I*n silico* techniques have become the first choice to explore these changes, which due to their atomic scale cannot be assessed with wet-lab techniques. A lot has been said about psychedelics and their potential therapeutic effects. In the last years, the number of articles about the action of these compounds on the brain suffered an enormous crescendo due to the rising interest in the usage of these drugs as therapeutic agents in brain disorders such as depression, anxiety, or PTSD. At the end of 2020, a paper came out presenting a non- hallucinogenic analog of ibogaine (Cameron et al, 2021). Ibogaine is a naturally occurring alkaloid that has anti-addictive properties in both humans and animals, having the potential to treat addiction to various substances, including opiates, alcohol, and psychostimulants. However, it presents a very long hallucinogenic component and other complications, such as nausea and cardiac complications, which can be seen as a step-back when considering it as a therapeutic agent. On the other hand, tabernanthalog (TBG), the non-hallucinogenic produced by the authors of the paper, was shown to maintain ibogaine´s therapeutic effects in mice without the associated risks. Comparing the effects of both compounds, TBG and ibogaine, may allow the answer to a hot question on psychedelic research: can therapeutic benefits occur without the subjective effects? If classical psychedelics increase neuroplasticity and decrease inflammation leading to an antidepressant effect, is the trip necessary?

Computational methods enable the study of atomistic changes, occurring in receptor-ligand binding, that should be taken into consideration when assessing these systems. This novel *in silico* methodology enables the design and development of better-tuned pharmaceuticals for the treatment of a variety of brain disorders. Besides, they reveal a previously unknown territory that enables the correlation between specific pharmacophores and cell responses.

3. Methods

3.a. Homology Modelling

During the past years the number of known protein structures increased a lot, However, only a fraction of these structures was studied at atomic resolution (Y, W, Chen 2014). Computational techniques can help close the existent gap between sequence and structure in protein modelling. Protein structure models derived from these *in silico* techniques provide valuable working sets to generate testable hypotheses. These models can be produced with the aid of comparative modelling methods, which rely on structural information from related proteins to guide the modelling procedure, or free modelling techniques (also known as *ab initio* or *de novo* modelling), which do not rely on related proteins, but instead uses a variety of methods to combine physics with the known behaviors of protein structures (however extremely computationally expensive) (Fiser 2010). Comparative modelling consists of four main steps: 1) fold assignment that identifies overall similarity between the target sequence and at least one known structure (template); 2) alignment of the target sequence and the template, 3) building a model based on the alignment with the chosen template and 4) predicting the accuracy of the model. An example of a homology modelling program is the software MODELLER (Webb and Sali 2016). In the simplest case, the input is an alignment of a sequence to be modelled with the template structure, the atomic coordinates of the template, and a simple script file. With this, MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER can also be used to perform other auxiliary functions such as fold assignment, alignment of two protein sequences, multiple alignment of protein sequences and/or structures, clustering of sequences and/or structures, and ab initio modelling of loops in protein structures (Eswar et al, 2003). MODELLER implements comparative protein structure modelling by satisfaction of spatial restraints that include: i) homology derived restraints on the distances and dihedral angles in the target sequence, extracted from its alignment with the template structures; ii) stereochemical restraints such as bond length and bond angle and bond angle preferences, obtained from the CHARMM-22 molecular mechanics' forcefield; iii) statistical preferences for dihedral angles and nonbonded interatomic distances, obtained from a representative set of known protein structures and, finally; iv) optional manually curated restraints, such as those from nuclear magnetic resonance (NMR) spectroscopy (M,A, Marti-Renom, et al, 2000). At the end of the modelling procedure, the models obtained must be evaluated to choose the

best ones. Several scoring systems were developed, such as: i) DOPE score (Discrete optimized protein energy), which is a statistical potential calculated based on atomic distances from a sample of native protein structures (M,-Y, Shen and Sali 2006); ii) the Z-Score, presented as a quality measure of the difference between the energy of a specific protein structure with the energy distribution calculated from random conformations of the protein (Wiederstein and Sippl 2007); iii) the LGscore, that is calculated based on P-values of the protein (Cristobal et al,2001); and iv) MaxSub, that is also a quality measure and is based on the identification ofthe maximal subset of alpha Carbons which are most similar to the structure obtained experimentally (Siew et al, 2000).

3.b. Docking

Docking can offer theoretical calculations for target-ligand binding conformation as well as scores of binding affinity, making it useful for both initial hit compound screening and computational analysis of lead compound binding patterns (Zhu et al, 2021). Rigid docking refers to the classic drug-ligand "lock-and-key" model, while flexible docking stems from the later development of "induced-fit" and "conformational selection" models. Molecular docking is central in computer-aided drug design (CADD). In a general way, computational docking consists in the prediction of the best orientation and conformation of a small molecule (drug) when interacting with the target (larger macromolecule)(Khamis, Gomaa, and Ahmed 2015), forming a stable complex molecule. The stability of the formed complex can be inferred by the binding free energy, which is normallycalculated by a molecular mechanics (MM) force field. Obtained binding free energy, other important descriptive values can be inferred, such as IC50 (half-maximal inhibitory concentration), Ki (inhibitor constant) or Kd (dissociation constant). After this, these binding descriptors are verified using experimental techniques (Cuccioloni et al, 2020). As a result, the computational docking offers many possible solutions, differing in their poses (both the ligand and the amino acids present in the receptor). With this, scoring functions (SF) are employed to better choose which output is the best.

Every docking algorithm shares two main inputs: the tertiary structure of the target receptor, determined by biophysical or prediction techniques and secondly a database of potential ligands (small drugs). Similarly, two outputs are generated after the docking experiments: the first is the description of the novel ligand and the description of the most stable complex, or optimal binding pose (Torres et al, 2019). This last description consists of the relative orientation of the ligand, compared with the receptor, and the conformation of both ligand and receptor when bound together. Every docking protocol shares the same rationale. Starting with the input of the protein and ligand files (both in .pdb file format) into the docking program and followed

by a thorough cleaning of these files. These .pdb files format include the atom features, position, connectivity, among others. This step consists in assessing missing atoms, chain breaks, the removal of crystallographic waters and the protonation of both molecules. This protocol can be performed in any docking software, like AutoDock4, MOE (Molecular operator environment), among others (Pagadala, Syed, and Tuszynski 2017). The main objectives here is to correctly identify the best geometry of the ligand when inserted in the protein, by available methods such as Monte Carlo, Molecular Dynamics, Simulated Annealing or Genetic Algorithms, and to calculate the correspondent energy score and energy terms, functioning as descriptors for each binding pose, with scoring functions. The final step consists of the analyses of the docking results, through any graphical user interface (GUI), such as AutoDock4, AutoDock Vina or MOE.

3.c. Scoring Functions

Scoring functions are characterized as mathematical predictive models that score the binding free energy of each complex, evaluating them in a relational fashion. They exhibit three major functions: the first one being the determination of the best site for the binding and the correspondent binding poses. On the other hand, scoring functions are also necessary for the prediction of the absolute binding affinity between the ligand and the protein for lead optimizations. This search will lead to the build of the best lead for the design of novel pharmaceuticals, helping in the virtual screening step, identifying potentially novel drug leads for a given target, searching in a large ligand database (Schneider 2010). It is important to note that scoring functions aren't motivated for a high-level theory of the physics of the system. Instead, they make a compromise between speed and accuracy, doing various approximations. In a general perspective, scoring functions should satisfy three main capabilities: scoring power, ranking power, and docking power (Khamis, Gomaa, and Ahmed 2015).

The scoring functions are divided in three main classes: force field, empirical and knowledge-based SFs (Huang, Grinter, and Zou 2010). With the recent advances in computer science, this field has evolved a great deal, including a four class of scoring

functions, based in machine learning (ML). The first difference between these classifications is that the first three are based on the features evaluated and normally follow a linear regression model. While the fourth one deals with nonlinear regression ML methods.

The physics-based scoring functions include the ones based on the force-field, solvation models and quantum mechanics methods. The first scoring functions were mainly based on the binding energy, considering the enthalpic contribution, assessing both the van der Waals (E_{vdw}) and the electrostatic interaction (E_{elec}) shared in the complex, by the ligand and protein (J. Li, Fu, and Zhang 2019) (Equation 1).

$$E_{bind} = E_{vdw} + E_{elec} \quad (1)$$

However, lacking the contribution of entropy, the scoring functions generated lacked accuracy. As such, it was added as a factor to Equation 1 (ΔG_{solv}), describing the torsion entropy of ligands and the solvation/desolvation effect, induced by explicit and implicit solvent models (Equation 2) (Ross, Morris, and Biggin 2012).

$$E_{bind} = E_{vdw} + E_{elec} + \Delta G_{solv} \quad (2)$$

Scoring functions continued to lack accuracy, due to the lack of information on covalent interactions, polarization and charged transfer in docking, which led to the development of quantum-mechanical based scoring functions. This new method was more accurate than the previous ones but with greater computational cost. Due to this last reason, hybrid quantum mechanical/molecular mechanics (QM/MM) approaches were developed (Steinmann, Olsson, and Ryde 2018) (Equation 3).

$$E_{bind} = E_{QM/MM} + \Delta G_{solv} \quad (3)$$

The second type of scoring functions consist of empirical SFs, where the binding affinity of a given complex is given by the sum of energetic factors, such as hydrogen bonds, hydrophobic effects, steric clashes, and others. Normally it is used as a training set to optimize the importance of the energetic factors considered, by comparing to known binding affinities, with the aid of linear regression analysis. An example of an empirical scoring function is X-score (Guedes, Pereira, and Dardenne 2018) and can be expressed as Equation 4.

$$E_{bind} = w_0 + w_1 \Delta G_{vdw} + w_2 \Delta G_{Hbond} + w_3 \Delta G_{rot} + w_4 \Delta G_{hydro} \quad (4)$$

Two main problems arise when using these types of scoring functions. The first one deals with the quantity and quality of the training data to optimize the complexes built. The other sets with the choosing of the most important energetic terms regarding the in-study complex. Normally this is performed by docking programs, such as AutoDock Vina (Trott and Olson 2010). This second type of scoring functions arises as a good option, due to its low computational power, a result of the simpler energy terms employed. They perform well binding affinities, ligand poses and virtual screening. Though they aren't suited in describing how binding affinities relate to the crystal structures (Y. Li et al, 2014).

The third type of scoring function consists of knowledge-based scoring functions. They derive the potential pairs between the protein and ligand, present in the complexes, from the three-dimensional structures. This is done based on the inverse Boltzmann statistical principle (Gohlke, Hendlich, and Klebe 2000). This method tries to infer the distance and interaction (distance-dependent potential of mean force) between two atoms, by the frequency of each atom pair. The computational flow of this technique starts with the classification of atoms in both the receptor and the ligand. After forming the atom pairwise, it is computed the density of each pair (ρij(r)), which are then compared with the previously computed reference density of each atom pair (ρ*ij). With this it is possible to compute the relative density (Equation 5), leading to the atom pairwise potential (Equation 6) and finally obtaining the sum of all atoms pairwise potentials (Equation 7).

$$g(r) = \frac{\rho_{ij}(r)}{\rho^*_{ij}} \quad (5)$$

$$w_{ij}(r) = -k_b T ln[g(r)] \quad (6)$$

$$E_{bind} = \sum_{i=1}^{L} \sum_{j=1}^{R} w_{ij}(r) \quad (7)$$

with *ij* respective to a determined atom pair, r the distance present between the atoms considered, k_b is the Boltzmann constant and T the absolute temperature.

This class of scoring functions are less computationally demanding, and the predictive accuracy is also compromised, having these advantages when compared to the previous two. However, their main disadvantage centers on the localization of the reference state, Because of this step back, some methods were developed to overcome it, like the volume factor correction method (Huang and Zou 2006) and the physics-based iterative method (Huang and Zou 2014). They all intend to infer the reference state based on the distribution of the atomic pairs. This strategy is better employed to know binding positions, rather than binding affinities.

The last from the scoring functions presented in this thesis are ML-based scoring functions (ML-SF). The main difference between these novel strategies and the classical ones, is that the first ones employ well-known mathematical functions, while ML-SF apply a variety of machine-learning algorithms, like support vector machine, random forest, neural network, or deep-learning, among others. On the other hand, the use of this kind of scoring improves the accuracy of the prediction of the binding poses, since they are used as a method of rescoring, after the use of more classical docking software (Khamis, Gomaa, and Ahmed 2015).

The computational protocol typically is constituted by three major moments: the Data selection, the Data representation, and the Feature selection. Based on the Data selection, a training set is built that will lead the model training (where it can be applied to any ML model). The Data representation helps build the validation set, creating a model selection, with the output from the training moment. Finally, the feature selection aids on the test set, where the binding is predicted, giving the information of the model selection. All of this leads to the final step of performance evaluation (J, Li, Fu, and Zhang 2019). With this, ML-SFs stand as the most promising scoring functions in predicting the best drug candidates to be synthesized for a given molecular target.

3.d. Machine Learning in drug discovery

As it was stated previously, with the advances in ML techniques, scoring functions based on these procedures have occupied a central role in the drug discovery pipeline (C, Shen et al, 2020). The use of this novel scoring function came with the need of processing bigger amounts of data, without losing the accuracy of the predictions made. On the other hand, the traditional development of novel therapeutics is very time consuming, very expensive and with low yield. Because of this, computer-aided drug design (CADD)

surged as one of the best techniques for the development of new drugs. CADD is target specific, automatic, structure-based, fast and presents low cost with a higher success rate. With this, ML methods have evolved mainly because of the need to improve the prediction ability of the binding affinities, when compared to traditional ones (Khamis, Gomaa, and Ahmed 2015). Moreover, ML techniques are advantageous because they can predict the binding affinity based on some features of the in-study complex, like geometric features, physical force field energy terms and pharmacophore features. The main goal with this type of experiment is to learn the relationship between these features and the corresponding experimentally measured binding affinities, after the training set of complex molecules. We can then use this learned function to predict binding affinities of novel complexes, between a known protein and other drugs (Mao et al, 2021). ML techniques also improve the learning of non-linear dependency, giving more accurate predictions of the binding affinity. In a general way, ML approaches use pattern recognition algorithms to infer mathematical relations between empirical values. Because these novel models use nonlinear methods, the resulting scoring function shows a better performance (C, Shen et al, 2020). In a general perspective, to develop a machine learning based scoring function first it is necessary to design a training and a test set, However, it is still unclear the best way to develop them. After this, it is generated a set of features that describe the interactions between the protein and ligand for the complexes presented in the training and test sets.

3.d.1. Supervised and unsupervised machine learning

The first distinction made when working with ML is between supervised learning and unsupervised learning ("Supervised vs, Unsupervised Learning: What's the Difference?" 2021). The main difference between these two families of algorithms is that in supervised ML, the training dataset is labelled, while in unsupervised learning it is not (Alloghani et al, 2020). The fact that the initial data is labelled enables the use of these datasets in training of predictive models, used to classify predicted outcomes. This type of ML can be used for both Classification and Regression problems. Because of these, this type of ML learning is normally more accurate. The difference between these two types of problems will be reviewed in the next topic. On the other hand, unsupervised learning takes advantage of the ML algorithms to analyze and cluster data sets that are not labelled. These operate by looking for hidden patterns in the studied data, without the need of prior human intervention or labelling. Nevertheless, human intervention is still

needed to validate the output results. These models can be used in clustering, association, or dimensionality reduction. Another major difference between these two types of algorithms is that in supervised learning, the researcher normally knows the type of output to expect, while in unsupervised, it is the machine itself that highlights what part of the data is different or interesting. Concerning the complexity, supervised methods are normally simpler and can be applied with Python or R programming. While, on the other hand, unsupervised learning asks for more complex tools due to the necessary size of the initial working datasets, to produce valuable results. Moreover, the main drawback of supervised learning is the time-consuming step of training the model and the necessity of a level of expertise in the labelling of the dataset and its interpretation. Moreover, considering unsupervised learning, these can give inaccurate results, obligating the human validation of the output (Alloghani et al, 2020).

3.d.2. Classification vs Regression

As mentioned previously, in supervised machine learning there are two subtypes of problems: Classification and Regression. When considering a classification problem, the ML algorithm tries to attribute each entry of the testset into a specific class or category. Because of this, it is applied to discrete data. On the other hand, in a Regression problem it is predicted a continuous value, based on given inputs, or descriptors, where the prediction is based. Machine learning methods include random forest (RF), support vector machine (SVM), and gaussian processes, among others. However, in this work, these are the methods that will be treated.

The Random Forest (RF) method employs a bagging and subset strategy to multiple decision trees (DTs). A decision tree is a decision support tool that implements a tree-like model of decisions, and they can answer sequential questions leading to a certain route of the tree and consequently to a specific answer (Breiman 2001). They have the advantages of being easy to interpret, perform well on large datasets and can handle both numerical and categorical data. On the other hand, RF consists of a collection of decision trees whose results are joined into one result. The novel approach here is that each tree is fed with a randomly sampled subset instead of the original dataset, resulting in a consensus score, when integrating the multiple outputs from DTs. Some examples of scoring functions including RFs are RF-Score, B2BScore and SFCscore. Moreover, this RF methodology was applied in already existing classical scoring functions, such with Vina, AffiScore and X-Score (C, Shen et al, 2020). Also, most RF-based SFs are thought of as

the binding affinities between ligands and proteins, not being so suitable for virtual screening.

A second machine learning set of algorithms are support vector machines (SVM) (Figure 9). They are a group of supervised learning methods, including linear, polynomial, sigmoid and radial basis function (RBF), that stand out because of their capacity to treat high-dimensional variables in small datasets (Cortes and Vapnik 1995). They can be used both for classification and regression, due to its derivative support vector regression (SVR). The biggest advantage of this method is the implementation of kernel nonlinear functions that can classify data that does not have a linear representation. Kernel functions are defined as a method to map originally nonlinear observations into a higher-dimensional space, where they can be separated and assessed (Afonja 2017). This aids a lot in computational processing, since it facilitates calculus. Because of these improvements, there are a variety of different implementations of these kinds of functions.

A third ML method applies gaussian processes to large datasets. These are described as a model capable of distributing the probability over a given group of functions. They are a collection of stochastic processes, characterized as being random variables with time and space indexed, that follow a multivariate normal distribution (Rasmussen and Williams 2005). This last concept is defined as a generalization of the one-dimensionality of a normal distribution to higher dimensions. This kind of method enables a good data treatment when the sampling is uneven or if the variables present a non-linear distribution (C, Shen et al, 2020).

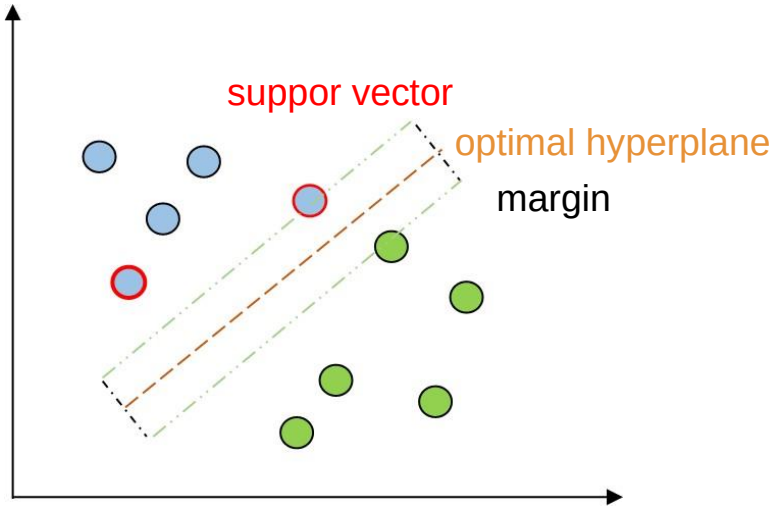


Figure 9. Simple representation of SVM, an optimal hyperplane separating two input groups, (based in (Badillo et al, 2020)).

3.d.3. Performance evaluation

Concerning the evaluation of these models, a variety of statistical tools are available. Here I will review the main statistical operators used in this project. For the assessment of the predictions made with a regression rationale, MAE (Mean Absolute Error) , MSE (Mean Square Error), R^2 (Coefficient of Determination), Pearson's Correlation Coefficient and Spearman's Correlation Coefficient can be used (Kumar and Dogra 2022). First, MAE (Equation 8) is one of the most used measures of accuracy when fitting ML models and measures the error between paired observations that explain the same phenomenon. In a prediction scenario is the error between the real value and the one predicted by the model.

$$MAE = \frac{\sum_{i=1}^{n} |y_{ob} - y_{pd}|}{n} = \frac{\sum_{i=1}^{n} |e^i|}{n} \quad (8)$$

Next, the MSE or Mean Square Error (Equation 9) is a measure based on MAE, with the difference between the real value and the predicted squared. This makes MSE more sensitive to outliers than MAE (Trevisan 2022) and turn bigger distances heavier when assessing the predictive power of the model. This presents a step back since the unit of MSE is also squared, losing its "real-life" meaning.

$$MSE = \frac{\sum_{i=1}^{n} (y_{ob} - y_{pd})^2}{n} \quad (9)$$

One of the most known and used measures in the determination of goodness of a specific predictive model is the Coefficient of Determination or R^2. This is computed as shown in Equation 10 and can vary between 0 and 1. The bigger the value of R^2 the better is the predictive model (Kumar and Dogra 2022).

$$R^2 = 1 - \frac{RMSE^2}{Var(y_{ob})} \quad (10)$$

with, $Var(y_{ob}) = \sum_{i=1}^{n} (y_{ob} - \underline{y}_{ob})^2$

and $\underline{y}_{ob}$ stands for the mean of observable values.

Because it surges in Equation 10, it is important to define RMSE also. This stands for Root Mean Square Error and corresponds to the square root of MSE (Equation 11). It is also known as the standard deviation of the prediction errors; it is scale dependent and it is a good measure of accuracy (Kumar and Dogra 2022).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_{ob} - y_{pd})^2}{n}} \quad (11)$$

Another metric used is the Pearson's Correlation Coefficient (*CCp*) (Equation 12.) and evaluates the variability between the observed and predicted values, or the consistency of the fitted model (Kumar and Dogra 2022).

$$CCp = \frac{\sum_{i=1}^{n} (y_{ob} - \underline{y}_{ob})(y_{pd} - \underline{y}_{pd})}{\sqrt{\sum_{i=1}^{n} (y_{ob} - \underline{y}_{ob})^2} \sqrt{\sum_{i=1}^{n} (y_{pd} - \underline{y}_{pd})^2}} \quad (12)$$

Lastly, it is also normally applied the Spearman's Correlation Coefficient (*CCs*) (Equation 13,), that stands as a non-parametric computation of *CCp*, meaning that the statistical operation doesn't take into consideration the characteristics of the sample, such as its parameters or if it is a qualitative or quantitative type of data (Hayes 2008). This measure indicates the strength and direction of the relationship between the observed and predicted values.

$$CCs = CCp\,(R(y_{ob}), R(y_{pd})) \quad (13)$$

with $R(y_{ob})$, as the rank value of observed values, and the $R(y_{pd})$ as the rank of the predicted scores.

On the other hand, when facing a classification problem, the statistical tools used to evaluate the build models are different, Here are included measures of Precision, Sensitivity (or Recall), Specificity, Accuracy and Cohen's Kappa, Precision (*PR*) is defined as the mathematical measure of the positive predictive value, given by Equation 14. It considers the number of false positives (*FP*), computing the negative tuples or the values incorrectly predicted by the model, and true positives (*TP*), which corresponds to the positive tuples, or the values correctly predicted by the model, This measure can vary between 0 and 1 (Kumar and Dogra 2022). Yet, tuples are defined as a finite ordered sequence of elements ("Tuple" 2002).

$$PR = \frac{TP}{TP + FP} \tag{14}$$

Secondly, Sensitivity (*Sen*) or recognition rate rationalizes the quantity of positive tuples per total predictions and is given by Equation 15. Instead of computing the false positives, it uses the number of false negatives, which are described as the positive tuples that were incorrectly predicted by the model (Kumar and Dogra 2022).

$$Sen = \frac{TP}{TP + FN} \tag{15}$$

A third measure used in classification problems is Specificity (*Spf*) or true negative rate. Here it is the proportion of negative tuples that are rightly predicted by the model that is evaluated as is mathematically represented as Equation 16.

$$Spf = \frac{TN}{TN + FP} \tag{16}$$

where *TN* is the number of true negatives.

A different type of evaluator is Accuracy (*Acc*), or how much the predicted values fall apart from the observed ones, in percentage, taking into consideration the acceptable error (ε) when in the problem assumptions. This is defined as Equation 17.

$$Acc = \frac{\sum_{i=1}^{n} Df_i}{n} \times 100 \tag{17}$$

where $Df_i = \{1, if\ |y_{ob} - y_{pd}| \leq \varepsilon \vee 0, otherwise\}$

Lastly, the Cohen's Kappa (κ) is very similar to the Accuracy with the exception that is normalized with the baseline of random events of the dataset in study. This is a better accuracy measure when in the presence of imbalanced classified datasets. This can be defined as in Equation 18.

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (18)$$

with, p_0 as the overall accuracy of the mode and p_e is the measure of the agreement between the values predicted by the model and the observed values, if happening by chance.

3.d.4. Applying Machine Learning

Besides, with the crescent inclusion of informatics in life sciences, a variety of on-line servers were made available by investigation groups all over the world (Kern, Fehlmann, and Keller 2020). One of these servers is Artificial Intelligence based Scoring Function Platform (ASFP) (http://cadd,zju,edu,cn/asfp/), developed by Zhang et al, (2021), including AffiScore , AutoDock, DPOCKET, DSX, NNscore and SMoG2016 (Zhang et al, 2021). It is important to explain what type of features each tool gives, Affiscore (Equation 19) is an energetic term type descriptor, based on empirical scoring functions (Jain 1996). It includes a total of thirteen descriptors. These are built based on a hydrophobic complementarity term, a polar term, and an unsatisfied polar term. The firstly mentioned polar term can be described as the weighted sum of the different types of bonds, including protein-ligand H-Bonds, protein-ligand salt-bridges and metal-ligand bonds. Onthe other hand, the unsatisfied polar term is respected to the weighted sum of the number of polar atoms that are not bound to any other atom, both related to their identification or charge.

$$\Delta G_{bind} = E_{hydrophobic} + E_{polar} + E_{unsat} \quad (19)$$

$$= w_1 N_{protein-ligand\ hydrophobic\ contacts} + w_2 N_{protein-ligand\ H-bonds}$$

$$+ w_1 N_{protein-ligand\ salt-bridges} + w_4 N_{metal-ligand\ bonds}$$

$$+ w_5 N_{interfacial\ unsatisfied\ polar\ atoms} +$$

$$w_6 N_{interfacial\ unsatisfied\ charged\ atoms}$$

The second tool analyzed was Autodock (Equation 20.) (Morris et al, 2009). This is a force field-based scoring function that includes six pairwise evaluations (V) and a term that estimates the conformational entropy, after binding (ΔSconf). These pairwise terms are built with contributions from dispersion/repulsion, hydrogen bonding, electrostatics and desolvation.

$$\Delta G = (V^{L-L}_{bound} - V^{L-L}_{unbound}) + (V^{P-P}_{bound} - V^{P-P}_{unbound}) + (V^{P-L}_{bound} - V^{P-L}_{unbound} + \Delta S_{conf}$$

with, $V = W_{elec}\sum_{i,j}\frac{q_i q_j}{e(r_{ij})r_{ij}}$ sol $\sum_{i,j} W_i W_j$ $(S^{(-r_{ij}^2/2\sigma^2)}$

$$+ W_{vdw}\sum_{i,j}(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}) + W_{hbond}\sum_{i,j}E(t)(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}) \quad (20)$$

Next it also used the features obtained with DSX (Neudert and Klebe 2011). This is a knowledge-based scoring function that includes pair potential distance dependent, potential of novel torsion angles and potentials based in the solvent accessible surface. This score is calculated for the whole complex (protein-ligand) with Equation 21.

$$score_{total} = W_p score_{pair} + w_t score_{tors} + w_s score_{SR} \quad (21)$$

with, $score_{pair} = \sum_{aigP}\sum_{ajgL} score^{DSX}_{pair}(c(ai,aj), r(ai,aj))$

$$score_{tors} = \sum \sum^{tors}_{b} {}^{score^{DSX}(t(T),\phi(t))}_{Tgb_{nT}}$$

$$score_{SR} = \sum_{agP} score^{DSX}_{SR}(c(a), SR(a)) + \sum_{agL} score^{DSX}_{SR}(c(a), SR(a))$$

The fourth type of tool utilized was NNScore (Durrant and Andrew McCammon 2010), which is a knowledge-based ML-based scoring function, This method uses the same atom types as Autodock. The tool used in this project corresponds to the latest version, NNscore2.0 (Durrant and McCammon 2011b), which includes 348 features. These descriptors include identifiers of close contacts between protein and ligand (with a maximum distance of 2,5Å apart), characterization of the electrostatic interaction between atom types, the frequency of each atom type in the ligand and the number of rotatable bonds present in the small molecule. This recent version of NNscore also includes terms prevenient from AutoDock Vina and from BINANA (Durrant and McCammon 2011a). Another computational metric used was SMoG2016 (Debroise, Shakhnovich, and Chéron 2017). This is also a hybrid scoring function, both knowledge-

based and empirical for protein-ligand interaction. It is composed by KBP2016 which is a knowledge-based potential and three additional terms representing repulsion effects, the number of rotatable bonds and ligand mass (Equation 22).

$$KBP = \sum_p \sum_l \Delta F(\sigma_p, \sigma_l) = \sum_p \sum_l -RTln\frac{f(\sigma_p,\sigma_l)}{f_{ref}}$$
$$SMoG2016 = KBP2016 + 0.535\sum_{i,j}\frac{A_{ij}}{r_{ij}^{12}} + 1.913 \times Rotor - 21.974ln(m_L) \quad (22)$$

Concerning the terms in the KBP equation, the p and l index stands for the atom types of protein and ligand, respectively. While, $f(\sigma_p, \sigma_l)$ stands for the frequency of contacts between the atoms in the pair considered, and this is compared with f_{ref} that stands for the same frequency in a reference state. Where A_{ij} stands por the potential term derived from the Lennard-Jones potential equation, computed with Amber van der Waals parameters. Yet R stands for the Boltzmann constant, T for absolute temperature and m_L for ligand mass (Debroise, Shakhnovich, and Chéron 2017).

The last descriptors used among the available by ASFP were the ones assessed by Dpocket (Schmidtke et al, 2010). These are presented as descriptors of the binding site, including measures of the ligand and binding pocket such as their respective volumes and flexibility, polarity score, measures of density concerning the content of alpha spheres and their respective characteristics, hydrophobicity, and charge scores, druggability among others.

Besides reviewing the features assessed in this study is also important to describe the R package caret, used to train, predict, and build ML based models, with the objective of study, interpret and design a possible novel scoring function, specific for the evaluation and characterization of the specific biological system inspected in this thesis, the receptor 5-HT2A when bound to ligands with different reported agonism. The caret (Classification and regression training) package is described as a set of functions for training and plotting classification and regression models (Kuhn 2008). It has in total 233 available models, that deeply facilitates the job of modelling for researchers. There are a variety of models for both classification and regression, but in this project, we only used models suited for both types of modelling. The ones applied include Bagged MARS (bagEarth), Gaussian Process with Polynomial Kernel (gaussprPoly), k-Nearest Neighbors (kknn), Random Forest (ranger, rf), Regularized Random Forest (RRF, RRFglobal), Subtractive Clustering and Fuzzy c-Means Rules (SBC), Support Vector Machines with Polynomial Kernel (svmPoly) and Support Vector Machines with Radial Basis Function Kernel (svmRadial,

svmRadialCost). All these models have different tuning parameters that enable a different train and consequently different predictions.

4. Materials

4.a. Homology Modelling Protocol

For the modelling experiments we used the software MODELLER. Firstly, the sequence of 5-HT2AR was screened with PSIPRED (Buchan et al, 2013) that enables the prediction of secondary structures, based on the sequence of the protein. First it was chosen the best template for each state of activation of the receptor. This was done with the inspection of each structure available in the PDB (Protein Data Bank) server (https://www,rcsb,org/). After this the structure 6WHA (Kim et al, 2020) was used as template for the modelling of the 5-HT2AR in its active form and its inactive form was modelled using the structure 6A93 (Kimura et al, 2019) as template. Additionally in the active model a restriction was added in the part of the sequence corresponding to the extracellular loop 2 (ECL2), for the MODELLER to build it as a helix, For both states 1000 models were built. These were evaluated with DOPE, Z-score, LGscore, MaxSub and the RMSD (root-mean-square deviation of atomic positions) between the models created and the structure that served as template. The representations of the structures were prepared with PyMol.

4.b. Docking Protocol

For the docking procedure we used the AutoDock4.0 software (Morris et al, 2009) and it followed a methodology like the one reported by Moreira et al. (2021) (Rosário-Ferreira et al. 2021). In total five different trials of docking were performed. The methodology was the same for every DOCK trial, For the active structures we selected the residues W151(3.28), I152(3.29), D155(3.32), V156(3.33), S159(3.36), L229(ECL2), V235(5.39), G238(5.42), S242(5.46), F332(6.44), W336(6.48), F339(6.51), F340(6.52), N343(6.55), V366(7.39) and G369(7.42). While for the inactive structures the flexible residues were W83, I84, D87, V88, S91, L161, V167, G170, S174, F221, W225, F228, F229, N232, V255, G258 (corresponding to the same Ballesteros Weinstein numbering) (Isberg et al, 2015), as flexible residues. The grid box used in all trials had 70 as dimension in all directions of the axis (x,y,z). The center of the box was located so that all the flexible residues would stay inside the same. For the docking computation it was applied a

Genetic Algorithm (Bursulaya et al, 2003) with 200 runs, with a population size of 200, evaluated with a maximum of 10000000 and with 27000 as the maximum number of generations. After the docking experiments ended, the 200 obtained conformations were first visualized with the AutoDock Tools visualizer. The resulting complexes were treated with python scripts, developed by the group. The assessment and evaluation of the resulting complexes were performed with the ASFP (http://cadd.zju.edu.cn/asfp/).

4.c. ML Protocol

After obtaining the respective table for each complex, grouped by DOCK, the results were summed in a single table (sum_dockings_features_final,xlsx). This table included all the features available from the following tools: AffiScore, AutoDock, DPOCKET, DSX, NNscore, SMoG2016, in addition to two more columns. One corresponds to the values of RMSD, measured by a self-made python script (Euclidian_distances.py, C2 in Additional information). These RMSD values were calculated compared with the crystallographic structure of 5-HT2AR in its active state (6WHA, PDB:ID), when bound with the agonist 25CN-NBOH. The second was a novel column created afterwards including a factorial classifier with four levels: 1, for complexes with RMSD lower or equal than five; 2, if the RMSD was between five and six; 3, for RMSD between six and eight; and 4 if RMSD was bigger than eight. In Rstudio, the variables with near zero variance were removed, which led to the complete elimination of columns containing NAs. Also, variables that exhibited very low variance were removed: number_of_interfacial_unsatisfied_charged_atoms, polar_componet_term, lat_OA and rot_bonds. With this the final working table was left with 61 features. The working dataset was partitioned with a p = 0,8, making the trainData with 800 rows and the testData with 199 rows. The X and Y data were saved, where the column referent to RMSD and RMSD2 were assigned as the y_data and y1_data, respectively, for both train and testData. Moreover, both datasets were preprocessed, being centered (by the mean subtraction) and scaled (divided by the standard deviation), transforming the variables. Finally, the previously stored Y columns were added to the tables. Before the modelling trials it was created a control dataset, with the repeatedcv method, three repeats with five numbers each. The models tested in this thesis were: avNNET, bagEarth, bayesglm, brnn, BstLm, cforest, ctree, ctree2, cubist, earth, enet, gaussprLinear, gaussprPoly, gaussprRadial, gcvEarth, glm, glmboost, glmnet, icr, kernelpls, kknn, knn, krlsRadial, lars, lasso, leapForward, leapSeq, lm, mlpWeightDecay, partDSA, pcaNNet, pcr,

plsRglm, ppr, qrf, ranger, rbfDDA, rf, ridge, rqlasso, rqnc, RRF, RRFlobal, rvmRadial, SBC, simples, spls, svmLinear2, svmPoly, svmRadial, svmRadialCost. The results were compared based on available performance evaluation metrics, different for the regression and for classification. Next, the best models were applied for both regression and classification. The regression problem was divided in two trials, the first considering 61 features, while the second trial also assessed dummy variables. On the other hand, for the classification problem, it was only considered the sixty-one features available. With this the complexes were classified with a first class (RMSD2), that aided on the second classification, between active and non-active structure. Yet, the results of both the classification problem and the most important features in the predictive model were plotted together to better visualize their co-dependence.

5. Results and Discussion

5.a. Modelling results

As stated in the protocol, the GPCRdb server (https://gpcrdb,org/) was inspected to gather the available models of 5-HT2A receptor. By the time it was assessed, there were five available crystallographic structures, with the following PDB-IDs: 6WGT (Kim et al, 2020) , 6WH4 (Kim et al, 2020), 6A94 (Kimura et al, 2019) , 6A93 (Kimura et al, 2019) and 6WHA (Kim et al, 2020) (Figure 10.). The first four being inactive structures and the fifth the only active structure of the in-study receptor, available in the entire database. This represents one of the main motivations to develop the presented predictive model.

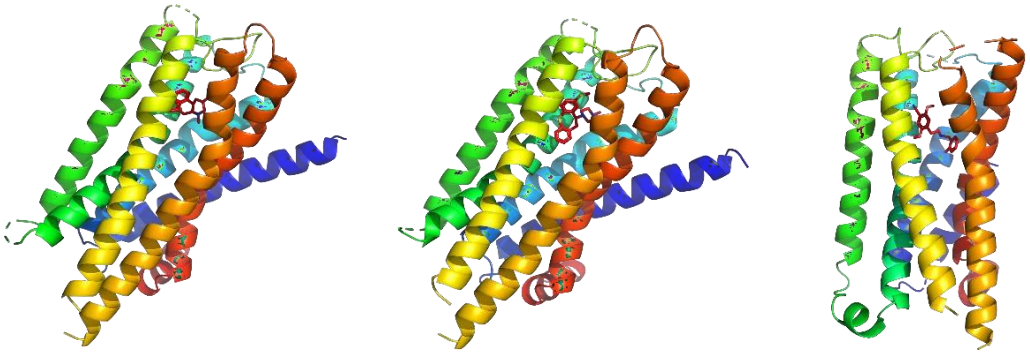


Figure 10. Structures obtained through crystallography of the receptor 5-HT2A, bound to different ligands (in red): 6WGT (with LSD), 6WHA (with 25CN-NBOH) and 6WH4 (with methiothepin) (in order).

Although being extremely helpful in understanding the receptor structure, due to the process of crystallization, some of this structural information was missing. For this reason, it was necessary to model both activation states of the 5-HT2AR. To build the most reliable model it was considered the most recent advances in the knowledge about the structure of this receptor. With that in mind, both models included the H8, an intracellular amphipathic helix (Zięba et al, 2021). After inspecting the results obtained with PSIPRED it was possible to point sequence areas that should be modelled as helix (Figure 11.). According to the Uniprot information three regions have been identified, comprising the 155-160 positions, here denominated as Region 1 with affinity for agonist binding, the Region 2 comprising the residues between 336-340, also reported as a possible intermediate for agonist binding and the Region 3 with a disordered presentation,

between the residues 450 and 471. Moreover, three domains are identified: Motif 1, between the residues 172-174 and described as a DRY motif, important in ligand induced conformational changes; the Motif 2 , which is a NPxxY motif also involved in conformational changes induced by the ligand and possibly in signaling associated alterations, defined from the residue 376 to the residue 380; and the Motif 3 between the residue 469 and 471, characterized as a PDZ domain, important in binding. Comparing these reports, it highlights the importance of the secondary structure for the receptor to interact or induce a certain response. During the Homology Modelling protocol, it was assured that all these secondary structures were modelled to obtain more trustworthy results. The built models were assessed with more than one score to assure an optimal structure to follow with, For the data treatment, all scores were summed in tables, as shown in Table 1. and 2. The first ten chosen models were selected based on the value of DOPE score, present in a crescent manner. For this score, the lower the value, the lower the energy value from the model created (M,-Y, Shen and Sali 2006), corresponding to a more stable theoretical model. The same rationale is adopted when evaluating the values in Z-score (Wiederstein and Sippl 2007). Concerning both LGscore and MaxSub, both are quality measures based on statistics, and so we looked for the biggest values (Cristobal et al, 2001), (Siew et al, 2000). Considered all these the model #811 was chosen as active model and the model #697 as the inactive model, marked in yellow on the tables below.

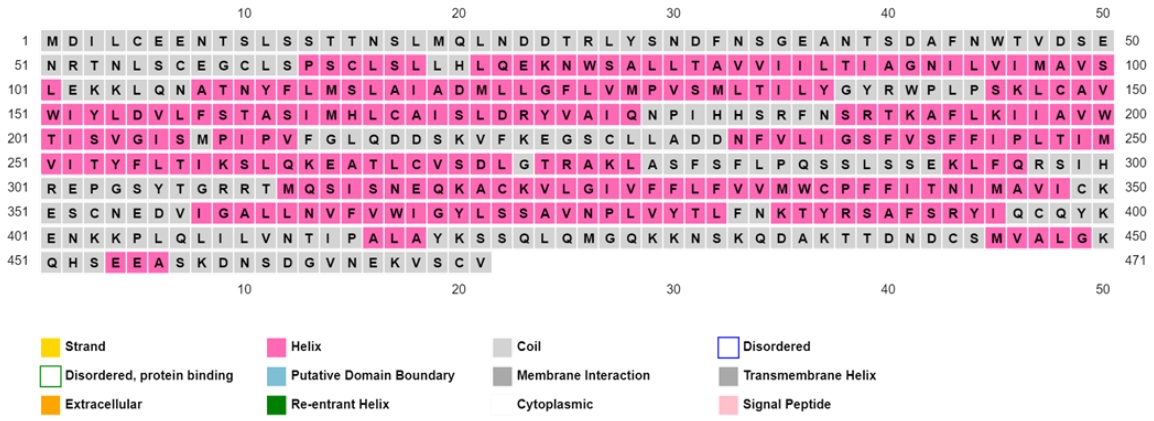


Figure 11, PSIPRED results, and prediction of secondary structures present in the 5-HT2A receptor

#model	molpdf	DOPE	RMSD w/6wha	Z-score	Lgscore	MaxSub
#811	1687,53	-46434,27	0,17	-2,57	9,34	-0,38
#423	1712,94	-46429,50	0,13	-2,55	9,22	-0,37
#109	1537,62	-46379,99	0,11	-2,66	9,21	-0,38
#929	1660,29	-46333,35	0,19	-2,72	9,27	-0,36
#78	162846	-46287,25	0,13	-2,66	9,12	-0,37
#794	1609,40	-46258,04	0,13	-2,75	9,28	-0,36
#467	1740,53	-46168,50	0,14	-2,77	9,34	-0,37
#822	1772,48	-46141,45	0,11	-2,35	9,25	-0,37
#542	1662,86	-46127,05	0,17	-2,76	9,21	-0,39
#813	1706,77	-46052,36	0,16	-2,62	9,20	-0,37

Table 1, Table with ten best active models for 5-HT2A assessed with the scores displayed,

model#	molpdf	DOPE	RMSD w/6a93	Zscore	Lgscore	MaxSub
#388	834,64	-40765,69	1,47	0,12	-4,08	9,11
#46	849,10	-40731,52	1,47	0,10	-4,14	9,13

#697	849,41	-40705,43	1,50	0,09	-4,19	9,21
#715	827,52	-40687,52	1,47	0,12	-4,07	9,11
#291	832,69	-40684,30	1,58	0,09	-4,18	9,07
#294	882,53	-40672,17	1,49	0,10	-4,16	9,24
#189	850,84	-40651,59	1,51	0,11	-4,06	9,29
#671	826,38	-40639,30	1,50	0,08	-4,07	9,18
#576	868,10	-40637,94	1,52	0,11	-4,12	9,15
#763	860,81	-40635,65	1,48	0,11	-4,03	9,25

Table 2, Table with ten best inactive models for 5-HT2A assessed with the scores displayed,

5.b. Docking results

The first dock (DOCK1) corresponds to the binding of 25CN-NBOH, which is an agonist of 5-HT2AR, and therefore docked against the model previously built of the active state of the receptor. The DOCK2 corresponds to the docking between LSD, which is a partial agonist of the in-study receptor and was bound to the inactive state of the built model. Moreover, the third dock (DOCK3) binds again LSD but, for a proof-of-concept rationale, to the structure with the PDB-ID 6WGT. The DOCK4 bound methiothepin, an inverse agonist of 5-HT2AR, to the PDB structure 6WH4, with the receptor in its inactive state. Finally, the DOCK5 also docks methiothepin but with the inactive built model. The protocol started with the upload of both the ligand and protein file, both in .pdb file format (Figure 12, 13 and 14).

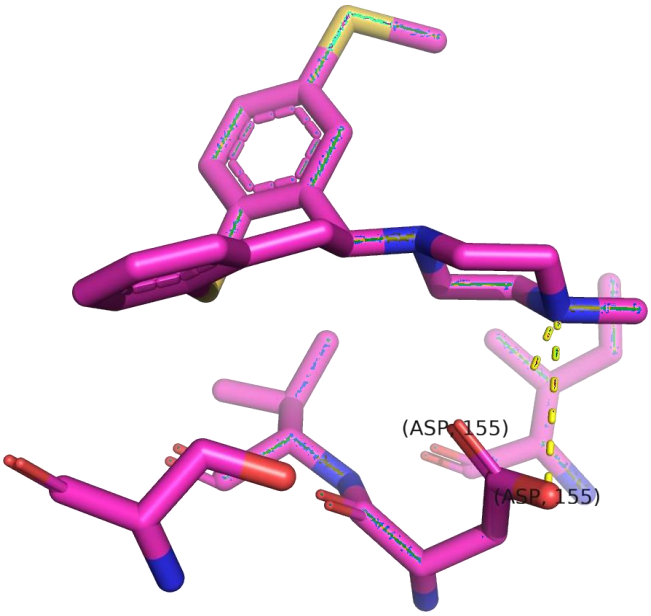


Figure 12. Representation of the binding pocket present in the crystallized structure 6WH4 (PDB,ID), with 5-HT2A in its inactive site, bound with methiothepin (pink). It also represented the residues (also in Pink) present within 5 Å from methiothepin, establishing polar contacts with the ligand, (in PyMOL).

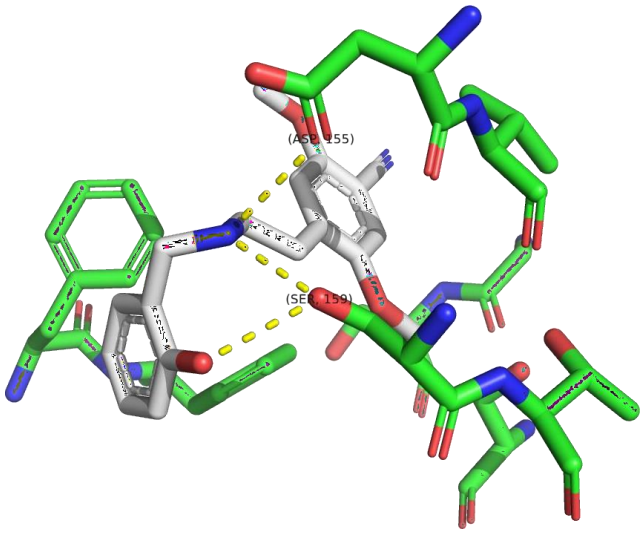


Figure 13. Representation of the binding pocket present in the crystallized structure 6WHA (PDB,ID), with 5-HT2A in its active site, bound with 25CN-NBOH (white). It also represented

the residues (green) present within 5Å from 25-CN-NBOH, establishing polar contacts (in PyMOL).

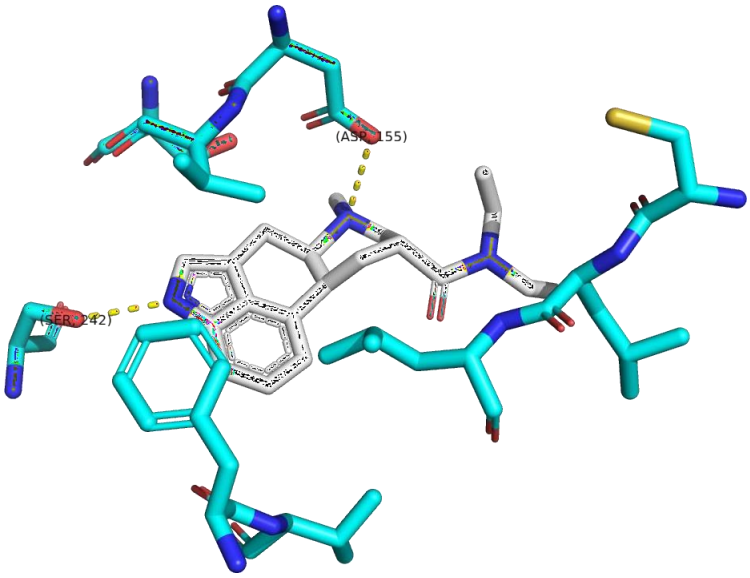


Figure 14. Representation of the binding pocket present in the crystallized structure 6WGT (PDB,ID), with 5-HT2A in its active site, bound with LSD (white). It also represented the residues (blue) present within 5 Å from LSD, establishing polar contacts (in yellow) with PyMOL.

After the upload of the ligand, it was necessary to determine the number of rotatable bonds present in the molecule. This is an important step to obtain results nearer to reality, concerning the ligand pose after the binding. A novel step in up-coming protocols would be to select individual bonds as routable and assess the different poses of both the ligand and protein, and how they influence the final complex conformation. These would give important data for a more sensitive drug design. For all the ligands present in this study, all the rotatable bonds were kept on the compounds (Figure 15.). Afterwards, it was necessary to prepare the uploaded macromolecule. First, all the residues of the protein were selected for the addition of hydrogens to the available spots. This step is important for the correct attribution of Gasteiger partial charges (Morris and Lim-Wilby 2008). Then it was necessary to select the residues that were thought to be flexible. These residues were chosen based on the structural information provided by Kimura et al (2019) and Kim et al(2020) (Kimura et al, 2019), (Kim et al, 2020). The complexes, containing the docking information were prepared by adding the rigid struture of the receptor with the docking conformation file, containing the positional information of the flexible residues

and the pose of the docked ligand, with a self-designed python script (cut_66.py, C.S.1 in additional information). With the complexes built, it was necessary to rank them by similarity with experimentally reported structures. For this it was calculated the RMSD of each complex, in comparison with the PDB structure 6WHA, the receptor in its active state. This step was carried out with a self-made python script (euclidian_distance.py, C.S.2 in additional information), which facilitated the process of attributing this value to the 200 complexes created, per DOCK. Because it was intended to gather the biggest number of descriptors about the interaction in this study, each complex was edited to obtain two files, one with the receptor (with its bound conformation) and the other with the positional information of the ligand. This again was facilitated by the development of a python script (scissors.py, C.S.3 in additional information). This was an essential step because we needed to load the complexes in the server provided by Zhang et al. (2021) available at http://cadd,zju,edu,cn/asfp/ (Zhang et al. 2021). The complexes were then evaluated by some tools present on the online server: AffiScore, AutoDock, DPOCKET, DSX, NNscore, and SMoG2016 (Zhang et al. 2021). These correspond to the tools available without license restriction.

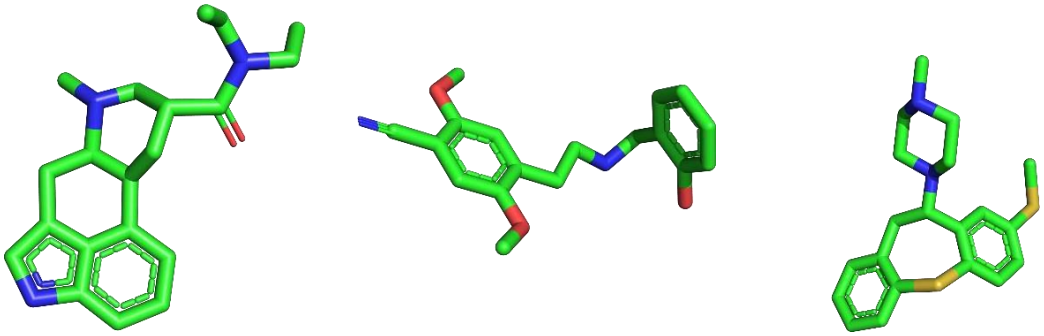


Figure 15. Ligand structures: LSD, 25CN-NBOH and methiothepin (in order) (in PyMOL).

5.c. ML results

Before the modelling trial per se, it was necessary to visualize which features were discrete or continuous. This was an important step to choose the features that should be transformed into dummy variables. With this it was decided to make two trials, one without these categorical variables and the other with these features as dummy variables. For that, all the features were plotted in a Scatterplot (SI.2.). By visual inspection of the mentioned graph, the features “number_of_interfacial_unsatisfied_charged_atoms”,

"Polar_component_Term", "lat_OA" and "rot_bonds" were eliminated, for the first trial. To choose the best model for prediction, a first evaluation with the models described in ML Protocol was conducted. The results from these first trials, for both the train and testData are represented in Table 3. For this exploratory first trial it was employed the "boot632" method of resampling, which is an improved cross-validation method (Efron 1983). When analyzing the following table, we first inspected the values of R^2. Moreover $R^2 > 0,90$ was selected as the criteria to choose the best models to continue the modeling trials. The models selected are marked as yellow in Table 3, being bagEarth, gausprPoly, kknn, qrf, ranger, rf, RRF, RRFglobal, SBC, svmPoly, svmRadial and svmRadialCost This first selection of the models to use, was the first step in tuning the predictive model,

TRAIN				**TEST**		
Caret_model	**RMSE**	**R2**	**MAE**	**RMSE**	**R2**	**MAE**
bagEarth	1,44	0,94	1,00	1,74	0,92	1,10
gausprPoly	1,63	0,92	1,16	1,76	0,91	1,20
gausprRadial	2,00	0,89	1,23	2,39	0,86	1,54
kknn	1,72	0,91	0,97	1,52	0,93	0,98
qrf	1,18	0,96	0,79	1,34	0,95	0,89
ranger	1,07	0,97	0,76	1,39	0,94	0,97
rf	1,22	0,96	0,83	1,40	0,94	0,94

RRF	1,24	0,95	0,84	1,87	0,94	0,96
RRFglobal	1,23	0,96	0,83	1,34	0,95	0,92
SBC	1,53	0,93	0,89	1,66	0,92	1,03
svmPoly	2,30	0,85	1,55	1,80	0,91	1,24
svmRadial	1,59	0,93	1,04	1,88	0,90	1,30
svmRadialCost	1,55	0,93	1,03	1,90	0,90	1,32

Table 3. Statistical metrics on the first trials of training and prediction data on the enunciated models.

The chosen models were then used to train and test the same dataset. However, for the results presented in Table 4., the method used for resampling the control data was "repeatedcv", which stands for repeated random sub-sampling validation, meaning that the division between train and test is done in a random fashion (Efron 1983; "Cross- Validation (statistics)" 2003). In order to best choose a possible model for prediction, the values were statistical assessed with mean absolute error (MAE), mean square error (MSE), coefficient of determination (R2), root mean square error (RMSE), Pearson's correlation coefficient (CCp) and Spearman's correlation coefficient (CCs). The results are presented in Table 4, for the computation without dummy variables and the data in Table 5, counts with this type of variable.

Model	MAE	MSE	RMSE	R2	CCp	CCs
bagEarth	0,83	0,86	1,40	0,94	0,97	**0,94**
gaussprPoly	0,67	0,86	0,93	0,97	0,99	**0,92**
kknn	0,78	1,94	1,39	0,94	0,97	**0,94**
qrf	0,57	1,04	1,02	0,97	0,98	**0,96**
ranger	0,57	1,04	1,02	0,97	0,99	**0,96**
rf	0,60	1,10	1,05	0,97	0,98	**0,96**
RRF	0,65	1,33	1,15	0,96	0,98	**0,96**
RRFglobal	0,60	1,10	1,05	0,97	0,98	**0,96**
svmPoly	0,81	1,46	1,21	0,95	0,98	**0,91**
svmRadial	0,81	1,44	1,20	0,96	0,98	**0,89**
svmRadialCost	**0,81**	**1,44**	**1,20**	**0,96**	**0,98**	**0,87**

Table 4. Statistical assessment on the predictions made by the predictive regression models, in the first trial.

Firstly, inspecting the R^2 value for each model, the RF based models were the ones presenting the best correlation between predicted and real values. Secondly, when looking at the lowest value of MAE, *qrf* surges as the most accurate model. To better understand which were the features with biggest importance in the modelling processing, it was plotted the *varImp* information, a function available in the caret package, as in Figure 16. Yet, it was only plotted the importance of the features with more that 50 of importance in the model building, to better visualize the data. The most important feature

reported was vina_guass_2, computed by the NNScore2.0 algorithm. According to the authors, Trott et al, (Trott and Olson 2010), this is computed as presented in Equation 23, and it represents the second steric term available in the Autodock Vina algorithm.

gauss1(d) =e(-d/0,5A)2 (23)

with, repulsion(d) = d2, if d < 0 ∨ 0, if d ≥ 0 , where d defines the distance between the atoms of the pair being assessed,

Secondly, SMoG2016_KBP2016 was evaluated as the second most important feature on the prediction procedure. This function computes the interaction between protein and ligand atom types, calculating the binding energy characteristic of the binding. This can be used to infer enthalpic and entropic factors, by relation with repulsion effects, the number of rotatable bonds and ligand mass.

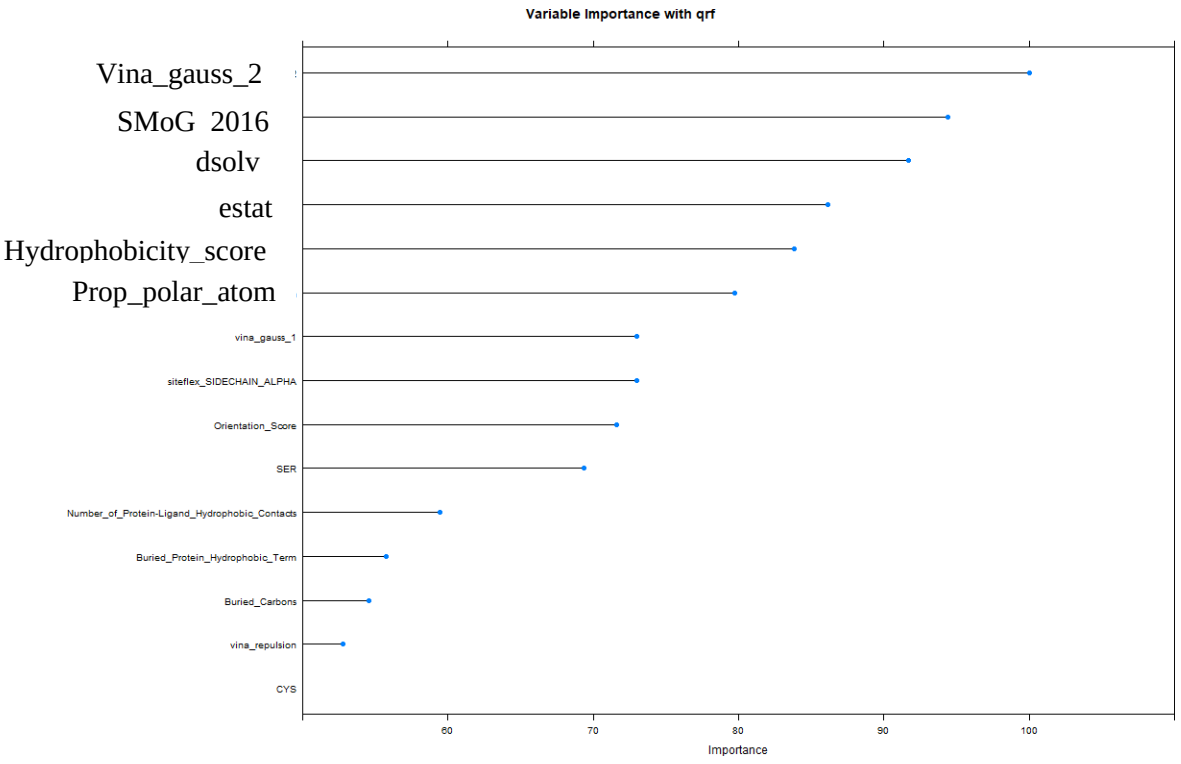


Figure 16. Features with more than sixty of importance, when computing the prediction model with *qrf*.

The next two features considered valuable in the prediction were two variables given by AutoDock, estat and dsolv. This represents the electrostatic and desolvation contribution respectively, as shown in Equation 9. Moreover, characteristics from the binding pocket

were also important in the prediction of the RMSD presented by the complexes. These include the proportion of polar atoms (prop_polar_atm), the hydrophobicity score, that takes into consideration the mean hydrophobicity score for all residues presented in the binding pocket and the number of serine residues present in the binding site, was also considered important for the RMSD prediction. Lastly, with a smaller importance, the orientation_score, the siteflex_sidechain_alpha and vina_gauss_1 were also taken into consideration. Considering the other best predictive model, ranger, it exhibits a bigger Pearson's correlation coefficient. When inspecting the variables deemed as most important (Figure 17.), in this case the model only considered the number of serine residues present in the binding pocket, as influential in the RMSD prediction.

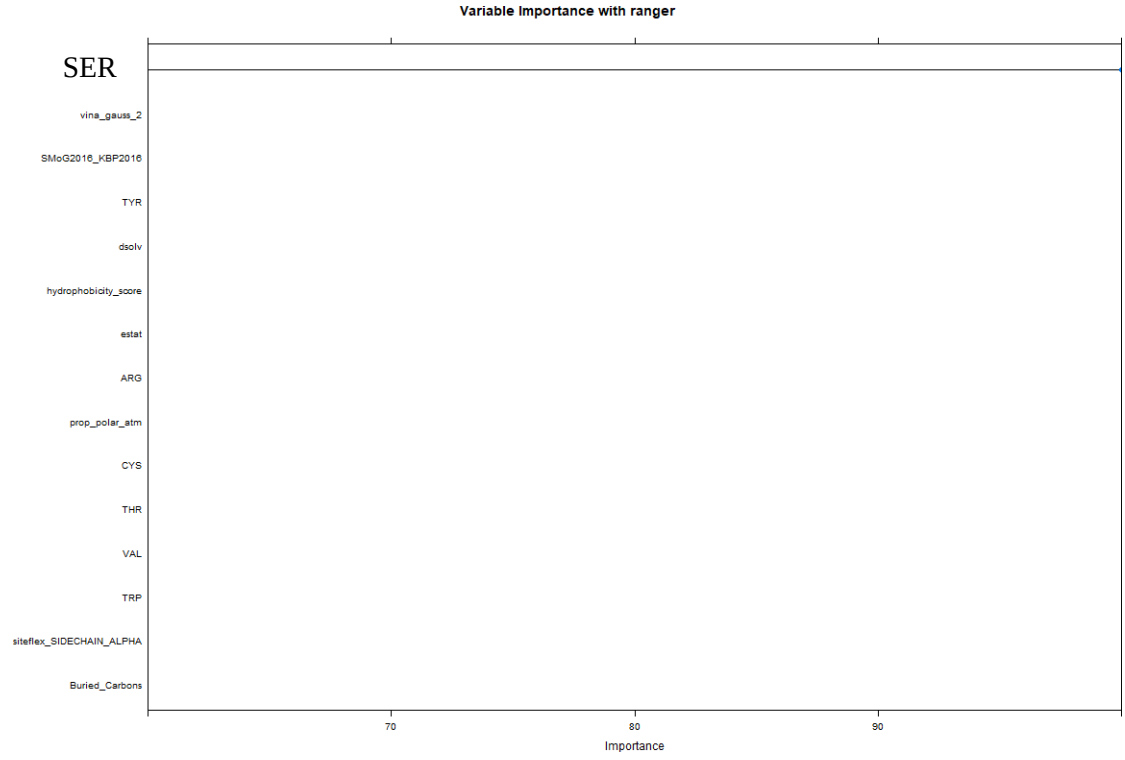


Figure 17, Features with more than sixty of importance, when computing the prediction model with *ranger*.

Although showing a bigger Pearson's correlation coefficient, the fact that ranger only took into consideration the number of Serines present in the active site diminishes the predictive power of the model. Though, it indicates that the ratio of certain residues present in active sites is, probably, the most decisive feature when trying to classify protein-ligand interaction. Meaning that these can become good measures in predicting a possible interaction between the receptor in study and small molecules.

To obtain a visual inspection on the predictive power of the model, the predicted values were plotted against the real, observed values, for both models (Figure 18 and 19).

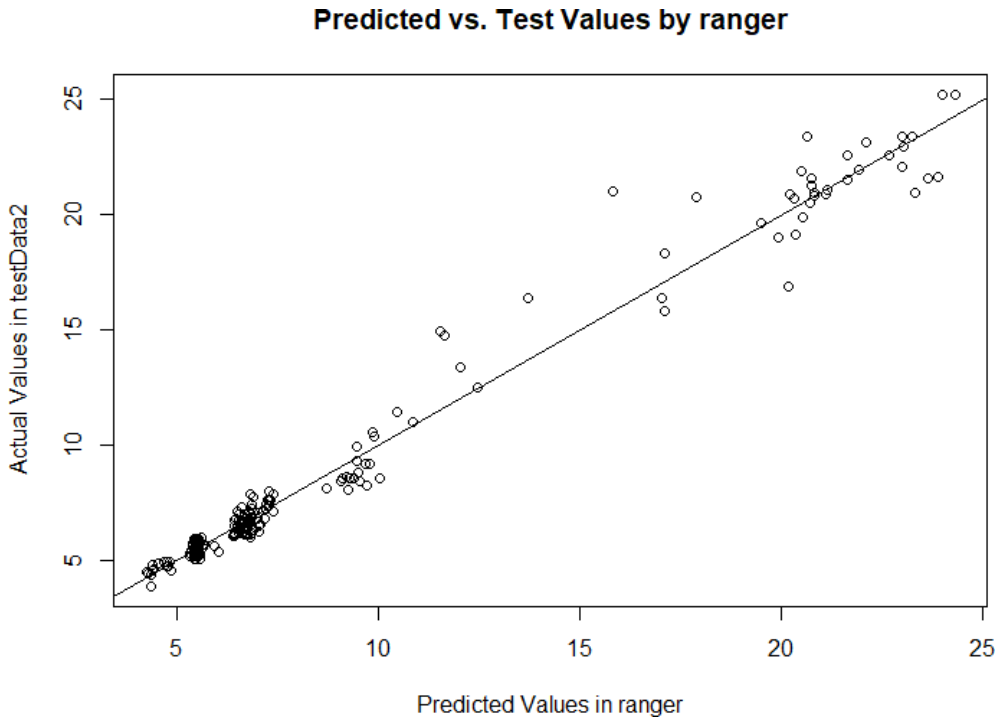


Figure 18. Predicted values plotted against real values, in the first trial for the regression problem, with the *ranger* model.

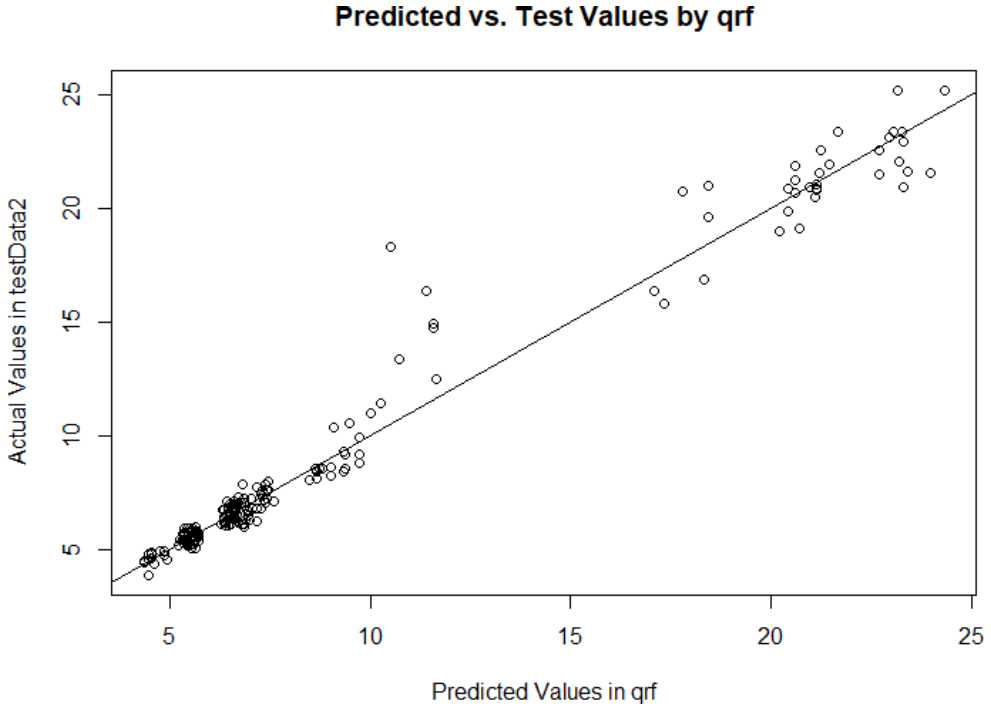


Figure 19. Predicted values plotted against real values, in the first trial for the regressionproblem, with the *qrf* model.

When analyzing the results from the second trial, with the computed dummies, the model ranger presents the biggest R^2 value. When assessing the other predictive evaluators, such as MAE, MSE and RMSE, ranger also exhibits a lesser error rate when compared to other good models, such as qrf (Table 5.).

Model	MAE	MSE	RMSE	R2	CCp	CCs
bagEarth	0,79	1,71	1,31	0,95	0,97	0,92
gaussprPoly	0,68	1,18	1,09	0,97	0,98	0,92
kknn	0,78	1,63	1,27	0,95	0,98	0,88
qrf	0,53	0,99	0,99	0,97	0,99	0,96
ranger	0,50	0,72	0,85	0,98	0,99	0,96
rf	0,56	1,00	1,00	0,97	0,99	0,96
RRF	0,57	0,99	0,99	0,97	0,99	0,96
RRFglobal	0,58	1,03	1,02	0,97	0,98	0,96
svmPoly	0,72	1,26	1,12	0,96	0,98	0,90
svmRadial	0,78	1,42	1,19	0,96	0,98	0,89
svmRadialC	0,78	1,40	1,18	0,96	0,98	0,89

Table 5. Statistical assessment on the predictions made by the predictive regression models, in the second trial with dummies.

Similarly, to the first trial, without the dummy variables, we plotted the most important variables in the construction of the predictive model. Assessing the qrf model first, it is notorious that mostly the same features, as in the first trial, were selected. Because the dummy variables transform discrete values into the probability of assuming the same value, when compared to the other assessments relative to the same variable, the TYR.0 was also plotted in this graph. Meaning that the model used the build dummy variable, relative to the presence of tyrosines, to predict RMSD With this said, vina_gauss_2, SMoG2016_KBP2016, estat, dsolv, hydrophobicity_score and prop_polar_atm continue to be the most important features in the prediction of the complexes' RMSD (Figure 20).

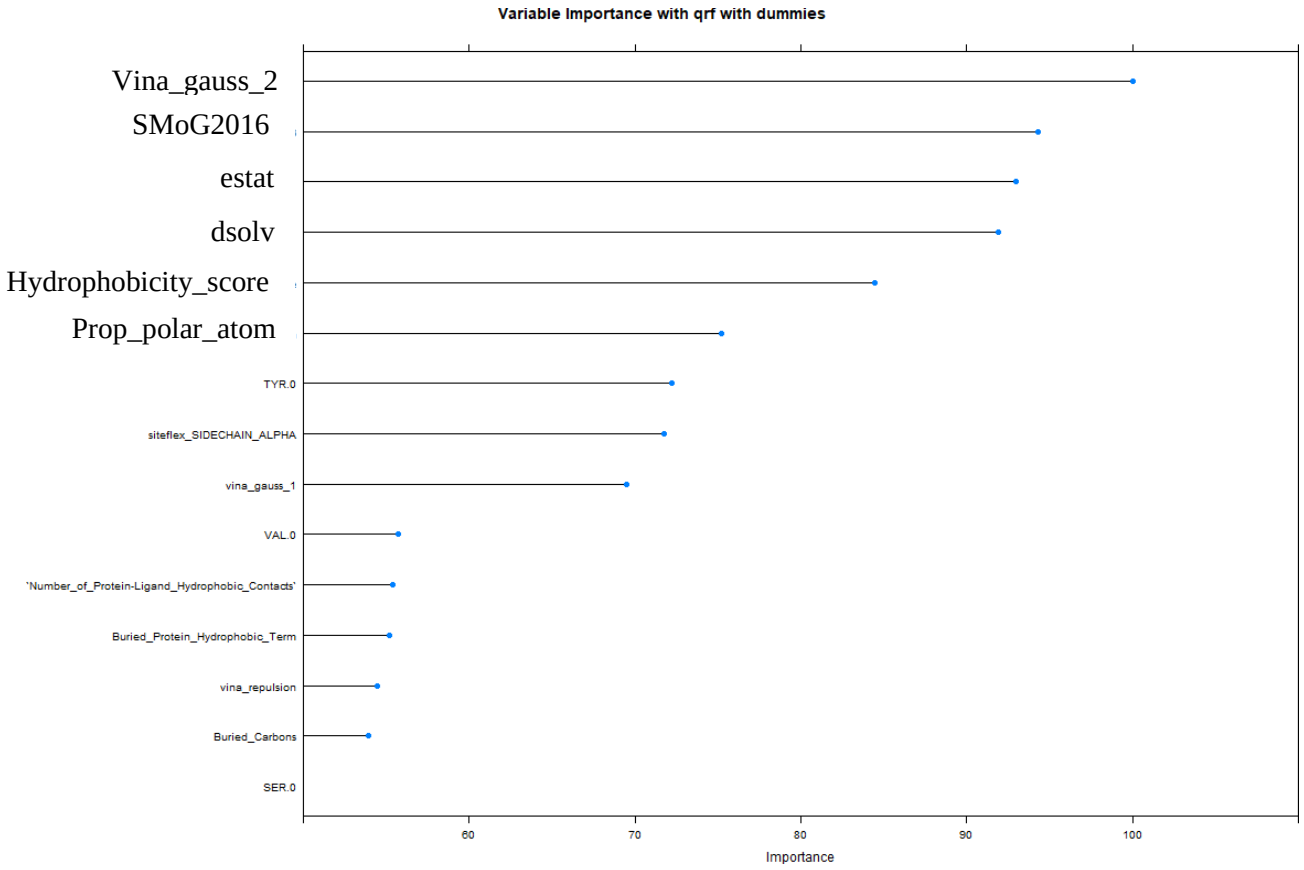


Figure 20. Features with more than sixty of importance, when computing the prediction model with *qrf*, during the second trial, with dummy variables.

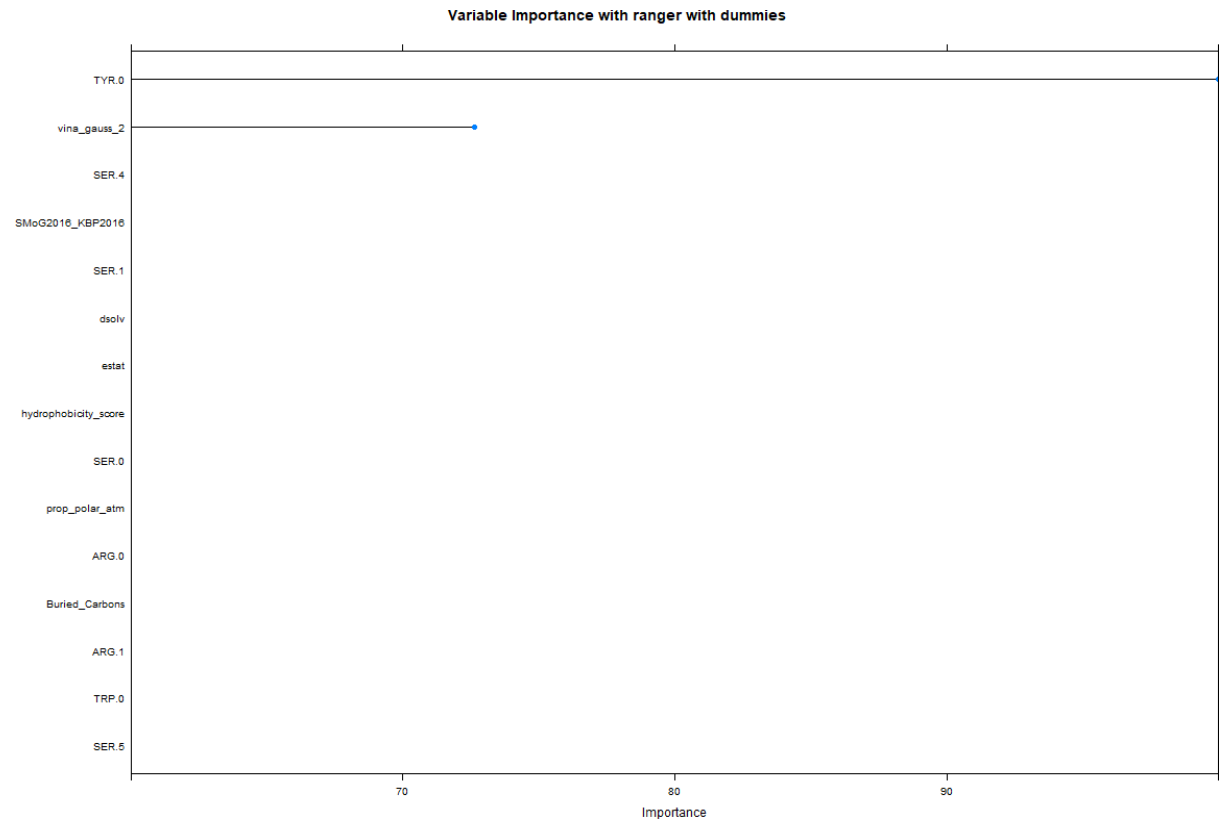


Figure 21. Features assessed with bigger importance, when computing the prediction model with *ranger,* during the second trial, with dummy variables.

The same happens for the ranger model. Though this last model attributes more importance to the number of tyrosines and serines present in the binding pocket (Figure 21.). This supports the previous stated hypothesis, stating that the number of specific residues in the active site is a good predictor on where the receptor might bind, and how it can influence the activation state of the receptor, after the binding with a family of ligands. This can give some leads on possible ligands' pharmacophores, that would potentiate or induce a conformational shift on areas of the receptor with a bigger ratio of specific residues, such as serine or tyrosine.

On the other hand, from a classification perspective, five models arise with an accuracy of 1: bagEarth, ranger, rf, RRF and RRFglobal. For the first trial on the class prediction, it was considered four classes: 1, 2, 3, 4. Since the RMSD was calculated with the structure 6WHA as reference, the values obtained can be considered as a measure of the "activeness" of the receptor, where complexes with a RMSD lower that 5 Å are included in class 1, complexes reported with a RMSD lower than 6 Å are stored as class 2, if the RMSD is between 6 Å and 8 Å it belongs to class 3 and, lastly, if it shows a RMSD bigger than 8 Å it is considered a class 4 complex. With this said, complexes classified

with class 1 are considered active, and the more they fall into class 4, the more inactive the structure should be (Figure 22.).

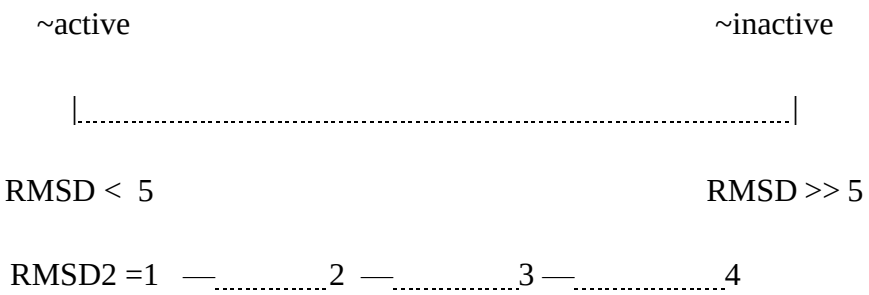


Figure 22. Scheme of the classification assessment.

On the other hand, the second trial of classification takes advantage from the previously created classes, and inserts a new class: 1, if it is an Active structure, only considered if the complex is located within class 1, and 0, if it is a Non-active structure, considered if the complex is evaluated with a class different than 1.

To get a better sense on how the different variables influence the class attribution to each complex, the features defined as most important in the modelling procedure were plotted against the classes available. To get a better visualization, values were normalized. In Figure 22, the comparison is made against the four classes created, while in Figure 23,the features are presented against the predicted state. By visual inspection of Figure 23, it is possible to conclude that complexes classified as “Active” are more concentrated near Y (value) = 0. The same happens for complexes inserted in class 1 and class 2. From this, it is notorious the sensitivity of the predictive model presented in assessing the activation state from the studied complexes. Although this can also mean that the model needsbetter tuning, since it can be biased toward the Active class. Moreover, this can help building thresholds for the features assessed in future predictive models, setting boundaries for the same, to choose best possible complexes, when studying a big starting set of structures.

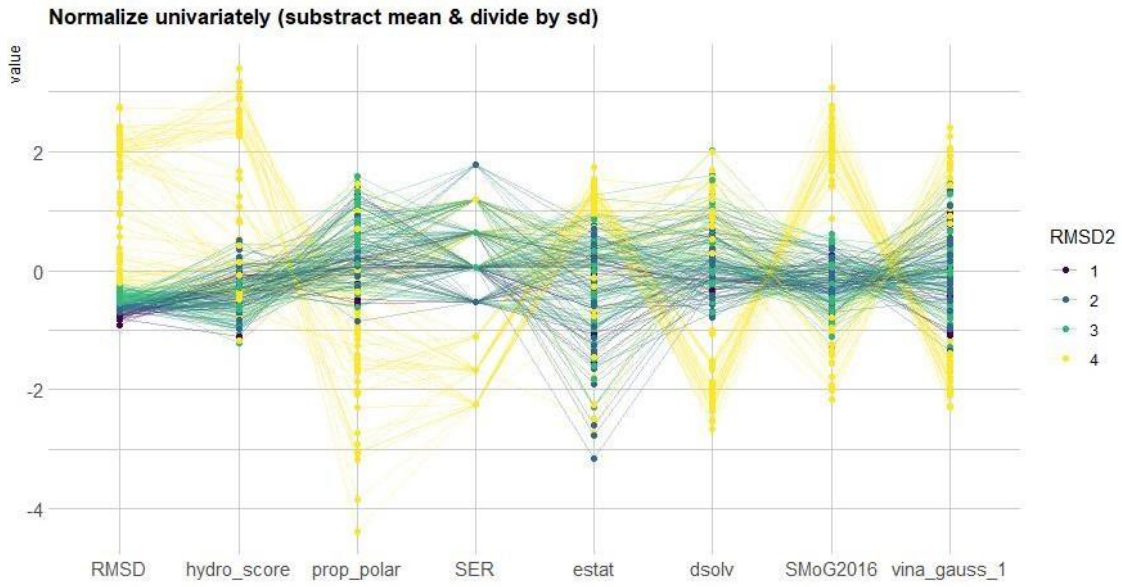


Figure 22. Parallel coordinates chart between the features with bigger importance in the predictive model and the class RMSD2.

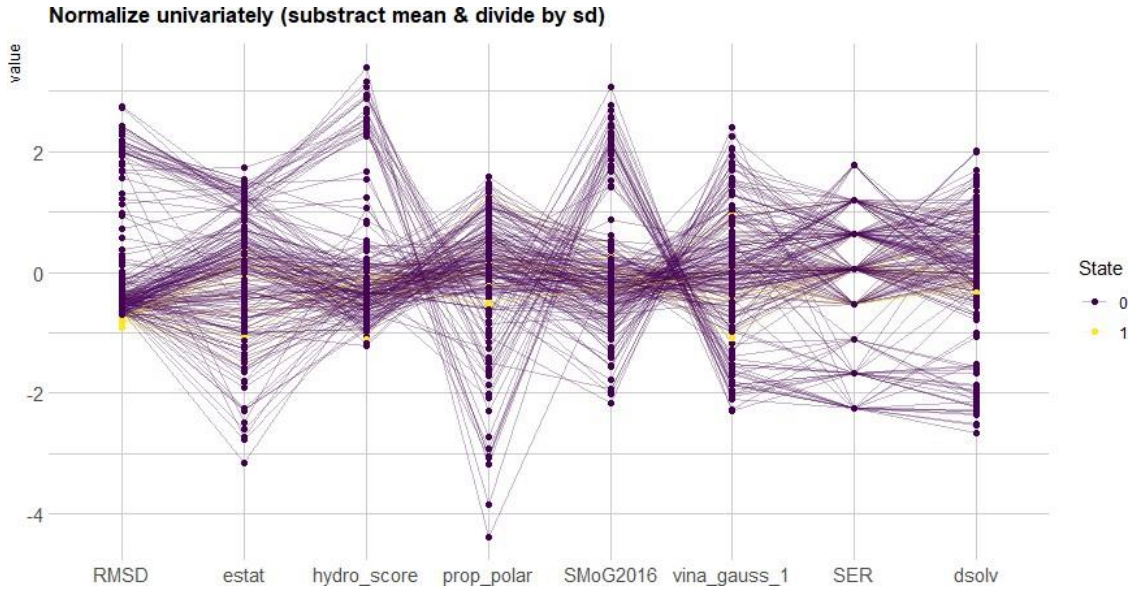


Figure 23. Parallel coordinates chart between the features with bigger importance in the predictive model and the class State.

To validate the predictive RMSD values by the best model, qrf, it was visually inspected the complex labelled with the lowest value of RMSD, with a value of 4.42. This complex corresponded to the one produced during DOCK2, indexed as complex #41. This second trial in the docking experiment was made between LSD, reported as a partial agonist of 5-

HT2A (Kim et al, 2020), and the self-made model of the inactive state of the in-study receptor. With the aid of PyMOL software it was inspected the residues that had their flexible chain within 5 Å from the ligand. After this the polar contacts between these sidechains and the small molecule were represented, shown in yellow in Figure 20. The same protocol was followed with the 6WHA structure, being this the reference structure for the RMSD's observable values computation. Both binding pockets (BP) are represented in Figure 24, where the BP from complex #41 is colored in green, while 6WHA is painted in blue. By labelling the interacting residues it is concluded that in both structures the residue ASP155(3.32) (in the inactive structure corresponding to ASP87) interacts with both ligands, by its acidic carboxylic group. This highlights the conservative action of this specific residue on the binding between 5-HT2A and small ligands. Moreover, in the complex #41 BP the residue TYR259 was reported as an interacting residue, although not included in the flexible residues list while in the Docking experiments, indicating a possible role for this residue in the interaction of 5- HT2A with LSD. Though, when binding 25CN-NBOH, it is the residue SER159 that pulls the ligand into the binding pocket, already reported as a flexible residue. On the other hand, this result validates the predictive capacity of qrf-based model in computing a trustable RMSD value. Because, although in DOCK2 it was docked an inactive structure of the receptor, after the binding (docking) with a partial agonist, the model correctly labelled the complex #41 as active (RMSD2 = 1, RMSD = 4.42). Moreover, it is important to highlight that this classification was based on physical features used as predictors.

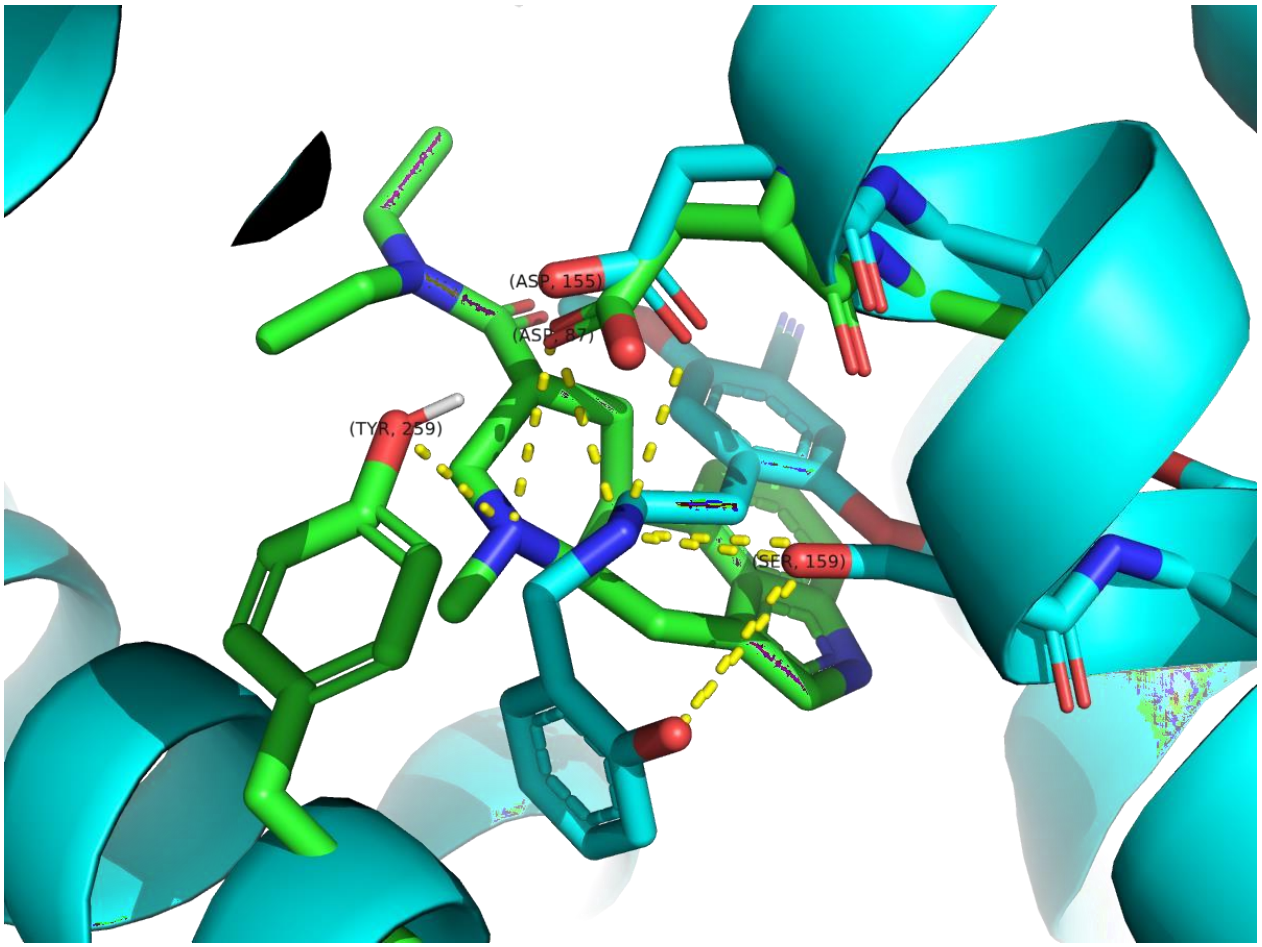


Figure 24. Comparison between the binding pocket of complex #41 (in green), calculated in DOCK2, with the binding pocket of the 6WHA structure (in blue), used as reference in the calculus of the observable RMSD, (in PyMOL).

6. Conclusions and Future Perspectives

With this project it was possible the solidification of some knowledge and the initial setting for the development of a predictive classification model on the state of activeness of the 5-HT2A receptor, when bound to small ligands, such as the psychedelics reviewed in this work. Firstly, with the performance evaluation of the used models, the ones based in RF, such as *ranger* and *qrf*, exhibited the most accurate results, when comparing the predictive values with the observed ones. Moreover, and the most valuable information produced in this work, it was possible to highlight the most important features used by the ML algorithms when predicting the RMSD values of the tested complexes. These include *vina_gauss_2*, a repulsion/steric term; *SMoG2016*, that represents the binding energy between protein and ligand and considers the interaction between the protein and ligand atoms, by computing enthalpic and entropic factors, related with repulsion effects, rotatable bonds and ligand mass; *estat* and *dsolv*, defined as measures of electrostatic and desolvation contributions, respectively; *prop_polar_atm*, presented as a value of the proportion of polar atoms in the binding pocket; hydrophobicity_score, resulting from the mean assessed hydrophobicity score of each residue present in the active site of the protein, and considered as a global measure of the pocket; and finally the number of certain types of residues, like serine (SER) and tyrosine (TYR), present in the interaction site. All these points to the importance of hindrance interactions between ligand and receptor in the binding decision. Moreover, these results enable the design of predictive ML models, computing these features, delimited by defined thresholds. This asks for a better characterization of these features and how they are influenced by other molecular characteristics, such as type of atoms/residues, kinetics, and types of intervenient bonds. Yet, more refined computational and mathematical tools are necessary, such as the exploration of other ML algorithms, such as Deep Neural Networks (DNN) and Artificial Neural Networks (ANN), or the computation of quantum mechanical factors. This line of methodology presents itself as a valuable endeavor, since a novel predictive model on the activeness state of the 5HT2A receptor can help design novel drugs, based on their physical characteristics, without the necessity of synthesizing them and test them in living systems, lowering both the cost and time duration of future projects. Yet, this rationale in drug design and development enables a new categorization of novel ligands, based on

their agonism with the studied receptor. This would be an approach to apply in other family receptors.

On the other hand, as it was highlighted in this thesis, a lot of novel work has been developed concerning the therapeutic effect of psychedelics. It is mandatory a functional reassessment of these substances, taking into consideration these recent findings (McCartney, McGovern, and De Foe 2022).

At the same time, the advances in computational techniques enabled a better understanding of structural modulations in the receptors with affinity for these compounds. This brings more unbiased and quantitative data on how these drugs might function in the human brain, opening the opportunity for a restructuralization on their usability, and consequently, how they can be used by our society (Ballentine, Friedman, and Bzdok 2022). Moreover, *in silico* techniques enable an atomistic level of study, highlighting the conformational shifts that occur after the ligand binding, on specific motifs present in the receptor. Yet, it is intended to relate this receptor conformation alterations with well-known pharmacophores included in the structure of classical psychedelics or novel compounds, designed with the aid of these atomistic information. This level of sensitivity aids in the construction of a novel classification system of these drugs, being more specific on how they act and influence certain cell and/or system responses.

In order to relate the reported mechanical alterations, present in the receptors with the activation of certain cell signaling pathways, it is necessary to couple the referenced *in silico* techniques with both *in vitro* and *in vivo* methodology. To validate the results brought by computational tools, the use of cell and rodent systems is mandatory when assessing the activity of novel drugs. One of the future goals is the development of 5-HT2AR mutants, based on the brought structural information and their expression in living systems. This would allow the correlation between specific motifs movements in the receptor, activated by a ligand interaction, with the activation of signaling pathways, propelled by the recruitment of a molecular intermediate that has affinity for the complex formed (biased signaling). This can lead to the dissection of the effects reported by psychedelics. Yet, it is necessary a better understanding on how the activity of these receptors lead to changes in the connectivity and communication between neurons, leading to mood and cognitive alterations. The bigger picture of how psychedelics function as a neuromodulator is still very shady. Studies on receptor desensitization by the action of agonists or antagonists with 5-HT receptor family are mandatory to relate the action of these receptors when interacting with hallucinogens. Moreover, studies concerning neuron-to-neuron communication are necessary to understand the role of psychedelics in the brain and how they can open a door for therapy. Other experiments

should evolve the understanding about the modulation of the serotonergic system by these compounds, relating their effects with specific neurological disorders phenotypes. With all of these said, the path in the renaissance of psychedelics still has many missing blocks, asking for better tuned protocols and implementation of novel techniques, inorder to make sense of a yet black box to neuroscience. Moreover, it is clear the necessity of diverse methods necessary to study these compounds, ranging from the atomistic size, with the synthesis of novel analogs, passing through molecular and cell communication, on how these drugs bind to receptors and activate specific cell responses, to a systemic perspective, trying to understand the holistic effects of entheogens.

7. References

"5-Hydroxytryptamine Receptor 2A." n.d. Accessed June 15. 2022a. https://www.uniprot.org/uniprot/P28223.

———. n.d. Accessed June 15. 2022b. https://www.uniprot.org/uniprot/P28223.

Abela. Andrew R.. Caleb J. Browne. Derya Sargin. Thomas D. Prevot. Xiao Dong Ji. Zhaoxia Li. Evelyn K. Lambe. and Paul J. Fletcher. 2020. "Median Raphe Serotonin Neurons Promote Anxiety-like Behavior via Inputs to the Dorsal Hippocampus." *Neuropharmacology*. https://doi.org/10.1016/j.neuropharm.2020.107985.

Adham. N.. H. T. Kao. L. E. Schecter. J. Bard. M. Olsen. D. Urquhart. M. Durkin. P. R. Hartig. R. L. Weinshank. and T. A. Branchek. 1993. "Cloning of Another Human Serotonin Receptor (5-HT1F): A Fifth 5-HT1 Receptor Subtype Coupled to the Inhibition of Adenylate Cyclase." *Proceedings of the National Academy of Sciences of the United States of America* 90 (2): 408–12.

Afonja. Tejumade. 2017. "Kernel Functions." Towards Data Science. January 2. 2017. https://towardsdatascience.com/kernel-function-6f1d2be6091.

Alloghani. Mohamed. Dhiya Al-Jumeily. Jamila Mustafina. Abir Hussain. and Ahmed J. Aljaaf. 2020. "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science." *Unsupervised and Semi-Supervised Learning*. https://doi.org/10.1007/978-3-030-22475-2_1.

Andalib. S.. M. R. Emamhadi. S. Yousefzadeh-Chabok. S. K. Shakouri. P. F. Høilund-Carlsen. M. S. Vafaee. and T. M. Michel. 2017. "Maternal SSRI Exposure Increases the Risk of Autistic Offspring: A Meta-Analysis and Systematic Review." *European Psychiatry*. https://doi.org/10.1016/j.eurpsy.2017.06.001.

Ansanay. H.. A. Dumuis. M. Sebben. J. Bockaert. and L. Fagni. 1995. "cAMP-Dependent. Long-Lasting Inhibition of a K Current in Mammalian Neurons." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.92.14.6635.

Askari. Neda. Mahdieh Moin. Mohammad Sanati. Masih Tajdini. Seyed-Mohammad-Reza Hosseini. Amirhossein Modabbernia. Babak Najand. et al. 2012. "Granisetron Adjunct to Fluvoxamine for Moderate to Severe Obsessive-Compulsive Disorder: A Randomized. Double-Blind. Placebo-Controlled Trial." *CNS Drugs* 26 (10): 883–92.

Azam. Shofiul. Md Ezazul Haque. Md Jakaria. Song-Hee Jo. In-Su Kim. and Dong-Kug Choi. 2020. "G-Protein-Coupled Receptors in CNS: A Potential Therapeutic Target for Intervention in Neurodegenerative Disorders and Associated Cognitive Deficits." *Cells* 9 (2). https://doi.org/10.3390/cells9020506.

Aznar. Susana. and Mona El-Sayed Hervig. 2016. "The 5-HT2A Serotonin Receptor in Executive Function: Implications for Neuropsychiatric and Neurodegenerative Diseases." *Neuroscience and Biobehavioral Reviews* 64 (May): 63–82.

Bachmann. Claude. 2002. "Mechanisms of Hyperammonemia." *Clinical Chemistry and Laboratory Medicine*. https://doi.org/10.1515/cclm.2002.112.

Badillo. Solveig. Balazs Banfai. Fabian Birzele. Iakov I. Davydov. Lucy Hutchinson. Tony Kam-Thong. Juliane Siebourg-Polster. Bernhard Steiert. and Jitao David Zhang. 2020. "An Introduction to Machine Learning." *Clinical Pharmacology and Therapeutics* 107 (4): 871–85.

Baldo. Brian A.. and Michael A. Rose. 2020. "The Anaesthetist. Opioid Analgesic Drugs. and Serotonin Toxicity: A Mechanistic and Clinical Review." *British Journal of Anaesthesia* 124 (1): 44–62.

Ballentine. Galen. Samuel Freesun Friedman. and Danilo Bzdok. 2022. "Trips and Neurotransmitters: Discovering Principled Patterns across 6850 Hallucinogenic Experiences." *Science Advances* 8 (11): eabl6989.

Bandelow. Borwin. Sophie Michaelis. and Dirk Wedekind. 2017. "Treatment of Anxiety Disorders." *Dialogues in Clinical Neuroscience* 19 (2): 93–107.

Bélanger. Stacey A.. Debbi Andrews. Clare Gray. and Daphne Korczak. 2018. “ADHD in Children and Youth: Part 1—Etiology. Diagnosis. and Comorbidity.” *Paediatrics & Child Health*. https://doi.org/10.1093/pch/pxy109.

Berg. Kelly A.. John A. Harvey. Umberto Spampinato. and William P. Clarke. 2008. “Physiological and Therapeutic Relevance of Constitutive Activity of 5-HT 2A and 5-HT 2C Receptors for the Treatment of Depression.” *Progress in Brain Research* 172: 287–305.

Blattner. Kevin M.. Daniel J. Canney. Douglas A. Pippin. and Benjamin E. Blass. 2019. “Pharmacology and Therapeutic Potential of the 5-HT Receptor.” *ACS Chemical Neuroscience* 10 (1): 89–119.

Boadle-Biber. Margaret C. 1993. “Regulation of Serotonin Synthesis.” *Progress in Biophysics and Molecular Biology*. https://doi.org/10.1016/0079-6107(93)90009-9.

Bockaert. Joël. Sylvie Claeysen. Carine Bécamel. Aline Dumuis. and Philippe Marin. 2006. “Neuronal 5-HT Metabotropic Receptors: Fine-Tuning of Their Structure. Signaling. and Roles in Synaptic Modulation.” *Cell and Tissue Research* 326 (2): 553–72.

Boess. F. G.. R. Beroukhim. and I. L. Martin. 1995. “Ultrastructure of the 5-hydroxytryptamine3 Receptor.” *Journal of Neurochemistry* 64 (3): 1401–5.

Bolaños. Carlos A.. Matthew D. Willey. Melissa L. Maffeo. Kyle D. Powers. Daniel W. Kinka. Katie B. Grausam. and Ross P. Henderson. 2008. “Antidepressant Treatment Can Normalize Adult Behavioral Deficits Induced by Early-Life Exposure to Methylphenidate.” *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2007.06.024.

Bostancıklıoğlu. Mehmet. 2020. “Optogenetic Stimulation of Serotonin Nuclei Retrieve the Lost Memory in Alzheimer’s Disease.” *Journal of Cellular Physiology* 235 (2): 836–47.

Bouhelal. Rochdi. Larbi Smounya. and Joël Bockaert. 1988. “5-HT1B Receptors Are Negatively Coupled with Adenylate Cyclase in Rat Substantia Nigra.” *European Journal of Pharmacology*. https://doi.org/10.1016/0014-2999(88)90799-6.

Bradley. P. B.. G. Engel. W. Feniuk. J. R. Fozard. P. P. Humphrey. D. N. Middlemiss. E. J. Mylecharane. B. P. Richardson. and P. R. Saxena. 1986. “Proposals for the Classification and Nomenclature of Functional Receptors for 5-Hydroxytryptamine.” *Neuropharmacology* 25 (6): 563–76.

Brady. Scott T.. and George J. Siegel. 2012. *Basic Neurochemistry: Principles of Molecular. Cellular and Medical Neurobiology*. Academic Press.

Breiman. Leo. 2001. “Machine Learning.” https://doi.org/10.1023/a:1010933404324.

Bruhn. Jan G.. Peter A. G. M. De Smet. Hesham R. El-Seedi. and Olof Beck. 2002. “Mescaline Use for 5700 Years.” *The Lancet* 359 (9320): 1866.

Buchan. Daniel W. A.. Federico Minneci. Tim C. O. Nugent. Kevin Bryson. and David T. Jones. 2013. “Scalable Web Services for the PSIPRED Protein Analysis Workbench.” *Nucleic Acids Research* 41 (Web Server issue): W349–57.

Buchborn. Tobias. Helmut Schröder. Daniela C. Dieterich. Gisela Grecksch. and Volker Höllt. 2015. “Tolerance to LSD and DOB Induced Shaking Behaviour: Differential Adaptations of Frontocortical 5-HT(2A) and Glutamate Receptor Binding Sites.” *Behavioural Brain Research* 281 (March): 62–68.

Burghardt. Nesha S.. David E. A. Bush. Bruce S. McEwen. and Joseph E. LeDoux. 2007. “Acute Selective Serotonin Reuptake Inhibitors Increase Conditioned Fear Expression: Blockade With a 5-HT2C Receptor Antagonist.” *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2006.11.023.

Bursulaya. Badry D.. Maxim Totrov. Ruben Abagyan. and Charles L. Brooks III. 2003. “Comparative Study of Several Algorithms for Flexible Ligand Docking.” *Journal of Computer-Aided Molecular Design*. https://doi.org/10.1023/b:jcam.0000017496.76572.6f.

Cameron. Lindsay P.. Robert J. Tombari. Ju Lu. Alexander J. Pell. Zefan Q. Hurley. Yann Ehinger. Maxemiliano V. Vargas. et al. 2021. “A Non-Hallucinogenic Psychedelic Analogue with Therapeutic Potential.” *Nature* 589 (7842): 474–79.

Castelhano. João. Gisela Lima. Marta Teixeira. Carla Soares. Marta Pais. and Miguel Castelo-Branco. 2021. “The Effects of Tryptamine Psychedelics in the Brain: A Meta-Analysis of

Functional and Review of Molecular Imaging Studies." *Frontiers in Pharmacology*. https://doi.org/10.3389/fphar.2021.739053.
Chanrion. Benjamin. Clotilde Mannoury la Cour. Sophie Gavarini. Mathieu Seimandi. Laurent Vincent. Jean-François Pujol. Joël Bockaert. Philippe Marin. and Mark J. Millan. 2008. "Inverse Agonist and Neutral Antagonist Actions of Antidepressants at Recombinant and Native 5-hydroxytryptamine2C Receptors: Differential Modulation of Cell Surface Expression and Signal Transduction." *Molecular Pharmacology* 73 (3): 748–57.
Chen. Xiaoning. Ran Ye. J. Jay Gargus. Randy D. Blakely. Kostantin Dobrenis. and Ji Ying Sze. 2015. "Disruption of Transient Serotonin Accumulation by Non-Serotonin-Producing Neurons Impairs Cortical Map Development." *Cell Reports* 10 (3): 346–58.
Chen. Yu Wai. 2014. *Structural Genomics: General Applications*.
Ciarleglio. C. M.. H. E. S. Resuehr. and D. G. McMahon. 2011. "Interactions of the Serotonin and Circadian Systems: Nature and Nurture in Rhythms and Blues." *Neuroscience* 197 (December): 8–16.
Cohen. Jeremiah Y.. Mackenzie W. Amoroso. and Naoshige Uchida. 2015. "Serotonergic Neurons Signal Reward and Punishment on Multiple Timescales." *eLife* 4 (February). https://doi.org/10.7554/eLife.06346.
Collins. Gregory T.. and Charles P. France. 2018. "Effects of Lorcaserin and Buspirone. Administered Alone and as a Mixture. on Cocaine Self-Administration in Male and Female Rhesus Monkeys." *Experimental and Clinical Psychopharmacology* 26 (5): 488–96.
Conn. P. J.. A. Janowsky. and E. Sanders-Bush. 1987. "Denervation Supersensitivity of 5-HT-1c Receptors in Rat Choroid Plexus." *Brain Research*. https://doi.org/10.1016/0006-8993(87)90641-x.
Cortes-Altamirano. Jose Luis. Adriana Olmos-Hernandez. Herlinda Bonilla Jaime. Paul Carrillo-Mora. Cindy Bandala. Samuel Reyes-Long. and Alfonso Alfaro-Rodríguez. 2018. "Review: 5-HT1. 5-HT2. 5-HT3 and 5-HT7 Receptors and Their Role in the Modulation of Pain Response in the Central Nervous System." *Current Neuropharmacology* 16 (2): 210–21.
Cortes. Corinna. and Vladimir Vapnik. 1995. "Support-Vector Networks." *Machine Learning*. https://doi.org/10.1007/bf00994018.
Cour. Clotilde Mannoury la. Clotilde Mannoury la Cour. Salah El Mestikawy. Naïma Hanoun. Michel Hamon. and Laurence Lanfumey. 2006. "Regional Differences in the Coupling of 5-Hydroxytryptamine-1A Receptors to G Proteins in the Rat Brain." *Molecular Pharmacology*. https://doi.org/10.1124/mol.106.022756.
Cowen. Philip. and Ann C. Sherwood. 2013. "The Role of Serotonin in Cognitive Function: Evidence from Recent Studies and Implications for Understanding Depression." *Journal of Psychopharmacology* 27 (7): 575–83.
Cristobal. S.. A. Zemla. D. Fischer. L. Rychlewski. and A. Elofsson. 2001. "A Study of Quality Measures for Protein Threading Models." *BMC Bioinformatics* 2 (August): 5.
"Cross-Validation (statistics)." 2003. Wikimedia Foundation. Inc. December 31. 2003. https://en.wikipedia.org/wiki/Cross-validation_(statistics).
Cuccioloni. Massimiliano. Laura Bonfili. Valentina Cecarini. Filippo Cocchioni. Dezemona Petrelli. Elena Crotti. Raffaella Zanchi. Anna Maria Eleuteri. and Mauro Angeletti. 2020. "Structure/activity Virtual Screening and in Vitro Testing of Small Molecule Inhibitors of 8-Hydroxy-5-deazaflavin:NADPH Oxidoreductase from Gut Methanogenic Bacteria." *Scientific Reports*. https://doi.org/10.1038/s41598-020-70042-w.
Curatolo. Paolo. Elisa D'Agati. and Romina Moavero. 2010. "The Neurobiological Basis of ADHD." *Italian Journal of Pediatrics*. https://doi.org/10.1186/1824-7288-36-79.
Dai. Y.. N. L. Dudek. T. B. Patel. and N. A. Muma. 2008. "Transglutaminase-Catalyzed Transamidation: A Novel Mechanism for Rac1 Activation by 5-hydroxytryptamine2A Receptor Stimulation." *The Journal of Pharmacology and Experimental Therapeutics* 326 (1): 153–62.
Dalva. Matthew B.. Andrew C. McClelland. and Matthew S. Kayser. 2007. "Cell Adhesion Molecules: Signalling Functions at the Synapse." *Nature Reviews. Neuroscience* 8 (3): 206–20.
Dargan. Paul I.. and David M. Wood. 2021. *Novel Psychoactive Substances: Classification.*

Pharmacology and Toxicology. Academic Press.
Debroise. Théau. Eugene I. Shakhnovich. and Nicolas Chéron. 2017. "A Hybrid Knowledge-Based and Empirical Scoring Function for Protein-Ligand Interaction: SMoG2016." *Journal of Chemical Information and Modeling* 57 (3): 584–93.
De Vivo. M.. and S. Maayani. 1986. "Characterization of the 5-hydroxytryptamine1a Receptor-Mediated Inhibition of Forskolin-Stimulated Adenylate Cyclase Activity in Guinea Pig and Rat Hippocampal Membranes." *The Journal of Pharmacology and Experimental Therapeutics* 238 (1): 248–53.
Dhonnchadha. B. Á. Nic. B. Á. Nic Dhonnchadha. R. G. Fox. S. J. Stutz. K. C. Rice. and K. A. Cunningham. 2009. "Blockade of the Serotonin 5-ht2a Receptor Suppresses Cue-Evoked Reinstatement of Cocaine-Seeking Behavior in a Rat Self-Administration Model." *Behavioral Neuroscience*. https://doi.org/10.1037/a0014592.
Dremencov. Eliyahu. Mostafa El Mansari. and Pierre Blier. 2007. "Noradrenergic Augmentation of Escitalopram Response by Risperidone: Electrophysiologic Studies in the Rat Brain." *Biological Psychiatry* 61 (5): 671–78.
Durrant. Jacob D.. and J. Andrew McCammon. 2010. "NNScore: A Neural-Network-Based Scoring Function for the Characterization of Protein−Ligand Complexes." *Journal of Chemical Information and Modeling*. https://doi.org/10.1021/ci100244v.
Durrant. Jacob D.. and J. Andrew McCammon. 2011a. "BINANA: A Novel Algorithm for Ligand-Binding Characterization." *Journal of Molecular Graphics & Modelling* 29 (6): 888–93.
———. 2011b. "NNScore 2.0: A Neural-Network Receptor-Ligand Scoring Function." *Journal of Chemical Information and Modeling* 51 (11): 2897–2903.
Efron. Bradley. 1983. "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation." *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1983.10477973.
Eswar. Narayanan. Bino John. Nebojsa Mirkovic. Andras Fiser. Valentin A. Ilyin. Ursula Pieper. Ashley C. Stuart. et al. 2003. "Tools for Comparative Protein Structure Modeling and Analysis." *Nucleic Acids Research* 31 (13): 3375–80.
Falsone. S. Fabio. Gerd Leitinger. Anita Karner. Andreas J. Kungl. Simone Kosol. Roberto Cappai. and Klaus Zangger. 2011. "The Neurotransmitter Serotonin Interrupts α-Synuclein Amyloid Maturation." *Biochimica et Biophysica Acta* 1814 (5): 553–61.
Fargin. A.. J. R. Raymond. M. J. Lohse. B. K. Kobilka. M. G. Caron. and R. J. Lefkowitz. 1988. "The Genomic Clone G-21 Which Resembles a Beta-Adrenergic Receptor Sequence Encodes the 5-HT1A Receptor." *Nature* 335 (6188): 358–60.
Fiser. Andras. 2010. "Template-Based Protein Structure Modeling." *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-60761-842-3_6.
Freedman. Daniel X. 1968. "On the Use and Abuse of LSD." *Archives of General Psychiatry*. https://doi.org/10.1001/archpsyc.1968.01740030074008.
Gaddum. J. H.. and Z. P. Picarelli. 1997. "Two Kinds of Tryptamine Receptor. 1957." *British Journal of Pharmacology* 120 (4 Suppl): 134–39; discussion 132–33.
Geddes. Sean D.. Saleha Assadzada. David Lemelin. Alexandra Sokolovski. Richard Bergeron. Samir Haj-Dahmane. and Jean-Claude Béïque. 2016. "Target-Specific Modulation of the Descending Prefrontal Cortex Inputs to the Dorsal Raphe Nucleus by Cannabinoids." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1522754113.
Goethem. N. P. van. R. Schreiber. A. Newman-Tancredi. M. Varney. and J. Prickaerts. 2015. "Divergent Effects of the 'biased' 5-HT_{1A}receptor Agonists F15599 and F13714 in a Novel Object Pattern Separation Task." *British Journal of Pharmacology*. https://doi.org/10.1111/bph.13071.
Gohlke. H.. M. Hendlich. and G. Klebe. 2000. "Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions." *Journal of Molecular Biology* 295 (2): 337–56.
Gray. J. A.. and B. L. Roth. 2001. "Paradoxical Trafficking and Regulation of 5-HT(2A) Receptors by Agonists and Antagonists." *Brain Research Bulletin* 56 (5): 441–51.
Gray. John A.. Beth A. Compton-Toth. and Bryan L. Roth. 2003. "Identification of Two Serine Residues Essential for Agonist-Induced 5-HT2A Receptor Desensitization." *Biochemistry* 42

(36): 10853–62.
Griebel. Guy. and Andrew Holmes. 2013. “50 Years of Hurdles and Hope in Anxiolytic Drug Discovery.” *Nature Reviews. Drug Discovery* 12 (9): 667–87.
Grof. Stanislav. 2008. *LSD Psychotherapy*.
Guedes. Isabella A.. Felipe S. S. Pereira. and Laurent E. Dardenne. 2018. “Empirical Scoring Functions for Structure-Based Virtual Screening: Applications. Critical Aspects. and Challenges.” *Frontiers in Pharmacology* 9 (September): 1089.
Gupta. Deepali. Visakh Prabhakar. and Mahesh Radhakrishnan. 2016. “5HT3 Receptors: Target for New Antidepressant Drugs.” *Neuroscience and Biobehavioral Reviews* 64 (May): 311–25.
Hamati. Rami. Mostafa El Mansari. and Pierre Blier. 2020. “Serotonin-2B Receptor Antagonism Increases the Activity of Dopamine and Glutamate Neurons in the Presence of Selective Serotonin Reuptake Inhibition.” *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 45 (12): 2098–2105.
Handbook of the Behavioral Neurobiology of Serotonin. 2020. Academic Press.
Hannon. Jason. and Daniel Hoyer. n.d. “Molecular Biology of 5-HT Receptors.” *Serotonin and Sleep: Molecular. Functional and Clinical Aspects*. https://doi.org/10.1007/978-3-7643-8561-3_6.
Hayes. Adam. 2008. “Nonparametric Method.” Investopedia. April 29. 2008. https://www.investopedia.com/terms/n/nonparametric-method.asp.
Heinricher. M. M.. I. Tavares. J. L. Leith. and B. M. Lumb. 2009. “Descending Control of Nociception: Specificity. Recruitment and Plasticity.” *Brain Research Reviews*. https://doi.org/10.1016/j.brainresrev.2008.12.009.
He. Jianzheng. Franziska Hommen. Nina Lauer. Sophia Balmert. and Henrike Scholz. 2020. “Serotonin Transporter Dependent Modulation of Food-Seeking Behavior.” *PloS One* 15 (1): e0227554.
Hensler. Julie G. 2006. “Serotonergic Modulation of the Limbic System.” *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2005.06.007.
Hoffer. Abram. and Humphry Osmond. 1967. *The Hallucinogens*. New York : Academic Press.
Hou. Yu-Wei. Ping Xiong. Xue Gu. Xin Huang. Min Wang. and Jing Wu. 2018. “Association of Serotonin Receptors with Attention Deficit Hyperactivity Disorder: A Systematic Review and Meta-Analysis.” *Current Medical Science* 38 (3): 538–51.
Hritcu. Lucian. Monica Clicinschi. and Toshitaka Nabeshima. 2007. “Brain Serotonin Depletion Impairs Short-Term Memory. but Not Long-Term Memory in Rats.” *Physiology & Behavior*. https://doi.org/10.1016/j.physbeh.2007.03.028.
Hsieh. P. C.. K. C. Chen. T. L. Yeh. I. H. Lee. P. S. Chen. W. J. Yao. N-T Chiu. C-C Chen. M-H Liao. and Y. K. Yang. 2014. “Lower Availability of Midbrain Serotonin Transporter between Healthy Subjects with and without a Family History of Major Depressive Disorder - a Preliminary Two-Ligand SPECT Study.” *European Psychiatry: The Journal of the Association of European Psychiatrists* 29 (7): 414–18.
Huang. Sheng-You. Sam Z. Grinter. and Xiaoqin Zou. 2010. “Scoring Functions and Their Evaluation Methods for Protein–ligand Docking: Recent Advances and Future Directions.” *Physical Chemistry Chemical Physics*. https://doi.org/10.1039/c0cp00151a.
Huang. Sheng-You. and Xiaoqin Zou. 2006. “An Iterative Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions: I. Derivation of Interaction Potentials.” *Journal of Computational Chemistry* 27 (15): 1866–75.
———. 2014. “A Knowledge-Based Scoring Function for Protein-RNA Interactions Derived from a Statistical Mechanics-Based Iterative Method.” *Nucleic Acids Research* 42 (7): e55.
Hylin. John W.. and Donald P. Watson. 1965. “Ergoline Alkaloids in Tropical Wood Roses.” *Science*. https://doi.org/10.1126/science.148.3669.499.
Iidaka. Tetsuya. Norio Ozaki. Atsushi Matsumoto. Junpei Nogawa. Yoko Kinoshita. Tatsuyo Suzuki. Nakao Iwata. Yukiko Yamamoto. Tomohisa Okada. and Norihiro Sadato. 2005. “A Variant C178T in the Regulatory Region of the Serotonin Receptor Gene HTR3A Modulates Neural Activation in the Human Amygdala.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25 (27): 6460–66.

Iigaya. Kiyohito. Madalena S. Fonseca. Masayoshi Murakami. Zachary F. Mainen. and Peter Dayan. 2018. “An Effect of Serotonergic Stimulation on Learning Rates for Rewards Apparent after Long Intertrial Intervals.” *Nature Communications* 9 (1): 2477.

Isberg. Vignir. Chris de Graaf. Andrea Bortolato. Vadim Cherezov. Vsevolod Katritch. Fiona H. Marshall. Stefan Mordalski. et al. 2015. “Generic GPCR Residue Numbers - Aligning Topology Maps While Minding the Gaps.” *Trends in Pharmacological Sciences* 36 (1): 22–31.

Jain. A. N. 1996. “Scoring Noncovalent Protein-Ligand Interactions: A Continuous Differentiable Function Tuned to Compute Binding Affinities.” *Journal of Computer-Aided Molecular Design* 10 (5): 427–40.

Jensen. A.. P. Davies. H. Braunerosborne. and K. Krzywkowski. 2008. “3B but Which 3B? And That’s Just One of the Questions: The Heterogeneity of Human 5-HT3 Receptors.” *Trends in Pharmacological Sciences*. https://doi.org/10.1016/j.tips.2008.06.001.

Kamata. Tooru. Munehiro Katagi. and Hitoshi Tsuchihashi. 2010. “Metabolism and Toxicological Analyses of Hallucinogenic Tryptamine Analogues Being Abused in Japan.” *Forensic Toxicology*. https://doi.org/10.1007/s11419-009-0087-9.

Kern. Fabian. Tobias Fehlmann. and Andreas Keller. 2020. “On the Lifetime of Bioinformatics Web Services.” *Nucleic Acids Research* 48 (22): 12523–33.

Khamis. Mohamed A.. Walid Gomaa. and Walaa F. Ahmed. 2015. “Machine Learning in Computational Docking.” *Artificial Intelligence in Medicine*. https://doi.org/10.1016/j.artmed.2015.02.002.

Kim. Kuglae. Tao Che. Ouliana Panova. Jeffrey F. DiBerto. Jiankun Lyu. Brian E. Krumm. Daniel Wacker. et al. 2020. “Structure of a Hallucinogen-Activated Gq-Coupled 5-HT2A Serotonin Receptor.” *Cell*. https://doi.org/10.1016/j.cell.2020.08.024.

Kimura. Kanako Terakado. Hidetsugu Asada. Asuka Inoue. Francois Marie Ngako Kadji. Dohyun Im. Chihiro Mori. Takatoshi Arakawa. et al. 2019. “Structures of the 5-HT2A Receptor in Complex with the Antipsychotics Risperidone and Zotepine.” *Nature Structural & Molecular Biology*. https://doi.org/10.1038/s41594-018-0180-z.

Kirby. L. G.. F. D. Zeeb. and C. A. Winstanley. 2011. “Contributions of Serotonin in Addiction Vulnerability.” *Neuropharmacology* 61 (3): 421–32.

Kroeze. Wesley K.. and Bryan L. Roth. n.d. “Molecular Biology and Genomic Organization of G Protein-Coupled Serotonin Receptors.” *The Serotonin Receptors*. https://doi.org/10.1007/978-1-59745-080-5_1.

Kuhn. Max. 2008. “Building Predictive Models in*R*Using the**caret**Package.” *Journal of Statistical Software*. https://doi.org/10.18637/jss.v028.i05.

Kumar. Vijay. and Nitin Dogra. 2022. “A Comprehensive Review on Deep Synergistic Drug Prediction Techniques for Cancer.” *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-021-09617-3.

Kurrasch-Orbaugh. Deborah M.. Jason C. Parrish. Val J. Watts. and David E. Nichols. 2003. “A Complex Signaling Cascade Links the serotonin2A Receptor to Phospholipase A2 Activation: The Involvement of MAP Kinases.” *Journal of Neurochemistry* 86 (4): 980–91.

Kurrasch-Orbaugh. Deborah M.. Val J. Watts. Eric L. Barker. and David E. Nichols. 2003. “Serotonin 5-Hydroxytryptamine 2A Receptor-Coupled Phospholipase C and Phospholipase A2 Signaling Pathways Have Different Receptor Reserves.” *The Journal of Pharmacology and Experimental Therapeutics* 304 (1): 229–37.

Labasque. Marilyne. Eric Reiter. Carine Becamel. Joël Bockaert. and Philippe Marin. 2008. “Physical Interaction of Calmodulin with the 5-Hydroxytryptamine$_{2C}$ Receptor C-Terminus Is Essential for G Protein-Independent. Arrestin-Dependent Receptor Signaling.” *Molecular Biology of the Cell*. https://doi.org/10.1091/mbc.e08-04-0422.

Lagerström. Malin C.. and Helgi B. Schiöth. 2008. “Structural Diversity of G Protein-Coupled Receptors and Significance for Drug Discovery.” *Nature Reviews. Drug Discovery* 7 (4): 339–57.

Lanzenberger. Rupert R.. Markus Mitterhauser. Christoph Spindelegger. Wolfgang Wadsak. Nikolas Klein. Leonhard-Key Mien. Alexander Holik. et al. 2007. “Reduced Serotonin-1A Receptor Binding in Social Anxiety Disorder.” *Biological Psychiatry* 61 (9): 1081–89.

Li. Jin. Ailing Fu. and Le Zhang. 2019. “An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking.” *Interdisciplinary Sciences: Computational Life Sciences*. https://doi.org/10.1007/s12539-019-00327-w.

Lin. Shih-Hsien. Lan-Ting Lee. and Yen Kuang Yang. 2014. “Serotonin and Mental Disorders: A Concise Review on Molecular Neuroimaging Evidence.” *Clinical Psychopharmacology and Neuroscience: The Official Scientific Journal of the Korean College of Neuropsychopharmacology* 12 (3): 196–202.

Lin. Stanley L.. Nadine N. Johnson-Farley. David R. Lubinsky. and Daniel S. Cowen. 2003. “Coupling of Neuronal 5-HT7 Receptors to Activation of Extracellular-Regulated Kinase through a Protein Kinase A-Independent Pathway That Can Utilize Epac.” *Journal of Neurochemistry*. https://doi.org/10.1046/j.1471-4159.2003.02076.x.

Lissemore. Jennifer I.. Debbie Sookman. Paul Gravel. Alexandre Berney. Amir Barsoum. Mirko Diksic. Thomas E. Nordahl. et al. 2018. “Brain Serotonin Synthesis Capacity in Obsessive-Compulsive Disorder: Effects of Cognitive Behavioral Therapy and Sertraline.” *Translational Psychiatry*. https://doi.org/10.1038/s41398-018-0128-4.

Liu. Pei-Pei. Yi Xie. Xiao-Yan Meng. and Jian-Sheng Kang. 2019. “History and Progress of Hypotheses and Clinical Trials for Alzheimer’s Disease.” *Signal Transduction and Targeted Therapy*. https://doi.org/10.1038/s41392-019-0063-8.

Li. Yan. Zhihai Liu. Jie Li. Li Han. Jie Liu. Zhixiong Zhao. and Renxiao Wang. 2014. “Comparative Assessment of Scoring Functions on an Updated Benchmark: 1. Compilation of the Test Set.” *Journal of Chemical Information and Modeling*. https://doi.org/10.1021/ci500080q.

Lorman. William J. 2018. “Pharmacology Update: The Selective Serotonin Reuptake Inhibitors.” *Journal of Addictions Nursing* 29 (4): 260–61.

Mao. Jiashun. Javed Akhtar. Xiao Zhang. Liang Sun. Shenghui Guan. Xinyu Li. Guangming Chen. et al. 2021. “Comprehensive Strategies of Machine-Learning-Based Quantitative Structure-Activity Relationship Models.” *iScience* 24 (9): 103052.

Marler. Sarah. Bradley J. Ferguson. Evon Batey Lee. Brittany Peters. Kent C. Williams. Erin McDonnell. Eric A. Macklin. et al. 2016. “Brief Report: Whole Blood Serotonin Levels and Gastrointestinal Symptoms in Autism Spectrum Disorder.” *Journal of Autism and Developmental Disorders* 46 (3): 1124–30.

M.A. Marti-Renom. A. Stuart. A. Fiser. R. Sánchez. F. Melo. A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29. 291-325. 2000.

McCartney. Annie M.. Hugh T. McGovern. and Alexander De Foe. 2022. “Psychedelic Assisted Therapy for Major Depressive Disorder: Recent Work and Clinical Directions.” *Journal of Psychedelic Studies*. https://doi.org/10.1556/2054.2022.00211.

McCorvy. John D.. Aubrie A. Harland. Rebecca Maglathlin. and David E. Nichols. 2011. “A 5-HT2C Receptor Antagonist Potentiates a Low Dose Amphetamine-Induced Conditioned Place Preference.” *Neuroscience Letters*. https://doi.org/10.1016/j.neulet.2011.07.036.

McGrew. L.. M. S. S. Chang. and E. Sanders-Bush. 2002. “Phospholipase D Activation by Endogenous 5-Hydroxytryptamine 2C Receptors Is Mediated by $G\alpha_{13}$and Pertussis Toxin-Insensitive Gβγ Subunits.” *Molecular Pharmacology*. https://doi.org/10.1124/mol.62.6.1339.

Mdawar. Bernadette. Elias Ghossoub. and Rita Khoury. 2020. “Selective Serotonin Reuptake Inhibitors and Alzheimer’s Disease.” *Neural Regeneration Research* 15 (1): 41–46.

Mendez-David. Indira. Denis J. David. Flavie Darcet. Melody V. Wu. Saadia Kerdine-Römer. Alain M. Gardier. and René Hen. 2014. “Rapid Anxiolytic Effects of a 5-HT4 Receptor Agonist Are Mediated by a Neurogenesis-Independent Mechanism.” *Neuropsychopharmacology*. https://doi.org/10.1038/npp.2013.332.

“Method for Comparative Protein Structure Modeling by MODELLER.” n.d. Accessed June 15. 2022. https://salilab.org/modeller/manual/node11.html.

Millan. Mark J.. Philippe Marin. Joël Bockaert. and Clotilde Mannoury la Cour. 2008. “Signaling at G-Protein-Coupled Serotonin Receptors: Recent Advances and Future Research Directions.” *Trends in Pharmacological Sciences* 29 (9): 454–64.

Mnie-Filali. Ouissame. Céline Faure. Laura Lambás-Señas. Mostafa El Mansari. Hassina Belblidia. Elise Gondard. Adeline Etiévant. et al. 2011. “Pharmacological Blockade of 5-

HT7 Receptors as a Putative Fast Acting Antidepressant Strategy." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 36 (6): 1275–88.

Molliver. M. E. 1987. "Serotonergic Neuronal Systems: What Their Anatomic Organization Tells Us about Function." *Journal of Clinical Psychopharmacology* 7 (6 Suppl): 3S – 23S.

Moreno. José L.. Terrell Holloway. Vinayak Rayannavar. Stuart C. Sealfon. and Javier González-Maeso. 2013. "Chronic Treatment with LY341495 Decreases 5-HT(2A) Receptor Binding and Hallucinogenic Effects of LSD in Mice." *Neuroscience Letters* 536 (March): 69–73.

Moroz. Leonid L.. John R. Edwards. Sathyanarayanan V. Puthanveettil. Andrea B. Kohn. Thomas Ha. Andreas Heyland. Bjarne Knudsen. et al. 2006. "Neuronal Transcriptome of Aplysia: Neuronal Compartments and Circuitry." *Cell* 127 (7): 1453–67.

Morris. Garrett M.. Ruth Huey. William Lindstrom. Michel F. Sanner. Richard K. Belew. David S. Goodsell. and Arthur J. Olson. 2009. "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility." *Journal of Computational Chemistry* 30 (16): 2785–91.

Morris. Garrett M.. and Marguerita Lim-Wilby. 2008. "Molecular Docking." *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-59745-177-2_19.

Muller. Christian P.. and Barry Jacobs. 2009. *Handbook of the Behavioral Neurobiology of Serotonin*. Academic Press.

Naber. Dieter. and Monika Bullinger. 2018. "Should Antidepressants Be Used in Minor Depression?" *Dialogues in Clinical Neuroscience* 20 (3): 223–28.

Nash. Jon R.. Peter A. Sargent. Eugenii A. Rabiner. Sean D. Hood. Spilios V. Argyropoulos. John P. Potokar. Paul M. Grasby. and David J. Nutt. 2008. "Serotonin 5-HT1A Receptor Binding in People with Panic Disorder: Positron Emission Tomography Study." *The British Journal of Psychiatry: The Journal of Mental Science* 193 (3): 229–34.

Nazeer. Ahsan. Finza Latif. Aisha Mondal. Muhammad Waqar Azeem. and Donald E. Greydanus. 2020. "Obsessive-Compulsive Disorder in Children and Adolescents: Epidemiology. Diagnosis and Management." *Translational Pediatrics* 9 (Suppl 1): S76–93.

Neudert. Gerd. and Gerhard Klebe. 2011. "DSX: A Knowledge-Based Scoring Function for the Assessment of Protein-Ligand Complexes." *Journal of Chemical Information and Modeling* 51 (10): 2731–45.

Nichols. David E. 2016. "Psychedelics." *Pharmacological Reviews* 68 (2): 264–355.

———. 2018. "Chemistry and Structure-Activity Relationships of Psychedelics." *Current Topics in Behavioral Neurosciences* 36: 1–43.

Nichols. David E.. and Charles D. Nichols. 2008. "Serotonin Receptors." *Chemical Reviews* 108 (5): 1614–41.

Nikiforuk. Agnieszka. Małgorzata Hołuj. Tomasz Kos. and Piotr Popik. 2016. "The Effects of a 5-HT5A Receptor Antagonist in a Ketamine-Based Rat Model of Cognitive Dysfunction and the Negative Symptoms of Schizophrenia." *Neuropharmacology*. https://doi.org/10.1016/j.neuropharm.2016.01.035.

Nowak. Mateusz. Marcin Kołaczkowski. Maciej Pawłowski. and Andrzej J. Bojarski. 2006. "Homology Modeling of the Serotonin 5-HT1A Receptor Using Automated Docking of Bioactive Compounds with Defined Geometry." *Journal of Medicinal Chemistry* 49 (1): 205–14.

Oades. Robert D. 2007. "Role of the Serotonin System in ADHD: Treatment Implications." *Expert Review of Neurotherapeutics* 7 (10): 1357–74.

Oblak. Adrian. Terrell T. Gibbs. and Gene J. Blatt. 2013. "Reduced Serotonin Receptor Subtypes in a Limbic and a Neocortical Region in Autism." *Autism Research*. https://doi.org/10.1002/aur.1317.

Ohmura. Yu. Iku Tsutsui-Kimura. Hitomi Sasamori. Mao Nebuka. Naoya Nishitani. Kenji F. Tanaka. Akihiro Yamanaka. and Mitsuhiro Yoshioka. 2020. "Different Roles of Distinct Serotonergic Pathways in Anxiety-like Behavior. Antidepressant-Like. and Anti-Impulsive Effects." *Neuropharmacology*. https://doi.org/10.1016/j.neuropharm.2019.107703.

Pagadala. Nataraj S.. Khajamohiddin Syed. and Jack Tuszynski. 2017. "Software for Molecular Docking: A Review." *Biophysical Reviews*. https://doi.org/10.1007/s12551-016-0247-1.

Paulus. Erin V.. and Eric M. Mintz. 2016. "Circadian Rhythms of Clock Gene Expression in the Cerebellum of Serotonin-Deficient Pet-1 Knockout Mice." *Brain Research* 1630 (January):

10–17.
Peroutka. S. J.. and T. A. Howell. 1994. “The Molecular Evolution of G Protein-Coupled Receptors: Focus on 5-Hydroxytryptamine Receptors.” *Neuropharmacology* 33 (3-4): 319–24.
Pierce. P. A.. and S. J. Peroutka. 1989. “Hallucinogenic Drug Interactions with Neurotransmitter Receptor Binding Sites in Human Cortex.” *Psychopharmacology* 97 (1): 118–22.
Pollak Dorocic. Iskra. Daniel Fürth. Yang Xuan. Yvonne Johansson. Laura Pozzi. Gilad Silberberg. Marie Carlén. and Konstantinos Meletis. 2014. “A Whole-Brain Atlas of Inputs to Serotonergic Neurons of the Dorsal and Median Raphe Nuclei.” *Neuron* 83 (3): 663–78.
Pourhamzeh. Mahsa. Fahimeh Ghasemi Moravej. Mehrnoosh Arabi. Elahe Shahriari. Soraya Mehrabi. Richard Ward. Reza Ahadi. and Mohammad Taghi Joghataei. 2021. “The Roles of Serotonin in Neuropsychiatric Disorders.” *Cellular and Molecular Neurobiology*. https://doi.org/10.1007/s10571-021-01064-9.
Preto. A. J.. Carlos A. V. Barreto. Salete J. Baptista. José Guilherme de Almeida. Agostinho Lemos. André Melo. M. Nátalia D. Cordeiro. Zeynep Kurkcuoglu. Rita Melo. and Irina S. Moreira. 2020. “Understanding the Binding Specificity of G-Protein Coupled Receptors toward G-Proteins and Arrestins: Application to the Dopamine Receptor Family.” *Journal of Chemical Information and Modeling*. https://doi.org/10.1021/acs.jcim.0c00371.
Provenzi. Livio. Monica Fumagalli. Giunia Scotto di Minico. Roberto Giorda. Francesco Morandi. Ida Sirgiovanni. Paola Schiavolin. Fabio Mosca. Renato Borgatti. and Rosario Montirosso. 2020. “Pain-Related Increase in Serotonin Transporter Gene Methylation Associates with Emotional Regulation in 4.5-Year-Old Preterm-Born Children.” *Acta Paediatrica* 109 (6): 1166–74.
Rabin. Richard A.. Meredith Regina. Mireille Doat. and J. C. Winter. 2002. “5-HT2A Receptor-Stimulated Phosphoinositide Hydrolysis in the Stimulus Effects of Hallucinogens.” *Pharmacology. Biochemistry. and Behavior* 72 (1-2): 29–37.
Rasmussen. Carl Edward. and Christopher K. I. Williams. 2005. “Gaussian Processes for Machine Learning.” https://doi.org/10.7551/mitpress/3206.001.0001.
Ray. Thomas S. 2010. “Psychedelics and the Human Receptorome.” *PloS One* 5 (2): e9019.
Riba. Jordi. Ethan H. McIlhenny. Marta Valle. José Carlos Bouso. and Steven A. Barker. 2012. “Metabolism and Disposition of N.N-Dimethyltryptamine and Harmala Alkaloids after Oral Administration of Ayahuasca.” *Drug Testing and Analysis*. https://doi.org/10.1002/dta.1344.
Rosário-Ferreira. Nícia. Salete J. Baptista. Carlos A. V. Barreto. Filipe E. P. Rodrigues. Tomás F. D. Silva. Sara G. F. Ferreira. João N. M. Vitorino. et al. 2021. “In Silico End-to-End Protein-Ligand Interaction Characterization Pipeline: The Case of SARS-CoV-2.” *ACS Synthetic Biology* 10 (11): 3209–35.
Ross. Gregory A.. Garrett M. Morris. and Philip C. Biggin. 2012. “Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites.” *PloS One* 7 (3): e32036.
Roth. B. L.. T. Nakaki. D. M. Chuang. and E. Costa. 1984. “Aortic Recognition Sites for Serotonin (5HT) Are Coupled to Phospholipase C and Modulate Phosphatidylinositol Turnover.” *Neuropharmacology* 23 (10): 1223–25.
Roth. Bryan L. 2011. “Irving Page Lecture: 5-HT(2A) Serotonin Receptor Biology: Interacting Proteins. Kinases and Paradoxical Regulation.” *Neuropharmacology* 61 (3): 348–54.
Ruat. M.. E. Traiffort. R. Leurs. J. Tardivel-Lacombe. J. Diaz. J. M. Arrang. and J. C. Schwartz. 1993. “Molecular Cloning. Characterization. and Localization of a High-Affinity Serotonin Receptor (5-HT7) Activating cAMP Formation.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.90.18.8547.
Ruck. C. A.. J. Bigwood. D. Staples. J. Ott. and R. G. Wasson. 1979. “Entheogens.” *Journal of Psychedelic Drugs* 11 (1-2): 145–46.
Sahli. Zeyad T.. and Frank I. Tarazi. 2018. “Pimavanserin: Novel Pharmacotherapy for Parkinson’s Disease Psychosis.” *Expert Opinion on Drug Discovery* 13 (1): 103–10.
Sahu. Apeksha. Lathika Gopalakrishnan. Nayana Gaur. Oishi Chatterjee. Praseeda Mol. Prashant Kumar Modi. Shobha Dagamajalu. Jayshree Advani. Sanjeev Jain. and T. S. Keshava Prasad. 2018. “The 5-Hydroxytryptamine Signaling Map: An Overview of Serotonin-Serotonin Receptor Mediated Signaling Network.” *Journal of Cell Communication and*

Signaling 12 (4): 731–35.
Schmid. Cullen L.. Kirsten M. Raehal. and Laura M. Bohn. 2008. “Agonist-Directed Signaling of the Serotonin 2A Receptor Depends on Beta-Arrestin-2 Interactions in Vivo.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (3): 1079–84.
Schmidtke. Peter. Vincent Le Guilloux. Julien Maupetit. and Pierre Tufféry. 2010. “Fpocket: Online Tools for Protein Ensemble Pocket Detection and Tracking.” *Nucleic Acids Research* 38 (Web Server issue): W582–89.
Schneider. Gisbert. 2010. “Virtual Screening: An Endless Staircase?” *Nature Reviews. Drug Discovery* 9 (4): 273–76.
Schultes. Richard Evans. Albert Hofmann. and Christian Rätsch. 2001. *Plants of the Gods: Their Sacred. Healing. and Hallucinogenic Powers*. Healing Arts Press.
Sebben. Michele. Herve Ansanay. Joel Bockaert. and Aline Dumuis. 1994. “5-HT6 Receptors Positively Coupled to Adenylyl Cyclase in Striatal Neurones in Culture.” *NeuroReport*. https://doi.org/10.1097/00001756-199412000-00037.
Selvaraj. Sudhakar. Danilo Arnone. Alessandra Cappai. and Oliver Howes. 2014. “Alterations in the Serotonin System in Schizophrenia: A Systematic Review and Meta-Analysis of Postmortem and Molecular Imaging Studies.” *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2014.06.005.
Shahbazi. Fred. Victoria Grandi. Abhinandan Banerjee. and John F. Trant. 2020. “Cannabinoids and Cannabinoid Receptors: The Story so Far.” *iScience* 23 (7): 101301.
Sheffler. Douglas J.. Wesley K. Kroeze. Bonnie G. Garcia. Ariel Y. Deutch. Sandra J. Hufeisen. Patrick Leahy. Jens C. Brüning. and Bryan L. Roth. 2006. “p90 Ribosomal S6 Kinase 2 Exerts a Tonic Brake on G Protein-Coupled Receptor Signaling.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0600585103.
Shen. Chao. Junjie Ding. Zhe Wang. Dongsheng Cao. Xiaoqin Ding. and Tingjun Hou. 2020. “From Machine Learning to Deep Learning: Advances in Scoring Functions for Protein–ligand Docking.” *WIREs Computational Molecular Science*. https://doi.org/10.1002/wcms.1429.
Shenker. A.. S. Maayani. H. Weinstein. and J. P. Green. 1987. “Pharmacological Characterization of Two 5-Hydroxytryptamine Receptors Coupled to Adenylate Cyclase in Guinea Pig Hippocampal Membranes.” *Molecular Pharmacology* 31 (4): 357–67.
Shen. Min-Yi. and Andrej Sali. 2006. “Statistical Potential for Assessment and Prediction of Protein Structures.” *Protein Science: A Publication of the Protein Society* 15 (11): 2507–24.
Shulgin. Alexander Theodore. and Ann Shulgin. 1991. *Pihkal: A Chemical Love Story*.
———. 1997. *Tihkal: The Continuation*.
Siew. N.. A. Elofsson. L. Rychlewski. and D. Fischer. 2000. “MaxSub: An Automated Measure for the Assessment of Protein Structure Prediction Quality.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/16.9.776.
Simon. Leslie V.. and Michael Keenaghan. 2022. “Serotonin Syndrome.” In *StatPearls*. Treasure Island (FL): StatPearls Publishing.
Sitaram. Balvant R.. Lynn Lockett. Rerngjit Talomsin. Graeme L. Blackman. and William R. McLeod. 1987. “In Vivo Metabolism of 5-Methoxy-N. N-Dimethyltryptamine and N.N-Dimethyltryptamine in the Rat.” *Biochemical Pharmacology*. https://doi.org/10.1016/0006-2952(87)90118-3.
Smith. Alexander H.. Jonathan Ott. and Jeremy Bigwood. 1979. “Teonanacatl: Hallucinogenic Mushrooms of North America.” *Mycologia*. https://doi.org/10.2307/3759089.
Smith. David E.. and Alan J. Rose. 1968. “The Use and Abuse of LSD in Haight-Ashbury.” *Clinical Pediatrics*. https://doi.org/10.1177/000992286800700605.
Stahl. Stephen M. 2018. “Beyond the Dopamine Hypothesis of Schizophrenia to Three Neural Networks of Psychosis: Dopamine. Serotonin. and Glutamate.” *CNS Spectrums* 23 (3): 187–91.
Steinmann. Casper. Martin A. Olsson. and Ulf Ryde. 2018. “Relative Ligand-Binding Free Energies Calculated from Multiple Short QM/MM MD Simulations.” *Journal of Chemical Theory and Computation* 14 (6): 3228–37.
Stępnicki. Piotr. Magda Kondej. and Agnieszka A. Kaczor. 2018. “Current Concepts and

Treatments of Schizophrenia." *Molecules* 23 (8). https://doi.org/10.3390/molecules23082087.
Stiedl. Oliver. Elpiniki Pappa. Åsa Konradsson-Geuken. and Sven Ove Ögren. 2015. "The Role of the Serotonin Receptor Subtypes 5-HT1A and 5-HT7 and Its Interaction in Emotional Learning and Memory." *Frontiers in Pharmacology* 6 (August): 162.
"Supervised vs. Unsupervised Learning: What's the Difference?" 2021. March 12. 2021. https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning.
Takumi. Toru. Kota Tamada. Fumiyuki Hatanaka. Nobuhiro Nakai. and Patrick F. Bolton. 2020. "Behavioral Neuroscience of Autism." *Neuroscience and Biobehavioral Reviews* 110 (March): 60–76.
Tao. Zhuo-Ying. Pei-Xing Wang. Si-Qi Wei. Richard J. Traub. Jin-Feng Li. and Dong-Yuan Cao. 2019. "The Role of Descending Pain Modulation in Chronic Primary Pain: Potential Application of Drugs Targeting Serotonergic System." *Neural Plasticity* 2019 (December): 1389296.
Teissier. Anne. Mariano Soiza-Reilly. and Patricia Gaspar. 2017. "Refining the Role of 5-HT in Postnatal Development of Brain Circuits." *Frontiers in Cellular Neuroscience* 11 (May): 139.
Törk. I. 1990. "Anatomy of the Serotonergic System." *Annals of the New York Academy of Sciences* 600: 9–34; discussion 34–35.
Torres. Pedro H. M.. Ana C. R. Sodero. Paula Jofily. and Floriano P. Silva-Jr. 2019. "Key Topics in Molecular Docking for Drug Design." *International Journal of Molecular Sciences* 20 (18). https://doi.org/10.3390/ijms20184574.
Trevisan. Vinícius. 2022. "Comparing Robustness of MAE. MSE and RMSE." Towards Data Science. January 11. 2022. https://towardsdatascience.com/comparing-robustness-of-mae-mse-and-rmse-6d69da870828.
Trigo. José Manuel. Thibault Renoir. Laurence Lanfumey. Michel Hamon. Klaus-Peter Lesch. Patricia Robledo. and Rafael Maldonado. 2007. "3.4-Methylenedioxymethamphetamine Self-Administration Is Abolished in Serotonin Transporter Knockout Mice." *Biological Psychiatry* 62 (6): 669–79.
Trott. Oleg. and Arthur J. Olson. 2010. "AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function. Efficient Optimization. and Multithreading." *Journal of Computational Chemistry* 31 (2): 455–61.
Tsuchioka. Mami. Minoru Takebayashi. Kazue Hisaoka. Natsuko Maeda. and Yoshihiro Nakata. 2008. "Serotonin (5-HT) Induces Glial Cell Line-Derived Neurotrophic Factor (GDNF) mRNA Expression via the Transactivation of Fibroblast Growth Factor Receptor 2 (FGFR2) in Rat C6 Glioma Cells." *Journal of Neurochemistry*. https://doi.org/10.1111/j.1471-4159.2008.05357.x.
"Tuple." 2002. Wikimedia Foundation. Inc. January 24. 2002. https://en.wikipedia.org/wiki/Tuple.
Twarog. Betty M.. and Irvine H. Page. 1953. "Serotonin Content of Some Mammalian Tissues and Urine and a Method for Its Determination." *American Journal of Physiology-Legacy Content*. https://doi.org/10.1152/ajplegacy.1953.175.1.157.
Vattano. Anthony J. 1981. "Psychedelic Drugs Reconsidered. By Lester Grinspoon and James B. Bakalar. New York: Basic Books. 1979. 343 Pp. $15.95 Cloth." *Children & Schools*. https://doi.org/10.1093/cs/4.1.71.
Volle. Julien. Tatiana Bregman. Brian Scott. Mustansir Diwan. Roger Raymond. Paul J. Fletcher. José N. Nobrega. and Clement Hamani. 2018. "Deep Brain Stimulation and Fluoxetine Exert Different Long-Term Changes in the Serotonergic System." *Neuropharmacology*. https://doi.org/10.1016/j.neuropharm.2018.03.005.
Wacker. Daniel. Sheng Wang. John D. McCorvy. Robin M. Betz. A. J. Venkatakrishnan. Anat Levit. Katherine Lansu. et al. 2017. "Crystal Structure of an LSD-Bound Human Serotonin Receptor." *Cell* 168 (3): 377–89.e12.
Wasson. R. Gordon. and R. Gordon Wasson. 1971. "The Soma of the Rig Veda: What Was It?" *Journal of the American Oriental Society*. https://doi.org/10.2307/600096.
Webb. Benjamin. and Andrej Sali. 2016. "Comparative Protein Structure Modeling Using

MODELLER." *Current Protocols in Protein Science / Editorial Board. John E. Coligan ... [et Al.]* 86 (November): 2.9.1–2.9.37.
White. Kellie J.. Crystal C. Walline. and Eric L. Barker. 2005. "Serotonin Transporters: Implications for Antidepressant Drug Development." *The AAPS Journal* 7 (2): E421–33.
Wiederstein. Markus. and Manfred J. Sippl. 2007. "ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins." *Nucleic Acids Research* 35 (Web Server issue): W407–10.
Williams. R. L. 1999. "Mammalian Phosphoinositide-Specific Phospholipase C." *Biochimica et Biophysica Acta* 1441 (2-3): 255–67.
Yamashita. Paula Sm. Daiane S. Rosa. Christopher A. Lowry. and Helio Zangrossi Jr. 2018. "Serotonin Actions within the Prelimbic Cortex Induce Anxiolysis Mediated by Serotonin 1a Receptors." *Journal of Psychopharmacology* . December. 269881118817384.
Yuen. Eunice Y.. and Zhen Yan. 2007. "Serotonergic Regulation of NMDA Receptor Trafficking and Function in Prefrontal Cortex." *Monoaminergic Modulation of Cortical Excitability*. https://doi.org/10.1007/978-0-387-72256-6_5.
Zhang. Xujun. Chao Shen. Xueying Guo. Zhe Wang. Gaoqi Weng. Qing Ye. Gaoang Wang. et al. 2021. "ASFP (Artificial Intelligence Based Scoring Function Platform): A Web Server for the Development of Customized Scoring Functions." *Journal of Cheminformatics* 13 (1): 6.
Zhou. Lei. and Laura M. Bohn. 2014. "Functional Selectivity of GPCR Signaling in Animals." *Current Opinion in Cell Biology*. https://doi.org/10.1016/j.ceb.2013.11.010.
Zhu. Siyu. Meixian Wu. Ziwei Huang. and Jing An. 2021. "Trends in Application of Advancing Computational Approaches in GPCR Ligand Discovery." *Experimental Biology and Medicine*. https://doi.org/10.1177/1535370221993422.
Zięba. Agata. Piotr Stępnicki. Dariusz Matosiuk. and Agnieszka A. Kaczor. 2021. "Overcoming Depression with 5-HT2A Receptor Ligands." *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms23010010.
Żmudzka. Elżbieta. Kinga Sałaciak. Jacek Sapa. and Karolina Pytka. 2018. "Serotonin Receptors in Depression and Anxiety: Insights from Animal Studies." *Life Sciences* 210 (October): 106–24.
Zugliani. Morena M.. Mariana C. Cabo. Antonio E. Nardi. Giampaolo Perna. and Rafael C. Freire. 2019. "Pharmacological and Neuromodulatory Treatments for Panic Disorder: Clinical Trials from 2010 to 2018." *Psychiatry Investigation* 16 (1): 50–58.

8. Supplementary Information

Generic Number (Ballesteros Weinstein Numbering)	Index in active structure	Index in inactive structure
3.28	W151	W83
3.29	I152	I84
3.32	D155	D87
3.33	V156	V88
3.36	S159	S91
ECL2	L229	L161
5.39	V235	V167
5.42	G238	G170
5.46	S242	S174
6.44	F332	F221

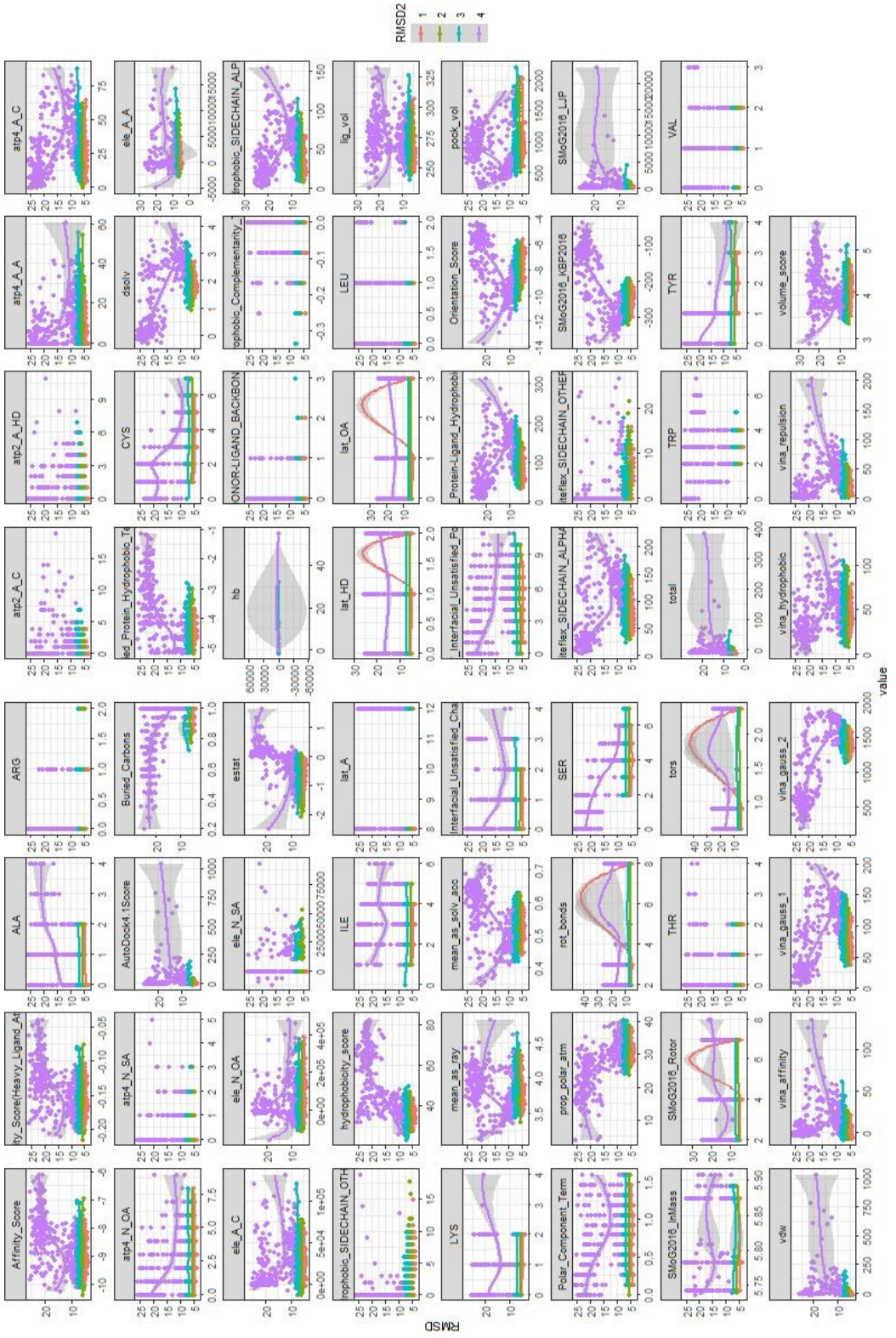
6.48	W336	W225
6.51	F339	F228
6.52	F340	F229
6.55	N343	N232
7.39	V366	V255
7.42	G369	G258

SI.1. Table relating the indexation present in the models build by Homology Modeling. both for the active and inactive structure. with the Ballesteros Weinstein numbering (Isberg et al. 2015).

```
[1] 999  65
'data.frame':   999 obs. of  65 variables:
 $ RMSD                                          : num  23.44 6.34 5.92 8.58 21.45 ...
 $ Orientation_Score                             : num  -8.77 -8.25 -6.36 -7.27 -4.69 ...
 $ Affinity_Score(Heavy_Ligand_Atoms)            : num  -0.14 -0.149 -0.135 -0.104 -0.083 -0.15 -0.138 -0.157 -0.
 $ Affinity_Score                                : num  -8.44 -8.64 -8.33 -7.61 -7.12 ...
 $ Buried_Protein_Hydrophobic_Term               : num  -4.13 -4.12 -3.92 -3.87 -2.31 ...
 $ Hydrophobic_Complementarity_Term              : num  -0.166 -0.249 0 0 0 -0.083 -0.083 -0.249 0 0 ...
 $ Polar_Component_Term                          : num  1.072 0.938 0.804 1.474 0.402 ...
 $ Number_of_Protein-Ligand_Hydrophobic_Contacts : num  122 86 56 72 17 100 91 52 57 28 ...
 $ Number_of_Interfacial_Unsatisfied_Polar_Atoms : num  7 6 6 10 3 5 7 3 8 3 ...
 $ Number_of_Interfacial_Unsatisfied_Charged_Atoms: num  1 1 0 1 0 1 1 1 1 0 ...
 $ Buried_Carbons                                : num  1 1 0.944 1 0.556 1 1 0.944 1 0.556 ...
 $ AutoDock4.1Score                              : num  31.17 2.206 -3.454 -2.322 0.425 ...
 $ estat                                         : num  0.607 -0.464 -0.23 -0.147 0.195 ...
 $ hb                                            : num  -0.4479 -0.7895 -0.0007 -0.0268 0 ...
 $ vdw                                           : num  26.03 -1.51 -7.26 -7.14 -2.17 ...
 $ dsolv                                         : num  2.5904 2.5868 1.6585 2.6076 0.0134 ...
 $ tors                                          : num  2.39 2.39 2.39 2.39 2.39 ...
 $ lig_vol                                       : num  247 292 293 274 277 ...
 $ pock_vol                                      : num  1489 1627 2032 1483 969 ...
 $ mean_as_ray                                   : num  3.93 4.24 4.41 3.85 3.96 4.25 4.16 4.2 4.26 3.77 ...
 $ mean_as_solv_acc                              : num  0.56 0.59 0.57 0.54 0.61 0.58 0.58 0.58 0.58 0.59 ...
 $ hydrophobicity_score                          : num  47.2 33.6 29.3 42.1 81.5 ...
 $ volume_score                                  : num  3.76 3.6 3.64 4 5.62 3.81 3.91 3.86 3.78 5.6 ...
 $ prop_polar_atm                                : num  29.2 34.7 34.2 35.5 22.2 ...
 $ ALA                                           : num  1 1 1 1 2 1 1 1 1 2 ...
 $ ARG                                           : num  0 0 0 0 0 0 0 0 0 0 ...
 $ CYS                                           : num  1 1 1 2 0 2 1 1 1 0 ...
 $ ILE                                           : num  3 2 2 5 4 3 4 3 2 5 ...
 $ LEU                                           : num  1 0 0 1 0 0 0 0 0 0 ...
 $ LYS                                           : num  0 0 1 1 0 1 1 1 0 1 ...
 $ SER                                           : num  1 4 5 4 0 4 4 4 2 0 ...
 $ THR                                           : num  3 1 1 0 0 1 1 1 1 0 ...
 $ TRP                                           : num  3 3 3 3 6 3 3 3 3 6 ...
 $ TYR                                           : num  0 1 1 1 0 0 1 1 1 0 ...
 $ VAL                                           : num  3 1 1 1 0 1 1 1 1 0 ...
 $ vina_affinity                                 : num  13.895 5.887 -3.723 -0.239 -1.914 ...
 $ vina_gauss_1                                  : num  109.8 88 54.8 80.8 12.9 ...
 $ vina_gauss_2                                  : num  1242 1346 1152 1441 448 ...
 $ vina_repulsion                                : num  39.944 25.648 4.168 13.454 0.513 ...
 $ vina_hydrophobic                              : num  62 28 27.7 34.6 11.9 ...
 $ atp2_A_C                                      : num  2 0 0 0 0 0 0 0 0 0 ...
 $ atp2_A_HD                                     : num  2 0 0 0 0 0 4 0 0 0 ...
 $ atp4_A_C                                      : num  36 18 22 18 3 23 32 15 14 4 ...
 $ atp4_A_A                                      : num  17 9 1 11 0 9 10 0 6 0 ...
 $ atp4_N_SA                                     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ atp4_N_OA                                     : num  2 3 0 2 0 1 3 3 1 0 ...
 $ lat_A                                         : num  12 12 12 12 12 12 12 12 12 12 ...
 $ lat_HD                                        : num  2 2 2 2 2 2 2 2 2 2 ...
 $ lat_OA                                        : num  3 3 3 3 3 3 3 3 3 3 ...
 $ ele_A_C                                       : num  68298.9 17360.4 10820.8 17214.7 84.6 ...
 $ ele_A_A                                       : num  -1057 -1891.8 -135.4 82.6 0 ...
 $ ele_N_SA                                      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ ele_N_OA                                      : num  74661 10078 0 89913 0 ...
 $ rot_bonds                                     : num  8 8 8 8 8 5 8 8 8 8 ...
 $ siteflex_SIDECHAIN_OTHER                      : num  24 0 0 0 0 0 0 0 0 0 ...
 $ siteflex_SIDECHAIN_ALPHA                      : num  135 96 41 90 8 79 116 86 48 15 ...
 $ hbond_HDONOR-LIGAND_BACKBONE_ALPHA            : num  0 0 0 0 0 0 0 0 0 0 ...
 $ hydrophobic_SIDECHAIN_OTHER                   : num  14 0 0 0 0 0 0 0 0 0 ...
 $ hydrophobic_SIDECHAIN_ALPHA                   : num  50 35 21 38 7 33 43 19 22 10 ...
 $ total                                         : num  5.14 -5.4 -6.76 -8.8 -5.04 ...
 $ SMoG2016_KBP2016                              : num  -228.5 -221 -203.4 -232.5 -79.9 ...
 $ SMoG2016_LJP                                  : num  937 308 196 131 66 ...
 $ SMoG2016_Rotor                                : num  7 7 7 7 7 4 7 7 7 7 ...
 $ SMoG2016_lnMass                               : num  5.74 5.74 5.74 5.74 5.74 ...
 $ RMSD2                                         : chr  "4" "3" "2" "4" ...
```

SI.2: dim() and part of str(). in Rstudio. of the initial working table : "sum_dockings_features_final"

SI.2. ScatterPlots made with Rstudio. plotting the distribution of the features considered in the characterization of the complexes in study.



```
import os

end_file = ".pdbqt"
start_file = "dock_conf"
for current_file in os.listdir(os.getcwd()):
      if current_file.endswith(end_file) and \
                  current_file.startswith(start_file):
            pdb_number = current_file.split(start_file)[1].split(".")[0]
            pdb_name = current_file.split(".")[0] + ".pdb"
            full_command = "cat protein_rigid.pdb " + \
                  pdb_name + " | grep -v '^END$' > complex" + \
                  str(pdb_number) + ".pdb"
            os.system(full_command)
```

C.S. 1: cut_66.py

```
# -*- coding: utf-8 -*-
"""
Created on Wed Apr  6 23:53:50 2022

@author: guilh
"""

import os
import pandas as pd
import numpy as np
path =
"C:/Users/guilh/Downloads/DOCKINGS/DOCK13_active_proof8_ligand1_mymodel"
os.chdir(path)

def process_pdb(input_pdb):
        ###open pdb file for crystal
        opened_file = pd.read_fwf(input_pdb. \
            widths = [6. 6. 4. 1. 4. 1. 4. 4. 8. 8. 8. 6. 6. 10. 2. 2].
header = None)
        ###change column names for crystal
        crystal_edited = opened_file.rename(columns = {0:'ATOM'.  2:'TYPE'.
8:'X'. 9:'Y'. 10:'Z'})
        ###make subset ATOM from crystal
        atom_crystal = crystal_edited.loc[crystal_edited["ATOM"] == "ATOM"]
        ###make subset HETATM from crystal
        hetatm_crystal = crystal_edited.loc[crystal_edited["ATOM"] ==
"HETATM"]
        ###remove hydrogens from complex arrays
        hetatm_crystal_lessH = hetatm_crystal.drop( \
            hetatm_crystal.index[(hetatm_crystal["TYPE"] == "H")] \. axis=0)
        x_hetatm_crystal = hetatm_crystal_lessH[['X']].to_numpy()
        y_hetatm_crystal = hetatm_crystal_lessH[['Y']].to_numpy()
        z_hetatm_crystal = hetatm_crystal_lessH[['Z']].to_numpy()
        ##convert arrays object to float
        x_hetatm_crystal_float = x_hetatm_crystal.astype(dtype = float . \
                order='K'. casting='unsafe'. subok=True. copy=True)
        y_hetatm_crystal_float = y_hetatm_crystal.astype(dtype = float . \
                order='K'. casting='unsafe'. subok=True. copy=True)
        z_hetatm_crystal_float = z_hetatm_crystal.astype(dtype = float . \
                order='K'. casting='unsafe'. subok=True. copy=True)
        return x_hetatm_crystal_float. y_hetatm_crystal_float.
z_hetatm_crystal_float
```

C.S. 2.1.: Euclidean_distance.py part 1

```
x_hetatm_crystal_float. y_hetatm_crystal_float. z_hetatm_crystal_float =
process_pdb("6wha_edited.pdb")
end_file = ".pdb"
start_file = "complex"
source_dir = os.getcwd()
output_list = []
for current_file in os.listdir(os.getcwd()):
    if current_file.endswith(end_file) and
current_file.startswith(start_file):
        x_hetatm_complex_float. y_hetatm_complex_float.
z_hetatm_complex_float = \
            process_pdb(current_file)
        ###make difference array
        x_difference_array = np.subtract(x_hetatm_crystal_float.
x_hetatm_complex_float)
        y_difference_array = np.subtract(y_hetatm_crystal_float.
y_hetatm_complex_float)
        z_difference_array = np.subtract(z_hetatm_crystal_float.
z_hetatm_complex_float)
        ###square difference arrays
        x_square = np.square(x_difference_array)
        y_square = np.square(y_difference_array)
        z_square = np.square(z_difference_array)
        ###sum squares
        sum_array = np.array([x_square . y_square . z_square])
        ###sum inside array.division per entries on sum_array
        division_array = np.divide(sum_array.sum(). 23)
        #####root-square on division = RMSD value
        root_square_array = np.sqrt(division_array
        ########create new variable for each complexe's RMSD
        file_name = current_file.split(start_file)[1].split(".")[0]
        output_list.append([file_name. root_square_array])

output_dataframe = pd.DataFrame(output_list. columns = ["Complex". "RMSD"])
output_dataframe.to_csv("rmsd.csv". index = False)
```

C.S. 2.2.: Euclidean_distance.py part 2

```
import os
import shutil
from pymol import cmd
import pymol

end_file = ".pdb"
start_file = "complex"
source_dir = os.getcwd()
pymol.finish_launching(['pymol'. '-qi'])
for current_file in os.listdir(source_dir):
   if current_file.endswith(end_file) and
current_file.startswith(start_file):
      cmd.load(current_file)
      cmd.select("resi_1". "resi 1")
      cmd.select("protein" . "!(resi_1)")
      dir_name = current_file.split(start_file)[1].split(".")[0]
      os.mkdir(dir_name)
      cmd.save(str(source_dir + "/" + dir_name + "/protein.pdb"). "protein")
      cmd.save(str(source_dir + "/" + dir_name + "/ligand.mol2"). "resi_1")
      cmd.delete("complex" + dir_name)
cmd.quit()
```

C.S. 3.: Scissors.py

Model	RMSE	R^2	MAE
rbfDDA	10.76	0.01	9.11
	10.76	0.01	9.11
	10.76	0.02	9.11

SI3. Table showing the worst evaluated model. for both classification and regression. and its performance evaluation.

#complex	241
Orientation_Score	0.735281
Affinity_Score(Heavy_Ligand_Atoms)	-0.20228
Affinity_Score	-0.2306
Buried_Protein_Hydrophobic_Term	-0.07083
Hydrophobic_Complementarity_Term	0.447936
Number_of_Protein-Ligand_Hydrophobic_Contacts	-0.73689
Number_of_Interfacial_Unsatisfied_Polar_Atoms	-0.0392
Buried_Carbons	0.056242
AutoDock4.1Score	-0.27662
estat	-0.69209
hb	-0.04825
vdw	-0.26606
dsolv	-0.24693
tors	-0.25139
lig_vol	-1.16584
pock_vol	-0.09813
mean_as_ray	-0.02466

mean_as_solv_acc	0.495172
hydrophobicity_score	-0.44712
volume_score	0.35452
prop_polar_atm	0.079772
ALA	-0.75685
ARG	1.178982
CYS	0.855925
ILE	-0.15693
LEU	-0.58343
LYS	1.52607
SER	0.062017
THR	1.850503
TRP	0.074597
TYR	0.816394
VAL	1.87965
vina_affinity	-0.70804
vina_gauss_1	-0.75336
vina_gauss_2	0.151617
vina_repulsion	-0.70347
vina_hydrophobic	-0.37802
atp2_A_C	-0.38845
atp2_A_HD	-0.50868
atp4_A_C	-1.2498
atp4_A_A	0.206916

atp4_N_SA	-0.46681
atp4_N_OA	0.544409
lat_A	-1.26309
lat_HD	0.284903
ele_A_C	0.552381
ele_A_A	-0.34344
ele_N_SA	-0.44596
ele_N_OA	0.009155
siteflex_SIDECHAIN_OTHER	1.402693
siteflex_SIDECHAIN_ALPHA	-0.6934
hbond_HDONOR-LIGAND_BACKBONE_ALPHA	-0.25072
hydrophobic_SIDECHAIN_OTHER	0.736381
hydrophobic_SIDECHAIN_ALPHA	-0.76698
total	-0.25894
SMoG2016_KBP2016	0.042885
SMoG2016_LJP	-0.26063
SMoG2016_Rotor	0.138565
SMoG2016_lnMass	-0.55857
RMSD2	1
RMSD	4.499917
Predicted_by_qrf	4.422436

SI.4. Table exhibiting the physical characteristics (features) of the binding pocket present in the complex with the lowest value of RMSD. complex #241. by the predictive model build with *qrf*.