Tutor

Dr. Jaime Rubio Martínez
Departament de Ciència de Materials i Química Física

Grau de Química

Treball Final de Grau

Ranking of inhibitors of SARS-CoV-2 Mpro protease with end-point methodologies
Clasificación de inhibidores de la Mpro proteasa del SARS-CoV-2 con metodología de punto final

Carlos Cruz Marín

June 2022

UNIVERSITAT DE BARCELONA

B:KC Barcelona Knowledge Campus
Campus d'Excel·lència Internacional

Me niego a pensar que la Luna no está ahí cuando no la miro.

Albert Einstein

Gracias a mi tutor Dr. Jaime Rubio por la orientación y guía, y en otro plano, a todas las personas que me acompañaron en mi camino y me han visto crecer con orgullo, como son, mi familia, mi madre, mi padre, mi hermana, amigos y sobre todo mi pareja, Claudia.

Gracias también, a mis profesoras Adaya y Pilar de Bachillerato por ser esas profesoras que siembran en ti vocación y curiosidad.

REPORT

IDENTIFICATION AND REFLECTION ON THE SUSTAINABLE DEVELOPMENT GOALS (ODS)

The Sustainable Development Goals are a global set of 17 interconnected global goals designed to achieve a better and more sustainable future for everybody, which are collected in the Agenda 2030. They were established by the UN (United Nations). They are the following ones:

This work could be classified under point 3:

Within the 5 major groups, the 5Ps , which are the following ones:

Placed in People's group. This work can provide knowledge in the drug's design field and, therefore, in the discovery of new drugs.

The objective of this work is to find a new calculation method able to rank a group of ligands which are, in addition, inhibitors of the main SARS-CoV-2 protease. Moreover, given the severity and incidence of COVID-19 in today's world, knowledge generated on this work has much greater value as it can help to stop the COVID's pandemic which changed the social and global paradigm.

Growing in this aspect, both in scientific research and in drug design, would help to alleviate any further disease or health problem in a much more effectively, quickly and economically, a part of as curbing other current diseases that cost so many human lives, such as cancer. Research into the affectation and any disease mechanism will provide exceptional information to become able to eradicate it. This would be the ultimate goal and the ultimate philosophy, fulfilling any of the goals proposed by the UN and the WHO.

https://www.un.org/sustainabledevelopment/health/

CONTENTS

1. SUMMARY

The aim of the following work is to find the computational calculation method capable of quantitatively ranking a list of inhibitors from natural products[1] of the main protease of SARS-CoV-2, the protein in charge of reproducing the virus related to the recent COVID-19 pandemic (WHO), according to its binding energy with the protein (receptor). Given the complexity of both the system and the reaction conditions, this simulation is complicated and it is necessary to evaluate it in different ways in order to find the method that most closely matches, thus understanding its behaviour and being able to develop different drugs or remedies against this disease and against other upcoming diseases with similar characteristics. Such inhibitors are present in nature and have been found by different molecular dynamics assays, starting from a database of approximately 2000 compounds providing a total of 11 inhibitors, of which 5 have inhibitory activity in vivo. Finding the calculation method that ranks them can help to expand this number and thus increase the chances of finding the substance with the highest inhibitory capacity. To rank the inhibitors, we will assess how well different methods calculate their binding energy to the receptor (as we will call the protein) relative to the inhibitory activity they have demonstrated in in vivo assays. This binding energy is the difference in energies of the individual receptor and inhibitor and the ligand-receptor complex.

The ideal stage of this work would be to obtain energies close to 0 for ligands with no inhibitory activity and a low energy for those with inhibitory activity. Results obtained do not give this exact order but give the first steps to achieve the final method. Previous order[1] give 2 true positive compounds and the final method in this work can give 3, a part of obtaining interesting conclusions.

Keywords: SARS-CoV-2, methods, molecular dynamics, inhibitor, ligand, receptor.

2. RESUM

El objetivo del siguiente trabajo es encontrar el método de cálculo computacional capaz de ordenar cuantitativamente una lista de inhibidores procedentes de productos naturales[1] de la proteasa principal de la SARS-CoV-2, proteína encargada de reproducir el virus relacionado con la reciente pandemia de la COVID-19 (OMS), según su energía de enlace con la proteína (receptor). Dada la complejidad tanto del sistema como de las condiciones de reacción, esta simulación se complica y es necesario evaluarla de diferentes maneras por tal de encontrar el método que más se aproxime, así entendiendo su comportamiento y pudiendo desarrollar diferentes fármacos o remedios contra esta enfermedad y contra otras venideras de características similares. Dichos inhibidores son presentes en la naturaleza y han sido encontrados mediante diferentes ensayos de dinámica molecular, partiendo de una base de datos de aproximadamente 2000 compuestos proporcionando un total de 11 inhibidores, de los cuales 5 tienen actividad inhibidora in vivo. Encontrar el método de cálculo que los ordene puede ayudar a ampliar este número y así aumentar las probabilidades de encontrar la sustancia con una mayor capacidad inhibidora. Por tal de ordenar los inhibidores, se evaluará la veracidad con la que diferentes métodos calculan su energía de enlace con el receptor (así denominaremos a la proteína) respecto a la actividad inhibidora que han demostrado en los ensayos in vivo. Esta energía de enlace es la diferencia de energías del receptor y el inhibidor individuales y el complejo ligando-receptor.

El estadio ideal de este trabajo sería obtener unas energías cercanas a 0 para los ligandos sin actividad inhibidora y, una energía baja para los que sí que la presentan. Los resultados obtenidos no dan este orden exacto pero dan los primeros pasos para conseguir el método final. El orden previo[1] da 2 verdaderos positivos, mientras que el método final de este trabajo proporciona 3, a parte de las conclusiones interesantes que se han obtenido.

Palabras clave: SARS-CoV-2, métodos, dinámica molecular, inhibidor, ligando, receptor.

3. INTRODUCTION

The remarkable progress in theoretical and computational chemistry has made the possibility to include its methods in the discovery of new drugs or the study of diseases, making them faster, cheaper and, above all, more effective. This has resulted in the computational discovery of successful drugs. Apart from the notable improvement that computational chemistry has brought to drug discovery, it has been urgently demanded after the appearance of the SARS-CoV-2 virus to stop the pandemic or at least mitigate its effects on the population. The aim of the following work, extensively detailed below, is to find a computational method able to properly rank a list of inhibitors of the main protein of the before mentioned virus.

3.1.SARS-CoV-2

The SARS-CoV-2 virus has been the main character in the recent pandemic of COVID-19, a highly transmissible disease that mainly affects the respiratory system, although it can also cause damage to other organs, including the heart, kidneys, and brain[2]. Although most people recover without problems, some sufferers may have long-lasting[3] symptoms such as respiratory problems, heart complications, chronic kidney failure or in the most extreme cases death.

Since its irruption on the global scene in December 2019 in China, more than 500 million people have been infected and the total number of deaths has reached 6.22 million people, giving a relatively low mortality (1.23%) compared to other diseases such as Ebola[4] with a 50% mortality rate, for example.

The virus has an RNA chain structure very similar to SARS-CoV and MERS-CoV (Middle East respiratory syndrome coronavirus) which belong to the betacoronavirus gene of the Coronaviridae family[5].

3.2. MAIN PROTEASE (MPRO)

Mpro (7cam) protease is the main replication-related protein of the virus. This protease has the function of catalysing proteolysis (cutting the peptide chain). A representation of this molecule is obtained from the RCSB PDB6 (Research Collaboratory for Structural Bioinformatics Protein Data Bank) to understand its structure.

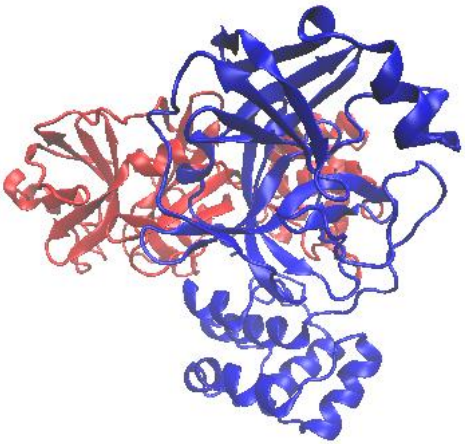

Figure 1. 3D representation of Mpro protease of SARS-CoV-2

This protease has a dimer structure in which each chain is coloured in a different colour. For the sake of computational efficiency, it will be considered a monomer[7], rather than a dimer, because of the reduced number of atoms and residues and because it gives similar results, saving time without losing precision.

The dimer consists of 4706 atoms and a molecular weight of 67.77 kDa, a fact that justifies the need to consider it a monomer from this point onwards. In the present work, the monomer (which we will refer to as the receptor) consists of 304 residues.

3.3. INHIBITORS

According to the preceding study[1], the inhibitors to be studied in the present work are the following ones. They are listed according to its inhibitory activity and ranked from the lowest binding energy to the highest[1]. This is a primary order, the final one will be extracted from this work.

3.3.1. Compounds with inhibitory activity in *in vivo* tests

Compounds in order of increasing free binding energy:

- (-) epigallocatechin gallate: It is the major active compound in green tea. It also exists in other food sources such as white tea, black tea, kiwis, strawberries, pears, apples ... It is a strong anti-inflammatory and antioxidant and is the most researched catechin[8]. IC_{50}= 22 µM
- Amentoflavone: Isolated from *Selaginella tamariscina*. It is a multifunctional molecule. It is an anticarcinogen, as well as being considered as a promising therapeutic agent for clinical research, in fields such as oncology, Alzheimer or diabetes[9 10]. IC_{50}= 28 µM
- Vitexin-2-O-rhamnoside: It is one of the major flavonoid constituents of *Crataegus pinnatifida* plants. It has several biological activities and pharmacological activity, as an antioxidant and as a treatment for heart disease[11]. IC_{50}= 65 µM
- Aloin: Isolated from the Aloe plant, it is often used to treat digestive problems due to its laxative action and to deal with constipation[12]. IC_{50}= 96 µM
- Rhoifolin: It is one of the most common compounds in the family of flavonoids, found in foods such as oranges, grapefruit, lemons, tomatoes, artichokes, and many plants. It has excellent anti-inflammatory, hepatoprotective and highly selective cytotoxicological effects[13]. IC_{50}= 230 µM

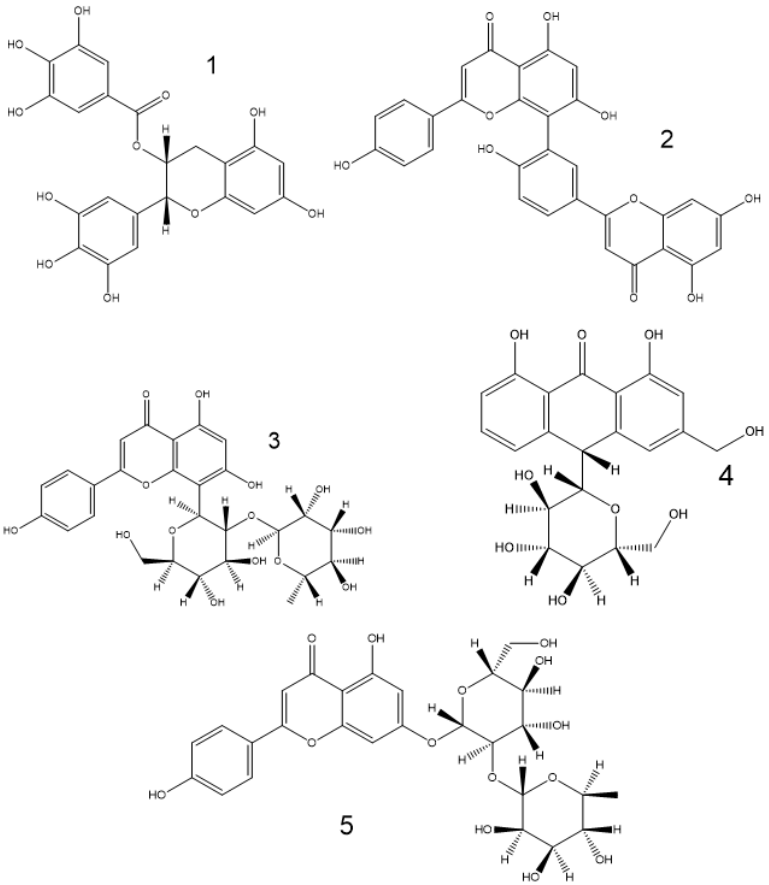

Figure 2. Chemical structure of ligands with inhibitory avtivity

3.3.2. Compounds without inhibitory activity in *in vivo* tests

Compounds in order of increasing free binding energy:

- Proanthocyanidins: Provided from berries and fruits, like lingonberry, cranberry or persimmon, banana, for example. They are produced at the end of flavonoid biosynthetic pathway and have some pharmacological properties like antioxidant and anticancer activity, antidiabetic and neuroprotective[14].
- Narirutin: Flavanone present in citrus fruits like oranges or grapefruit which has anti-inflammatory and antioxidant effects. It is also interesting because his inhibition activity on BACE-1 and Aβ, pharmacological targets on Alzheimer's disease investigation[15 16].
- Ziyu-glycoside I: Is one of the major active ingredients in *Sanguisorba officinalis* and it has been reported as the terpene with strongest haemostatic activity[17 18].
- Luteoloside: Flavone derived from luteolin. It can be found in Dandelion and in artichoke (Cynara Scolymus). In some experiments it has exhibited promotive effects on human hepatocyte cells, becoming a potential hepatoprotective compound[19 20].
- Linarin: It has been abundantly characterised in species such as Cirsium, Micromere or Buddleia. Its main physiological activities are as remedial effect on central nervous system disorders, sleeping enhancing and sedative effects[21].
- Corilagin: Is one of the major active components of many ethnipharmacological plants. Isolated from *Caesalpinia coriaria* in 1951. It has anti-tumor, hepatoprotective and anti-inflammatory activity particularly as an anti-tumor agent candidate[22].

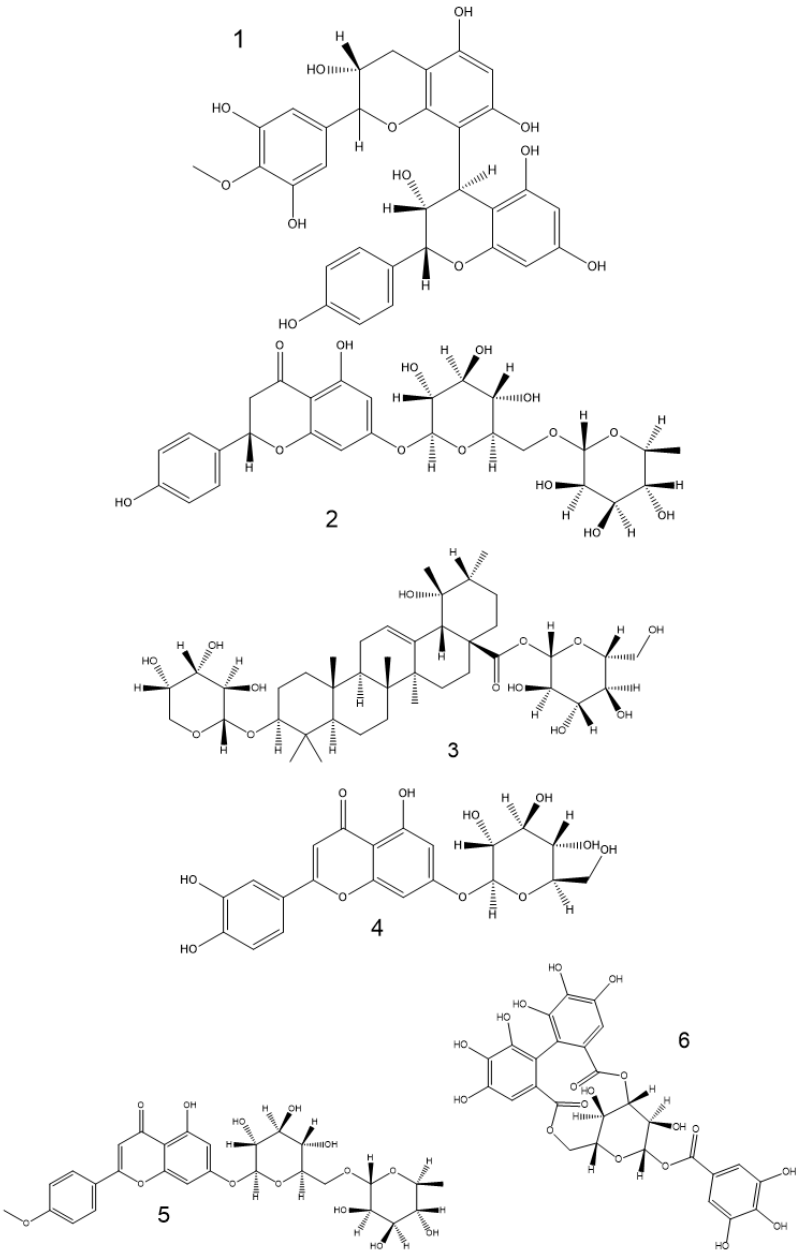

Figure 3. Chemical structure of ligands without inhibitory activity

3.4. DRUG DESIGN

Thanks to advances, the process of finding a new drug or therapeutic compound is no longer slow and costly (both in health and economic terms), thanks to computational advancement it is much faster, more accurate and efficient. By locating the active centre of the pharmaceutically targeted molecules (in this case, the Mpro protease from SARS-CoV-2), the search for an effective inhibitor begins easily.

This significantly accelerates the clinical phase, as well as providing much more advanced compounds for this phase than other methods. The process (more exactly, the process followed in the previous study[1]) can be subdivided in the following scheme:

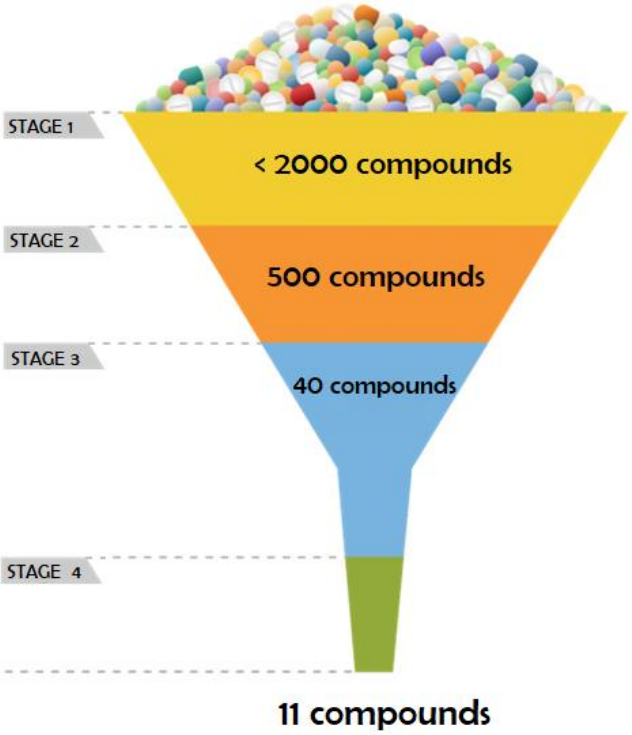

Figure 4. Process to obtain the list of 11 potential inhibitors

3.4.1. Stage 1: Database

In the case of this work, as we are looking for inhibitors present in natural products, we will use the Selleck Database of Natural Compounds to start the process. This database consists of approximately 2000 compounds.

3.4.2. Stage 2: Docking

Next, to continue the process of obtaining potential inhibitors of a given substance is docking, which consists of testing all possible compounds from the database in different forms and poses (a pose is a rearrangement of the molecule in the space, different orientation or distribution of ramifications) to see whether or not they bind to the active centre of the receptor, in this case, the Mpro protease of the SARS-CoV-2 virus.

For this purpose, a first free binding energy is approximately calculated of each compound and each pose with the receptor. This process, called scoring, is a first approximation, as the functions used are not very accurate, as they have the following cons:

- Consider the compound without solvent (water)
- System with stiffness without ligand adaptation

3.4.3. Stage 3: Molecular Dynamics

The next step is to perform molecular dynamics. This operation is extensively used and is based on the calculation and simulation of the behaviour of a substance (in this case the potential inhibitor and the receptor) over time, seeing if and how binding takes place.

This will require solving Newton's equations of motion for each instant and for each molecule and pose, under different conditions and over different times. This is the most computationally expensive and time-consuming part.

From this new calculation we will extract a very small group of compounds that fulfil the binding of the ligand to the receptor throughout the molecular dynamics in conditions similar to physiological ones (temperature between 0 and 300 K). For each instant t (time) along the molecular dynamics, a "picture" is obtained, i.e., the coordinates and velocities of the atoms and of the set of atoms for a given t (time).

3.4.4. Stage 4: Free binding energy calculation

Once we have generated the different favourable "poses" and "structures", it will be necessary to calculate the energies of each of the images, to obtain a quantification of the binding, to order them according to their ligand (inhibitor)-receptor binding energy. To do this, we will make use of different computational methods of calculation, which we will combine and test different parameters and methods until we find the combination capable of ordering them correctly.

This correct order comes from a previous study where the list we are going to use was obtained and its energy was initially calculated, and its inhibitory activity was studied in in vivo assays1. From this work, we will consider the last 100 nanoseconds of simulation, and it will be with respect to which we will work. These 100 ns represent 5000 "structures".

The binding energy is calculated by the following formula:

$$\Delta G_{binding} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand} \qquad \text{Equation 1}$$

Where each ΔG is calculated as follows:

$$\Delta G_i = \Delta H - T\Delta S + \Delta G_{solv} \qquad \text{Equation 2}$$

Since the complex, receptor and ligand are all biomolecules and therefore have biological action, the medium is a factor to be considered, but it has been tested during the molecular dynamic's simulations and therefore, disregarding it in the calculations is not so wrong, since it has already been considered before. This is an error principle, but computationally speaking considering the solvent molecules is not efficient and the approximation is very close.

The methodology to be followed is an end-point calculation because only the positions of the atoms are considered, neither the path they have taken nor their past is considered, simply the place they occupy and the energy they have at the moment of interest.

3.5. FREE BINDING ENERGY FACTORS

As described in 3.4.4, free binding energy has enthalpic factor, entropic factor and the free binding solvation energy, which are developed below:

3.5.1. Enthalpic factor

The enthalpy factor, ΔH will be considered from two points of view: a treatment according to Molecular Mechanics and a path according to Quantum Mechanics.

3.5.1.1. Molecular Mechanics

According to the Molecular Mechanics (MM)[23], particles are considered to be point-like and their motions follow the laws of classical physics, i.e., concerted and concrete. In addition, the bonds are harmonic springs and the interactions between particles are fundamentally divided into two: the electrostatic interaction and the Van der Waals interactions, dividing this enthalpic factor into two contributions:

$$\Delta H_i = \Delta H_{elec}^{gas} + \Delta H_{VdW}^{gas} \text{º} \qquad \text{Equation 3}$$

3.5.1.1.1. Electrostatic interaction

Related to the electric force, responsible for the attraction or repulsion between charged objects. This force came through Coulomb, with his famous law:

$$F = -k \frac{q_1 q_2}{r^2} \qquad \text{Equation 4}$$

K being the Coulomb constant and depends on the relative permittivity, as follows: $k = \frac{1}{4\pi\varepsilon}$ and in the IS (International System) it has units of Nm^2/C^2. Likewise, q is the charge on each particle and r is the distance separating these charges. Like all forces, it will have units of N (Newton). According to this law, two particles of the same sign will repel each other, and two particles of different signs will attract each other.

3.5.1.1.1. Van der Waals interaction

These forces (attractive or repulsive) occur between different particles (i.e. they are intermolecular forces) depending mainly on the distances that separate them. These forces are weak and very susceptible, in fact, at long distances they are imperceptible, only existing at very short distances[24]. This force is calculated following the Lennard-Jones equation, which gave the principle of molecular interactions, giving a repulsion/attraction potential dependent on the distance of the two particles. This equation has several forms, but the most used computationally is:

$$V_{LJ}(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \qquad \text{Equation 5}$$

Where A and B are parameters to be optimised according to the system and r is the distance between the particles. The first term refers to Pauli repulsion, while the second quantifies the attractive force, represented by the dispersion forces.

Each of these interactions must be calculated for each pair of particles present in the system, and given the large number of particle pairs involved, this will be one of the main reasons for the high computational demand required for each of these calculations.

From a philosophical point of view, this methodology is not close to reality, but computationally speaking it works very well, since its functions are optimised to represent experimental phenomena giving great accuracy. Therefore, it is a valid point of view, which will certainly be considered in the present work.

3.5.1.2. Quantum Mechanics

The other philosophy that can be adopted to carry out this calculation is a much more modern theory (20th century) which was born from the contradictions generated by classical physics[25]. It is characterised by treating particles as waves, i.e. they do not follow a Newtonian movement as have wave properties, being particles in themselves as they have mass. The main characteristics that we are going to consider with regard to our system are as follows:

- The wave considers multiple possible positions for the same particle, each with an associated probability, so exact position is unknown. This is related to Heisenberg's uncertainty principle.

- The particles do not have a radius, because as their positions are not concerted, the radius of an atom cannot be defined. It exists a space region where the particle can be found, with the probability defined before.

Therefore, to find the energy of the system and, consequently, the enthalpy factor, it will be necessary to calculate for each atom its energy, which is obtained by solving the Schrödinger equation, which will be presented in its most general form:

$$E\psi = \hat{\mathrm{H}}\psi \qquad \text{Equation 6}$$

From the time-independent Schrödinger equation the energy of each particle will be obtained. This equation can be solved in many ways and with many methods, the one that will be used in the following work is the Hartree-Fock method. The computational method by which this energy will be quantified is PM6 (Parameterization Method 6)[26]. This is a semi-empirical method that follows traditional "Dewar-style" methodology such as the MNDO, AM1, PM3, SAM1 and PM5 methods. While it is focused and parameterised with more emphasis on biological systems, it works well for general systems.

Some features that make it different from previous methods are:

- Careful examination of the experimental data to ensure quality and consistency.

- When data is lacking, use ab initio and DFT.

- New experimental data is included in parameterization

- Core repulsion function is adjusted by addition of new pairwise interactions.

- Molecular mechanics correction terms for certain difficult cases.

This method has been tested and has given better results than other methods.

Given the complexity, this method cannot be applied for the whole system, as there are thousands of particles and the calculation would be unachievable. Therefore, to be computationally efficient without losing precision, several different considerations will be made:

- The ligand, being a small molecule, can itself be considered as quantic (QM).

- The receptor is where the problem lies and we will consider the possible interactions that occur at the active centre, i.e. where the ligand binds to the receptor. It is decided that these interactions will occur between residues that are at a certain distance from the ligand. This will be studied at two different distances: 5 Å (angstrom) and 4 Å (angstrom). These distances have been chosen taking into account that interactions normally occur at about 3 angstroms, therefore, taking extra residues to cover possible interferences and get as close as possible to reality. In the following image it can be seen which area of the receptor will be considered QM (with a higher thickness) and in yellow the potential inhibitor, in this case, Amentoflavone.

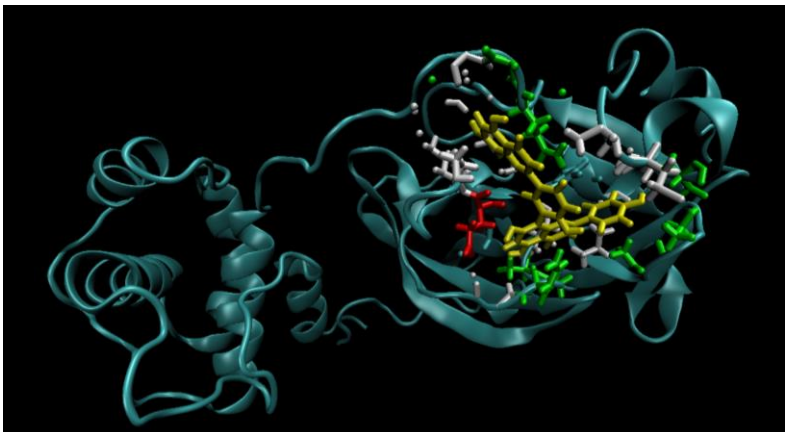

Figure 5. 3D representation of QM receptor and ligand

The receptor will be considered as QM only in calculations using the PM6-DH+ method, an improved PM6 method that includes corrections due to the polarisation of the receptor, considering the residues that are QM and the charge they generate in the quantum part (residues listed in Appendix 1 and 2).

3.5.2. Entropic factor

In order to calculate the enthalpy factor[27], it will be necessary to calculate the normal modes of vibration for each fragment, optimising the structures. Once these vibration frequencies have been calculated, the entropy will be calculated using the following equation:

$$S_{vib} = -R\ln\left(1 - e^{-h\upsilon_o/kT}\right) + \frac{N_A h\upsilon_o e^{-h\upsilon_o/kT}}{T\left(1-e^{-h\upsilon_o/kT}\right)} \qquad \text{Equation 7}$$

This entropy will be necessary, since it is a magnitude that must always be considered, even though functions that calculate it are very expensive, computationally speaking. It must be borne in mind that the entropy will be closely related to the torsional and rotational freedom of the different molecules under study.

3.5.3. Solvation correction

It is the correction related to the energy of solvating the atoms and the different residues of the molecule. This factor can be very computationally expensive, so different methodologies have been developed that are close enough to make computational savings cost-effective. This solvation energy will be broken down into two contributions:

$$\Delta G_{solv} = \Delta G_{polar} + \Delta G_{no\ polar}$$

For these methods it is important to define the concept of effective radius, which will be used to calculate the different energies.

3.5.3.1. Radius

When calculating these parameters using the AMBER program, it is necessary to introduce what radius we are going to treat for each atom. The effective Born radius of an atom reflects the degree of integration of an atom within the molecule, i.e. how tightly it is surrounded within the molecule. For an isolated ion, this radius is equal to the Van der Waals radius (ρ_i).

In the present work we are going to deal with different radii established within the AMBER programme, such as the bondi, mbondi, mbondi2, mbondi3.

3.5.3.2. Polar solvation

It is the free energy associated with first eliminating all the charges in the vacuum and then adding them in the continuum of the medium. The calculation method traditionally used is by solving the Poisson-Boltzmann equation. The great disadvantage of this equation is that it is very costly to solve, since it is a differential equation and the conditions needed to calculate it vary according to the conformation of the molecule.

$$\nabla\varepsilon(r)\nabla\phi(r) - \varepsilon(r)K^2\sinh[\phi(r)] + \frac{4\pi\rho(r)}{kT} = 0 \qquad \text{Equation 8}$$

$$K^2 = 8\pi q^2 I/\varepsilon(r)kT$$

On the other hand, AMBER's designers have generated a computational model approximating the Poisson-Boltzmann method, with changes to make it more computationally efficient. In the new Generalised Born (GB) model, each atom in the molecule is represented as a sphere of radius ri with a charge qi at its centre. In each atom is assumed to be uniformly filled with a material of dielectric constant (ε) equal to 1 and the solvent surrounding the molecule is given a high dielectric constant value of 80. The formula of the GB model follows the following analytical formula:

$$\Delta G_{polar} \approx -\frac{1}{2}\sum\frac{q_i q_j}{f_{GB}(r_{ij},R_i,R_j)}\left(1 - \frac{\exp[-kf_{GB}]}{\varepsilon}\right) \qquad \text{Equation 9}$$

Where r_{ij} is the distance between atoms *i* and *j*, r_i is what we defined earlier as the effective Born radius, and f_{GB} is a parameter-dependent function. The electrostatic effects of the salt are incorporated by means of the Debye-Huckel screening parameter *k*. From the diversity of parameters that can be chosen and that are incorporated into the AMBER programme, the different GB are born. In the present work we will test GB 1, GB 2, GB 5, GB 7 and GB 8[28].

- GB1: to prepare this method, you will have to put in the input igb = 1. By default, it will consider a value of 0.0072 for α and a radiii=mbondi. These parameters were described by Tsui and Case[28].

- GB2: radii is re-scaled to account for the interstitial spaces between atom speheres missed by the last method. Then, to develop this method, igb must be changed from 1 to 2, and radii = mbondi2.

- GB5: In some tests, this method showed a better approximation to the Poisson- Boltzmann calculation for a list of proteins. The difference between this method and the last one is α, which is 0,005 and igb is changed into igb=5.

- GB7: In this method, the radii was optimized, using a correction term that can eliminate interstitial regions of high dielectric smaller than a solvent molecule. Then, the only change from last method is igb=7 and radii=bondi.

- GB8: a list of new parametres were described for a better approximation of the behaviour of H,C,N,O,S and P elements. Then, in this method igb = 8 and radii = mbondi3.

3.5.3.3. Non-polar solvation

This polar ΔGnon comes from the combined effect of favourable van der Waals attraction between solute and solvent molecules and the unfavourable cost of breaking the solvent structure, in this case water, around the solute. In AMBER, this contribution is proportional to the SASA (Solvent Accessible Surface Area) computed by an LCPO (Linear Combination of Pairwise Overlaps)[29] algorithm that will give an analytical approximation to the surface area.

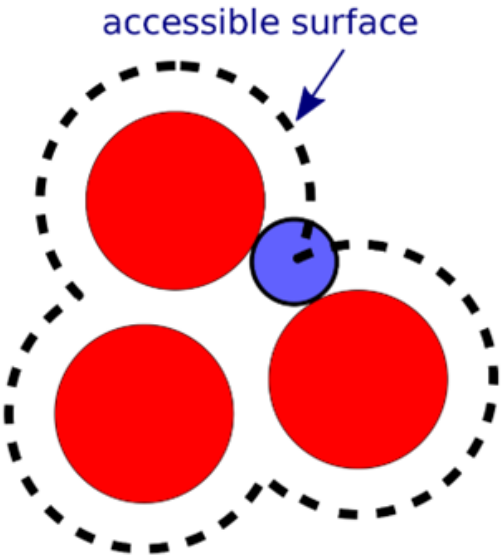

Figure 6. Representation of SASA

This ratio shall be indicated as follows:

$$\Delta G_{no\ polar} = \alpha SASA + \beta$$

Where α and β are parameters to be optimised. In the present work, α varies between 0,0072 and 0,005[28].

4. OBJECTIVES

The main objective of the present work is to find the calculation method and the appropriate parameters able to order a series of inhibitors (from natural products) of SARS-CoV-2 main protease considering their activity shown in vivo by QM-MM/GB-PB/SA methods, thus being able to find new compounds and giving a potential solution to this disease and future compounds.

If no such combination is found, draw different conclusions based on the parameters that are permuted and give a method that orders them in a better way than the starting one.

5. EXPERIMENTAL SECTION

5.1. PROGRAMMES USED

5.1.1. AMBER

The term AMBER[30] refers to two concepts; firstly to a set of force fields applied in molecular mechanics in the simulation of biomolecules, and the second term refers to a set of molecular simulation programs. In this work we will use version 18, i.e. AMBER18. Within this software package we will use the one known as AmberTools, which is the software related to data analysis. The AmberTools packages used will be:

- Ante-MMPBSA. This program will be able to produce the molecule topologies under the conditions required. In this work, it will create the topology files according to the radii (mentioned above) of both the receptor, the ligand, and the receptor-ligand complex without water, since, as mentioned above, it will facilitate the calculation without losing veracity in practice.

- MMPBSA.py; main data analysis tool, as it will provide the results of the different energies and will allow to collect and analyse them.

- CPPTRAJ; this tool will be useful when generating the different pdb files of the molecules, as its multiple options allow you to adjust the files to your needs.

5.2. RESULTS

5.2.1. Tests

With all this explained, and to reduce the quantity of calculations to be carried out, a series of calculations have been carried out as tests, with which it has been decided how to study the different variations in the most efficient way possible and without leaving out any possible combination. These tests have been done with 4 of the inhibitors, which are:

- Epigallocatechin gallate
- Amentoflavone
- Proanthocyanidins
- Narirutin

In these tests, both the differences caused in the binding energy by the different GBs or PB , different factors α of the SASA or the different radii of the atoms have been checked. The feasibility of the QM methods compared to MM ones and the combinations between each parameter have also been tested.

To sum up, the following variables are considered:

- GB; can be 1, 2, 5, 7 y 8.
- PB / GB
- α (SASA) = 0,0072-0,005
- radii = bondi, mbondi, mbondi2, mbondi3
- QM only ligand and ligand + receptor (5 and 4 angstroms) with PM6 and PM6-DH+

The list of all the tests/methods (which are giving the results listed below) realized is in Appendix 3.

The order desired for the free binding energy in these tests should be:

$$\Delta G_{b\ epigallocatechin\ gallate} < \Delta G_{b\ amentoflavone} < \Delta G_{b\ proanthocyanidins} < \Delta G_{b\ narirutin}$$

as explained in the Introduction (Section 3.3). Every number calculated has an error of 0.1 kcal/mol.

5.2.1.1. α test (SASA)

The first parameter tested was the variation of the free binding energy induced by the SASA, which can vary between 0,007 and 0,005. Then tests 2-10 and 1-34 give the following data:

Table 1. Energy for each ligand in tests 1 and 34

Ligands	Free binding energy [kcal/mol]	
	Test 1 (GB1 mbondi α=0.0072)	Test 34 (GB1 mbondi α=0.005)
Epigallocatechin gallate	-36.8	-34.9
Amentoflavone	-48.2	-46.1
Proanthocyanidins	-40.8	-39.0
Narirutin	-45.7	-43.5

Table 2. Energy for each ligand in tests 2 and 10

Ligands	Free binding energy [kcal/mol]	
	Test 2 (GB2 mbondi2 α=0.005)	Test 10 (GB2 mbondi2 α=0.0072)
Epigallocatechin gallate	-52.8	-54.6
Amentoflavone	-46.4	-48.5
Proanthocyanidins	-51.5	-53.3
Narirutin	-48.0	-50.3

As we can see in Table 1 and 2, there is no difference between using α = 0,0072 or α = 0,005 because the order is not the correct in any case. The differences between the free binding energies are approximately the same, so any error is corrected changing this parameter.

5.2.1.2. Poisson-Boltzmann test

The second test realized shows the yield of using a Poisson-Boltzmann calculation instead of a Generalized Born. Tests 9-11-12 give the following information:

Table 3. Energy for each ligand in tests 9, 11 and 12

Ligands	Free binding energy [kcal/mol]		
	Test 9 (PB mbondi)	Test 11 (PB bondi)	Test 12 (PB mbondi2)
Epigallocatechin gallate	-22.4	-46.8	-47.5
Amentoflavone	-19.1	-29.8	-30.3
Proanthocyanidins	-33.9	-52.1	-53.4
Narirutin	-27.4	-41.3	-41.8

The difference between each test is the radii set for every calculation; mbondi for test 9, bondi for test 11 and mbondi2 for test 12. The results obtained of these methods show no approximation from the ideal order of free binding energies, even changing the radii parameter.

5.2.1.3. Number of structures test

As explained in the Introduction, in the present work 5.000 structures are present, giving one energy per picture. To have a more efficient computationally talking calculation, this number of structures is reduced to 2500 in general. The following data is comparing the difference between having this 2500 structures (test 23) or its half, 1250 (test 24):

Table 4. Energy for each ligand in tests 22 and 23

Ligands	Free binding energy [kcal/mol]	
	Test 23 (2500 structures)	Test 24 (1250 structures)
Epigallocatechin gallate	-43.0	-43.2
Amentoflavone	-30.9	-30.9
Proanthocyanidins	-45.3	-45.1
Narirutin	-40.6	-40.8

As may be observed in the table, the number of structures is not making any difference on the results, the order of energies is the same and the differences between each other is almost equal. This test is important, as time is a clutch factor in computational calculation, and reducing the number of structures from 2.500 to 1.250 is saving half of the time to do other things.

5.2.1.4. Radii test

Another important factor is the one related to the radii. This parameter will determine GB and PB methods and it is very important to see its impact. Tests 33 and 34 give the following data:

Table 5. Energy of each ligand in tests 33 and 34

Ligands	Free binding energy [kcal/mol]	
	Test 33 (GB1 mbondi2 α=0.0072)	Test 34 (GB1 mbondi α=0.005)
Epigallocatechin gallate	-61.7	-34.9
Amentoflavone	-56.1	-46.1
Proanthocyanidins	-61.1	-39.0
Narirutin	-57.8	-43.5

In the previous tests, even changing the SASA parameters (which were tested before and gave the conclusion that do not vary the results), 2 radii were tested the order is wrong anyways and differences are almost equal. So, radii is not a relevant factor and is not inducing any variation.

5.2.1.5. Quantum Mechanics test

In the introduction (3.5.1) the QM option was introduced. Calculations can be developed considering the ligand following quantum mechanics, or the ligand and part of the receptor as was explained in the introduction. Test 22 is considering the residues within 5 angstroms of the inhibitor and test 36 is doing the same job but within 4 angstroms. Both test are using PM6 method, and ligand is considered QM too.

Table 6. Energy of each ligand in tests 22 and 36

Ligands	**Free binding energy [kcal/mol]**	
	Test 22 (QM residues (5 A))	**Test 36 (QM residues (4 A))**
Epigallocatechin gallate	2.7	-15.9
Amentoflavone	13.8	3.4
Proanthocyanidins	2.8	-37.3
Narirutin	-32.5	-40.7

From this test we can assume that these methods are not able in this complex simulation, as some of inhibitory active ligands have a positive free binding energy, and Narirutin, which is not active, is giving a very low free binding energy, which is wrong conceptually talking. So QM methods which consider the receptor following QM theory must be calculated with PM6-DH+, having the polarity correction, which is compulsory as Tests above showed.

All the Generalized Born methods were tested and gave different orders of binding energy, so every GB is going to be calculated for all the ligands. For every GB, the radii and α are set to its determined parameters, as were explained in the Introduction.

5.2.1.6. Entropy test

Entropy is a factor not usually calculated in computational methods, because is a very expensive calculation because of the quantity of parameters that must be determined and because its functions are not accurate, then it is not worth it. Even that, entropy has been quantified (the TΔS factor) in another test.

Table 7. Entropy factor for each ligand tested

Ligands	Entropy (TΔS) [kcal/mol]
Epigallocatechin gallate	-28.5
Amentoflavone	-29.5
Proanthocyanidins	-27.8
Narirutin	-27.7

This test shows a better entropic factor for those ligands which are not inhibitory talking. This must, as was explained before, the non-accurate methods and because other thermodynamic factors which are not controllable.

5.2.1.7. PM6-DH+ test

The only method which is giving conceptually talking good results when considering part of the receptor as QM is the PM6-DH+, but it has a considerable con, it is expensive computationally talking and it consumes a big part of the available time. Then all the PM6-DH+ GBX will be tested, for only developing the best ones. Tests 40, 24, 41, 42 and 43 are tabled below:

Table 8. Energy of each ligand in tests 33 and 34

Ligands	Free binding energy [kcal/mol]				
	Test 40 (GB1)	Test 24 (GB2)	Test 41 (GB5)	Test 42 (GB7)	Test 43 (GB8)
Epigallocatechin gallate	-30.1	-43.2	-47.0	-48.3	-42.4
Amentoflavone	-33.4	-30.9	-32.3	-29.0	-30.6
Proanthocyanidins	-34.5	-45.1	-47.9	-46.8	-37.4
Narirutin	-40.8	-40.8	-43.5	-42.1	-37.9

As results shown, all the methods are giving wrong orders (Amentoflavone is below the other ligands in almost every test). Because of this and considering the computational cost of doing this calculations, only PM6-DH+ GB1 and PM6-DH+ GB2 are going to be developed.

5.2.2. Final results

After all the tests carried out, some methodologies are used to calculate this free binding energy and are listed and tabled below. All the calculations have considered the results obtained in 5.2.1.3, because reducing the number of structures accelerates the process of obtaining the results and it is computationally more efficient. In all the tables, the ligands are listed following the order of inhibitory activity.

5.2.2.1. Generalized Born method

Conclusions extracted in 5.2.1.1 and 5.2.1.4 models these methods so, every GB will be calculated using its default parameters, which was explained in 3.5.3.2. Method 10 was the one carried out in the previous work[1].

Table 9. Energies obtained in GB methods

Ligands	Free binding energy [kcal/mol]				
	Test 1 (GB1)	Test 10 (GB2)	Test 5 (GB5)	Test 7 (GB7)	Test 8 (GB8)
Epigallocatechin gallate	-36.8	-54.6	-57.7	-58.1	-44.9
Amentoflavone	-48.2	-48.5	--48.7	-44.3	-41.1
Vitexin-2-o-rhamnoside	-33.2	-41.0	-44.0	-45.7	-35.3
Aloin	-31.2	-38.9	-40.8	-42.2	-32.9
Rhoifolin	-39.1	-37.0	-37.1	-34.0	-29.9
Proanthocyanidins	-40.8	-53.5	-57.2	-55.1	-44.4
Narirutin	-45.7	-50.3	-51.3	-51.0	-46.5
Ziyu-glycoside I	-47.0	-47.1	-47.6	-48.8	-43.0
Luteoloside	-42.0	-43.9	-44.2	-42.9	-37.5
Linarin	-34.8	-39.0	-40.7	-42.8	-34.2
Corilagin	-30.0	-33.3	-33.8	-31.2	-31.2

5.2.2.2. Poisson-Boltzmann method

Even results obtained in 5.2.1.2 two methods has been carried out to have the results anyways and test the overall availability.

Table 10. Energies obtained in PB methods

Ligands	Free binding energy [kcal/mol]	
	Test 9 (PB mbondi)	Test 11 (PB bondi)
Epigallocatechin gallate	-22.4	-46.8
Amentoflavone	-19.1	-29.8
Vitexin-2-o-rhamnoside	-21.1	-33.6
Aloin	-19.9	-31.5
Rhoifolin	-27.2	-32.2
Proanthocyanidins	-33.9	-52.1
Narirutin	-27.4	-41.3
Ziyu-glycoside I	-31.6	-38.3
Luteoloside	-29.8	-39.3
Linarin	-28.9	-37.7
Corilagin	-20.6	-34.1

5.2.2.3. PM6 method

PM6 method has been carried out as a more accurate methodology than the other ones, which is considering the ligand as QM, related to conclusions extracted in 5.2.1.5.

Table 11. Energies obtained in PM6 methods

Ligands	Free binding energy [kcal/mol]				
	Test 35 (GB1)	Test 27 (GB2)	Test 37 (GB5)	Test 38 (GB7)	Test 39 (GB8)
Epigallocatechin gallate	-36.2	-49.5	-52.7	-52.2	-41.1
Amentoflavone	-44.1	-43.7	-45.4	-39.7	-37.1
Vitexin-2-o-rhamnoside	-31.1	-36.8	-41.7	-42.4	-33.6
Aloin	-27.1	-30.4	-32.8	-33.5	-26.5
Rhoifolin	-33.0	-28.3	-30.3	-26.4	-22.7
Proanthocyanidins	-36.6	-46.1	-50.3	-47.6	-39.0
Narirutin	-42.6	-43.6	-46.0	-44.5	-41.2
Ziyu-glycoside I	-40.5	-34.8	-33.9	-42.8	-38.3
Luteoloside	-40.1	-39.1	-41.0	-38.3	-33.4
Linarin	-30.3	-31.0	-33.6	-34.8	-27.8
Corilagin	-31.1	-32.9	-34.1	-32.1	-30.5

5.2.2.4. PM6-DH+ method

PM6-DH+ it is the theoretically most close to the real behaviour of particles in the system. It has been done as explained in 5.2.1.7, considering receptor's residues from Appendix 2 as QM.

Table 12. Energies obtained in PM6-DH+ methods

Ligands	Free binding energy [kcal/mol]	
	Test 40 (GB1)	Test 24 (GB2)
Epigallocatechin gallate	-30.1	-43.2
Amentoflavone	-33.4	-30.9
Vitexin-2-o-rhamnoside	-25.6	-30.4
Aloin	-22.6	-25.3
Rhoifolin	-29.7	-24.4
Proanthocyanidins	-34.5	-45.1
Narirutin	-40.8	-40.8
Ziyu-glycoside I	-21.4	-16.3
Luteoloside	-38.3	-39.7
Linarin	-21.4	-21.5
Corilagin	-19.9	-21.0

6. DATA ANALYSIS

To analyse the results obtained in every method another test is developed. Considering there are 5 active compounds and 6 inactive compounds, results are divided in two groups , the 5 best energy score and the 6 worst ones. Then, this table is developed:

Table 13. Analyse method

	Experimental active	Experimental inactive
Predicted active	True Positive (TP)	False Positive (FP)
Predicted inactive	False Negative (FN)	True Negative (TN)

A TP is a compound within the 5 lowest energies, and which is an active compound (in in vivo tests (3.3.1)). A FN is a compound between the 6 highest energies even being experimentally active. A FP is a compound between the 5 lowest energies when it has no activity in *in vivo* tests (3.3.2). A TN is a compound within the 6 highest energies and experimentally inactive. The ideal stage is having 5 TP and 6 TN.Methods developed are scored following these rules (*Appendix 4)*:

Table 14. Precision score of every test

Method		TP	TN
Generalized Born (GB)	Test 1 (GB1)	1	2
	Test 10 (GB2)	2	3
	Test 5 (GB5)	2	3
	Test 7 (GB7)	2	3
	Test 8 (GB8)	2	3
Poisson-Boltzmann (PB)	Test 9 (PB mbondi)	0	1
	Test 11 (PB bondi)	1	2
PM6	Test 35 (GB1)	1	2
	Test 27 (GB2)	2	3
	Test 37 (GB5)	3	4
	Test 38 (GB8)	2	3
	Test 39 (GB9)	2	3
PM6-DH+	Test 40 (GB1)	2	3
	Test 24 (GB2)	2	3

Following the rule of TP and TN, method 37 (PM6 GB5 QM ligand mbondi2 α = 0,005) is the best one, but it is also important to explain what method 24 is bringing. Both results are listed below.

Table 15. Final order of ligands with its energies, using Method 37 (left) and Method 24 (right)

Compound	$\Delta G_{binding}$ [kcal/mol]	Compound	$\Delta G_{binding}$ [kcal/mol]
(-) epigallocatechin gallate*	-52.7	Proanthocyanidins	-45.1
Proanthocyanidins	-50.3	(-) epigallocatechin gallate*	-43.2
Narirutin	-46.0	Narirutin	-40.8
Amentoflavone*	-45.4	Luteoloside	-39.7
Vitexin-2-O-rhamnoside*	-41.7	Amentoflavone*	-30.9
Luteoloside	-41.0	Vitexin*	-30.4
Corilagin	-34.1	Aloin*	-25.3
Ziyu-glycoside I	-33.9	Rhoifolin*	-24.4
Linarin	-33.6	Linarin	-21.5
Aloin*	-32.8	Corilagin	-21.0
Rhoifolin*	-30.3	Ziyu-glycoside I	-16.3

*Inhibitory active

The table on the left is giving a better TP/TN score, but there are active compounds at the bottom part of the list. The table on the right is conceptually better because there are more inactive compounds at the bottom part of the list.

7. CONCLUSIONS

Finding the ultimate method who gives the exact order of ligands considering its inhibitory activity has not been achieved. But some good conclusions can be extracted from the work; method 37 gives a better result than the original method/order. Method 24 gives an important order even not being the best one, probably the final method is coming from a derivative method from that one.

There are many factors that explain why the final method is not achieved. For example, entropy is not considered because is not giving a good result. All methods have many approximations, which is inducing error, and only positive variables are considered.

Also, philosophically talking thermodynamics are always related to probability, and atomic behaviours are quite random because of atomics principles[31].

8. REFERENCES AND NOTES

1. Rubio-Martínez, J. et al. Discovery of Diverse Natural Products as Inhibitors of SARS-CoV-2 M pro Protease through Virtual Screening. Journal of Chemical Information and Modeling 61, 6094–6106 (2021).
2. Carfì, A., Bernabei, R. & Landi, F. Persistent Symptoms in Patients After Acute COVID-19. JAMA 324, 603 (2020).
3. https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html 11/06/2022
4. https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease 11/06/2022
5. Pirzada, R. H., Haseeb, M., Batool, M., Kim, M. & Choi, S. Remdesivir and Ledipasvir among the FDA-Approved Antiviral Drugs Have Potential to Inhibit SARS-CoV-2 Replication. Cells 10, 1052 (2021).
6. https://www.rcsb.org/structure/7CAM 11/06/2022
7. Tam, N. M. et al. Binding of inhibitors to the monomeric and dimeric SARS-CoV-2 Mpro. RSC Advances 11, 2926–2934 (2021).
8. https://www.healthline.com/nutrition/egcg-epigallocatechin-gallate#benefits 11/06/2022
9. Peterson, L. COVID-19 and Flavonoids: In Silico Molecular Dynamics Docking to the Active Catalytic Site of SARS-CoV and SARS-CoV-2 Main Protease. SSRN Electronic Journal (2020)
10. Xiong, X. et al. Insights Into Amentoflavone: A Natural Multifunctional Biflavonoid. Frontiers in Pharmacology 12, (2021).
11. https://www.selleckchem.com/products/vitexin-2-o-rhamnoside.html11/06/2022
12. https://en.wikipedia.org/wiki/Aloin 11/06/2022
13. John Refaat et al 1. RHOIFOLIN: A REVIEW OF SOURCES AND BIOLOGICAL ACTIVITIES. International Journal of Pharmaceutical Sciences and Research.
14. Rauf, A. et al. Proanthocyanidins: A comprehensive review. Biomedicine & Pharmacotherapy 116, 108999 (2019).
15. https://www.selleckchem.com/datasheet/narirutin-S924200-DataSheet.html 11/06/2022
16. Naturally Occurring Chemicals Against Alzheimer's Disease. (Elsevier, 2021).
17. https://www.selleckchem.com/products/ziyu-glycoside-i.html 11/06/2022
18. Sun, W. et al. Terpene Glycosides from the Roots of Sanguisorba officinalis L. and Their Hemostatic Activities. Molecules 17, 7629–7636 (2012).
19. Zhao, Y. X. et al. Bioactive luteoloside produced by Myroides odoratimimus , solvent-tolerant bacterium form the rhizosphere of Lonicera japonica. Natural Product Research 33, 3559–3562 (2019).
20. https://en.wikipedia.org/wiki/Cynaroside 11/06/2022

21. Mottaghipisheh, J. et al. Linarin, a Glycosylated Flavonoid, with Potential Therapeutic Attributes: A Comprehensive Review. Pharmaceuticals 14, 1104 (2021).
22. Li, X. et al. Corilagin, a promising medicinal herbal agent. Biomedicine & Pharmacotherapy 99, 43–50 (2018).
23. https://en.wikipedia.org/wiki/Molecular_mechanics 11/06/2022
24. https://en.wikipedia.org/wiki/Van_der_Waals_force 11/06/2022
25. Paniagua, Juan Carlos et al. Química Quàntica, Universitat de Barcelona, (2017)
26. http://www.semichem.com/ampac/pm6.php 11/06/2022
27. Classical Normal Mode Analysis: Harmonic Approximation, Colby College (2020)
28. D.A. Case Case D.A. et al. AMBER 2020, University of California, (2020)
29. Huang, H. & Simmerling, C. Fast Pairwise Approximation of Solvent Accessible Surface Area for Implicit Solvent Simulations of Proteins on CPUs and GPUs. Journal of Chemical Theory and Computation 14, 5797–5814 (2018).
30. https://ambermd.org/ 11/06/2022
31. Ben-Naim, Arieh, La entropía desvelada, World Scientific Publishing, (2007).

9. ACRONYMS

MM: Molecular Mechanics

QM: Quantum Mechanics

GB: Generalized Born

PB: Poisson Boltzmann

EGCG: Epigallocatechin-gallate

SASA: Solvent Accessible Surface Area

LCPO: Linear Combination of Pairwise Overlaps

WHO: World Health Organization

UN: United Nations

PDB: Protein Data Bank

R: Residue

C: Charge

APPENDICES

APPENDIX 1: RESIDUES AT 5 ANGSTROMS

Vitexin-2-o-rhamnoside		EGCG		Rhoifolin		Corilagin		Ziyu-glycoside	
R	C	R	C	R	C	R	C	R	C
27LEU		39PRO		25THR		41HIP	1	40ARG	1
40ARG	1	40ARG	1	26THR		54TYR		41HIP	1
41HIP	1	41HIP	1	27LEU		164HIE		45THR	
44CYS		44CYS		41HIP	1	165MET		52PRO	
45THR		47GLU	-1	46SER		166GLU	-1	53ASN	
46SER		85CYS		49MET		167LEU		54TYR	
49MET		145CYS		118TYR		168PRO		54TYR	
145CYS		164HIE		119ASN		187ASP	-1	85CYS	
163HID		165MET	-1	142ASN		188ARG	1	164HIE	
164HIE		166GLU	-1	143GLY		189GLN		165MET	
165MET		167LEU		145CYS		190THR		166GLU	-1
166GLU	-1	168PRO		164HIE				167LEU	
181PHE		187ASP		165MET				168PRO	
186VAL		188ARG		166GLU	-1			187ASP	-1
187ASP	-1	189GLN		167LEU				188ARG	1
188ARG	1			168PRO				189GLN	
190THR				170GLY					
191ALA				189GLN					
192GLN									

Narirutin		Linarin		Luteoloside		Aloin		Amento-flavone		Proantho-cyanidie	
R	C	R	C	R	C	R	C	R	C	R	C
24THR		19GLN		25THR		145CYS		190THR		41HIP	1
25THR		26THR		26THR		165MET		191ALA		46SER	
26THR		27LEU		27LEU		164HIE		189GLN		47GLU	-1
27LEU		28ASN		28ASN		41HIP	1	50LEU		46SER	
41HIP	1	28ASN		41HIP	1	187ASP	-1	49MET		49MET	
43ILE		41HIP	1	42VAL		40ARG	1	45THR		50LEU	
45THR		47GLU	-1	45THR		188ARG	1	24THR		165MET	
46SER		119ASN		46SER		189GLN		168PRO		167LEU	
119ASN		120GLY		49MET		44CYS		165MET		171VAL	
142ASN		143GLY		118TYR		45THR		41HIP	1	172HIE	
143GLY		145CYS		119ASN		49MET		167LEU		173ALA	
145CYS		163HID		120GLY		54TYR		166GLU	-1	182TYR	
164HIE		164HIE		143GLY		46SER		42VAL		185PHE	
165MET		165MET		145CYS				25THR		187ASP	-1
166GLU	-1	166GLU	-1	164HIE				27LEU		188ARG	1
				165MET				143GLY		189GLN	
				166GLU	-1			140PHE		190THR	
								141LEU		191ALA	

APPENDIX 2: RESIDUES AT 4 ANGSTROMS

Vitexin		EGCG		Rhoifolin		Corilagin		Ziyu-glycoside	
R	C	R	C	R	C	R	C	R	C
41HIP	1	40ARG	1	26THR		41HIP	1	40ARG	1
46SER		41HIP	1	27LEU		165MET		41HIP	1
49MET		47GLU	-1	41HIP	1	166GLU	-1	44CYS	
145CYS		85CYS		46SER		167LEU		52PRO	
164HIE		164HIE		49MET		168PRO		53ASN	
165MET		165MET		119ASN		187ASP	-1	85CYS	
166GLU	-1	166GLU	-1	142ASN		188ARG	1	164HIE	
187ASP	-1	187ASP	-1	143GLY		189GLN		165MET	
189GLN				165MET		190THR		166GLU	-1
192GLN				166GLU	-1			167LEU	
				167LEU				168PRO	
				168PRO				181PHE	
				189GLN				186VAL	
								187ASP	-1
								188ARG	1
								189GLN	

Narirutin		Linarin		Lutheo-loside		Aloin		Amento-flavone		Proantho-cyanidie	
R	C	R	C	R	C	R	C	R	C	R	C
24THR		26THR		25THR		40ARG	1	24THR		41HIP	1
25THR		27LEU		26THR		41HIP	1	25THR		46SER	
26THR		28ASN		27LEU		44CYS		27LEU		47GLU	-1
27LEU		41HIP	1	28ASN		46SER		41HIP	1	49MET	
41HIP	1	47GLU	-1	41HIP	1	49MET		44CYS		165MET	
44CYS		119ASN		46SER		54TYR		45THR		167LEU	
45THR		120GLY		49MET		165MET		49MET		171VAL	
119ASN		143GLY		118TYR		167LEU		50LEU		173ALA	
142ASN		145CYS		119ASN		187ASP	-1	142ASN		183GLY	
143GLY		163HID		143GLY		188ARG	1	143GLY		184PRO	
168PRO		145CYS		165MET		145CYS		189GLN		185PHE	
181PHE		166GLU	-1	166GLU	-1	165MET				187ASP	-1
186VAL				192GLN		166GLU	-1			189GLN	
187ASP	-1									190THR	
188ARG	1									192GLN	
189GLN										194ALA	

APPENDIX 3: LIST OF METHODS

Legend			
Number	**Method**	**Radii**	**α**
1	MM GB1	mbondi	0,0072
2	MM GB2	mbondi2	0,005
5	MM GB5	mbondi2	0,005
7	MM GB7	bondi	0,005
8	MM GB8	mbondi3	0,005
9	MM PB	mbondi	-
10	MM GB2	mbondi2	0,0072
11	MM PB	bondi	-
12	MM PB	mbondi2	-
20	PM6* GB2 LIG**	mbondi2	0,005
21	PM6-DH+ GB2 LIG	mbondi2	0,005
22	PM6 GB2 REC 5A*** + LIG	mbondi2	0,0072
23	PM6-DH+ GB2 REC 5A + LIG 2500 structures	mbondi2	0,0072
24	PM6-DH+ GB2 REC 5A + LIG 1250 structures	mbondi2	0,0072

25	PM6 GB2 REC 5A + LIG	mbondi2	0,005
26	PM6 GB1 REC 5A + LIG	mbondi	0,0072
27	PM6 GB2 LIG	mbondi2	0,0072
28	PM6 GB1 LIG	mbondi2	0,0072
29	PM6 GB5 LIG	mbondi2	0,0072
30	PM6 GB7 LIG	mbondi2	0,0072
31	PM6 GB8 LIG	mbondi2	0,0072
32	PM6-DH+ GB1 REC 5A + LIG 2500 structures	mbondi2	0,0072
33	MM GB1	mbondi2	0,0072
34	MM GB1	mbondi	0,005
35	PM6 GB1 LIG 1250 structures	mbondi	0,0072
36	PM6 GB2 REC 4A + LIG 1250	mbondi2	0,0072
37	PM6 GB5 LIG	mbondi2	0,005
38	PM6 GB7 LIG	bondi	0,005
39	PM6 GB8 LIG	mbondi3	0,005
40	PM6-DH+ GB1 REC 5A + LIG 1250 structures	mbondi	0,0072
41	PM6-DH+ GB5 REC 5A + LIG 1250 structures	mbondi2	0,005

42	PM6-DH+ GB7 REC 5A + LIG 1250 structures	bondi	0,005
43	PM6-DH+ GB8 REC 5A + LIG 1250 structures	mbondi3	0,005

*QM method

**LIG means only LIG is considered QM

*** 5A = QM on residues at 5 angstroms

4A = QM on residues at 4 angstroms

APPENDIX 4: METHODS' RANKING

Generalized-Born									
Ligand	Test 1 energy [kcal/mol]	Ligand	Test 10 energy [kcal/mol]	Ligand	Test 5 energy [kcal/mol]	Ligand	Test 7 energy [kcal/mol]	Ligand	Test 8 energy [kcal/mol]
443	-48,2	**106**	-54,6	**106**	-57,7	**106**	-58,1	**1373**	-46,5
1312	-47,0	**850**	-53,3	**850**	-57,2	**850**	-55,1	**106**	-44,9
1373	-45,7	**1373**	-50,3	**1373**	-51,3	**1373**	-51,0	**850**	-44,4
1486	-42,0	**443**	-48,5	**443**	-48,7	**1312**	-48,8	**1312**	-43,0
850	-40,8	**1312**	-47,1	**1312**	-47,6	**1044**	-45,7	**443**	-41,1
1253	-39,1	**1486**	-43,9	**1486**	-44,2	**443**	-44,3	**1486**	-37,5
106	-36,8	**1044**	-41,0	**1044**	-44,0	**1486**	-42,9	**1044**	-35,3
1409	-34,8	**1409**	-39,0	**15477**	-40,8	**1409**	-42,8	**1409**	-34,2
1044	-33,2	**15477**	-38,9	**1409**	-40,7	**15477**	-42,2	**15477**	-32,9
15477	-31,2	**1253**	-37,0	**1253**	-37,1	**1253**	-34,0	**1258**	-31,2
1258	-30,0	**1258**	-33,3	**1258**	-33,8	**1258**	-31,2	**1253**	-29,9

Poisson-Boltzmann			
Ligand	Test 9 energy [kcal/mol]	Ligand	Test 11 energy [kcal/mol]
850	-33,9	850	-52,1
1312	-31,6	106	-46,8
1486	-29,8	1373	-41,3
1409	-28,9	1486	-39,3
1373	-27,4	1312	-38,3
1253	-27,2	1409	-37,7
106	-22,4	1258	-34,1
1044	-21,1	1044	-33,6
1258	-20,6	1253	-32,2
15477	-19,9	15477	-31,5
443	-19,1	443	-29,8

PM6-DH+			
Ligand	Test 40 energy [kcal/mol]	Ligand	Test 24 energy [kcal/mol]
1373	-40,8	850	-45,1
1486	-38,3	106	-43,2
850	-34,5	1373	-40,8
443	-33,4	1486	-39,7
106	-30,1	443	-30,9
1253	-29,7	1044	-30,4
1044	-25,6	15477	-25,3
15477	-22,6	1253	-24,4
1312	-21,4	1409	-21,5
1409	-21,4	1258	-21,0
1258	-19,9	1312	-16,3

PM6									
Ligand	Test 35 energy [kcal/mol]	Ligand	Test 27 energy [kcal/mol]	Ligand	Test 37 energy [kcal/mol]	Ligand	Test 38 energy [kcal/mol]	Ligand	Test 39 energy [kcal/mol]
443	-44,1	106	-49,5	106	-52,7	106	-52,2	1373	-41,2
1373	-42,6	850	-46,1	850	-50,3	850	-47,6	106	-41,1
1312	-40,5	443	-43,7	1373	-46,0	1373	-44,5	850	-39,0
1486	-40,1	1373	-43,6	443	-45,4	1312	-42,8	1312	-38,3
850	-36,6	1486	-39,1	1044	-41,7	1044	-42,4	443	-37,1
106	-36,2	1044	-36,8	1486	-41,0	443	-39,7	1044	-33,6
1253	-33,0	1312	-34,8	1258	-34,1	1486	-38,3	1486	-33,4
1044	-31,1	1258	-32,9	1312	-33,9	1409	-34,8	1258	-30,5
1258	-31,1	1409	-31,0	1409	-33,6	15477	-33,5	1409	-27,8
1409	-30,3	15477	-30,4	15477	-32,8	1258	-32,1	15477	-26,5
15477	-27,1	1253	-28,3	1253	-30,3	1253	-26,4	1253	-22,7

Active Compounds (NATSEL code) = Green background

Inactive Compounds (NATSEL code) = Pink background