INSTITUTO UNIVERSITARIO DE BIO-ORGÁNICA "ANTONIO GONZÁLEZ".
UNIVERSIDAD DE LA LAGUNA

MEMORIA DEL TRABAJO FIN DE GRADO

DISEÑO DE INHIBIDORES SELECTIVOS DE CK1ε

END OF DEGREE PROJECT TITLE

DESIGN OF SELECTIVE INHIBITORS OF CK1ε

GRADO EN FARMACIA
Curso 2016/2017

Jonay Rodríguez Chinea
Tutor: Dr. José M. Padrón

La presente memoria de investigación ha sido realizada por el alumno Jonay Rodríguez Chinea durante el curso académico 2016-2017 en las instalaciones del Instituto Universitario de Bio-Orgánica "Antonio González" (IUBO-AG) y bajo la dirección del Dr. José M. Padrón Carrillo. El trabajo forma parte de la línea de investigación *Diseño, descubrimiento y evaluación de fármacos anticancerígenos.*

La Laguna, a 4 de septiembre de 2017.

José M. Padrón

2

AGRADECIMIENTOS

A mis tutores Dr. José M. Padrón Carrillo y Dr. Miguel X. Fernandes, muchas gracias por su dedicación, por su pasión investigadora y por descubrirme un campo tan interesante como la química computacional.

A mis amigos y compañeros de la carrera, en especial a Elena que me ha apoyado desde el principio e instigado a no abandonar nunca.

A mi familia, que me ha apoyado incondicionalmente toda mi vida, en especial a mi tía Luisa, por la ayuda y el apoyo incondicional en mi vida, tanto personal como académica.

ABSTRACT

Computational techniques play an important role in early stages of new drug discovery by means of virtual screening. In this process of analysis, diverse compound collections are selected for a later experimental screening, getting to optimize and identify new molecules with biological activity (hits) in a much faster way than conventional methods. For this reason, series of compounds whose inhibitory activity is known are selected to create a pharmacophore by molecular modelling based on the different binding sites of casein kinase 1 epsilon (CK1ε).

Firstly, we collected information from previous studies to create a database of compounds with known activity against CK1ε. Then, we optimized these compounds by means of an energy minimisation progress.

The pharmacophore model integrates the Quantitative Structures Activity Relationships (QSAR) to get valuable information about the steric and electronic influence as well as hydrophilic and hydrophobic features, allowing to give a better estimation in the search of the molecules with unknown activity to CK1ε. This model was subsequently refined through a retrospective analysis with a test database. In order to create this test database, we selected a second set of molecules with inhibitory activity to casein kinases and they were merged with another database composed by inactive molecules and active molecules for other targets.

 With the pharmacophore model, once established and refined, we made a screening of the BioLab's database to filter in order to discard the molecules that do not interact with the active site of casein kinase. The resulting hits are calculated from the scaffold similarities to our pharmacophoric groups present in the pharmacophore model.

As a result, 48 candidate molecules to bind and inhibit the casein kinase 1 epsilon were found in BioLab's database. These results make it possible to establish an initial approach the discovery of new drugs with biological activity. However, being a theoretical model based on structural predictions, post-approval studies with specific bioassays with diverse cell lines must be done. For this reason, we proposed a set of five candidates proposed for conducting these tests.

RESUMEN

Las técnicas computacionales juegan un importante papel en las primeras etapas del descubrimiento de nuevos fármacos por medio del cribado virtual. En este proceso analítico diversas moléculas son seleccionadas para un posterior cribado experimental, logrando optimizar e identificar moléculas con actividad biológica (*hits*) de manera mucho más rápida respecto a los métodos convencionales. Por este motivo, se seleccionan una serie de sustancias con actividad inhibidora conocida para crear un modelo de farmacóforo por moldeado molecular basado en los sitios de unión de la caseina quinasa 1 épsilon (CK1ε).

Primero, hemos recopilado información de la literatura para crear una base de datos de componentes activos frente la CK1ε. A continuación, optimizamos estos componentes por medio de un proceso de minimización de la energía.

El modelo de farmacóforo integra las *"Quantitative Structures Activity Relationships"* (QSAR) para obtener información sobre la influencia electrónica y estérica, como las fuerzas hidrofóbicas e hidrofílicas, permitiendo ofrecer una mejor estimación en la búsqueda de las moléculas con actividad desconocida para CK1ε. Posteriormente, el modelo fue refinado mediante un análisis retrospectivo valiéndose de una base de datos de prueba. Esta base fue creada a partir de una segunda tanda de moléculas con actividad inhibidora conocida para las caseina quinasas y se fusionaron posteriormente con una base de datos compuesta por sustancias inactivas o activas para otras dianas.

 El modelo de farmacóforo, una vez creado y corregido, hace un cribado en la base de datos del BioLab para filtrar las moléculas que no interaccionen con el sitio de unión de la caseína quinasa. Las posteriores moléculas resultantes son calculadas mediante las similitudes estructurales con nuestros grupos farmacofóricos presentes en el modelo de farmacóforo.

Como resultado, el modelo encontró 48 moléculas candidatas a unirse e inhibir la caseina quinasa 1 épsilon en la base de datos de BioLab. Este resultado permite establecer una primera aproximación al descubrimiento de nuevas moléculas con actividad biológica. Sin embargo, al ser un modelo teórico basado en predicciones estructurales se deben realizar posteriores estudios biológicos con distintas líneas celulares de forma específica. Por este motivo, en el trabajo se proponen las cinco mejores moléculas candidatas para realizar dichos ensayos.

CONTENTS

List of Symbols and Acronyms

CK1ε: Casein Kinase 1 epsilon
MOE: Molecular Operating Environment
SMILES: Simplified Molecular-Input Line-entry System
K_D: Dissociation constant
IC_{50}: Half maximal inhibitory concentration
PES: Potential Energy Surface
τ: Degree of torsion
OPLS: Optimized Potentials for Liquid Simulations
IUPAC: International Union of Pure and Applied Chemistry
NBCI: National Center for Biotechnology Information
RMS: Root Mean Square
RMSD: Root Mean Square Deviation

1. INTRODUCTION

The drug discovery and development path from a new molecule is long and expensive. In order to increase the efficiency of this process, new technologies such as chemoinformatics and computational techniques are used to identify hits molecules with a given biological activity. The combination of high throughput screening and the information created from computational methods provide an opportunity to speed the process of identifying new hits and therefore eliminate laboratory workload.

Casein kinase family of serine/threonine kinases has six human isoforms (α, δ, ε, γ1, γ2 and γ3) that are involved in cell regulation and disease pathogenesis phosphorylating serine, threonine and tyrosine residues. The inhibition of CK1ε has been involved in the circadian clock [1] and plays an important role in cellular growth and survival processes in Wnt signalling and apoptosis in pancreatic and breast cancer [2,3].

When the information about biological targets such as X-ray crystal structure and the information of inhibition concentration-response curves (e.g., an IC_{50}) are available to determine the Quantitative Structure-Activity Relationships (QSARs) to correlate molecular structure with chemical or biochemical activity [4,5].

These properties are used to generate a pharmacophore model that defines the minimal requirements for a molecule to be active and perform virtual screening to find new hits in databases [4,5].

2. HYPOTHESIS

Among the 1423 molecules that form BioLab's database there are molecules with inhibitory activity to the casein kinase 1 epsilon.

3. OBJETIVES

The objective of this study is the identification of compounds from BioLab's database with inhibitory activity for casein kinase 1 epsilon. To achieve this, we have created a list of substances with inhibitory activity at CK1ε found in different archives of biomedical and life sciences. Taking into account the diverse structures we have created a pharmacophore model with which we have searched the best candidates to later evaluate the activity with biological assays.
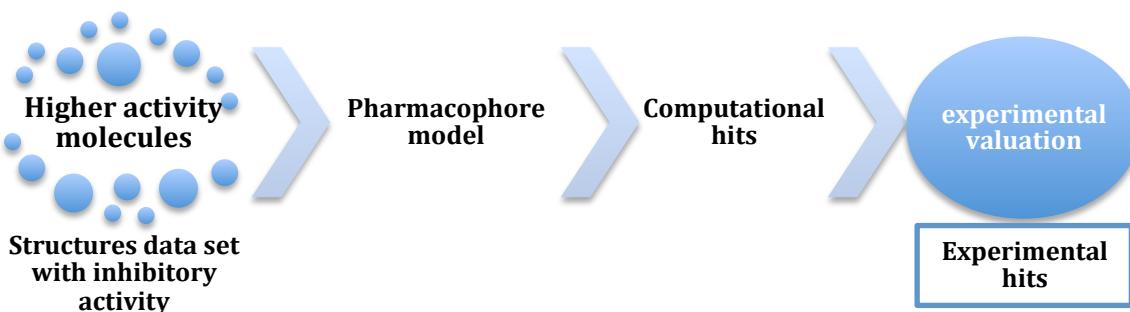
Figure 1. A schematic view of the virtual screening process.

4. MATERIALS

The pharmacophore is not a fragment of a molecule or functional groups of molecules; it is only a three-dimensional arrangement of features separated by geometric constraints. To generate the pharmacophore model, we selected the five compounds that showed the best Ki values against the CK1ε and MOE program was used. In addition to MOE, other computer tools were used to simplify the notation for describing the structure of the substances in SMILES strings (OpenBabel) and draw molecules and optimize their structures (HyperChem).

- Open Babel is a free, open-source version of the Babel chemistry file translation program. Open Babel is a project to facilitate the inter-conversion between many files formats used in molecular modelling, computational chemistry, and many related areas. [5]

- HyperChem is a complete set of molecular modelling that provides all the necessary tools for chemical simulation, calculation and visualization of molecular properties. HyperChem explores with a quantum or classical model the potential energy surfaces with a single point (geometry optimization) or transition state search calculations. [6]

- MOE is a software tool for development and deployment of chemical computation. MOE integrates flexible alignment of multiple molecules, molecular mechanics force fields, virtual screening of conformational database for compound conformations to satisfy a pharmacophore query, etc. [7]

Additionally, we used different search engines to collect activity molecules with inhibitory activity to CK1ε.

- PubMed provides free access via the internet to more than 23 million records, of which over 19 million are from the MEDLINE database of journal articles and also to other databases, such as the NCBI Bookshelf. [8]

- BindingDB is a public, web-accessible database of measured binding affinities, focused chiefly on the interactions of proteins considered to be candidate drugs-targets with ligands that are small, drug-like molecules. [9]

5. METHODOLOGY

To achieve a therapeutic target of interest to inhibit CK1ε, we will follow a series of strategy steeps:

5.1 Create a data set

To generate an effective pharmacophore model, we searched the database BindingDB [9] and PubMed [8] inhibitors to CK1ε and collected 61 structures (tab.1) ranked by potency (Ki and IC_{50}). In this work, we used the Open Babel program to convert all entry formats like SMILES (2D coordinates) and subsequently in Hin format to generate 3D coordinates (Fig 2).

Structures	Measurement of activity	Search engine
8	IC_{50}	Pubmed
5	Ki	BindingDB
48	IC_{50}	

Table 1. Different sources of databases that we used to collect the molecules with inhibitory activity to create the pharmacophore model.
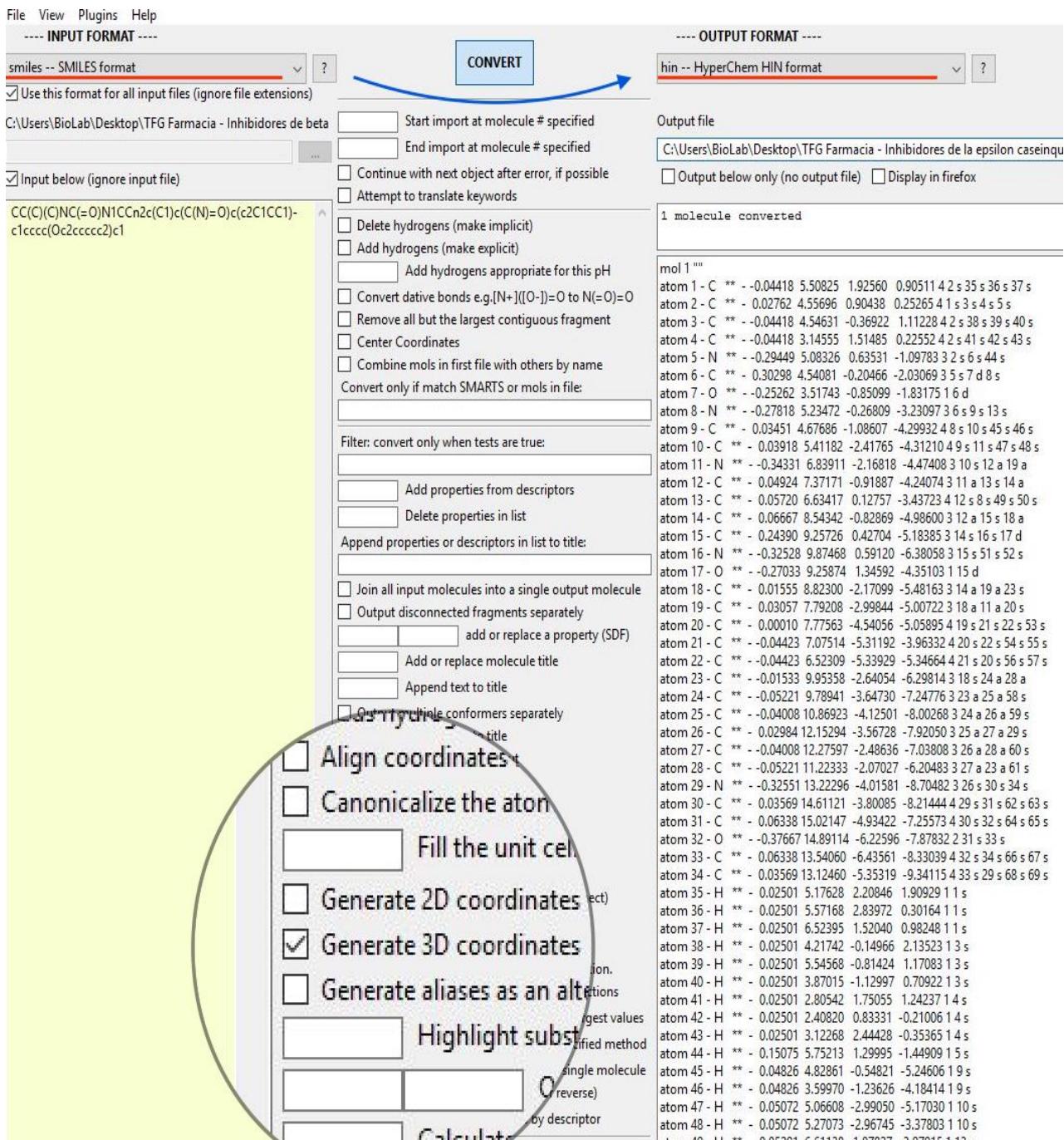
Figure 2. Inter-conversion between SMILES and Hn formats that we used to generate 3D coordinates. Adapted from Open Babel.

5.2 Geometry optimization

The process of finding the most stabile conformation that minimizes the energy is known as geometry optimization. We used the HyperChem program to find the global minimum in the PES and therefore the convergence of each structure. We selected a conjugate gradient method, Polak-Ribière Algorithm.
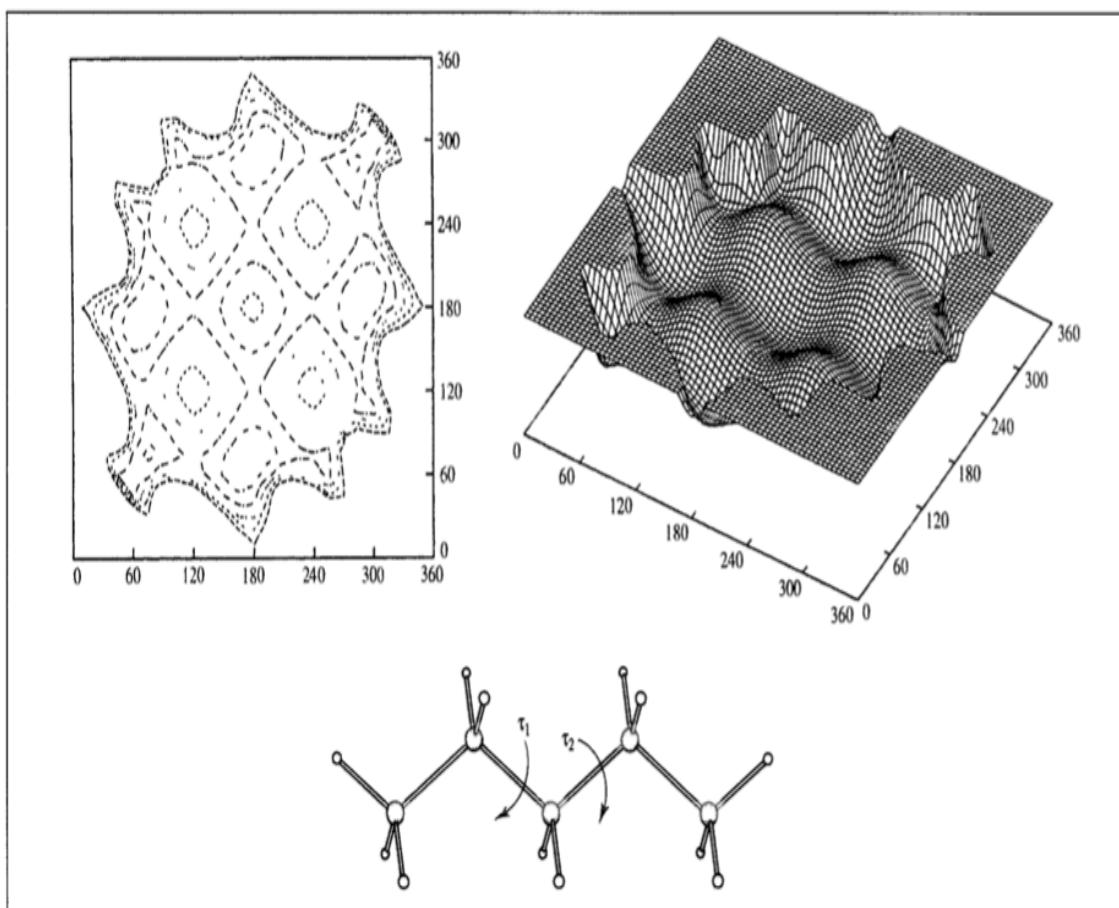
Figure 3. Variation in the energy of pentane with the two torsion angles indicated. This energy map shows how those changes in structure rotations determine the global minimum of PES. Adapted from ref. [4]

5.3 Pharmacophore modelling

To create the CK1ε pharmacophore, we selected the five compounds that showed the best Ki (nM) and IC_{50} (nM) values in the CK1ε's data set (Figure 3). MOE program creates a flexible alignment of the different conformations for each compound and generates diverse superimposed models using molecular mechanics parameters (Table 2) to maximize the overlap of common features.

Force field	OPLS
Algorithm	Polak-Ribière
Termination condition	RMS gradient of 0.001 Kcal/(Åmol)
	1005 maximum cycles
Screen refresh period	10

Table 2. Molecular mechanics parameters
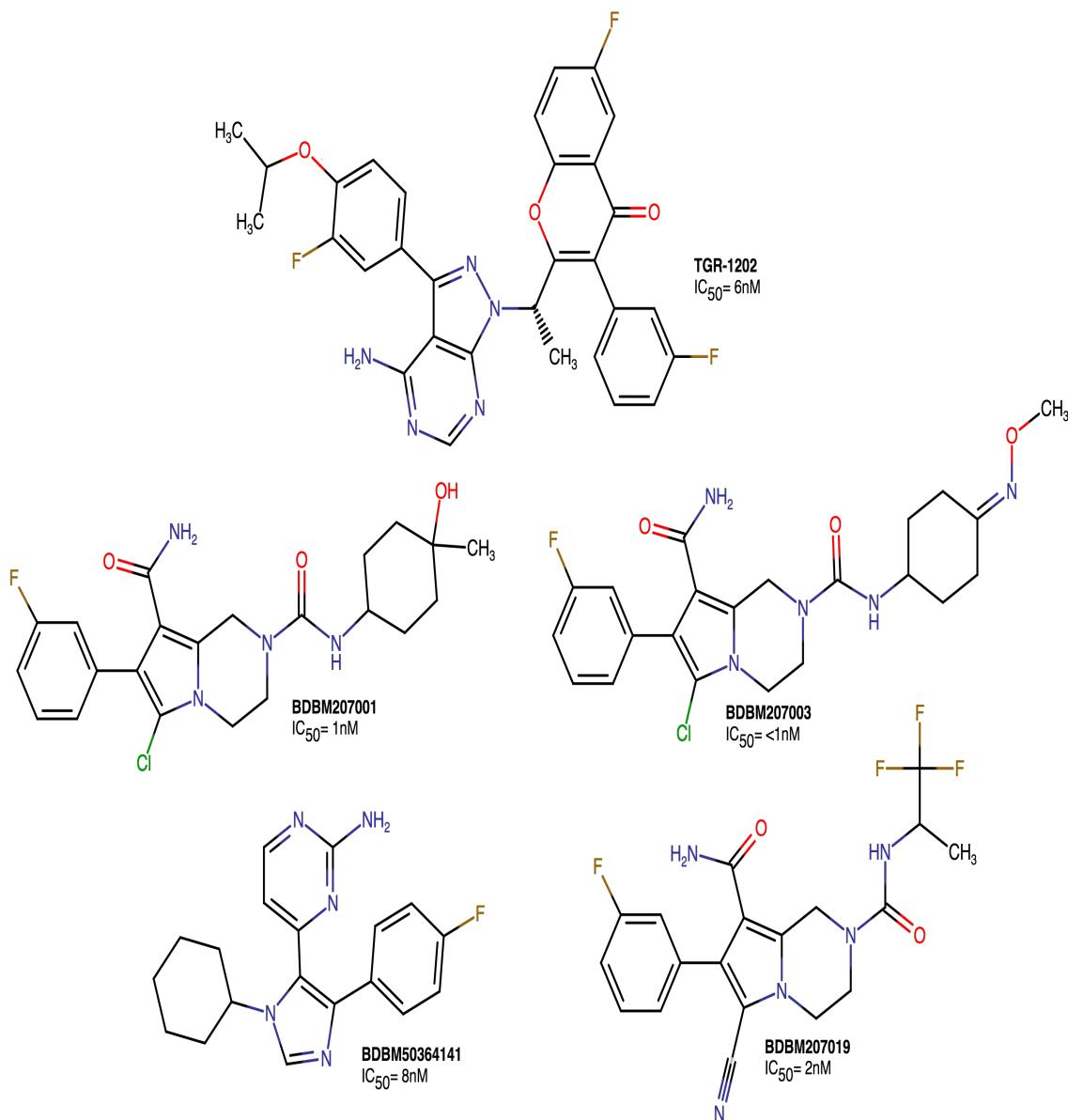
Figure 4. Structure of CK1ε used to create the pharmacophore model.

6. RESULTS AND DISCUSSION

The molecular mechanics describes the molecules like collections of atoms bound together controlled by a set of potential mechanic functions, therefore the total energy of the molecule is obtained as a series of contributions that rely on the spatial coordinates of the cores. The energy potential functions and the parameters used to create the PES are generally referred to as force fields. For this study we worked with the OPLS force field in accordance with the model requirements.

To a computational chemist, the PES is a surface that can be generated point by point by using computational methods which determine a molecular energy for each point's structure. The positions of the energy minima along different

13
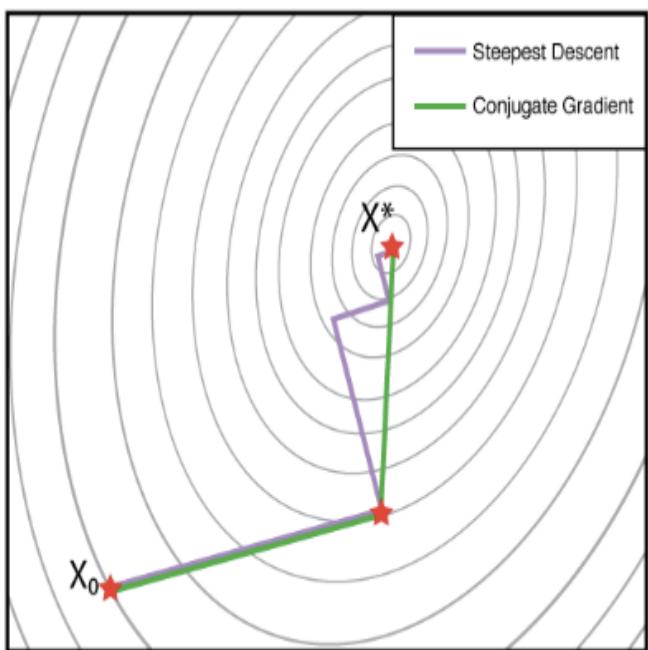
Figure 5. Conjugate gradient methods can be seen to converge far more rapidly than standard gradient descent methods. Adapted from ref. [11]

conformations give the equilibrium structures of the molecules. However, the position of the energy maximum gives the structure of the transition state. For this reason, we carry out the geometry optimization of the molecules before the generation of the pharmacophore model [10]. The geometry optimization depends on the force field that we have used and the convergence criteria (table 2). We used a conjugate gradient descent algorithm (Polak-Ribierè) to find the minimum of a function. In a conjugate gradient, a step down is iteratively taken along a direction of the previous step. [11]

In 1998, IUPAC defined pharmacophore as *"a set of steric and electronic characteristics that are necessary to ensure optimal supramolecular interactions with a specific biological target and trigger (or block) their biological response"* [12]. The creation of the pharmacophore needs to calculate a previous flexible alignment of the best five structures (Fig. 4). MOE creates different overlays to maximize the overlap of common features between the different possibilities obtaining five aligned conformations.

mol	U	F	S	dU	dF
1	37.3432	13.8046	51.1478	2.0895	0.0000

Figure 6. Best overlap of CK1ε inhibitors generated by MOE. The parameters evaluated are similarity (F), energy (U) and score (S)

We selected the first overlap (Fig. 5) to create the CK1ε pharmacophore based on MOE values for the score (S). The three-dimensional (3D) pharmacophore is generated from a set of features present in specific regions of the five structures selected (Fig.4), these common regions are designated as

'pharmacophoric groups'. The diverse features to create the 3D pharmacophore that specifies the spatial relationships between the groups were selected in the MOE program (Tab. 3). Parallel to that, the similarity criteria by means of two parameters (Threshold and Tolerance) were established.

	Ani: Anionic atom
	O$_2$: CO$_2$ centroid
	Cat: Cationic atom
	CN$_2$: NCN+ center
	Acc: H-bond acceptor
	Don: H-bond donor
	Aro: Aromatic center
	PiR: Pi ring center
	Hyd: Hydrophobic centroid

Table 3. Features and parameters selected to create the 3D pharmacophore

To refine the initial pharmacophore model for screening BioLab's database, a number of actions were taken.

1. The creation of a test database with 17996 conformations of inactive structures to casein kinases and 148 conformations of known active structures to casein kinase 1 epsilon that were added subsequently. We used this test database to maximize the efficiency of the screening and rectify the pharmacophore structure.
2. Change the parameters of the search such as threshold and tolerance enable to find pharmacophoric groups in a less restrictive way. Thus, the more pharmacophoric groups are involved in the screening, the more precise and easier is to eliminate interferences.
3. MOE program permits to choose the essential and no-essential pharmacophoric groups obtaining a flexible screening.
4. The used exclusion spheres allow indicating locations in the 3D pharmacophore structure where no part of a candidate structure is permitted to be positioned.

With this in mind, the refined pharmacophore has seven pharmacophoric groups and has been constructed with five essentials of them: one H-bond donor (F1), one H-bond acceptor (F2), two aromatic regions (F3, F6) and one hydrophobic region (F4). During screening test, the F5 and F7 groups were considered no-essential because we realised during the search that these groups were dispensable. We established an 80% threshold and a 1.2 tolerance in the screening. In other matters, we have positioned in the pharmacophore structure 53 exclusion spheres in total, adjusting the spheres radio to minimize the appearance of inactive molecules (fig. 7).

To sum up, we generated the refined pharmacophore with a retrospective process using the test database with active and inactive substances. However, this process has limitations because false positives have been found in the screening. Despite using exclusion spheres, the false positives identified in the last screening with the refined pharmacophore were impossible to discard because of their tiny structure.

Number of screen	Pharmacophoric groups		Exclusion spheres	Σ Identified molecules	False positives	% Identified molecules
	Essential	No essential				
1	7	0	12	347	342	17,53
2	7	0	25	205	200	10,47
3	7	0	39	67	62	3,42
4	7	0	47	54	49	2,76
5	5	2	51	52	47	2,66
6	5	2	53	50	45	2,55

Table 4. Results after each screen in the test database (1957 molecules).

It should be noted that modifications proposed in the model have been made reviewing that all search results contain the five molecules with known activity to CK1ε present in the test database.
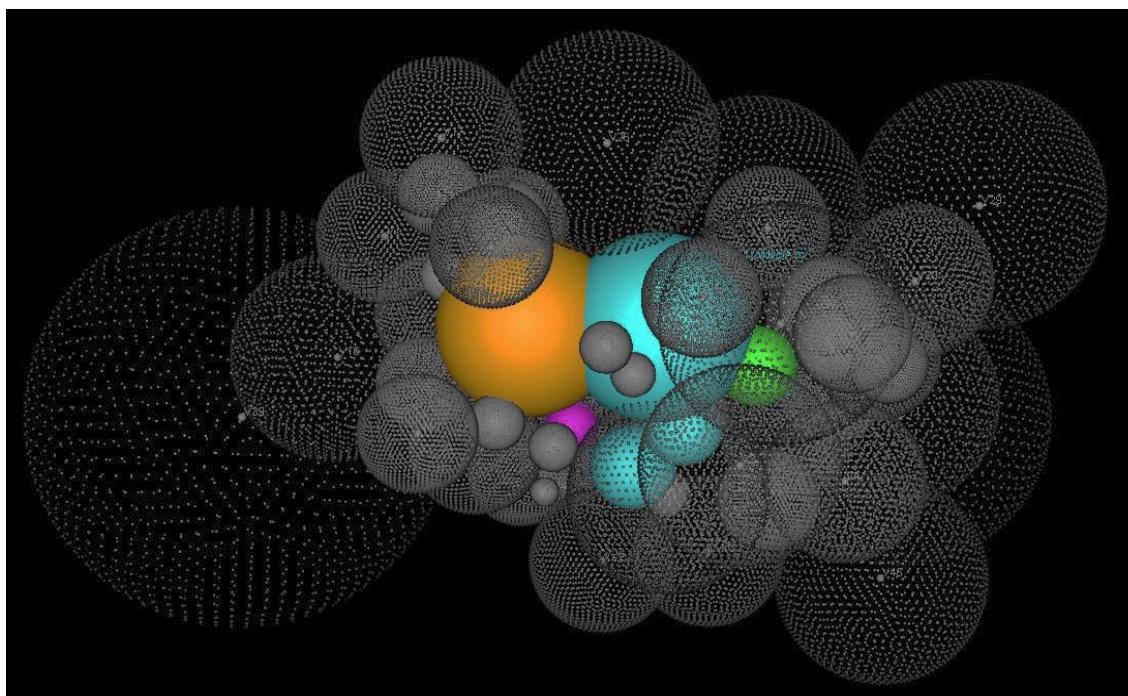

Figure 7. Refined pharmacophore with 7 pharmacophoric groups (coloured areas) and 53 exclusion spheres (grey areas).

Finally, we made the last screening in BioLab's database with the refined pharmacophore model obtaining 38 hits of the 1423 molecules that BioLab's database contains. The following table shows the five compounds with the most favourable predicted binding affinity for CK1ε based on RMSD. This parameter is a quantitative measure of similarity between two or more structures.
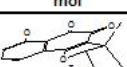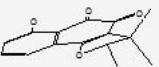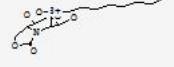

Figure 8. Structures with the most favourable predicted binding affinity for CK1ε based on RMSD.

7. CONCLUSIONS

To conclude, we can say that the main objective of this work and of the computational chemistry has been tackled, minimizing the quantity of molecules to submit in the experimental evaluation with biological assays. All in all, we obtained 38 molecules candidate to inhibit CK1ε of the 1423 molecules present in BioLab's database, reducing the search field over 97%. Undoubtedly, these proposals may continue to optimize and reduce the results with other computational techniques for instance docking screening or generating more pharmacophore models.

8. REFERENCES

[1] Walton KM, Fisher K, Rubitski D, Marconi M, Meng QJ, Sladek M, et al. Selective inhibition of casein kinase 1 epsilon minimally alters circadian clock period. J Pharmacol Exp Ther, 2009, 330:430–439.

[2] Brockschmidt C, Hirner H, Huber N, Eismann T, Hillenbrand A, Giamas G, et al. Anti-apoptotic and growth-stimulatory functions of CK1 delta and epsilon in ductal adenocarcinoma of the pancreas are inhibited by IC261 in vitro and in vivo. Gut, 2008. 57:799–806.

[3] Kim SY, Dunn IF, Firestein R, Gupta P, Wardwell L, Repich K, et al. CK1epsilon is required for breast cancers dependent on beta-catenin activity. PLoS One, 2010. 5:e8979.

[4] Leach A R 1996. Molecular Modelling: Principles and applications. Pretice Hall: Second edition, 2001. 648-661.

[5] HyperChem release 7: Tools for molecular modelling; Publication HC70-00-01-00; Hypercube Inc.; January 2002.

[6] Hutchison RG, Morley C, Craig J, Swain C, Winter H, Vandermeersch T, et al. Open Babel Documentation. First edition, 2011; 3-5.

[7] MOE, Molecular Operating Enviroment, Version 2008, Chemical Computing Group Inc.

[8] Binding database, Skaggs School of Pharmacy and Pharmaceutical Sciences at the University of California, San Diego, U.S.,https://www.bindingdb.org (accessed June 1,2017)

[9] National Center for Biotechnology Information, U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pubmed/ (accessed June 1, 2017)

[10] Cramer JC. Essentials of Computational Chemistry: Theories and Models. Wyley: Second Edition, 2004.17-50.

[11] Dannenberg M. Pattern Recognition in High-Dimensional Data. Harvey Mudd College, 2016. 15-17.

[12] Langer T, Hoffman RD. Pharmacophores and Pharmacophore Searches; Volume 32. Wyley, 2006. ISBN:978-3-527-60872-0