Computational Biology - 3rd Assignment

Nicolás Espinoza Muñoz

November 20, 2020

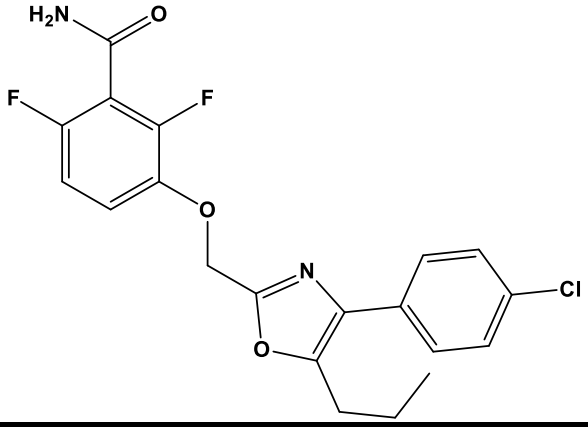
Introduction

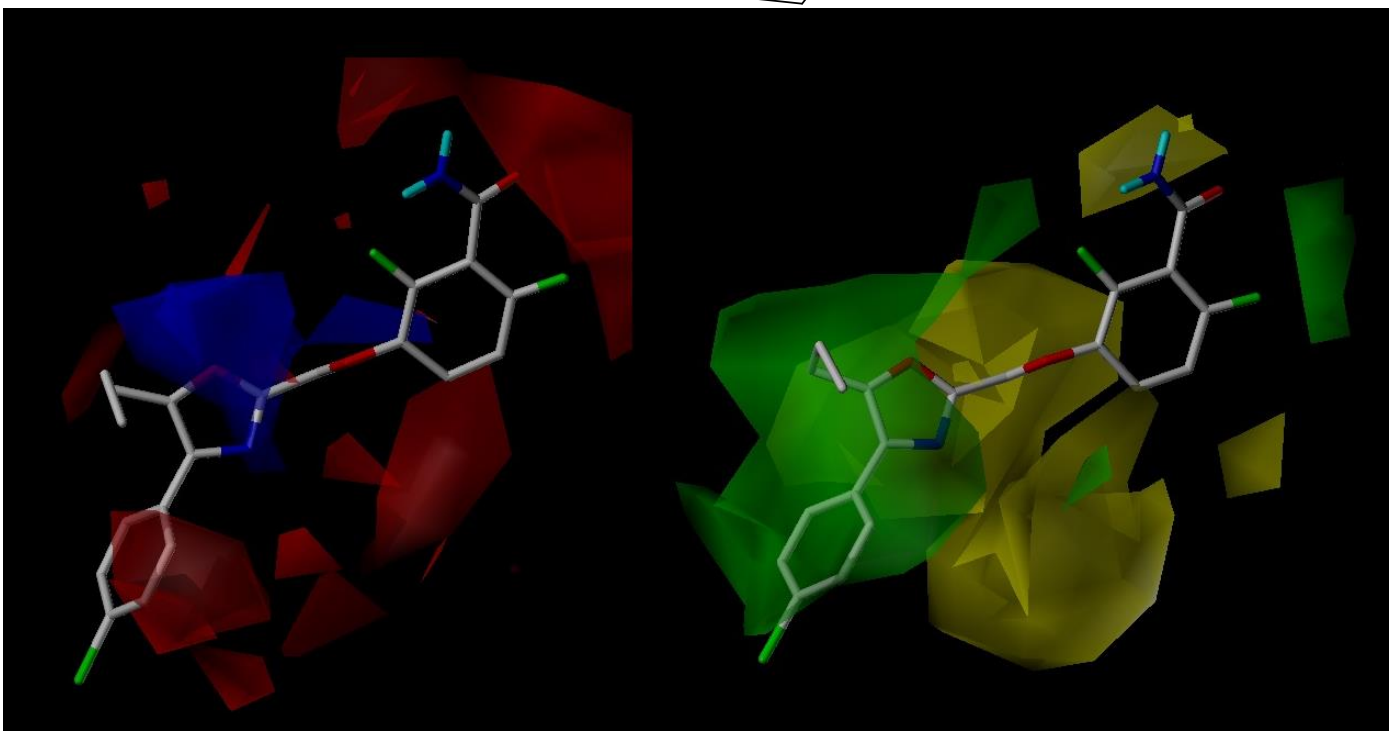
When it comes to drug design, computational methods are becoming more important as time passes. The Quantitative Structure-Activity Relationship analysis, or QSAR, has been at the front in the field, since it proves to be both of easy implementation and very powerful. It offers a quantification of the influence of a particular bioactive molecule or family of molecules on the biological activity of the target system, usually a cell. In the following we will answer the questions presented as part of the third assignment of the Computational Biology course, which focuses on the 2D QSAR methods of Free-Wilson and Hansch analysis, as well as a 3D QSAR study of electrostatic and steric effects.

Questions

We will directly include the questions of the assignment into this report, and in the next section proceed to answer them, as it's simpler this way.

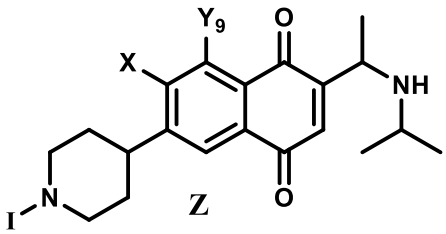
1. Considere los siguientes mapas de contornos en torno al siguiente antibacteriano prototipo, el cual está reportado en la literatura como un inhibidor de la proteína bacterial FtsZ: (3 ptos c/u, 15 ptos. en total)





A. Haga una descripción detallada del mapa de potencial electrostático
B. Ídem para el mapa estérico
C. Considerando la estructura del prototipo, señale e indique claramente qué tipo de interacciones intermoleculares con el receptor debería presentar este compuesto
D. ¿Qué otro tipo de sustituciones utilizaría en vez del patrón 2,6-difluoro en la benzamida y el conector oxígeno de ambos anillos?
E. ¿Por qué otro tipo de fragmento podría reemplazar el benceno conectado al oxazol? De al menos 3 variaciones posibles.

2. De un análisis de Hansch de una serie de compuestos diseñados como antineoplásicos de estructura general "Z" se obtuvieron las siguientes correlaciones: (Preguntas 2.1-2.3). 2 ptos c/u, 6 ptos en total.



$$\log 1/IC_{50} = 5.31(\pm 3.77)CMR - 0.26(\pm 0.16)CMR^2 - 21.60(\pm 19.34)$$

$$\log 1/IC_{50} = -1.32(\pm 0.61)\text{Clog } P + 9.52(\pm 1.18)$$

$$\log 1/IC_{50} = 0.43\text{Clog } P - 0.43\sigma_X - 0.89MR_9 + 1.11I + 6.37$$

$$\log 1/IC_{50} = -0.30(\pm 0.19)\pi_X - 1.37(\pm 0.42)\pi_Y + 5.87(\pm 0.28)$$

Donde *I* = 1 ó 0 indica la presencia o ausencia de un grupo etilo sobre el nitrógeno. MR_9 es la refractividad molar del sustituyente Y_9. Por otra parte π y σ son las constantes de lipofilia de Hansch y electrónica de Hammet respectivamente para los sustituyentes X o Y según se indique.

2.1. Es correcto que:

I. El aumento de la lipofilia siempre es beneficioso para la actividad
II. El aumento de la refractividad (CMR) siempre aumenta la actividad pues el aporte de CMR^2 es despreciable.
III. La presencia de un grupo etilo no es beneficiosa para la actividad antitumoral pues aumenta la lipofilia
IV. Conviene que X e Y sean de reducida lipofilia

a) solo IV
b) solo III
c) I, III y IV
d) III, IV
e) Todas

2.2. Es correcto que:
I. X podría ser un sustituyente poco lipofílico y donor electrónico como –OH
II. El uso de un átomo de Yodo como sustituyente en Y_9 debiese ser más favorable para la actividad que el uso de un átomo de cloro
III. grupos como $–NO_2$ o –CN en X serían beneficiosos para la actividad
IV. Los efectos electrónicos son más importantes que los efectos hidrofóbicos en la actividad antitumoral

a) solo I
b) II y III
c) solo III
d) I, II, IV
e) II, III, IV

2.3. ¿Qué información **no** podría derivar usted de las ecuaciones QSAR acá presentadas?
I. Conviene que el sustituyente X sea un grupo atractor electrónico y de baja lipofilia
II. Es bueno para la actividad que Y sea un grupo donor electrónico y de baja lipofilia.
III. Es beneficiosa la presencia de un grupo voluminoso en el nitrógeno piperidínico (*I*).
IV. El incremento en la refractividad molar es beneficioso pero tiene un límite como valor óptimo.

a) II y III
b) I y III
c) I, II y III
d) I y II
e) solo IV

3) En las siguientes dos tablas se presenta la actividad antitumoral (EC_{50}) de dos series de compuestos isósteros estrechamente relacionados con el mismo mecanismo de acción. En la primera tabla se entregan los valores de lipofilia para los sustituyentes en X e Y. En la Tabla 2 se muestra la refractividad del grupo Y (MR_Y), así como la contribución de Free-Wilson de los fragmentos I (= 1 si es que hay imidino en Y, ó 0 si no lo hay) e I_1 (=1 si es que hay metil en X ó 0 si es que hay H o etil). Para cada tabla halle mediante regresión multilineal simple una ecuación 2D-QSAR. Recuerde que puede además crear columnas adicionales con la lipofilia o refractividad elevado al cuadrado. En base a las ecuaciones hallas responda: (<u>3 ptos cada ecuación</u>)

a) ¿Cómo es conveniente que sea la lipofilia de los sustituyentes X e Y? **2 ptos.**

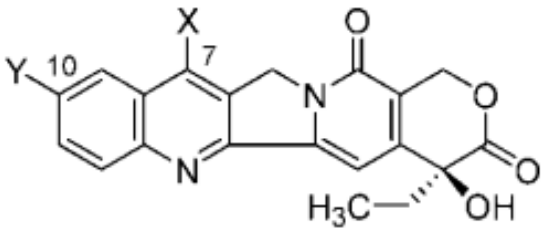
b) ¿Cómo es conveniente que sea la refractividad molar de Y? **2 ptos.**

c) ¿Conviene que en X halla metil? **2 ptos.**

d) conviene que en Y halla imidino? **2 ptos.**

e) ¿Cuáles son las propiedades o fragmentos que más contribuyen a la actividad en cada caso? **2 ptos.**

f) Proponga 4 moléculas nuevas en base a la información analizada. Recuerde utilizar los criterios de relación estructural vistos en clases. **2 ptos.**

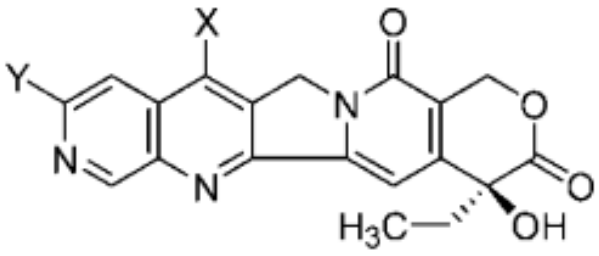


XVII

Table 1. Biological (EC_{50} or IC_{50}; mol L^{-1})[127] and Physicochemical Parameters

			log 1/EC_{50} (eq 10)		
No.	X	Y	obsd.	π_X	π_Y
1	CH_3	H	5.97	0.56	0.00
2	CH_2CH_3	H	5.35	1.02	0.00
3[a]	$(CH_2)_2CH_3$	H	5.14	1.55	0.00
4[a]	$(CH_2)_3CH_3$	H	5.14	2.13	0.00
5	H	OH	6.46	0.00	−0.67
6	CH_3	OH	6.89	0.56	−0.67
7	CH_2CH_3	OH	6.49	1.02	−0.67
8	$(CH_2)_2CH_3$	OH	6.22	1.55	−0.67
9	$(CH_2)_3CH_3$	OH	6.28	2.13	−0.67
10	H	OCH_3	6.14	0.00	−0.02
11	CH_3	OCH_3	5.62	0.56	−0.02
12	CH_2CH_3	OCH_3	5.55	1.02	−0.02
13	$(CH_2)_2CH_3$	OCH_3	5.35	1.55	−0.02
14	$(CH_2)_3CH_3$	OCH_3	5.55	2.13	−0.02

[a] Not included in the derivation of QSAR **22** [b] ND = not determined.



XXI

Table 2. Biological (IC_{50}; mol L^{-1}),[128] Physicochemical, and Structural Parameters

			log 1/IC_{50}			
No.	X	Y	obsd.	MR_Y	*I*	I_1
1	H	H	6.42	0.00	0	0
2	C_2H_5	H	6.47	0.00	0	0
3	CH_3	Br	7.55	0.78	0	1
4	C_2H_5	Br	6.91	0.78	0	0
5	CH_3	CN	6.84	0.48	0	1
6	C_2H_5	CN	6.57	0.48	0	0
7	CH_3	CH_2NH_2	6.63	0.83	0	1
8	C_2H_5	CH_2NH_2	6.78	0.83	0	0
9	C_2H_5	$C(NH_2)NOH$	7.15	1.48	1	0
10	C_2H_5	$C(NH_2)NH$	7.48	1.02	1	0
11	CH_3	$C{\equiv}CCH_2NH_2$	6.30	1.71	0	1
12	C_2H_5	$C{\equiv}CCH_2NH_2$	5.97	1.71	0	0
13	C_2H_5	$C{\equiv}CCH_2N(CH_3)_2$	5.88	2.64	0	0
14	CH_3	$C{\equiv}CCH_2(\text{-}NCH_2CH_2OCH_2CH_2\text{-})$	5.95	3.54	0	1
15	C_2H_5	$(CH_2)_3N(CH_3)_2$	5.83	2.69	0	0
16	C_2H_5	$COOC_2H_5$	6.02	1.58	0	0
17	C_2H_5	$CONH(CH_2)_2N(CH_3)_2$	5.03	3.09	0	0

I = imidino

I1 = metil / 0 H o etil

Nota: en este video puede ver como activar la opción de regresión multilineal en excell y como crear ecuaciones de regresión multilineal:

https://www.youtube.com/watch?v=Bye0ZBdd6iI

Answers

1.

A

The electrostatic potential reveals which sections of the molecule would benefit from the insertion of electron rich (red) or electron deficient (blue) atoms or groups. The carboxamido substituent from the benzamide is not sufficiently rich in electrons; the same appears to happen with one of the fluorine atoms, and therefore the potential map relative to them shows in red. Likewise, the electrostatic potential corresponding to the lower spatial portion of both six-membered benzene rings is also red, which reveals that currently those components of the molecule are electron deficient and should be replaced by electron-rich groups. The ring-connecting oxygen seems electron rich, as does the majority of the oxazole ring, because the spatial representation of the potential around those components is blue. The smaller polyhedrons are to be ignored, as they are representative of lesser electrostatic effects.

B

The steric map represents the volumetric information of the current molecule; green sections mean a bigger substituent, in terms of volume, is more beneficial. Conversely, yellow polyhedra correspond to groups or elements that have too big a volume. In the map presented, the oxazole and the benzene ring directly connected to it appear to be too small, and it is better to include bigger groups. Surrounding the oxygen that connects the oxazole to the benzamide, as well as part of the benzene ring of the benzamide, is a massive yellow polyhedron. This shows it is convenient to replace some elements in that section with smaller groups. It also looks like the volume of the NH_2 in the amide group is a bit too large, and one of the fluorine atoms connected to the benzamide could be replaced with something with a bigger volume.[1]

C

-

D

Based on the steric map, we should replace fluoride atoms by something with a bigger volume, but not too much, so an answer would be a chlorine in each position. As for the ring connector, a carbon is the only thing that comes to mind, because we have to replace the oxygen with an element with less electrons and less volume.

E

-

2

Only the explanations as to why a statement is considered true or false, together with the letter of the chosen alternative, are presented next. The statement itself is not quoted.

2.1

I. False: From the correlations of the Hansch analysis, π_X and π_Y have a negative factor, so a higher lipophilicity would lower the Biological Activity. This also happens for the second correlation with $C \log P$. The only exception would be $C \log P$ in the third correlation. Then again, when

[1] Bigger and smaller are used to refer to volume.

lipophilicity is too high, the molecule will likely just remain in the membrane and no actually enter the cell.

II. False: For high values of CMR, the term CMR^2 will be more important to Biological Activity, simply because it scales with the square of CMR.

III. False: The presence or absence of an ethyl group affects the third correlation, but that is a Free-Wilson of term, and not lipophilicity.

IV. True: The lower the lipophilicity of X and Y, the better, because they are factored by a negative term.

Answer: a) solo IV

2.2

I. True: A lower lipophilicity and electron donor group is better for the Biological Activity. σ_X and π_X are factored by negative terms, so a lower π_X and σ_X are both beneficial for these correlations.

II. False: As chlorine is smaller in terms of volume, its lipophilicity is lower, and since π_Y is factored by a negative term, that's the better scenario.

III. False: NO_2 and CN groups are electron deficient, so they have a positive Hammett's sigma value. Since the factor multiplying σ_X is negative, this would not be a beneficial substituent.

IV. False: We have to check the factors of the σ, π and $\log P$ terms. Those multiplying π_X, π_Y and $\log P$ are bigger than that multiplying σ_X.

Answer: a) solo I

2.3

I. We can know this; we have correlations with both σ_X and π_X. It is false though.

II. We can't know this; we have no σ_Y term in any of the correlations presented.

III. We can't know this; there is no information regarding the molar refractivity (MR) of *I*.

IV. We can know this; we have the corresponding term in at least one of the correlations.

Answer: a) II y III

3

For this exercise we added two columns, one per table, as there were columns with values that rose way past 1. So for the first table, we added a π_X^2 column, and for the second we added a MR_Y^2 column. We now present both equations, (1) for Table 1 of the exercise, and (2) for Table 2. These were obtained with Multiple Linear Regression. For Table 1 we used 10 out of the 14 lines of information available, and for Table 2 we used 12 out of 17. This accounts for roughly 70% of the data available in each case. We then used the remaining data to verify the correlations obtained.

$$\log\left(\frac{1}{EC_{50}}\right) = -0.398(\pm 0.398)\pi_X + 0.020(\pm 0.178)\pi_X^2 - 1.385(\pm 0.258)\pi_Y + 5.927(\pm 0.203) \quad (1)$$

$$\log\left(\frac{1}{EC_{50}}\right) = 1.057(\pm 0.502)\mathrm{MR}_Y - 0.749(\pm 0.258)\mathrm{MR}_Y^2 + 0.797(\pm 0.249)I + 0.253(\pm 0.188)I_1 + 6.406(\pm 0.184) \quad (2)$$

Before continuing with the answers requested, the point has to be made that the second correlation actually fails to return a good value of $\log p$ for some data in the Table 2 set. We tried many different variable combinations between MR_Y, MR_Y^2, I and I_1, sometimes excluding one or more of those terms, but failed to get results that were substantially better, so the decision was made to present correlation (2) as is.

a)

It is convenient, at least basing the answer in the median factor obtained for each π, that π_X and π_Y both be small. Given the correlation (1), it is better to have hydrophilicity rather than lipophilicity.

b)

Molar refractivity should be higher than 1 but not too much, because there is a point where the MR_Y^2 term, which has a negative factor, begins to dominate. This means there is a limit to how much a bigger volume is beneficial in terms of the $\log(1/EC_{50})$ indicator.

c)

It is convenient that there be a methyl group in the X group position, as shown by the positive factor accompanying I_1.

d)

According to our correlations, it is indeed good for the molecule to have an imide group in the Y position; the reason is the same as for letter c).

e)

For compound number 1, the main property is lipophilicity of the group occupying position Y. For compound number 2, the main factor would be the molar refractivity of the substituent that should go in position Y as well, although one could make the point that it is also important to consider the presence of imide in Y, as it has the second highest factor in the correlation.

f)

-

Conclusions

QSAR methods are a powerful tool to design bioactive molecules, as they allow us to develop and test molecules directly on the computer, before going onto live tests. It is also rather simple to formulate these correlations, and a researcher has a lot of flexibility when it comes to component replacements for testing different drugs, given there is data about the parameters of special interest for each case. The catch is, one has to be rather knowledgeable about the compatibility of each substituent being considered as part of the compound family studied, and has to be able to correctly interpret the correlations obtained. Also, this field actually integrates a good amount of disciplines, from statistics to electrostatics, and of course chemistry, so it is by no means a simple subject.

Here we tried to show, at least partially, how useful the QSAR techniques are for identifying relevant properties in the Biological Activity influence a specific molecule has, and went on to produce and interpret our own correlations. We also studied 3D QSAR methods, which allow us to study non-planar models of molecules, so one could argue they are in fact more precise in some ways. Nevertheless, 2D and 3D QSAR methods remain important as they immensely facilitate bioactive molecule design and reduce the costs associated with it.