

Màster Universitari en bioinformàtica i bioestadística

Biologia estructural

PAC 4

Roger Massaguer Carles

Preguntes inicials

1. Com s'originen les interaccions de van der Waals?

Les interaccions de van der Waals són forces atractives de rang curt entre grups químics. S'originen a causa de petits desplaçaments de càrrega que permeten que els electrons d'un àtom siguin atrets pels protons d'un altre àtom molt proper malgrat que tots dos no estiguin enllaçats covalentment.

2. Què són i perquè s'usen els camps de força?

En el context de la química i el modelatge molecular, un camp de força és un mètode computacional que és utilitzat per estimar les forces entre els àtoms dins de les molècules i també entre diferents molècules. De manera més precisa, el camp de força fa referència a la forma de la funció i els paràmetres utilitzats per calcular l'energia potencial d'un sistema d'àtoms o de partícules granulades mitjançant mecànica molecular, dinàmica molecular o el mètode de Montecarlo. Els paràmetres per a una funció d'energia seleccionada poden sorgir d'experiments físics o químics, càlculs de mecànica quàntica o una combinació de tots dos.

3. En quins medis es poden modelar les molècules?

Les molècules es poden modelar al buit (també anomenades simulacions de fase gas) o en presència d'un solvent com l'aigua (solvent explícit). En un altre tipus de simulacions, l'efecte del solvent s'estima fent servir una expressió matemàtica empírica i aquestes es coneixen com a simulacions de solvent implícit. L'elecció del medi on les molècules simulades es troben és molt important; segons el solvent que se seleccioni, el temps real de l'execució computacional es pot veure alterat significativament.

4. Quin tipus de receptor són les proteïnes G? Poseu 3 exemples de funcions que realitzen i els seus lligands.

Proteïnes transmembrana, com moltes hormones i neurotransmissors.

Elles activen o inactiven transduccions de senyal. Hi ha els receptors units a proteïna G i que afecten els enzims que controlen els canals de ions.

Els receptors de proteïna G, coneguts també com a GPCR (G protein-coupled receptor) són una àmplia família de receptors transmembrana (creuen la membrana fins a 7 vegades) associats a compostos que són receptors de la llum, l'olor, les feromones, hormones a general i neurotransmissors. Aquests receptors utilitzen la proteïna G a la qual estan units per amplificar els senyals.

5. Quines propietats fonamentals té un lligand?

La molècula candidata a actuar sobre el receptor s'anomena lligand. Necessitem tenir en compte dues propietats fonamentals d'un lligand per poder estudiar la unió receptor-lligand:

- Afinitat: es defineix com l'habilitat d'un lligand per unir-se a un receptor específic i formar un complex lligand-receptor.
- Activitat intrínsecoeficàcia: és l'habilitat d'un lligand per induir una resposta biològica després de la unió al receptor.

6. Quina és la diferència entre un fàrmac agonista i un fàrmac antagonista?

- Agonista: és un lligand amb gran afinitat al receptor i alta activitat intrínseca. Genera una resposta farmacològica similar al lligand natural.
- Antagonista: és un lligand amb gran afinitat al receptor, però sense activitat intrínseca (no té una resposta farmacològica). Aquests lligands disminueixen o inhibeixen, depenent de la dosi, l'efecte d'agonistes impedit les unions receptor-ligant o impedit la generació de reaccions secundàries per formar els complexos receptor-ligant.

7. En quines condicions faríem servir un model per homologia i com s'obté?

Un model per homologia és un model de proteïna amb una estructura 3D desconeguda i que es construeix a partir de les coordenades dels àtoms de proteïnes similars utilitzant tècniques d'alineament. Sempre que hi hagi més d'un 60% d'homologia de seqüència és possible fer un model útil i d'una qualitat adequada per a aquest tipus d'estudis. Quan un té el model del lloc d'unió del receptor, es poden conèixer les interaccions específiques que són responsables d'aquesta unitat.

8. Quin és el model clau-pany? Què no s'havia tingut en compte en desenvolupar aquest model?

Els mètodes d'acoblament molecular pioners es basaven en el principi de la clau-pany i es focalitzaven principalment en l'ús de criteris geomètrics per provar el grau de complementarietat entre el lligand i el lloc d'unió.

Però aviat es va veure que la complementarietat química s'havia de tenir en compte en aproximacions d'acoblament molecular per reduir el nombre de solucions físicament no realistes que només s'obtenien tenint en compte la forma.

9. Quins són els tipus de funcions de cribratge? Què mesura cadascuna d'elles?

Hi ha tres tipus de funcions de cribratge: basades en un camp de forces, empíriques i funcions de cribratge basades en el coneixement.

· Normalment, els camps de força quantifiquen la suma de dues energies, l'energia d'interacció proteïna-ligand i l'energia interna del lligand. Les interaccions entre lligand i receptor estan descrites usant l'energia de vdW i l'energia electrostàtica.

· Aproximacions empíriques: Aquestes estan basades en funcions d'energia simple o en la freqüència que tenen diferents parells de contacte àtom-àtom en complexos d'estructura coneguda. Podem assignar un valor a cada interacció fent servir paràmetres empírics. Després sumem els valors per a totes les interaccions i obtenim una estimació empírica de l'energia lliure d'unió. S'aproxima l'energia com a suma d'interaccions ponderades que es descriuen per les funcions geomètriques simples entre el lligand i les coordenades del receptor.

· Les funcions de cribratge basades en el coneixement previ estimen energies lliures d'interaccions moleculars a partir de bases de dades.

10. Què és un model QSAR? Què ens permet determinar un bon model QSAR?

En el mètode QSAR es tracta de correlacionar propietats estructurals moleculars (descriptors) amb la seva activitat biològica (propietats fisicoquímiques, activitats biològiques, toxicitat, etc.) per a un conjunt de compostos similars, utilitzant mètodes estadístics.

Els avantatges d'aquest model és que és molt informatiu i fàcilment interpretable. És fàcil suggerir substituents nous. El problema principal d'aquest model és que hi pot haver contribucions no additives i que fer servir models molt complexos pot portar a obtenir models no significatius.

Exercici 1: Dinàmiques moleculars.

En el temps que teniu per a realitzar la PAC és difícil, per no dir impossible, realitzar una dinàmica molecular i la seva corresponent anàlisi. Per tant, el que fareu serà familiaritzar-vos amb els paràmetres, els temps, etc....de dinàmiques moleculars gràcies a dinàmiques ja fetes, accessibles en una base de dades.

Entreu en la web de MODEL (Molecular Dynamics Extended Library)

• Què és MODEL?

Model (Molecular Dynamics Extended Library). Es tracta d'una gran base de dades de trajectòries de dinàmica molecular d'estructures de proteïnes representatives utilitzades amb tecnologies de última generació.

L'objectiu principal del MoDEL és proporcionar informació sobre la dinàmica a escala multinanosegon de proteïnes en condicions gairebé fisiològiques. Aquesta informació es pot utilitzar per a molts propòsits, que van des d'estudis evolutius fins a l'anàlisi biofísica i els processos de disseny de fàrmacs.

• Qui ho ha desenvolupat?

MODEL és un esforç massiu que implica, directa o indirectament, gran part del grup de Modelització Molecular i Bioinformàtica de l'IRB Barcelona i el BSC (Barcelona Supercomputing Center), l'equip de suport de MareNostrum.

• Trobeu l'article associat a aquesta base de dades i citeu-lo

<https://reader.elsevier.com/reader/sd/pii/S0969212610003539?token=42A98119EC34DD820601D3F71E2B3EB0276E5C89FBFC59A9DA625A3D9933A77620067DDD9541FD132D53C6F851C4D7A1&originRegion=eu-west-1&originCreation=20211227183301>

- Quantes estructures s'han simulat?

S'han obtingut més de 1700 trajectòries de proteïnes representatives d'estructures solubles monomèriques al banc de dades de proteïnes (PDB) mitjançant simulacions de dinàmica molecular atòmica d'última generació en condicions gairebé fisiològiques.

- Entre quin rang de temps de simulació s'han calculat les dinàmiques?

En el moment de redactar aquest informe, el MoDEL El magatzem de dades contenia més de 1700 trajectòries de proteïnes, que van des de 10 ns (el més curt) fins a 1 ms (el més llarg).

- Quins programes de dinàmica molecular s'han utilitzat en aquests càlculs? Anomeneu-los.

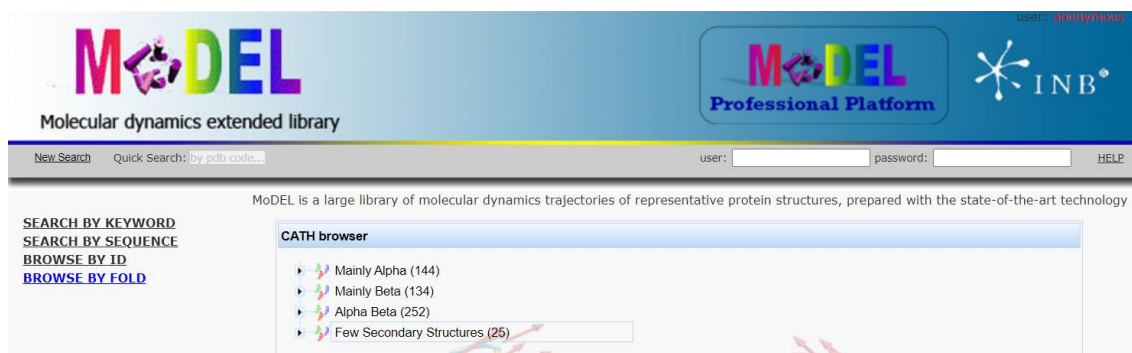
El pipeline permet a l'usuari llançar la simulació al final del procés, mitjançant diferents codis MD (actualment: **AMBER**, **NAMD** i **GROMACS**). A més, s'ha desenvolupat una aplicació web independent (MDWeb; AH, MO, JLG, dades no publicades) que inclou totes les funcionalitats com a producte secundari del projecte MoDEL per ajudar en la configuració automàtica (però flexible) de simulacions MD per a usuaris no experts.

- Amb quins recursos computacionals s'han pogut calcular totes aquestes dinàmiques moleculars?

L'esforç computacional requerit per a la derivació de MoDEL va requerir un ús massiu del super ordinador MareNostrum al Barcelona Supercomputing Center (www.bsc.es) i plataformes locals del nostre grup, i va trigar més de 4 anys a arribar al seu estat de finalització actual.

- Quina és la classe estructural de proteïnes que més s'ha simulat? I la que menys?

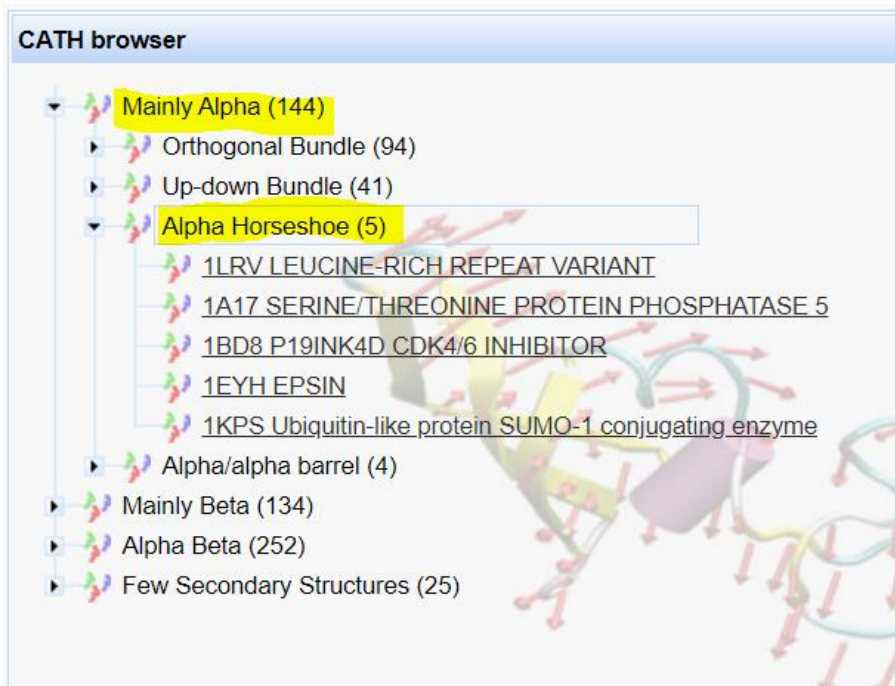
La classe estructural de proteïnes que més ha estat simulada és la Alpha Beta amb un total de 252 simulacions, en canvi few secondary structures (25 simul.) i mainly beta (134 simul.) són les que presenten menys simulacions.



Imatge 1 Resum de l'apartat Navegació per plecs. <http://mmb.irbbarcelona.org/MoDEL/>

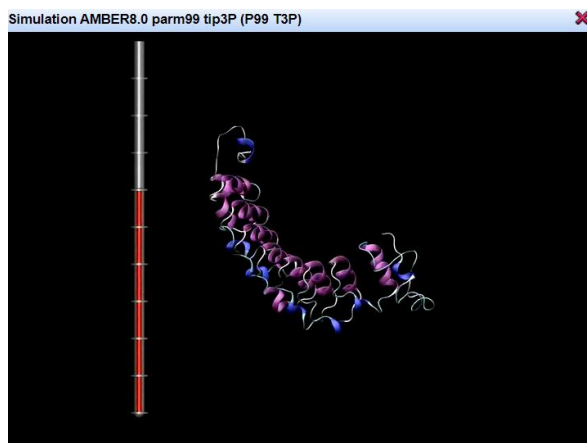
- Quantes estructures hi ha simulades amb el plegat alpha horseshoe? On heu trobat aquesta informació? Quina és la classe estructural més representada en aquesta base de dades?

Amb el plegat alpha horseshoe trobem 5 estructures. Aquesta informació ve donada en l'apartat BROWSE BY FOLD, concretament en l'estructura MAINLY ALPHA. La classe estructural més representada en aquesta base de dades és l'estructura ALPHA BETA amb un total de 252 estructures.



Imatge 2 Estructures Alpha Horseshoe.

- Ara triareu alguna de les estructures simulades amb el plegat anterior i visualitzarem la dinàmica molecular. Heu de clicar a Trajectory Vídeo i podeu fer una captura de pantalla de la imatge i plasmar-la en el vostre document.



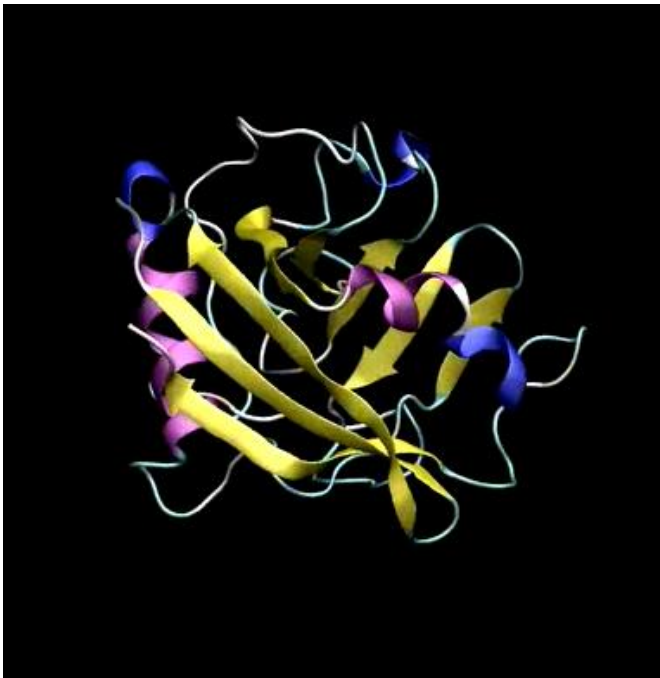
Imatge 3 Simulació de la trajectòria de 1LRV. A LEUCINE-RICH REPEAT VARIANT WITH A NOVEL REPETITIVE PROTEIN STRUCTURAL MOTIF

Ara buscareu en aquesta biblioteca de dinàmiques moleculars una dinàmica ja feta.

Busquem amb *Browse by ID* PDB: 2CPL.

- A quina classe estructural pertany aquesta proteïna? On trobeu aquesta informació? Pertany a la classe estructural més representada en aquesta base de dades?

Aquesta proteïna pertany a una estructura de barril beta, compressa per tant al grup de les estructures Alpha-beta que es tracta de la classe estructural més representada en la base de dades.



Imatge 4 Estructura barril-beta de la proteïna.

- Quantes simulacions per a aquesta proteïna hi ha a la base de dades? Quin és el temps de dinàmica de cadascuna d'elles?

Per aquesta proteïna apareixen dues simulacions:

AMBER 8.0 P99 (T3P), 80500
AMBER 8.0 P99 (T3P), 20000

Imatge 5 Simulacions per 2CPL

Simulation time 80500 ps
Integration time 0.0020 ps
Sampling time 1.0 ps

Imatge 6 El temps de dinàmica per la 80500

Simulation time 20000 ps
Integration time 0.0020 ps
Sampling time 1.0 ps

Imatge 7 El temps de dinàmica per la 2000.

Selecioneu la simulació de major temps de dinàmica i observeu la dinàmica amb el vídeo de la trajectòria (*trajectory vídeo*):

- Quines parts de la proteïna creieu que tenen més mobilitat?

La part proteica amb més mobilitat són la part dels enllaços per la qual té la capacitat de rotar.

- Podeu dir amb quin camp de força s'ha calculat?

Parm99 (standard amber forcefield).

- Quin programa de dinàmica molecular s'ha utilitzat per a realitzar la simulació?

AMBER 8.0.

- A quina temperatura s'ha simulat la dinàmica?

300.0 K

- Quantes molècules d'aigua (el solvent) s'han posat a la simulació?

Solvent molecules: 6665


- Quina és la durada de la dinàmica?









Simulation time 80500 ps
Integration time 0.0020 ps
Sampling time 1.0 ps

- Quants residus té la proteïna simulada?

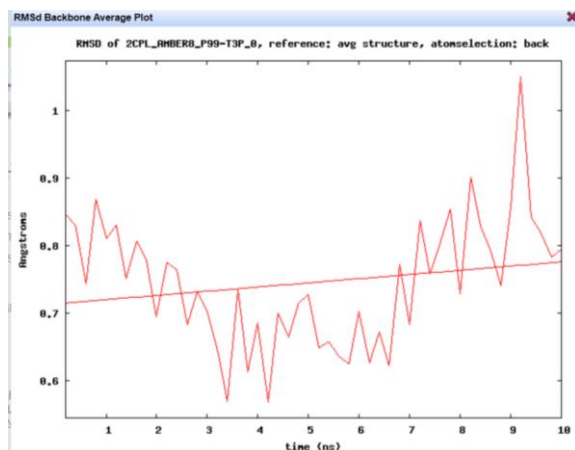
164 residus

- Podeu trobar la desviació del *backbone* de la proteïna respecte la mitjana de la dinàmica molecular (RMSD)?

RMSd 


Reference	Average		Experimental	
CA	0.723 ± 0.094 (0.535 - 1.026) Å		1.113 ± 1.139 (0.897 - 1.48) Å	
Backbone	0.745 ± 0.092 (0.569 - 1.049) Å		1.139 ± 0.116 (0.937 - 1.485) Å	
Heavy	1.322 ± 0.11 (1.13 - 1.693) Å		2.028 ± 0.148 (1.787 - 2.504) Å	
All	1.294 ± 0.108 (1.106 - 1.656) Å		1.983 ± 0.146 (1.746 - 2.45) Å	









Imatge 8 Desviació del backbone de la proteïna respecte la mitjana de la dinàmica molecular



Imatge 9 Representació de la desviació.

- Quin és el RMSD del *backbone* de l'estructura experimental? És comparable aquest valor amb el de la dinàmica molecular?

RMSd 

Reference	Average		Experimental	
CA	0.723 ± 0.094 (0.535 - 1.026) Å		1.113 ± 1.139 (0.897 - 1.48) Å	
Backbone	0.745 ± 0.092 (0.569 - 1.049) Å		1.139 ± 0.116 (0.937 - 1.485) Å	
Heavy	1.322 ± 0.11 (1.13 - 1.693) Å		2.028 ± 0.148 (1.787 - 2.504) Å	
All	1.294 ± 0.108 (1.106 - 1.656) Å		1.983 ± 0.146 (1.746 - 2.45) Å	

- Perquè creieu que és important comparar aquests valors? De què ens informa el valor RMSD?

RMSd és la magnitud estàndard per calibrar la desviació d'una estructura respecte a una conformació de referència.

Quan la suma s'estén pel nombre total d'àtoms/residus considerats en el càlcul i els vectors de posició de l'estructura considerada i la conformació de referència es calculen després de l'alineació de l'estructura amb la conformació de referència per minimitzar el valor del RMSd. MoDEL utilitza l'algorisme de Kabsch (Kabsch, 1976) per a l'alineació de les estructures.

L'anàlisi proporcionada inclou valors i gràfics mitjans RMSd al llarg del temps per a CA, Backbone, Heavy i Tots els àtoms en comparació amb l'estructura mitjana de la simulació i l'estructura experimental. Les tendències lineals o exponencials de les dades RMSd al llarg del temps s'inclouen a les gràfics per avaluar l'estabilitat de l'estructura.

Ara busqueu una seqüència (*Search by sequence*) per a veure si existeix a la base de dades la dinàmica de la proteïna que té aquesta seqüència. Per a dur a terme això, el programa realitza un alineament de seqüència amb BLAST per a buscar una seqüència semblant a la que introduïm.

```
PHSHPALTPEQKKELSDIAHRIVAPGKGILAADESTGSI AKRLQSIGTENTENRRFYRQLLLTADDRVNPCIGGVI
LFHETLYQKADDGRPFPPQVIKSKGGVVGIKVDKGVVPLAGTNGETTTQGLDGLSERCAQYKKDGADFAKWCVLKIG
EHTPSALAIMENANVLARYASICQQNGIVPIVEPEILPDGDHDLKRCQYVTEKVLAAVYKALSDHHIYLEGTLLKPN
MVTTPGHACTQKYSHEEIAMATVTALRRTVPPAVTGVTFLSGGQSEEEASINLNAINKCPLLKPWALTFSYGRALQAS
ALKAWGGKKENLKA AQEEYVKRALANSLACQKYTPSGQAGAAASESLFISNHAY
```

Podeu observar que coincideix en un 98% amb dues seqüències que hi ha en la base de dades. Per tant, aquesta seqüència és gairebé idèntica a la d'unes proteïnes que tenen una dinàmica ja calculada a la base de dades.

- Quines proteïnes (codi PDB) són les que ha alineat BLAST amb un 98%? A quins organismes corresponen?

Les dues proteïnes alineades amb BLAST amb un 98% (357/367 de identitat) són: 2ALD i 1ALD.

hit	e-value ▲	query from	query to	identity	description
2ALD chain A	0.0	1	363	357 / 363 (98%)	FRUCTOSE-BI...
1ALD chain A	0.0	1	363	357 / 363 (98%)	ALDOLASE A

Imatge 10 Proteïnes alineades amb BLAST.

La proteïna 2 ALD correspon a HUMAN MUSCLE ALDOLASE, és a dir una proteïna que prové de l'organisme *Homo sapiens*.

La proteïna 1ALD correspon a ACTIVITY AND SPECIFICITY OF HUMAN ALDOLASES, proteïna que prové de l'organisme de *Homo sapiens*.

- De quants nanosegons són les simulacions?

2ALD té un temps de simulació de 20.0 ns.

AMBER 8.0	2ALD	20.0 ns	0	FRUCTOSE-BISPHOSPHATE ALDOLASE
-----------	------	---------	---	--------------------------------

Imatge 11 Temps de simulació de 2ALD

1ALD té un temps de simulació de 20.0 ns.

AMBER 8.0	1ALD	20.0 ns	0	ALDOLASE A
-----------	------	---------	---	------------

Imatge 12 Temps de simulació de 1ALD

- A quina proteïna i organisme correspon la seqüència problema? (Podeu buscar “per seqüència” a PDB o a UNIPROT per a esbrinar-ho).

Un cop cercat per PDB la proteïna a partir de la seqüència facilitada podem concloure que aquesta es tracta de la proteïna 2OT0_1|Chains A, B, C, D|Fructose-bisphosphate aldolase A, provinent de *Oryctolagus cuniculus*, és a dir, el conill comú.

```
>2OT0_1|Chains A, B, C, D|Fructose-bisphosphate aldolase A|Oryctolagus cuniculus (9986)
PHSHPALTPPEQKELSDIAHRIVAPGKGLAADESTGSIARLQSIGTENTEENRRFYRQLLLTADDRVNPCIGGVILFHETLYQKAD
DGRPFQVIKSKGGVGVKVDKGVVPLAGTNGETTTQGLDGLSERCAQYKKDGADFAKWCVLKIGEHTPSALAIMENANVLARYASI
CQNGIVPIVEPEILPDGDHDLKRCQYVTEKVLAAVYKALSDHHIYLEGTLKPNMVTPGHACTQKYSHEEIAMATVTALRRTVPPAV
TGVTFSLGGQSEEEASINLNAINKCPLLKPWALTFSYGRALQASALKAWGGKKENLKAAQEEYVKRALANSLACQGYTPSGQAGAAA
SESLFISNHAY
```

Imatge 13 Seqüència completa de 2OT0_1, descarregada de PDB

- Quin tipus de relació existeix entre la proteïna simulada a MODEL i la que hem buscat?

Les dues proteïnes es tracten de aldolases, és a dir un enzim que participa en la glucòlisis. Catalitzant la escissió de la fructosa-1,6-bisfosfat en dos trioses, dihidroxiacetona fosfato y gliceraldehído3-fosfato.

- En base a això, creieu que la dinàmica de la seqüència problema serà semblant a les trobades a MODEL? Raoneu la resposta.

Si que serà semblant ja que es tracta pràcticament de la totalitat de semblança amb la proteïna MODEL, tot i tractar-se de organismes diferents.

Podem observar la tercera coincidència, amb un 72% d'identitat de seqüència:

- A quina proteïna i organisme correspon?

Correspon a la proteïna 1FBA (FRUCTOSE-1,6-BISPHOSPHATE ALDOLASE), aquesta proteïna prové de *Drosophila melanogaster*.

- Quina utilitat pot tenir una base de dades de dinàmiques com MODEL?

Les trajectòries i les anàlisis s'emmagatzemen en un gran magatzem de dades, que es pot consultar per obtenir informació dinàmica sobre proteïnes, incloses les interaccions. A part es proporcionen exemples de com es pot utilitzar per descriure les propietats de flexibilitat global de les proteïnes. Les anàlisis bàsiques i les trajectòries despallades de molècules de dissolvent a un nivell de resolució reduït estan disponibles a aquesta base de dades.

Ara busquem la seqüència:

MHVKERYKNFLNQHVGPDMVQRCNSEIGPNRKITLSGTDNGCKPVNTFILANK
RLIKTVCGRAGSPQGNMVRSNQFPVVKCVLNNGERHPYCEYRGTRSTRYIVLKCE
EGWPVHYHEDEVNNG

- Trobem una seqüència semblant a la base de dades? Quin valor d'identitat de seqüència té el primer resultat? A quin organisme correspon?

Amb aquesta seqüència trobem amb estructures no més elevades que un 38% de identitat (44/116) que correspon a la proteïna 1ANG provenint de l'organisme *Homo sapiens*.

UniProt or PDB id:

MHVKERYKNFLNQHVGPDMVQRCNSEIGPNRKITLSGTDNGCKPVNTFILANKRLIKTVCGRAG
SPQGNMVRSNQFPVVKCVLNNGERHPYCEYRGTRSTRYIVLKCEEGWPVHYHEDEVNNG

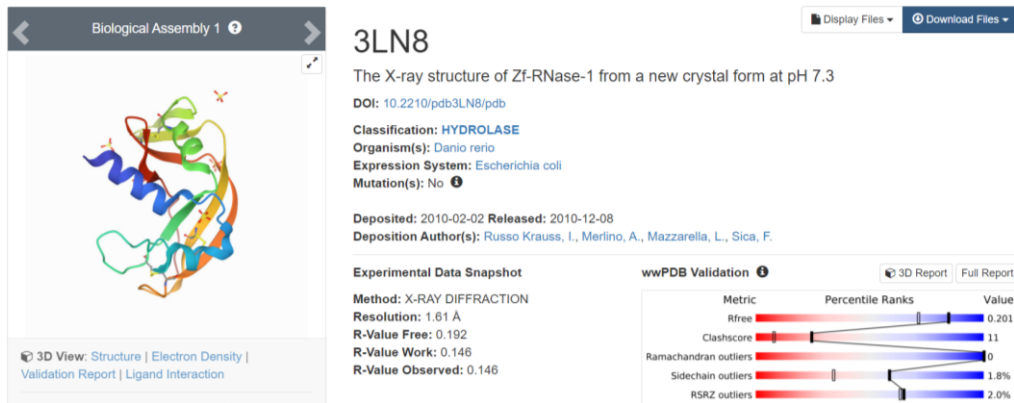
blast

hit	e-value	query from	query to	identity	description
2BWK chain A	9.09699E-17	6	116	43 / 116 (37%)	ANGIOGENIN
1ANG chain A	8.51453E-15	6	114	44 / 116 (38%)	ANGIOGENIN
1B1E chain A	2.47735E-14	6	114	43 / 116 (37%)	HYDROLASE A...
1AGI chain A	5.51895E-14	6	118	46 / 120 (38%)	ANGIOGENIN
1A5P chain A	1.11461E-6	10	120	29 / 115 (25%)	RIBONUCLEAS...

Imatge 14 Seqüències trobades a MODEL a partir de la seqüència facilitada.

- Ara busqueu a UNIPROT o PDB a quin organisme i proteïna correspon aquesta seqüència incògnita.

A partir de la cerca per seqüència a través de PDB trobem que aquesta correspon a 3LN8, proteïna que correspon al peix zebra (*Danio rerio*).



Imatge 15 Resultat de la cerca de la seq esmentada per PDB (<https://www.rcsb.org/structure/3LN8>)

- Expliqueu de manera raonada si creieu que podem utilitzar aquesta base de dades per a estudiar la dinàmica de la seqüència incògnita.

Si que podríem fer un estudi relatiu de la dinàmica molecular de la seqüència incògnita ja que aquest tipus de base de dades permetria simular la estructura i les propietats estàtiques de les molècules i els moviments relatius dels seus àtoms en context.

Exercici 2: Acoblament de fàrmacs (Docking): Ibuprofèn

La capacitat de l'ibuprofèn de suprimir la producció de prostaglandines (substàncies que informen el sistema nerviós del dolor) es deu a la inhibició de la ciclooxigenasa (COX), enzim necessari per a la síntesi d'aquestes molècules pro inflamatòries.

<https://es.wikipedia.org/wiki/PTGS1>

https://es.wikipedia.org/wiki/Antiinflamatorio_no_esteroideo

- Quin és el mecanisme d'acció de l'ibuprofèn?

Els antiinflamatoris no esteroïdals (abreujats AINE) són un grup variat i químicament heterogeni de fàrmacs principalment antiinflamatoris, analgèsics i antipirètics, per la qual cosa redueixen els símptomes de la inflamació, el dolor i la febre respectivament.

Els antiinflamatoris no esteroïdals disponibles al mercat inhibeixen l'activitat tant de la ciclooxigenasa-1 (COX-1) com a la ciclooxigenasa-2 (COX-2) i, per tant, la síntesi de prostaglandines i tromboxans. Es pensa que és la inhibició de la COX-2 la que en part causa l'acció antiinflamatòria, analgèsica i antipirètica dels AINE, però aquells que simultàniament

inhibeixen la COX-1 tenen la capacitat de causar hemorràgies digestives i úlceres, a especial l'aspirina. Per tant, s'emfatitzen els avantatges d'inhibidors selectius per a la COX-2.

L'AINE prototip és l'aspirina i l'acompanyen una gran varietat d'àcids orgànics, incloent-hi derivats de l'àcid propanoic (com l'ibuprofè i el naproxen)

Ara imaginem que no sabem en quin lloc (lloc d'unió o binding site) de la proteïna actua.

Necessitarem realitzar computacionalment un acoblament d'un fàrmac o lligand (en aquest cas l'ibuprofèn) a una molècula biològica (en aquest cas la ciclooxygenasa).

Per a realitzar un acoblament entrareu en la web de SwissDock:

<http://www.swissdock.ch/>

· Què és el que fa SwissDock?

SwissDock és un servidor web dedicat a realitzar simulacions d'acoblament de proteïnes-lligants de manera intuïtiva i exacte.

· Quin programa d'acoblament utilitza?

SwissDock es basa en el programa d'acoblament de proteïnes-lligants EADock DSS i té una interfície senzilla i integrada.

· Com es realitza l'acoblament molecular? Expliqueu breument tot el procés.

El SwissDock permet a l'usuari carregar fitxers d'estructura per a una proteïna i un lligand, i retorna els resultats per correu electrònic. Per facilitar la càrrega dels fitxers de proteïnes i lligands, podem preparar aquests fitxers d'entrada mitjançant el programa UCSF Chimera.

En aquest cas es buscaran els llocs d'unió de l'ibuprofèn a la ciclooxygenasa de l'ovella.

Aquesta és l'estructura de l'ibuprofèn tal i com es troba a la base de dades de ZINC:

<https://zinc.docking.org/substances/zinc000000002647/>

Al cap d'unes hores obtindreu el resultat de les prediccions dels modes d'unió de l'ibuprofèn a la proteïna.

• Quina és l'energia lliure d'unió (kcal/mol) estimada per a la primera estructura en el rànkung?

Show	Cluster	Element	FullFitness (kcal/mol)	Estimated ΔG (kcal/mol)
<input checked="" type="radio"/>	0	0	-2177.57	-9.73
<input type="radio"/>	0	1	-2174.95	-9.63

Imatge 16 Valors resultants de l'acoblament

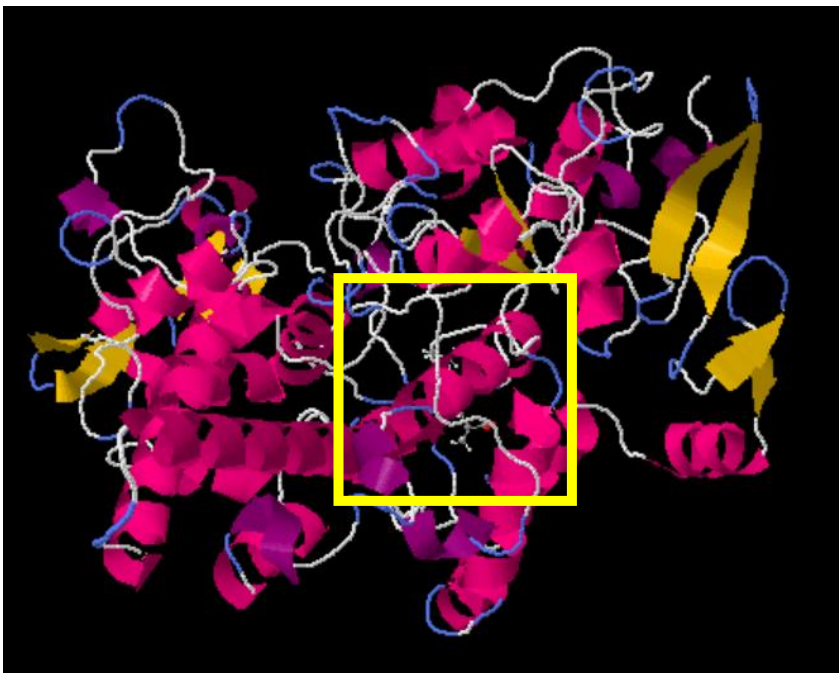
La energia lliure d'unió estimada per a la primera estructura en el rànding és -9,73 kcal/mol.

- Què representa l'energia lliure d'unió?

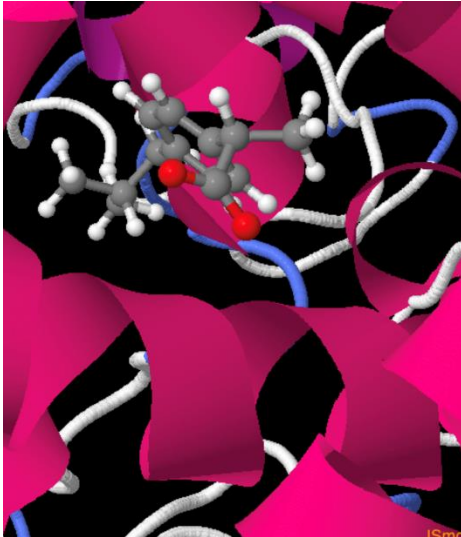
L'energia lliure d'unió és la suma de totes les interaccions intermoleculars que hi ha entre el lligand i la diana.

Selecioneu la primera solució del rànding amb Show per a veure l'estructura de l'ibuprofèn unida a la proteïna.

- Podeu fer una captura de pantalla per a visualitzar l'ibuprofèn a la ciclooxygenasa?



Imatge 17 Resultat de la interacció.



Imatge 18 Ampliació de la interacció.

- Les 5 primeres solucions se situen en la mateixa zona de la proteïna? Què és el que varia entre elles?

Si, les 5 primeres solucions estan situades gairebé a la mateixa localització de la proteïna, en aquestes 5 primeres solucions podem trobar diversitat en els valors de les energies lliures, tal com veiem representat en la següent imatge.

Show	Cluster	Element	FullFitness (kcal/mol)	Estimated ΔG (kcal/mol)
<input checked="" type="radio"/>	0	0	-2177.57	-9.73
<input type="radio"/>	0	1	-2174.95	-9.63
<input type="radio"/>	0	2	-2172.67	-9.51
<input type="radio"/>	0	3	-2168.21	-9.24
<input type="radio"/>	0	4	-2168.20	-9.24
<input type="radio"/>	0	5	-2151.24	-7.18

Imatge 19 Valors d'energia lliure assignats per cada element a la proteïna.

Ara llançareu el mateix càlcul, però tenint en compte la flexibilitat de les cadenes laterals de la proteïna (acoblament amb proteïna flexible). Seleccioneu Show extra paràmetres abans de llançar el càlcul i seleccioneu que hi hagi moviment de les cadenes laterals en els àtoms que es trobin a menys de 5 Àngstroms del lligand.

Hide extra parameters

Docking type

Accurate ▾

Definition of the region of interest

X center: Y center: Z center:

X size: Y size: Z size:

Flexibility

Allow flexibility for side chains within Å of any atom of the ligand in its reference binding mode - experimental

Imatge 20 Afegim paràmetres extra al càlcul.

- Compareu els resultats quant a valors d'energia lliure i quant a la posició del fàrmac dins de la proteïna.

Error:

- The ZINC database is currently not available.

No hit found!

<input type="button" value="Cancel"/>	<input type="button" value="Select all"/>	<input type="button" value="Toggle selection"/>	<input type="button" value="Select none"/>	<input type="button" value="Please select up to 1 ligand"/>
<input type="button" value="Cancel"/>	<input type="button" value="Select all"/>	<input type="button" value="Toggle selection"/>	<input type="button" value="Select none"/>	<input type="button" value="Please select up to 1 ligand"/>

Imatge 21 Error Zinc data base

A l'hora de calcular els paràmetres extra em genera un error que no em permet cercar Ligand selection Zinc2647.

- Creieu que és important que es tingui en compte la flexibilitat del lligand (fàrmac)? Raoneu la resposta.

Si que és important ja que l'activitat biològica d'un lligand també depèn de la flexibilitat. Molècules amb força enllaços rotals poden adoptar moltes conformacions diferents. Si una de les conformacions difereix de la conformació bioactiva o els àtoms interfereixen amb la unió, l'activitat biològica es pot destruir o com a mínim veure's afectada.

- A SwissDock, es té en compte la flexibilitat del lligand?

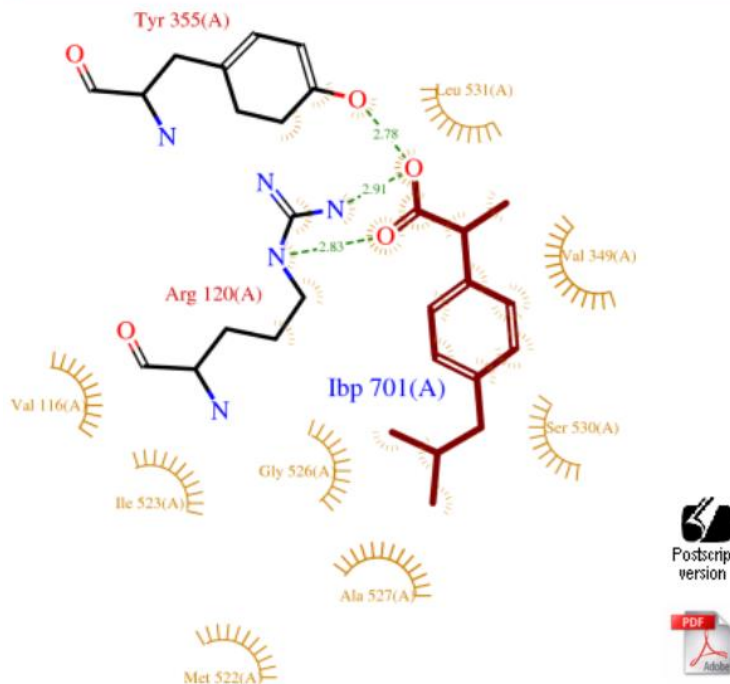
Si que és té en compte degut a la influència de la flexibilitat del lligand en la taxa d'èxit observada amb EADock DSS, el motor d'acoblament de SwissDock.

Ara busqueu a PDBSUM la nostra proteïna (1EQG) i trobeu a Ligands quines interaccions té el lligand ibuprofèn.

<https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/getpage.pl?pdbcode=1eqg&template=main.html>

- Realitzeu una captura de pantalla d'aquestes interaccions de l'ibuprofèn amb la proteïna. Amb quins residus interacciona?

1eqg: Ligplot



1eqg: Ligplot of interactions with IBP

Imatge 22 Interacció de Ibp amb Tyr 355 y Arg 120

- La nostra solució de Swissdock, se situa a la mateixa zona?

Si.

Exercici 3: Disseny computacional de fàrmacs

hERG és un gen que codifica per a una proteïna que és un canal d'ions potassi. Aquest canal de potassi és el millor conegut per la seva activitat elèctrica en el cor i coordina el seu batec.

<http://en.wikipedia.org/wiki/herg>

Quan el canal perd la seva habilitat per a conduir el corrent elèctric a través de la membrana cel·lular a causa d'algun fàrmac o alguna mutació en el gen,

es produeixen arrítmies i fins i tot la mort. Hi ha un gran nombre de fàrmacs en el mercat que inhibeixen aquest canal, cosa que suposa un greu efecte secundari per als pacients. La indústria farmacèutica posa un gran interès a detectar a priori quines estructures són susceptibles de bloquejar el canal hERG.

S'han publicat diversos treballs de models estructurals de fàrmacs que inhibeixen aquest canal, entre ells aquest:

Cavalli A, Poluzzi E, De Ponti F, Recanatini M. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K⁺ channel blockers. *J Med Chem* 2002; 45: 3844–53.

Trobareu l'article adjunt a l'enunciat de la PAC.

En el chart 1 de l'article (pàg. 2, 3 i 4) podeu observar una sèrie de fàrmacs (antibiòtics, antiinflamatoris, ...). Tots ells inhibeixen el canal hERG de potassi amb una certa activitat, mesurada amb el valor IC₅₀ (observeu la taula 1).

- Busqueu informació sobre què significa el valor IC₅₀ (per exemple, als apunts o a la Wikipedia) i expliqueu què representa.

El valor IC₅₀ representa el valor per assolir la concentració necessària per bloquejar el 50% de la resposta biològica. Aquest valor és útil per comparar la potència de fàrmacs amb eficàcies similars. Com més baix sigui IC₅₀, més gran és la potència de l'antagonista, és a dir, cal menys concentració de fàrmac per inhibir la resposta biològica. Per al disseny de fàrmacs cal tenir en compte que concentracions baixes de fàrmacs estan associades a menors efectes secundaris.

- Quin dels fàrmacs que ens anomena Cavalli a l'article bloqueja amb una activitat major el canal? I el que menys bloqueja el canal?

Astemizole i gatifloxacin són els fàrmacs que bloquegen una menor i major activitat del canal K.

Table 1. Observed and Calculated HERG K⁺ Channel Blocking Activity of Compounds 1–31

compound	IC ₅₀ (nM)	pIC _{50obsd}	pIC _{50fit} ^a	Δ
astemizole (1)	0.9 ^b	9.04	8.53	0.51
cisapride (2)	6.5 ^b	8.19	7.96	0.23
E-4031 (3)	7.7 ^b	8.11	7.85	0.26
dofetilide (4)	9.5–15 ^b	7.91	7.67	0.24
sertindole (5)	14 ^b	7.85	8.04	−0.19
pimozide (6)	18 ^b	7.74	7.80	−0.06
haloperidol (7)	28.1 ^b	7.55	7.58	−0.03
droperidol (8)	32.2 ^b	7.49	7.82	−0.33
thioridazine (9)	35.7 ^b	7.45	7.23	0.22
terfenadine (10)	56–204 ^b	6.89	7.22	−0.33
verapamil (11)	143 ^b	6.84	7.05	−0.21
domperidone (12)	162 ^b	6.79	6.88	−0.09
loratadine (13)	173 ^b	6.76	5.83	0.93
halofantrine (14)	196.9 ^c	6.70	6.81	−0.11
mizolastine (15)	350 ^b	6.45	6.65	−0.20
bepiridil (16)	550 ^d	6.26	6.30	−0.04
azimilide (17)	560 ^c	6.25	6.15	0.10
mibefradil (18)	1430 ^d	5.84	5.75	0.09
chlorpromazine (19)	1470 ^c	5.83	5.68	0.15
imipramine (20)	3400 ^c	5.47	5.98	−0.51
granisetron (21)	3730 ^b	5.42	5.64	−0.22
dolasetron (22)	5950 ^b	5.22	4.99	0.23
perhexiline (23)	7800 ^b	5.11	5.18	−0.08
amitriptyline (24)	10000 ^b	5.00	5.66	−0.66
diltiazem (25)	17300 ^b	4.76	5.02	−0.26
sparfloxacin (26)	18000–34400 ^c	4.58	4.39	0.19
glibenclamide (27)	74000 ^e	4.13	4.07	0.06
grepafloxacin (28)	50000–104000 ^c	4.11	4.35	−0.24
sildenafil (29)	100000 ^b	4.00	3.50	0.50
moxifloxacin (30)	103000–129000 ^c	3.93	3.82	0.11
gatifloxacin (31)	130000 ^c	3.89	4.16	−0.27

^a Calculated from the non-cross-validated CoMFA model. ^b In human embryonic kidney (HEK) cells. ^c In Chinese hamster ovary (CHO) cells. ^d In African green monkey kidney derived cell line COS-7. ^e In neuroblastoma cells.

Imatge 23 Llistat de fàrmacs bloquejadors del canal K.

• Quin és el valor de les activitats d'aquests fàrmacs?

El valor de les activitats d'aquests fàrmacs són, per astemizole 0,9nM i per gatifloxacin 130000nM.

Drugbank és una base de dades que proporciona dades sobre fàrmacs. Cada registre conté dades sobre l'estructura química i sobre les dianes terapèutiques del fàrmac.

<https://go.drugbank.com/>

Trieu per exemple el fàrmac thioridazine i l'introduïm a la base de dades (Browse---Drugs). En seleccionar l'entrada obtenim característiques del fàrmac en qüestió.

• Per a què estava indicat aquest fàrmac?

La tioridazina és un antipsicòtic de fenotiazina que s'utilitza per tractar l'esquizofrènia i el trastorn d'ansietat generalitzada.







- Què podeu dir sobre els efectes secundaris que té el fàrmac? què diu sobre el bloqueig d'hERG?

Somnolència, visió borrosa, boca seca, nàusees, vòmits, diarrea, restrenyiment, canvis en la gana, l'augment de pes, pell pàl·lida, enfosquiment de la pell o dels ulls, inflor dels braços, mans, peus, turmells o cames inferiors, expressió facial en blanc, inquietud, augment dels pits, dificultat per orinar.

La tioridazina (THIO) és un derivat de la fenotiazina que s'utilitza principalment per al tractament de trastorns psicòtics. Tanmateix, les arítmies cardíques, especialment la prolongació de l'interval QT associada a l'aplicació d'aquest compost, han rebut una atenció seriosa després de la seva introducció a la pràctica clínica, i els mecanismes subjacents a la cardiotoxicitat induïda per THIO no han estat ben definits.

Se sap que la tioridazina inhibeix els canals cardíacs hERG K(+).

- En quines proteïnes actua? El lligand és inhibidor, agonista o antagonista per a aquestes proteïnes? Quina d'elles està relacionada amb les arrítmies?

TARGET	ACTIONS	ORGANISM
 Dopamine D2 receptor	antagonist	Humans
 Dopamine D1 receptor	antagonist	Humans
 Alpha-1A adrenergic receptor	antagonist	Humans
 Alpha-1B adrenergic receptor	antagonist	Humans
 5-hydroxytryptamine receptor 2A	antagonist	Humans
 Potassium voltage-gated channel subfamily H member 2	inhibitor	Humans

Imatge 24 Targets en els que actua.

- En quin any es va retirar del mercat?

La tioridazina es va retirar a tot el món l'any 2005 a causa de la seva associació amb arrítmies cardíques.

- Quin article científic demostra la vinculació d'aquest fàrmac amb l'efecte secundari?

Thioridazine and severe cardiac arrhythmia. (<https://pubmed.ncbi.nlm.nih.gov/11824444/>)

Existeix un programa: Pred-hERG que busca el possible bloqueig o no d'hERG, una eina molt útil per a saber si la molècula pot o no tenir aquest efecte secundari:

<http://predherg.labmol.com.br/>

SMILES és un text amb un format que defineix inequívocament una molècula a partir d'una cadena de caràcters que la defineix químicament.

Busqueu astemizole a Drugbank i busqueu el seu codi SMILES.

Introduïm el codi SMILES en el predictor Pred-hERG.

Smiles


astemizole:

COC1=CC=C(CCN2CCC(CC2)NC2=NC3=CC=CC=C3N2CC2=CC=C(F)C=C2)C=C1

- És un potent inhibidor o, per contra, no inhibeix el canal? Amb quina probabilitat inhibeix al canal de potassi? Coincideix això amb les dades de l'article de Cavalli?

Endpoint:	hERG
Method:	RandomForest
Accuracy:	0.844
AUC:	0.879
CV fold:	5
Descriptors:	ndb, nsb, ncarb, nsulph, naro, ndonr, nhev, nacor, nta, nring, PC6, GMTIV, AW, Geto, BertzCT, J, MZM2, phi, kappa2, MATSv1, MATSv5, MATSe4, MATSe5, MATSe6, TPSA, Hy, LogP, LogP2, UI, QOss, SPP, LDI, Qass, QOmin, QNmax, Qmin, Mnc, EstateVSA7, EstateVSA0, EstateVSA3, PEOEVSa0, PEOEVSa6, MRVSA5, MRVSA4, MRVSA3, MRVSA6, slogPVSA1
Explain:	The Probability for the Category means probability of predicted category 1; The details about how to explain the results , please see here >>

Imatge 25 Resultat del model

Show	10	entries	Search:	<input type="text"/>
Identifier	Category	Probability	Structure	
0	<chem>COC1=CC=C(CCN2CCC(CC2)NC2=NC3=CC=CC=C3N2CC2=CC=C(F)C=C2)C=C1</chem>	1	0.986	

Showing 1 to 1 of 1 entries

Previous

1

Next

Imatge 26 Resultat del model

Ara fem el mateix per a gatifloxacin..

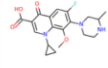
Smiles gatifloxacin: COC1=C2N(C=C(C(O)=O)C(=O)C2=CC(F)=C1N1CCNC(C)C1)C1CC1.

- Quins resultats obtenim en aquest cas? Què signifiquen les zones en color rosa?

Details of the model	
Endpoint:	hERG
Method:	RandomForest
Accuracy:	0.844
AUC:	0.879
CV fold:	5
Descriptors:	ndb, nsb, ncarb, nsulph, naro, ndonr, nhev, nacor, nta, nring, PC6, GMTIV, AW, Geto, BertzCT, J, MZM2, phi, kappa2, MATSv1, MATSv5, MATSe4, MATSe5, MATSe6, TPSA, Hy, LogP, LogP2, UI, QOss, SPP, LDI, Qass, QOmin, QNmax, Qmin, Mnc, EstateVSA7, EstateVSA0, EstateVSA3, PEOEVSa0, PEOEVSa6, MRVSA5, MRVSA4, MRVSA3, MRVSA6, slogPVSA1
Explain:	The Probability for the Category means probability of predicted category 1; The details about how to explain the results , please see here >>

Imatge 27 Resultats obtinguts.

Showing 1 to 1 of 1 entries

Identifier	Category	Probability	Structure
0 <chem>COC1=C2N(C=C(C(O)=O)C(=O)C2=CC(F)=C1N1CCNC(C)C1)C1CC1</chem>	0	0.21	

Showing 1 to 1 of 1 entries

Previous 1 Next

Download as CSV

Imatge 28 Resultats obtinguts