Este documento pretende recoger los contenidos de las reuniones con Rocío: preguntas, comentarios, anotaciones, tareas que hacer, cosas pendientes semana a semana, revisiones, y puntos a mejorar, entre otras cosas.

# Primera reunión (27/10/22)

## Objetivos de la reunión

* Presentación y primera toma de contacto con el tema del trabajo

## Preguntas para esta reunión

* ¿Estructura de la memoria? 🡪 ver tfgs antiguos
* Vías de comunicación 🡪 por correo
* ¿Llevar un registro de horas de cada cosa en lo que trabajo? 🡪 sí, pero no es necesario que sea muy exhaustivo

## Conclusiones de la reunión

Me enviará material de lectura, tfgs antiguos de referencia para que les eche un vistazo, un set de 100 moléculas con los que poder realizar el análisis, y código para probar en Python.

- Ir guardando la bibliografía que vea, y si es una página web apuntar el día de visita (ir guardando esto para luego no quedarme vendido)

- Páginas webs que usaré para las cadenas SMILES: sigmaaldrich.com / scifinder / pubchem (esto está pensado para investigación, los compuestos comerciales no los suele tener). Yo tengo que comparar scifinder y sigmaaldrich principalmente.

- [Rdkit va muy bien para cosas básicas, pero cuando se le pide un poco más, no funciona bien. Algunas moléculas de organometalica ni siqueira las acepta por ser químicamente inviables. Este paquete lo podemos usar para probar y hacer una introducción de que existe más de un paquete, pero vamos a trabajar con Openbabel. (no tengo que comparar los paquetes como tal, es simplemente para probarlo)

Si dibujamos la misma molécula con ambos paquetes, dan un dibujo distinto, que no tiene por qué estar mal.

Openbabel tiene un método ‘canonical’ que si le metes una cadena SMILEs, lo convierte a la secuencia canónica, con la idea que todos los sinónimos de la misma molécula, debería darte el mismo canónico. Esa es la idea, pero no funciona del todo bien.

Mi tarea es crear un método canonical que mejore aunque sea mínimamente eso, específicamente para moléculas organometalicas.

Cosas que hacen mal todos los paquetes: dibujar. Para las moléculas convencionales va muy bien, para las orgánicas/inorgánicas también; pero las organometalicas le cuesta. SciFinder por otro lado, hace los dibujos mucho mejor, pero es de pago y no tenemos acceso al código de como lo hacen.

Hay un paquete que usan los químicos para dibujar, que es chemdraw o algo así. Pero solo funciona con RDkit, y como rdkit no se traga las moléculas organometalicas, no podemos usar esto.]

## Tareas nuevas que hacer

* Leer papers y material 🡪Hecho
* Repaso de química básica 🡪 Hecho
* Estudiar y escoger metodología de trabajo (tradicional o agil) 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪**Por hacer**
* Analisis de riesgos, planes de contingencia 🡪**Por hacer**
* Ir guardando la bibliografía que vaya viendo 🡪 viene bien usar gestores como Zotero**.**
* Algo para contabilizar las horas que le echo al tfg🡪 Uso Clockify

## Próxima reunión

La próxima reunión está fechada para el 09/11/22

# Segunda reunión (16/02/23)

La reunión para el 09/11/22 quedó a la espera debido a las asignaturas del primer cuatrimestre y las vacaciones de Navidad.

## Objetivos de la reunión

* Repasar los objetivos del trabajo en líneas generales
* Establecer la metodología de desarrollo
* Resolver las dudas surgidas tras las lecturas iniciales
* Problemas instalando algunos paquetes de python para probar el código

## Preguntas para esta reunión

* ¿Hasta qué punto necesito aprender SMILES, es solo para describir compuestos, o también reacciones?
  + Respuesta: lo justo solo para describir compuestos y moléculas
* ¿Cómo busco artículos e información buena? ¿Cómo accedo a plataformas tipo ieeexplore o ACSPublications?
  + Respuesta: Normalmente la ugr tiene convenios para poder acceder a algunos papers. Pero lo más fácil es usar Google Scholar y buscar por ahí. Todo lo que vea/lea (aunque sea por curiosidad, o solamente el abstract, o directamente no lo use para la memoria) tengo que guardarlo.
* Dijiste que tenias una idea de como podría ser esa forma canónica. ¿Eso lo tengo que desarrollar yo, o los químicos con los que trabajas me lo proporcionan?
  + Respuesta: mirar conclusiones
* ¿Cómo accedo a SciFinder, necesito registrarme con una cuenta ugr?
  + Respuesta: en principio si, se supone que es gratis, pero tengo que usarlo desde la VPN de la ugr. 🡪 **No me funciona, necesito un formulario especial para hacer el registro.**

## Conclusiones de la reunión

- Hay varias representaciones lineales modernas. SMILES, InChI y SELFIES. SMILES es la más usada, pero con el problema de que una misma molécula se puede escribir de muchas formas distintas. InChI va muy bien para la máquina, pero no es nada legible por los humanos. Y SELFIES, tiene la ventaja de que una molécula se representa de una única manera, y todas las representaciones que genera son químicamente viables; cosa que SMILES no hace.

A la hora de dibujar las moléculas, específicamente en el campo de la organometalica, no se dibujan para nada bien. Si le meto un código SMILES determinado me lo dibuja de una manera; si lo modifico (reordenando la cadena) aun siendo el mismo compuesto, me lo dibuja de otra manera distinta. Como en cada página (base de datos) tienen un código diferente, no hay nada unificado.

- Por tanto, tengo que analizar esas diferencias entre la misma molécula dependiendo de donde la saque (webs).

- Determinar un sistema que me genere, a partir de cualquier sinónimo SMILES de la misma molécula, un único SMILES canonico. Internamente Openbabel monta un grafo. La idea es a partir del SMILES de entrada, montarlo en un grafo; y escribir ese grafo de manera canonica. Aquí la complejidad está en ver cómo recorro ese grafo. Esto me tendré que inventar algunas reglas. (p.ej, usando el nº atomico)

Cadena original 🡪 Grafo (esto ya lo hace openbabel) 🡪 Generar cadena canonica

## Tareas nuevas que hacer

* Mirar Google Colab e intentar montar ahí el entorno de pruebas si me sigue fallando el openbabel. 🡪 Hecho
* Hacer una tabla resumen con las diferencias entre los códigos SMILES y dibujos para una misma molécula en las distintas webs. Usar también Rdkit, hay algunas moléculas que no soporta. Esto para la introducción/motivación. 🡪 **Por completar**
* Escribir la introducción al trabajo entorno a lo anterior (química informática, sistemas de representación, organometalica…) 🡪 **Por completar**
* En general, ir rellenando las partes iniciales de la memoria.

## Tareas pendientes

* Estudiar y escoger metodología de trabajo (tradicional o agil) 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪 **Por hacer**
* Analisis de riesgos, planes de contingencia 🡪 **Por hacer**

## Tareas a futuro

* Una vez tenga la introducción medio que, lo siguiente sería ir cogiendo algunas moléculas del set de 30 (las que sean fáciles de entender, con un SMILES corto) y comenzar con el método de canonización.
* Para lo anterior, necesitaré entender y estudiar el código de openbabel, y su estructura de clases. Tendré que hacer también un diagrama de clases general (será bastante enorme) y quizás uno con resaltando las clases que yo vaya a utilizar/modificar/ampliar.

## Próxima reunión

La próxima reunión está fechada para el 01/03/23

# Tercera reunión (01/03/23)

## Objetivos de la reunión

* Revisión del sprint hasta el momento
* Solucionar algunas dudas

## Preguntas para esta reunión

* No son 100 moléculas, eran 30 los del excel. No sé si afectará mucho esto.
  + Respuesta: no importa, usamos este conjunto de datos.
* Preguntar sobre este documento de las actas, si lo ve bien.
  + Respuesta: sí, está bien. Luego esto lo puedo anexar como informes de sprints.
* En google colab: Preguntar por los distintos modos de dibujado ¿Qué paquete está dibujando cada molécula?
  + Respuesta: el nombre del import Chem.Draw de Rdkit coincide con el nombre del programa (que no tiene que ver) ChemDraw, por lo que lo dibuja Rdkit propiamente. Openbabel igual, tiene sus propios métodos de dibujar.
* Porque en algunos SMILES usan el “.” Y luego en el dibujo aparecen con enlaces. ¿No se supone que eso son moléculas desconectadas? P.ej el 16 del excel. O más fácil, el 4 del excel: en sigma aldrich tiene un “.”, pero el dibujo sale junto; en pubchem sale el “.” Y el dibujo sale separado; y en SciFinder no aparece “.” Y el dibujo sale junto.
  + Respuesta: se supone que el punto indica desconexión, pero no siempre. Puedes hacer enlaces a mano usando los números de los ciclos (no se supone que tengan que servir para eso, pero se puede)
* ¿Trabajas con químicos del ICIQ? ¿Con cuántos? ¿En qué área trabajan, específicamente en organometalica? Todo esto para meter un párrafo en la motivación de la memoria, para poner en contexto
  + Respuesta: si, trabaja con un grupo de investigación de Tarragona, rocío se comunica con la jefa del grupo. Sí, en concreto de organometalica.
* ¿Hasta qué punto les supone un problema a los investigadores/químicos que usan estas bases de datos, estas diferencias entre representaciones?
  + Respuesta: Su problema como tal no es que sean distintas, sino que no pueden dibujarlos, de ninguna de las maneras de las bases de datos. Por lo visto el código SMILES no es muy apto para organometálica, les dificulta el trabajar con estas moléculas, si no las pueden dibujar, no las pueden entender. El problema no es tanto de la estructura de la cadena, sino de que luego puedan trabajar con ella. Lo que necesitan ellos, es que SMILES codifique correctamente una molécula de organometalica para poder luego usar todas las herramientas. Hasta el momento, según comenta rocío, no hay ninguna notación que haga esto bien. De hecho, no tiene por qué ser SMILES, la cosa es obtener una nomenclatura que les sirva. Nos centramos en SMILES porque es la más usada e intentamos “parchearla”
* Preguntar por la estructura de la memoria: ¿estado del arte?
  + Respuesta: habrá que hacer algo si. Podré meter una revisión bibliográfica de la literatura existente (tampoco muy extenso esto). Tendré que hablar por aquí de los distintos paquetes que existen (rdkit, openbabel); las distintas formas de representar una molécula (SMILES, InChI, SELFIES) con una imagen y una explicación corta/media de cada una.

Los fundamentos teóricos sin pasarme tampoco, que sino se me hace muy largo este capitulo

* En la tabla comparativa inicial: yo lo he enfocado más entre Scifinder y sigma-aldrich ¿incluyo también PubChem? ¿inlcuyo tb los dibujos que yo mismo pueda generar con rdkit y openbabel?
  + Respuesta: con esas 2 va bien, en la motivación es solamente para mostrar que cada fuente de datos hace más o menos lo que quiere. Puedo hacer una tabla más pequeña como la que tengo para la motivación, a modo de ejemplo. Luego, con el dataset de moléculas, hacer una tabla más grande (en horizontal si es necesario, como anexo/apéndice).

## Conclusiones de la reunión

- Han sido todo prácticamente preguntas, lo importante ya lo he anotado

- Terminar los apartados por completar del primer capítulo

## Tareas nuevas que hacer

* Hacer diagrama de gantt de planificación, aunque sea grosso modo, para ver los tiempos que llevo y ver si voy lento o no.

## Tareas pendientes

* Hacer una tabla resumen con las diferencias entre los códigos SMILES y dibujos para una misma molécula en las distintas webs. Usar también Rdkit, hay algunas moléculas que no soporta. Esto para la introducción/motivación. 🡪 **Por completar**
* Escribir la introducción al trabajo entorno a lo anterior (química informática, sistemas de representación, organometalica…) 🡪 **Por completar**

Menos prioritario:

* Estudiar metodología de trabajo (tradicional o agil) 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪 **Por hacer**
* Analisis de riesgos, planes de contingencia 🡪 **Por hacer**

## Tareas a futuro

Sin cambios respecto a la anterior reunión.

## Próxima reunión

La próxima reunión está fechada para el 16/03/2023

# Cuarta reunión (16/03/23)

## Objetivos de la reunión

* Revisión del sprint hasta el momento
* Tareas para el siguiente sprint

## Preguntas para esta reunión

* Viendo las moléculas en la base de datos de SciFinder, en todas ponían que eran “coordination compound”. ¿Es lo mismo que organometálicos?
  + Respuesta: leer conclusiones.
* Dependiendo del uso que se le quiera dar a la molécula, ¿canonizar puede implicar codificarlo de una manera, y para otros químicos en otra área les puede venir mejor otra forma?
  + Respuesta: leer conclusiones

## Conclusiones de la reunión

* Con respecto a la tabla grande del apéndice: ver si puedo meter de alguna manera el nombre del compuesto. Si puedo incluirlo en otra columna bien. Si se me junta todo mucho, puedo ponerle un identificador y hacer otra tabla debajo con los nombres y los ids.

Mónica se unió a media reunión para resolver algunas dudas:

Los compuestos organometálicos son los que tienen un enlace metal-carbono. Los compuestos de coordinación solamente tienen un metal coordinado a ‘cosas’ (que no tienen que ser un carbono, son ‘restos’ y otras cosas, ligandos; supongo que resultado de alguna reacción), entendiendo coordinado como enlazado.

No todos los compuestos de coordinación son compuestos organometálicos, pero sí podemos considerar todos los organometálicos como de coordinación. En una reacción, la cosa es que exista una unión metal-carbono en algún sitio para considerar la reacción organometalica.

Considerando ‘metales’ los de transición porque tienen orbitales ‘d’ disponibles. (formalmente desde el grupo 3 hasta el 12, del escandio al zinc), pero para Monica, en la práctica sería del Titanio al Cobre (el zinc tiene muchos estados de oxidación, y no se comporta en las reacciones como el resto de metales muchas veces)

En principio, si es canónico es único, pero claro, puedes mover el orden de los átomos en el string y seguir teniendo la misma estructura. Bien es cierto que hay algunas reglas para la química orgánica a la hora de ordenar moléculas, trasladar estas reglas a organometalica es complicado porque en la mayoría de casos un mismo fragmento está enlazado por varios sitios al metal (y claro, por donde empiezas), hay muchas excepciones. La cosa es establecer un orden.

En los dibujos, comenta que salen muy pegados. Simplemente el hecho de separar un poco los átomos para que no salgan tan juntos, sería bueno. Pierden también la tridimensionalidad de los enlaces al salir tan aplastados.

* Ese conjunto de moléculas que escoja mandárselo a Rocío para ver como Mónica los dibujaría a mano (o como le gustaría que quedara dibujado) para ver si coinciden con alguno de las bases de datos [ver tareas]

## Tareas nuevas que hacer

* Elegir un subconjunto de unas 5 moléculas (que yo vea que son muy distintas en el dibujo, o que sean sencillas de seguir en el código) y trabajar con ellas.
  + Mándaselo a Rocío para ver como Mónica los dibujaría a mano (o como le gustaría que quedara dibujado) para ver si coinciden con alguno de las bases de datos.
  + Con estas moléculas, ver cómo el código de openbabel trabaja con ellas para codificarlas/almacenarlas/dibujarlas.
  + Probar tb a cambiar el código SMILES: escribirlo de una manera distinta, pero que siga siendo la misma molécula, y ver si cambia mucho la imagen.

## Tareas pendientes

Ir haciendo de la memoria:

* Estudiar metodología de trabajo (tradicional o agil) 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪 **Por hacer**
* Análisis de riesgos, planes de contingencia 🡪 **Por hacer**

## Tareas a futuro

* Lo siguiente sería ir cogiendo algunas moléculas del set de 30 (las que sean fáciles de entender, con un SMILES corto) y comenzar con el método de canonización. **-> En proceso**
* Para lo anterior, necesitaré entender y estudiar el código de openbabel, y su estructura de clases. Tendré que hacer también un diagrama de clases general (será bastante enorme) y quizás uno con resaltando las clases que yo vaya a utilizar/modificar/ampliar.
* Estado del arte y fundamentos teóricos, ir haciéndolo. Pero más tarde cuando entienda más del tema.

## Próxima reunión

La próxima reunión está fechada para el 30/03/2023

# Quinta reunión (28/03/23)

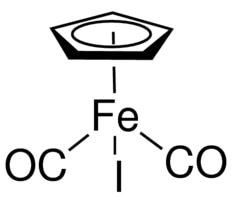
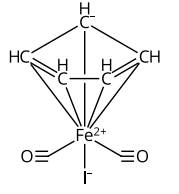
## Objetivos de la reunión

* Comentar varias cosas y problemas

## Preguntas y comentarios para esta reunión

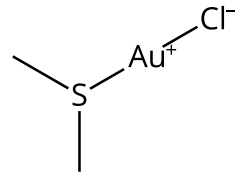
* La regla esta de que los átomos quieren intentar llenar sus orbitales ¿se aplica solo en química orgánica básica? ¿en organometálica es distinto? Por ejemplo: el compuesto Dicarbonylcyclopentadienyliodoiron(II) 🡪

SMILES: O#C[Fe+2]1234([I-])(C#O)[CH]= 5[CH]4=[CH]3[CH-]2[CH]51

Se supone que el Hierro+2 haría enlace con 6 átomos (Fe, tiene de normal configuración ) para llenar el orbital d completo. Pero hace 8 enlaces según el dibujo.

Con este compuesto igual: [Cl-][Au+][S](C)C. ¿Por qué el Cl- se va a enlazar a algo?



Lo pregunto porque probando RdKit, el error más común que da, y por el que parece que no trabaja bien con organometálicos, es algo del tipo: “Explicit valence for atom #0 Cl, 2, is greater than permitted” o “Explicit valence for atom #1 O, 3, is greater than permitted”, como si no permitiera esas valencias o algo así.

* + Respuesta: se resolvió por correo la duda.
* Openbabel como tal no soporta el dibujado en 3D de las moléculas, por lo que los únicos dibujos que puede hacer son en 2D. Openbabel tampoco soporta el uso de wedge ni hash bonds para el dibujado de moléculas en 2D con perspectiva (es curioso porque en el código sí que hay funciones dedicadas a esto, pero luego cuando le metes símbolos SMILES de @@ y demás, los medio ignora. Sigo ejemplos del tutorial de Daylight, pero no salen los mismos dibujos. Otros símbolos más dedicados a geometría que viene en Daylight o en OpenSMILES, tipo @SP, @TB, @OH, los ignora por completo). Pero sí es capaz de generar archivos .sdf con información 3D, que se pueden usar en otros softwares de dibujado específicos como Avogadro (y no están mal, mucho mejor que los 2D desde luego)
* Estoy viendo que los SMILES de SA son muy malos. Prácticamente todos usan el punto, por lo que los dibujos salen separados (no es muy útil esto). SF arregla esto con el uso de ciclos, haciendo que los enlaces estén en condiciones; pero openbabel sigue haciendo los dibujos muy pegados, y superponiendo símbolos (sobre todo en compuestos que a mano se dibujarían con perspectiva. Los compuestos que son planos los dibuja bastante bien (si tiene muchos atomos quizás se solapan un poco, pero se ve bien), y aquí tampoco hay mucho margen de mejora)

## Conclusiones de la reunión

## Tareas nuevas que hacer

* Entender bien el código de Openbabel, si no, no puedo trabajar en condiciones:
  + Averiguar cómo se construyen los vectores de átomos y enlaces dentro de OBMol (en qué orden se introducen en el vector) y si varía según el string de entrada que le meta.
  + Una vez eso, saber cómo recorre el vector de átomos a la hora de dibujarlo, para modificar ese orden; o bien, le doy los átomos en el orden necesario (de entrada supongo) para que me lo dibuje como yo quiera.

## Tareas pendientes

* Sin cambios

## Tareas a futuro

* Sin cambios

## Próxima reunión

La próxima reunión queda fechada para después de semana santa.

# Sexta reunión (15/05/23)

## Objetivos de la reunión

* Presentar avances y proponer ideas

## Preguntas y comentarios para esta reunión

* ¿Vendría bien documentar el proceso de cómo he buildeado el proyecto de Openbabel, y como puedo modificar el código para probar los cambios? ¿Lo meto en un apéndice quizás?
  + *Respuesta*: sí, podría estar bien hacer como un manual de usuario

Conclusiones debajo de lo que he visto de codigo (para más, ver documento de apuntes.docx)

* Como el algoritmo de gen2D es relativamente complejo modificarlo (es un plugin que se añadió después, no hay nada documentado), he pensado usar lo que genere, y luego yo modificar los atomos que me interesen para que quede mejor en según qué moléculas. He visto que las moléculas ‘planas’ las dibuja bastante bien (aunque hay margen de mejora en las más complejas). Para las moléculas con perspectiva, me he querido centrar en las de tipo ferroceno o que tienen Cp. Y en eso estoy.
* Ahora mismo, el progreso que llevo con el dibujado de los panes de Cp:
  + De momento solamente funciona ‘bien’ con Cp individuales.
  + Tengo que desarrollar el método que me permita separar varios Cp distintos y poder crear el enlace tangente a cada uno de ellos. ESTO ES IMPORTANTE (hay otras moléculas tipo ferroceno que se estropean si no completo esto)
  + De momento estoy dibujando los carbonos usando el sistema de dibujado existente, aprovechando que ya tengo los enlaces creados (quedando el dibujo como SF)
    - Si quisiera darle un aspecto a los Cp como los de SA, necesito más tiempo, pero puede llegar a quedar muy limpio.
    - Pero la percepción de anillos que tiene esto es según SSSR (lo cual complica un poco las cosas a la hora de operar con los anillos. Los anillos del SMILES a veces se dividen en varios más pequeños, sobre todo en organometalica; con anillos normales de benceno y esos no ocurre)
  + Me gustaría también, ser capaz de mover los enlaces que pudieran quedar en medio de los Cp, para que no se solapen mucho.
  + He visto que los ferrocenos se dibujan bastante bien. Pero los Cp individuales no.
  + Para seguir testeando esto, necesitaría algunos complejos que tengan Cps, porque del dataset actual solamente tienen 3, y 2 de ellos son relativamente complejos.
* Se me ocurre también, como tengo varias maneras de dibujar los Cp (con circulo, sin circulo, con enlace único central, o con todos los enlaces); intentar poner esto como una opción de línea de comandos (tipo -:”<smiles>” -O <archivo> -<opciones\_cp>)

## Conclusiones de la reunión

Mirar las tareas

## Tareas nuevas que hacer

* Seguir con la mejora de los dibujos (en concreto, intentar rotar los panes para darle perspectiva). A la espera de consultar con Mónica qué tipo de dibujo le resulta más cómodo o fácil de entender.
* Probar aunque sea 1 método de canonización. Lo de las ramas. Atomo central y luego el resto en ramas, ordenándolas por algún criterio (en sentido horario, o quizás por longitud de rama, o intercalados, algo así)
* Intentar desarrollar un método que separe el smiles como en bloques, según las ramas o como me sea más cómodo a mi para luego parsearlo con condiciones.
  + Hacer quizás algún testing con un set enorme de moléculas (de organometalica y orgánica normal) para ver el resultado.

## Tareas pendientes

Ir haciendo de la memoria lo que pueda:

* Estudiar metodología de trabajo (agil) y planificación temporal 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪 **Por hacer**
* Análisis de riesgos, planes de contingencia 🡪 **Por hacer**

Hay partes de la memoria que hasta no tener el código hecho, no puedo documentar obviamente:

* Métodos creados (explicarlos)
* Diagrama de clases con los cambios/adiciones/clases modificadas
* Experimentación y Resultados
* Método de canonización alcanzado/creado con sus reglas definidas

## Tareas a futuro

## Próxima reunión

La próxima reunión está prevista para el 29/05/2023