

Nous mètodes computacionals basats en grafs per a tractar la reactivitat química i la catàlisi

New graph-based computational methods for dealing with chemical reactivity and catalysis

Diego Garay-Ruiz,¹ Enric Petrus¹ i Carles Bo^{1,2}

¹ Institut Català d'Investigació Química (ICIQ)

² Universitat Rovira i Virgili (URV). Departament de Química Física i Inorgànica

Resum: El creixement exponencial del poder de la computació ha comportat un impuls decisiu en la capacitat de predicció de la química teòrica i computacional. Malgrat aquests avenços, l'estudi de la reactivitat en sistemes químics complexos que implica el tractament de xarxes de reaccions molt denses és difícil. En aquest article mostrem tres mètodes nous que permeten crear, tractar i processar d'una manera eficient i automàtica cicles catalítics i xarxes de reaccions complexes. En primer lloc presentem una eina per a predir la freqüència de recanvi en cicles catalítics, tant en la catàlisi homogènia com en l'heterogènia. En segon lloc, tractem el que hem anomenat POMSimulator, un mètode que permet determinar mapes de reacció, predir l'especiació en funció del pH i de les concentracions, i també els mecanismes de formació dels òxids moleculars. Finalment, presentem OntoRXN, una nova ontologia orientada a definir formalment els mecanismes de reacció, que facilita el tractament i el processament d'aquesta informació química complexa.

Paraules clau: Xarxes de reacció, química computacional, catàlisi, quimioinformàtica.

Abstract: The exponential growth of computing power has crystallized in a breakthrough in the predictive capability of theoretical and computational chemistry. In spite of this leap forward, the study of the reactivity of complex chemical systems, implying the treatment of very dense reaction networks, is still a challenge. In this article, we showcase three new methods that allow the generation, treatment and processing of catalytic cycles and complex reaction networks in an efficient automated manner. First, we present a tool for predicting the turnover frequency in catalytic cycles, both for homogeneous and heterogeneous catalysis. Then, we present what we have named the POMSimulator, a method for generating reaction maps and predicting speciation on the basis of pH and concentrations, as well as the formation mechanisms of complex molecular oxides. Lastly, we present OntoRXN, a new ontology for the formal definition of reaction mechanisms whose goal is to simplify the treatment and processing of complex chemical information of this type.

Keywords: Reaction networks, computational chemistry, catalysis, cheminformatics.

Introducció

La química computacional (QC), com molts camps en la ciència i l'enginyeria que depenen de la potència de la computació, s'ha vist marcada per l'increment espectacular d'aquesta potència al llarg de les darreres dècades. Aviat farà cent anys que el reconegut Paul Dirac va vaticinar el present i el futur d'aquest camp de la química: «Les lleis físiques necessàries per a la totalitat de la química són completament conegudes [...] i, per tant, cal que es desenvolupin mètodes pràctics que condueixin a una explicació de les

característiques principals dels sistemes atòmics complexos sense gaire computació» [1]. Al primer quart del segle XXI podem dir que hem arribat al punt en què, gràcies a la computació, els mètodes de la QC proporcionen solucions pràctiques en el disseny de les molècules i dels materials nous que la societat necessita per a trobar solucions als múltiples problemes que cal afrontar: clima, salut, energia, etc.

Un dels mètodes quàntics més utilitzats actualment és la teoria del funcional de la densitat (DFT) [2]. El baix cost computacional que té i l'elevada precisió dels resultats que permet obtenir l'han convertit en l'estàndard en molts casos. La precisió d'aquests mètodes i el seu poder predictiu han esperonat la recerca en noves metodologies per a la generació automàtica de mecanismes de reacció. Hi ha nous algoritmes que permeten explorar l'espai molecular i descobrir les espècies i les reaccions necessàries per a caracteritzar un meca-

Correspondència: Carles Bo
Institut Català d'Investigació Química (ICIQ)
Av. dels Països Catalans, 16. 43007 Tarragona
Tel.: +34 977 929 201
A/e: [cbo@iciq.cat](mailto: cbo@iciq.cat)

nisme de reacció. Entre les nombroses solucions a aquest problema [3], podem destacar mètodes com ADDF [4-6], AFIR [7-9], AutoMeKin [10-13], *ab initio* nanoreactor [14] o AARON [15], focalitzats en la localització d'intermediaris i estats de transició.

Construir grans col·leccions de dades de forma automàtica pot arribar a ser un obstacle quan cal interpretar-les, ja que els mètodes d'anàlisi emprats per a conjunts petits de dades deixen de ser aplicables. Per tant, hi ha la necessitat de desenvolupar mètodes de gestió i tractament de les dades que també siguin automàtics. Aquest paradigma està en línia amb un problema molt més general: el processament del que s'anomena *big data* [16, 17].

A l'hora de tractar el *big data*, una pràctica habitual consisteix a comprimir o simplificar la informació química. Una opció molt emprada es basa a representar les molècules com si fossin cadenes de text, els SMILES (*simplified molecular input line entry specification*), que encara són un referent en aquest camp [18]. Les estructures de les molècules també es poden representar com a grafs moleculars. Això es tradueix en la conversió dels àtoms en vèrtexs i dels enllaços químics en arestes del graf. Aquesta transformació és molt convenient perquè permet tractar la informació química d'una manera modular. Recentment, els grafs moleculars també s'han aplicat a l'àmbit de la intel·ligència artificial per a predir energies de dissociació [19] i de solvatació [20]. Aquest increment en el nombre d'aplicacions basades en grafs es deu en gran part a la publicació de llibreries d'accés obert com la NetworkX [21].

En el cas concret dels mecanismes de reacció, una manera d'estructurar la informació són les xarxes de reaccions químiques (*chemical reaction networks* o CRN). En contraposició amb altres representacions com els perfils d'energia de reacció, aquestes xarxes tenen com a punt focal la interconnexió entre espècies químiques mitjançant reaccions. En aquest marc, les CRN es defineixen com a *grafs*, on els vèrtexs són els intermediaris de la reacció i les arestes, les transformacions entre aquests. Aquesta formulació, doncs, ens permet aplicar els mètodes de la teoria de grafs a sistemes químics reactius i facilita el tractament de xarxes complexes. Addicionalment, les propietats generades amb les eines de la QC es poden incloure en el graf com a atributs associats als vèrtexs i les arestes corresponents.

Metodologia

Els grafs són objectes matemàtics formats per un conjunt de vèrtexs $V(G)$ i un conjunt d'arestes $E(G)$ que connecten parelles de vèrtexs. En relació amb això, un aspecte fonamental és la definició de grafs dirigits i no dirigits (figura 1a i 1b). En els primers, les arestes corresponen a parells ordenats i, per tant, es representen com a fletxes amb un sentit determinat. En canvi, en els grafs no dirigits, les arestes tan sols determinen que hi ha una connexió entre dos vèrtexs, sense especificar-ne la direcció.

Una propietat important dels grafs és l'isomorfisme. Així, la correspondència morfològica entre dos grafs pot ser completa, cas en què els grafs són equivalents, o parcial, fet que dona lloc al concepte de *subgraf* (figura 1c). Si ens focalitzem en grafs d'interès químic, la idea de subgrafs ens permet tractar tant fragments moleculars com seccions d'un mecanisme determinat i, de fet, és clau per a les eines que introduïrem al llarg d'aquest article. Finalment, un altre aspecte destacat és la noció de *travessar* un graf, és a dir, de trobar camins diferents (figura 1d) a través de la xarxa. Si treballem amb una CRN, aquests camins es relacionen amb les seqüències d'intermediaris i reaccions que es poden trobar en el sistema. D'aquesta manera, trobar i classificar camins diferents permet avaluar els processos més afavorits i, per tant, els més probables.

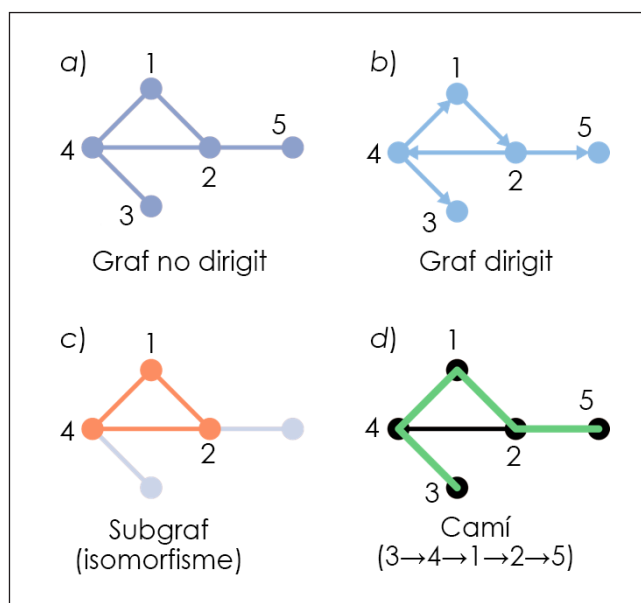


FIGURA 1. Representacions d'exemples corresponents a a) un graf no dirigit, b) un graf dirigit, c) un subgraf i d) un camí. Elaboració pròpia.

Resultats i discussió

En aquest article presentem tres mètodes computacionals que hem desenvolupat en el nostre grup de recerca en els darrers anys. En primer lloc, mostrem el gTOFfee, un codi que permet estimar d'una manera ràpida i acurada la freqüència de recanvi en la catàlisi homogènia i en l'heterogènia. En segon lloc, presentem el POMSimulator, un mètode que prediu l'especificació aquosa i els mecanismes d'autoassemblatge dels òxids moleculars. I, en tercer i darrer lloc, introduïm l'OntoRXN, una ontologia per a la definició de *grafs de coneixement* (*knowledge graphs* o KG) per tal d'organitzar i estandaritzar tota la informació relativa a una CRN.

gTOFfee

La caracterització computacional de mecanismes de reacció i cicles catalítics es fonamenta a proposar un conjunt d'intermediaris i d'estats de transició que justifiquin una transformació concreta, i, a partir d'aquí, determinar l'accessibilitat energètica del mecanisme.

Tanmateix, l'ús d'energies per a descriure la reactivitat dels sistemes troba limitacions a l'hora de comparar els resultats obtinguts amb estudis experimentals. En aquests estudis, per a mesurar l'activitat del sistema reactiu es tenen en compte unes altres propietats, com poden ser rendiments, selectivitats o constants de velocitat. En el cas de la catàlisi, la freqüència de recanvi (*turnover frequency* o TOF) s'empra habitualment per a determinar i comparar l'activitat dels catalitzadors.

Per tant, hi ha interès a desenvolupar estratègies que permetin obtenir d'una manera computacional propietats directament comparables amb les dades experimentals. En són un exemple les simulacions microcinètiques [22], per a predir l'evolució de les concentracions de totes les espècies implicades en un mecanisme al llarg del temps, així com les conversions o selectivitats. Pel que fa a la freqüència de recanvi en cicles catalítics, una altra aproximació, més senzilla matemàticament, és el model de l'abast energètic (*energy span model*, ESM) desenvolupat per S. Kozuch [23, 24]. Aquesta aproximació permet el càlcul de la TOF des del perfil d'energia lliure corresponent, tenint en compte a la vegada tots els altres intermediaris i estats de transició. A més del càlcul de la TOF, el model també pro-

porciona una formalització del concepte d'etapa determinant de la velocitat de la reacció, definint l'intermediari i l'estat de transició determinants del recanvi (TDI –*TOF determining intermediate* o intermediari determinant de la TOF– i TDTS –*TOF determining transition state* o estat de transició determinant de la TOF–), que és l'etapa que té més influència en l'activitat catalítica. La diferència d'energia entre el TDI i el TDTS és l'abast energètic que dona nom al model i, que per a CRN senzilles, es pot relacionar directament amb la TOF.

No obstant això, l'ús de perfils d'energia representa una limitació fonamental per a l'ESM, i és que, quan tractem amb sistemes complexos (cicles catalítics acoblats o espècies fora dels cicles) que no es poden definir com a perfils lineals, el model no és aplicable. En aquests casos, cal fer servir CRN en lloc dels perfils d'energia. El 2015, Kozuch va proposar una extensió teòrica [25] del model basada en el tractament de grafs, però no es va implementar fins al cap d'uns quants anys, gràcies al desenvolupament del codi gTOFfee [26, 27].

La variant de l'ESM basada en grafs defineix el concepte *meccanisme* com un *subgraf* de la xarxa de reacció, en què només hi hauria un sol cicle, que seria el cicle catalític. Així, d'una CRN complexa sorgirien múltiples mecanismes, cadascun dels quals amb diverses branques de reacció i, possiblement, diferents cicles productius que donarien lloc a productes diferents. Dependent de les branques concretes que apareguin en un mecanisme, aquest serà més o menys favorable des dels punts de vista cinètic i termodinàmic. A través del codi gTOFfee, es pot calcular la TOF per a cadascun d'aquests mecanismes possibles i, després, combinar-los en funció del cicle productiu corresponent. D'aquesta forma, per a una sola xarxa podem obtenir tantes TOF com tipus de reaccions hi ha codificats en aquesta xarxa, i caracteritzar-ne així la selectivitat corresponent. A més, a part de la TOF, vam introduir un altre descriptor, anomenat *abast energètic efectiu* (*effective energy span*), que transforma la freqüència de recanvi en unitats d'energia per tal de tenir una estimació de l'energia d'activació aparent del sistema, tant per a la xarxa en conjunt com per als tipus de reaccions o per als mecanismes individuals.

L'ESM, tant en la forma original com en la variant basada en grafs, també pot considerar l'efecte de les concentracions dels reactius i dels productes en l'activitat catalítica, que poden arribar a ser molt rellevants a l'hora de comparar resultats computacionals amb dades experimentals. Si bé altres mètodes

des com els models microcinètics permeten una descripció més completa dels efectes de concentració, l'ESM encara representa un avenç important, ja que aquests efectes habitualment s'ometen quan es fan anàlisis més tradicionals com la simple interpretació de perfils d'energia lliure.

En la catàlisi homogènia, una reacció especialment interessant és la hidroformilació d'alquens per a donar aldehyds, catalitzada per complexos de metalls de transició com ara el rodi o el cobalt. Nombrosos estudis, tant experimentals com teòrics, han considerat el mecanisme de la reacció. En aquest cas, vam utilitzar la proposta mecanística de Rush, Pringle i Harvey [28] per al procés catalitzat amb cobalt per tal de construir la xarxa de reacció, que vam tractar amb gTOFfee (figura 2). Aquesta CRN inclou tres possibles reaccions globals: la hidroformilació del propè a butiraldehid (2 a 9x), la hidrogenació del propè a propà (2 a 5x) i la possible reducció del butiraldehid ja format amb alliberament de propà i monòxid de carboni (9x a 5x).

Mitjançant el codi gTOFfee, vam calcular l'energia d'activació efectiva de cadascuna d'aquestes reaccions (taula 1) i vam observar que la producció de butiraldehid (hidroformilació) és clarament el procés més favorable des del punt de vista cinètic, tal com es manifesta en els experiments. La inspecció directa del perfil d'energia, en canvi, no mostra d'una manera tan evident aquesta preferència per l'aldehid.

Per aprofundir en l'anàlisi de la selectivitat per a generar l'aldehid o l'alcà, vam tenir en compte les concentracions inicials de CO i de H₂. Així, vam utilitzar el quocient entre les TOF per als dos productes com a descriptor per a la selectivitat, a fi de construir un mapa bidimensional variant els valors de la concentració entre 0,5 i 5,0 m en els dos reactius (figura 3). El mapa de selectivitat mostra la necessitat de tenir concentracions relativament grans dels dos reactius per tal d'afavorir la producció d'aldehid (regió groga), en línia amb les condicions emprades experimentalment i industrialment. Així mateix, quan hi ha un excés d'hidrogen, consegüentment s'afavoreix la producció d'alcà, tal com s'observa en el mapa en desplaçar-s'hi horitzontalment. A més, la selectivitat màxima per a l'aldehid (~10:1) concorda quantitativament amb els resultats de simulacions microcinètiques de la mateixa xarxa.

En resum, la combinació de la simplicitat matemàtica de l'ESM amb la generalitat del tractament basat en grafs per a

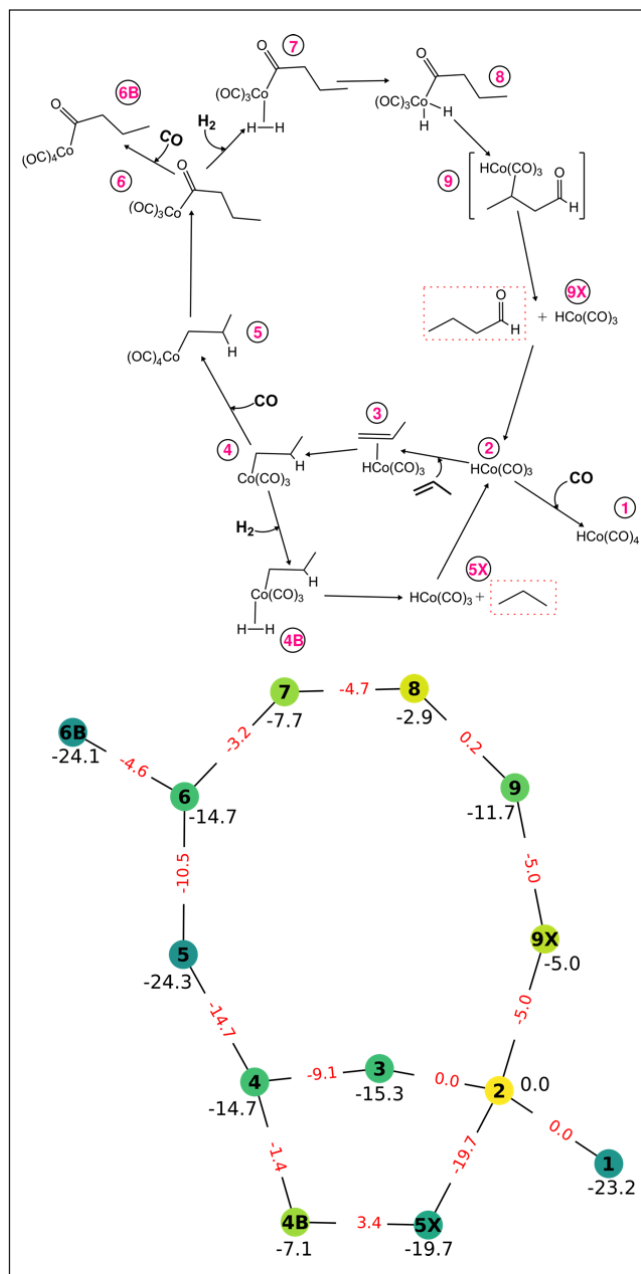


FIGURA 2. Cicle catalític i xarxa de reacció per a la hidroformilació de propè, catalitzada mitjançant cobalt, proposada per Rush, Pringle i Harvey. Elaboració pròpia.

TAULA 1. Energies d'activació efectives, en kcal · mol⁻¹, sense considerar els efectes de la concentració. Elaboració pròpia.

Reacció	∂E_{ef} (kcal · mol ⁻¹)
hidroformilació	25,2
hidrogenació	27,9
reducció aldehid	28,4

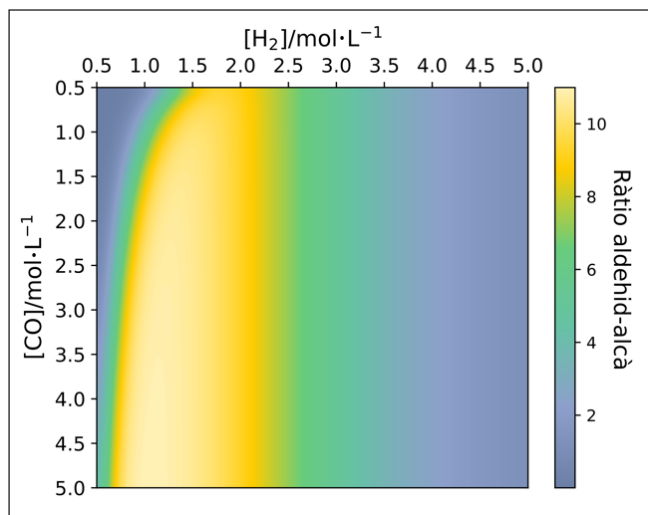


FIGURA 3. Selectivitat de la reacció en funció de les concentracions inicials de CO i H₂, en mol·L⁻¹. Elaboració pròpia.

xarxes complexes, tal com s'ha implementat a gTOFfee, proporciona una eina flexible i polivalent per al tractament de processos catalítics, que permet anar més enllà de l'exploració directa de perfils de reacció.

POMSimulator

Les reaccions d'autoassemblatge en la síntesi dels polioxometal·lats (POM) encara són un tema no resolt, tant des del punt de vista experimental com computacional. La complexitat d'aquesta qüestió prové del gran nombre de reaccions químiques simultànies que tenen lloc en dissolució. Tots els oxoclústers estan interconnectats amb altres espècies, ja sigui per reaccions àcid-base o per reaccions de nucleació. A més, el procés de nucleació depèn de diversos paràmetres, com el pH, la força iònica, la concentració total, la temperatura i la pressió. De fet, alguns òxids moleculars només es formen en un medi àcid, mentre que n'hi ha d'altres que són insolubles a un pH alt. Per tant, petits canvis en els paràmetres anteriors donen lloc a la formació de productes diferents. Malgrat aquesta dificultat, en les últimes dècades s'han fet molts avenços en aquest camp [29]; per exemple, a partir de la síntesi d'òxids moleculars gegants [30], de la descoberta de la reactivitat del niobi i el tàntal [31, 32] o de l'aplicació dels POM a la catàlisi i a les bateries [33]. D'una manera anàloga, la QC, arrelada en els principis de la mecànica quàntica, també ha millorat el coneixement dels POM [34]; per exemple, ha permès determinar mecanismes de reacció [35] i ha facilitat la interpretació de

les propietats electròniques [36], així com el refinament del desordre inherent en la tècnica de raigs X [37]. Tanmateix, hi havia una mancança: no existia cap mètode computacional que abordés el mecanisme i l'especiació simultàniament, per la qual cosa vam considerar que calia desenvolupar una nova metodologia que anés més enllà de l'enfocament clàssic i que vam anomenar POMSimulator, ja que és un mètode dissenyat per a simular les reaccions de nucleació dels polioxometal·lats [38-40].

La figura 4 mostra un esquema del funcionament d'aquesta metodologia. En primer lloc s'extreuen els resultats principals dels càlculs d'optimització de geometries i freqüències (l'energia de Gibbs, la càrrega i els punts crítics segons la teoria de Bader [41]). Amb aquests últims resultats podem deduir l'estructura molecular, així com tots els enllaços químics de cada molècula. El pas següent consisteix a convertir aquesta informació en grafs moleculars. Per tal de trobar les relacions morfològiques entre els grafs moleculars utilitzem la propietat anomenada *isomorfisme*, la qual és molt convenient ja que els isomorfismes estan relacionats amb les reaccions químiques. Per aquest motiu, el pas següent és assignar un tipus de reacció. Per exemple, si dos grafs són isomorfs i el balanç correspon a un àtom d'hidrogen, la relació correspon a una reacció àcid-base. Aquesta lògica és aplicada a les reaccions d'addició, de condensació, d'hidròlisi, d'isomerització i de dimerització. El resultat final és un mapa de reacció d'espècies en què els òxids moleculars són els vèrtexs i les reaccions són les arestes.

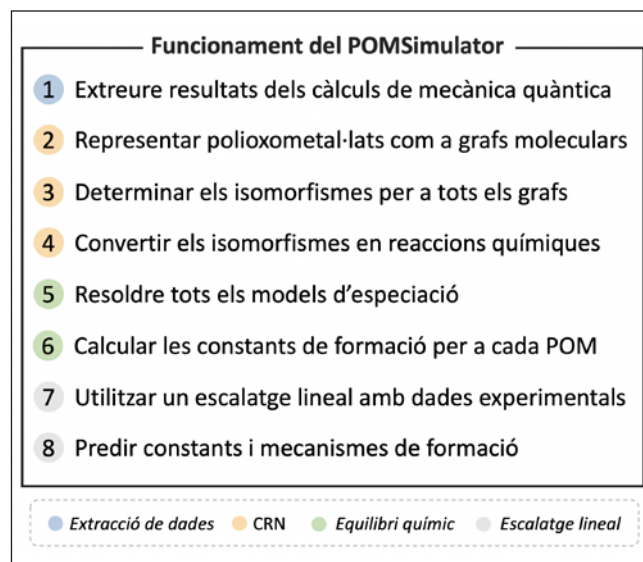


FIGURA 4. Esquema general del funcionament del POMSimulator. Elaboració pròpia.

Seguidament el mapa de reacció és utilitzat per a generar els models d'especiació. Aquests models consisteixen en sistemes d'equacions no lineals en què cada equació correspon a un equilibri químic més l'equació de balanç de massa. Les concentracions dels òxids són les variables dependents i el pH és la variable independent. A més, les concentracions són expressades com a activitats per tenir en compte l'efecte de la força iònica.

A l'hora de preparar els models d'especiació apareix un problema relacionat amb el nombre d'equacions i de variables dependents: hi ha més reaccions químiques que compostos químics. Així doncs, el sistema està sobredeterminat. A fi de resoldre aquest problema, utilitzem el coeficient binomial per a generar tants models com combinacions de reaccions químiques hi ha. Aquesta solució també té un inconvenient, ja que el nombre de combinacions creix de forma factorial. Cada model d'especiació es resol en un rang ampli de pH.

L'última part de la metodologia consisteix a realitzar un escalat lineal de les constants obtingudes (figura 5a). Aquest pas és necessari perquè les constants teòriques estan sobreestimades respecte de les experimentals, la qual cosa no és fortuïta i s'ha observat en múltiples estudis previs sobre la predicció de constants àcid-base en compostos orgànics [42, 43].

Cal remarcar que les constants teòriques predites pel POMSimulator estan correlacionades de forma molt acurada amb els valors experimentals. Una vegada s'han comparat les constants de cada model amb les experimentals, s'escull el model que té una desviació quadràtica mitjana (*root mean square error*, RMSE) més baixa. Utilitzem l'equació de la recta corresponent i escalem totes les constants del model, tant les reportades com les no reportades. D'aquesta manera podem predir constants que no estan disponibles en la literatura. Així mateix, el millor model d'especiació també proporciona informació sobre les reaccions que hi intervenen. L'últim pas de la metodologia no només proporciona noves constants de formació, sinó que també indica quin mecanisme termodinàmic és més probable.

La figura 5 mostra els resultats obtinguts per als òxids moleculars de vanadi. Per mitjà d'aquest sistema vam optimitzar la geometria de quaranta espècies d'òxids moleculars. Posteriorment, vam introduir els resultats al POMSimulator i vam obtenir les constants de formació teòriques. La figura 5a mostra

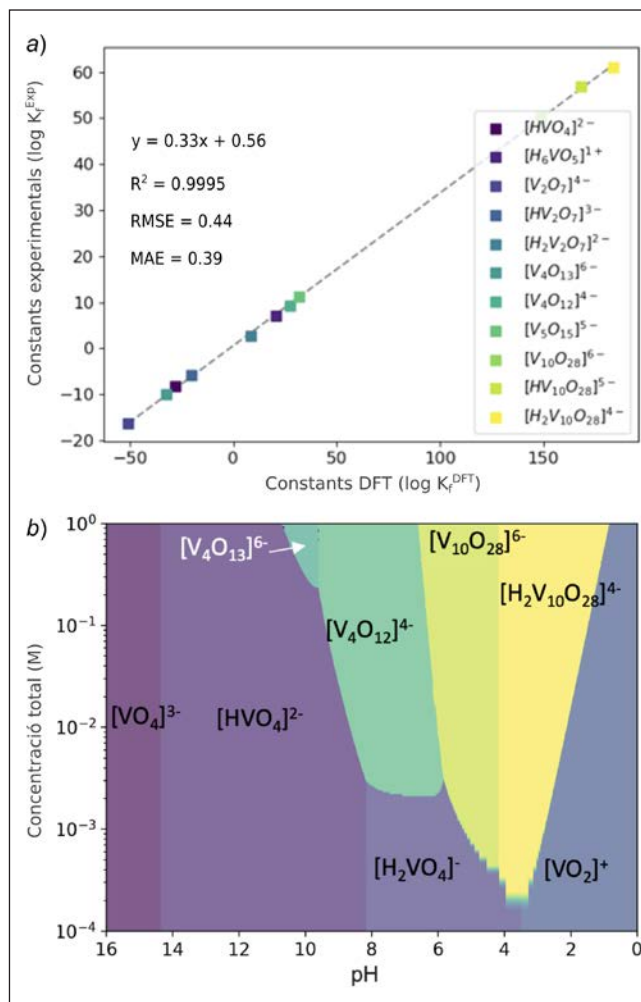


FIGURA 5. a) Escalat lineal entre les constants de formació experimentals i les obtingudes amb el POMSimulator. b) Diagrama de fases d'especiació per als isopolioxovanadats. Elaboració pròpia.

la regressió lineal més acurada respecte a les dades experimentals [44]. L'escalat és robust, ja que conté un ampli ventall de compostos amb un nombre d'àtoms diferents: des del monòmer, $[HVO_4]^{2-}$, al decavanadat, $[V_{10}O_{28}]^{6-}$. Així mateix, l'elevat valor de correlació lineal (0,9995) i la baixa desviació (0,44) ratifiquen l'excel·lent correlació.

La figura 5b mostra un diagrama de fases d'especiació. Aquesta representació permet analitzar quins compostos són més predominants a partir d'un valor de pH i de la concentració inicial total de monòmer. Com que els vanadats tenen una química aquosa molt rica, el diagrama conté un nombre elevat de fases. Tot i la complexitat d'aquest sistema, les fases calculades teòricament coincideixen amb els resultats experimentals [45]. En medi alcalí només predominen els monòmers en dos

estats de protonació: $[\text{VO}_4]^{3-}$ (pH 14–16) i el $[\text{HVO}_4]^{2-}$ (pH 8–14). La nucleació dels òxids de vanadi té lloc a pH més baixos que 8. El primer compost que apareix és el $[\text{V}_4\text{O}_{13}]^{6-}$, i només esdevé el més abundant a una concentració elevada ($\approx 1,0 \text{ mol}\cdot\text{L}^{-1}$). A un pH lleugerament més àcid (4–8) es troba la fase dels *metavanadats*, compostos cíclics amb fórmula general $[\text{V}_n\text{O}_{3n}]^{n-}$. Tant el $[\text{V}_5\text{O}_{15}]^{5-}$ com el $[\text{V}_4\text{O}_{12}]^{4-}$ són metavanadats importants. De fet, el $[\text{V}_4\text{O}_{12}]^{4-}$ és un compost rellevant biològicament ja que té propietats antibacterianes [46]. La fase següent correspon als decavanadats i es troba a la part més àcida del diagrama (pH 1–4). Aquests òxids moleculars també tenen propietats antibacterianes [47]. Per aquest motiu, és especialment important controlar l'especiació aquosa; en cas contrari, canvis lleugers en el pH fisiològic poden degradar el compost actiu a favor de la formació d'un altre òxid molecular.

La segona família de polioxometal·lats que tractem en aquest treball són els òxids moleculars de niobi. En les dues darreres dècades, el nombre de treballs sobre niobats ha augmentat notablement [48, 49], però encara és una química poc estudiada. En aquest context, el POMSimulator ha permès analitzar i predir constants de formació que no s'havien reportat prèviament. La figura 6a mostra un mapa de reacció que conté els quaranta-dos òxids de niobi considerats en aquest estudi. Com es pot observar, la major part de les reaccions són exergòniques. D'una manera especial, les reaccions de protonació (línies horitzontals) són molt favorables. Aquesta evidència es deu al fet que l'addició de protons rebaixa el caràcter aniònic de les molècules i les fa més estables.

La figura 6b mostra el diagrama d'especiació predit pel POMSimulator. Aquesta gràfica indica l'abundància dels niobats en funció del valor del pH i a una concentració inicial fixa. A diferència del diagrama de fases d'especiació, la figura 6b també recull els òxids moleculars que es troben en concentracions més baixes. El rang de pH estudiat és entre 4 i 14, ja que a valors inferiors o superiors l'especiació es manté constant. A un pH neutre el compost majoritari és el $[\text{Nb}_{10}\text{O}_{28}]^{6-}$. A mesura que el medi es basifica, la concentració del decaniobat decau i es forma el $[\text{H}_9\text{Nb}_{24}\text{O}_{72}]^{15-}$. Aquesta darrera molècula està en equilibri amb el $[\text{Nb}_{10}\text{O}_{28}]^{6-}$, i la seva formació es veu afavorida a concentracions de precursor més altes. A partir d'un pH alcalí de 10, l'especiació es veu dominada pels derivats dels hexanuclears. Concretament, a un pH 10 es troben les espècies tri- i diprotonades: $[\text{H}_3\text{Nb}_6\text{O}_{19}]^{5-}$ i $[\text{H}_2\text{Nb}_6\text{O}_{19}]^{6-}$. A un pH més alcalí encara trobaríem el monoprotonat,

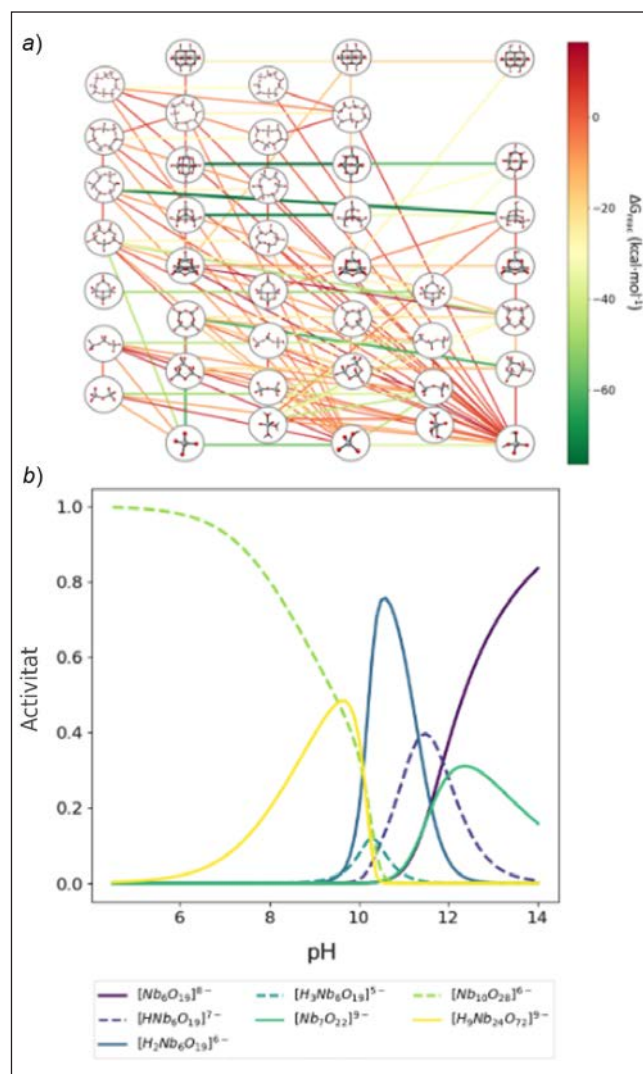


FIGURA 6. a) Mapa de reacció amb les energies lliures de Gibbs. b) Diagrama d'especiació a 298,15 K i 1 M de NaCl. Elaboració pròpia.

$[\text{HNb}_6\text{O}_{19}]^{7-}$, i no protonat, $[\text{Nb}_6\text{O}_{19}]^{8-}$. Cal remarcar que també en aquest rang de pH apareix un compost rellevant en la química dels niobats: l'heptaniobat, $[\text{Nb}_7\text{O}_{22}]^{13-}$. Aquesta molècula és estable a un pH 12, però té un paper important com a intermediari en l'equilibri entre el $[\text{Nb}_{10}\text{O}_{28}]^{6-}$ i el $[\text{H}_9\text{Nb}_{24}\text{O}_{72}]^{15-}$.

El POMSimulator representa una nova metodologia per a estudiar l'especiació en dissolució i els mecanismes de formació dels òxids moleculars *in silico*. Hem demostrat que el mètode és robust i pot ser aplicat de forma acurada a cinc sistemes de polioxometal·lats. Els bons resultats tant dels escalats lineals com dels diagrames d'especiació mostren la implementació correcta del mètode. Esperem que en el futur el POMSimulator pugui crear més sinergies amb la química experimental.

OntoRXN

Una qüestió clau en la gran majoria de les branques de la ciència és la gestió de les dades que es generen en la recerca, sobretot quan es tracta de grans volums de dades. Encara és més important que les dades estiguin vinculades al coneixement específic del camp i al context que dona a les dades un significat. Les dades, totes soles, no diuen res, i cal embolcallar-les amb informació addicional: qui les ha recollit i quan i com s'ha fet, ja sigui mitjançant l'observació, l'experimentació o la computació, entre d'altres; és el que en diem *metadades* i que, de fet, són tan importants com les dades. Per tant, cal desenvolupar estratègies per a organitzar les dades d'una manera adequada, per a simplificar-ne el processament i per a facilitar-ne la transformació en informació química.

Una d'aquestes aproximacions al tractament estructurat de la informació, és a dir, de les dades, és el concepte *dades semàntiques* (*semantic data*), introduït per Tim Berners-Lee [50], un dels fundadors principals de la World Wide Web, l'any 2001. Aquesta idea es fonamenta en l'addició d'estructura i significat a les dades «en cru» i que ja són presents a la web, amb l'objectiu de permetre una millor interacció de l'usuari amb aquestes dades. Aquesta nova estructura de les dades i el significat semàntic que aporta habiliten la possibilitat d'interrogar bases de dades enunciant qüestions complexes. Si bé aquesta implementació general de la *web semàntica* certament no va tenir gaire èxit, sí que hi ha iniciatives actives com la Wikidata [51], una font de dades creada amb base a projectes col·laboratius. L'organització semàntica es fa servir en dominis específics, en la indústria (J-Park Simulator [52], European Materials Modelling Consortium) i en les biociències [53-55].

Dos elements clau d'aquest sistema són el model d'organització de dades RDF (*resource description framework* o marc per a la descripció de recursos) [56] i les ontologies [57]. El model RDF permet la creació de bases de dades molt flexibles, codificant la informació en forma de triplets, que estan formats per un subjecte, un predicat i un objecte, anàlogament a les frases simples del llenguatge natural. Els predicats són el que donen significat a la frase. La concatenació d'aquests triplets produeix *grafs*, en què subjectes i objectes (les dades) formen un conjunt de vèrtexs connectats per arestes, que són els predicats. Així, és possible expandir el graf (i, per tant, la base de

dades) simplement afegint-hi més triplets per a generar nous nodes i connexions. Pot semblar que aquesta flexibilitat entra en col·lisió amb la idea central d'aportar *significat* a les dades, atès que els subjectes, els predicats i els objectes dels triplets poden ser immensament variats. Aquest problema es pot resoldre precisament amb la introducció d'ontologies: vocabularis que representen les diverses entitats d'una àrea del coneixement. Una ontologia defineix classes, que categoritzen els tipus d'entitats pròpies de l'àrea, i propietats, que són les relacions entre les classes. Així, les ontologies permeten estandarditzar l'estructura de la informació, les bases de dades d'un camp de coneixement, seguint els principis del *linked data* [58], per tal de compartir-la i reutilitzar-la. Partint, doncs, d'un conjunt de dades, l'organització segons una ontologia que segueixi el model RDF produeix un graf de coneixement on totes les dades estaran organitzades i ben etiquetades (figura 7). L'ús d'ontologies en química encara no està gaire estès. Hi ha esforços destacables per a processar dades de molècules d'interès biològic (ChEBI [59, 60]) o per a la formalització de protocols experimentals (Chemical Methods Ontology [61]), entre d'altres [62], però encara som lluny del grau d'estandardització que hi ha, per exemple, en les biociències.

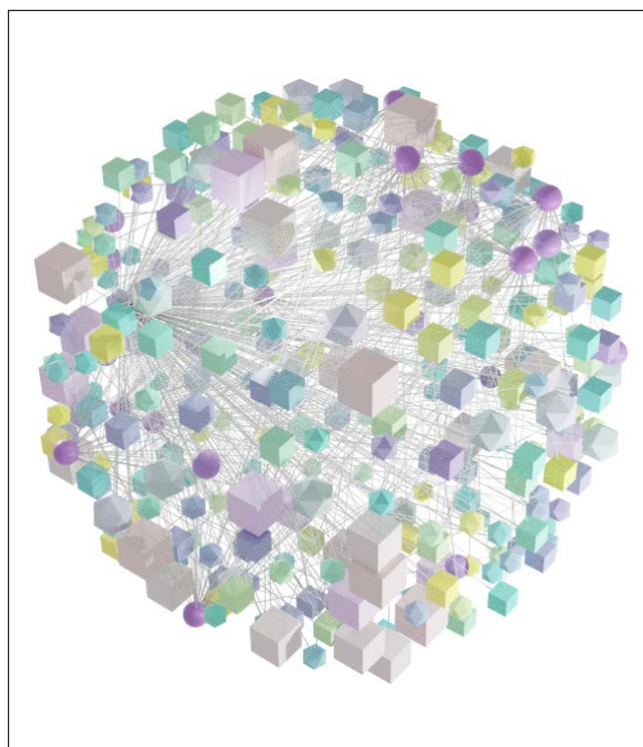


FIGURA 7. Representació tridimensional d'un graf de coneixement, utilitzant formes i colors diferents per a diversos tipus de dades i de propietats. Elaboració pròpia.

En el nostre cas, hem desenvolupat una nova ontologia que, encara que estigui dedicada a una part molt concreta de la química, pensem que té rellevància. L'objectiu va ser la definició d'una nova ontologia per a tractar semànticament totes les dades associades a mecanismes de reacció caracteritzats computacionalment, que hem anomenat OntoRXN [63, 64].

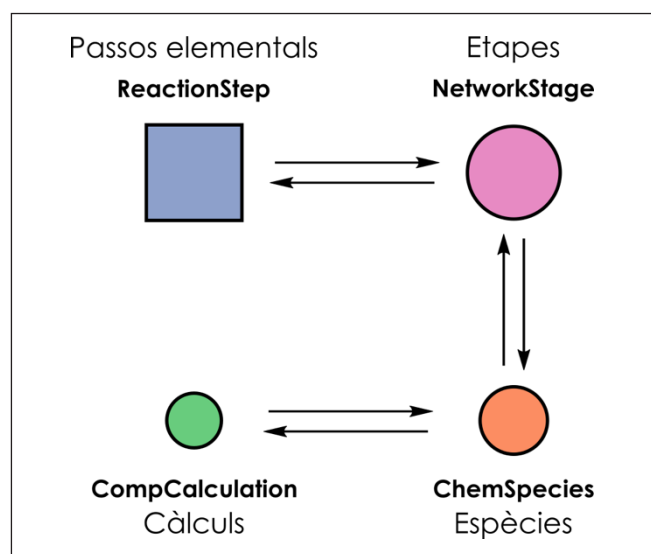


FIGURA 8. Classes principals introduïdes en OntoRXN. Elaboració pròpia.

L'entitat resultant, és a dir la nova base de dades, conté tota la informació d'una CRN, unificant l'estructura de la mateixa xarxa de reaccions amb les propietats calculades per a cada espècie, en el que es coneix com un *graf de coneixement*. Per exemple, en el graf es definiren quines espècies químiques hi intervenen i a través de quines reaccions es connecten amb totes les dades enllaçades. D'aquesta forma es facilita en gran manera l'accés a totes les dades associades a una xarxa de reaccions, de mecanismes de reacció o de cicles catalítics, que va més enllà del simple dibuix amb què representem avui el perfil d'energia d'un mecanisme de reacció. Per altra banda, la nova estructura de les dades permet desenvolupar protocols automatitzats per a processar i interpretar CRN, que són molt generalistes i àmpliament aplicables.

L'estructura de la nostra ontologia conté quatre classes principals per a modelitzar xarxes de reacció, tal com es mostra a la figura 8. En primer lloc, tenim els passos elementals de reacció. Com que considerem les CRN com a grafos no dirigits, no hi ha cap referència a la direccionalitat del pas, sinó que tan sols el definim com una connexió entre dues etapes, que poden tenir o no un estat de transició associat. Tant els inter-

mediaris com els estats de transició es defineixen com a *etapes*, en què hi ha una molècula o més que s'han considerat en conjunt per tal d'ajustar el nombre d'àtoms i així poder estimar les energies relatives de cada pas. Després, les altres dues classes serien ja les espècies químiques individuals, i els càlculs concrets que s'han emprat per a caracteritzar-les, juntament amb les propietats obtingudes. D'aquesta manera és possible modelitzar situacions en què per a una sola entitat química hi pot haver múltiples càlculs associats: anàlisi conformacional, combinació de nivells de teoria, etc.

A l'hora de definir un graf de coneixement a partir d'una CRN, a banda de l'ontologia mateixa, hem d'establir un protocol per a capturar l'estructura de la xarxa i les propietats dels càlculs. Amb vista a simplificar aquest pas, l'eina que hem desenvolupat s'ha connectat amb la plataforma ioChem-BD [65, 66], ja que permet disposar d'un format unificat, així com de la definició de perfils d'energia i xarxes de reacció dins de la mateixa plataforma, que engloba totes les peces necessàries per a construir el graf de coneixement final segons les «instruccions» expressades en l'ontologia.

L'objectiu final del marc semàntic per al tractament de dades és permetre la formulació de qüestions complexes, inferint respostes a partir de les dades disponibles i de la seva estructura. El procés de consulta d'un graf de coneixement implica la filtració de les dades, aportant tan sols la part del graf (o subgraf) que conté la informació sol·licitada. Idealment, seria possible combinar grafos de coneixement basats en diverses ontologies per a aspectes diferents, que permetessin fer preguntes del tipus «quin seria el catalitzador més eficaç per a una reacció X, sintetitzat en menys de Y passos, amb el cost més reduït possible?». Si bé aquesta situació implicaria un grau de desenvolupament d'ontologies molt per sobre del que tenim avui dia, el concepte de consultes semàntiques amb grafos de coneixement més específics que els proposats per OntoRXN pot ser interessant en si mateix.

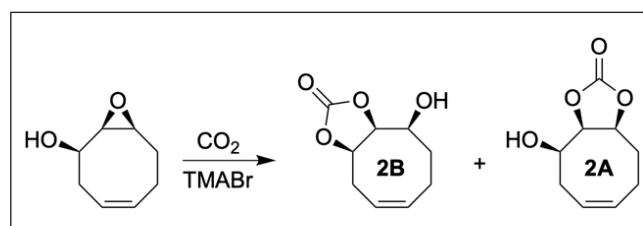


FIGURA 9. Reacció de fixació de CO₂ en epoxialcohol, amb dos productes (carbonats cíclics) diastereoisomèrics. Elaboració pròpia.

Una vegada establert el nostre marc de treball, presentem un exemple concret d'utilització de grafs de coneixement per al processament d'un mecanisme de reacció; concretament, un procés de fixació de diòxid de carboni amb un epòxid bicíclic [67] (figura 9). Aquesta reacció s'observa experimentalment amb molt bona selectivitat per al carbonat cíclic 2B, el producte termodinàmic. El mecanisme detallat, però, és relativament complex, a conseqüència de la disposició espacial de l'anell de ciclooctè, que permet la transferència del protó de l'alcohol a l'epòxid i facilita que pugui tenir lloc un procés concertat d'obertura de l'epòxid i inserció del CO_2 . Per tal de predir la conversió i la selectivitat del procés i confirmar la concordança del mecanisme proposat amb els experiments, vam portar a terme una simulació microcinètica del sistema. Així doncs, vam generar el graf de coneixement per al mecanisme i vam fer servir consultes semàntiques (figura 10) per tal de construir el model, que després vam simular mitjançant el programa COPASI [68].

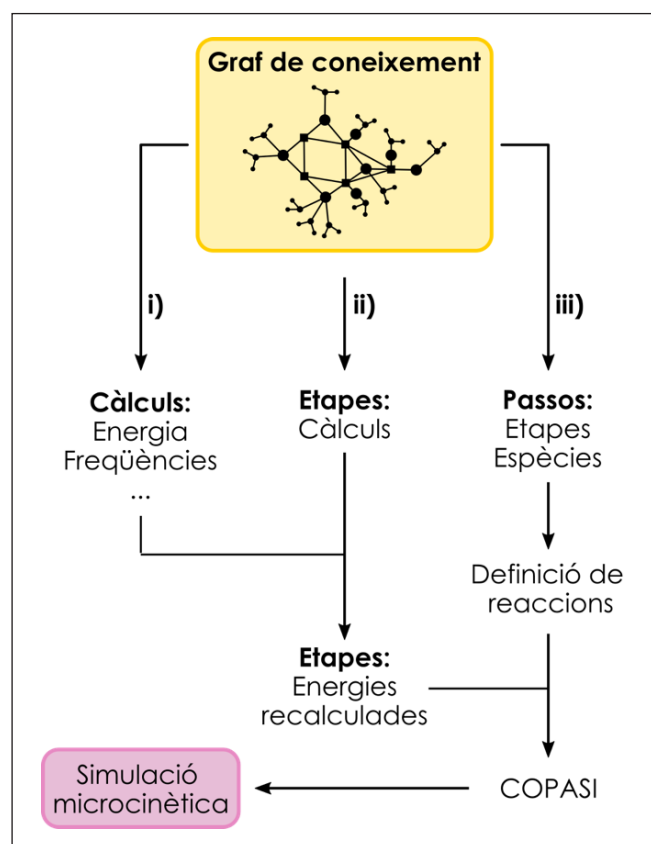


FIGURA 10. Esquema del protocol de consulta emprat per a la construcció automatitzada d'un model microcinètic. Elaboració pròpia.

El primer pas del protocol és l'extracció de les propietats necessàries per a modificar l'estat de referència de l'estudi ter-

moquímic del sistema dels 298,15 K i 1,0 atm utilitzats normalment per defecte a la temperatura de treball (353,15 K) i la concentració estàndard per a solucions (1,0 M). Després, la segona consulta associa totes les etapes del mecanisme (intermediaris i estats de transició) als càlculs corresponents, per tal d'obtenir energies relatives per a cadascuna. Finalment, la darrera consulta processa els passos elementals de la xarxa per extreure'n totes les reaccions úniques que s'hi han codificat, a partir de la determinació de les espècies que apareixen a cada banda i la generació de les estequiometries de tots els processos. A partir d'aquestes definicions i energies, podem calcular les constants de velocitat directa i inversa de cada reacció i completar així el model microcinètic.

L'avantatge principal d'aquesta aproximació és que el protocol proposat a la figura 10 és totalment general, i, per tant, aplicable a qualsevol xarxa de reacció definida com a graf de coneixement, amb la qual cosa se simplifica d'una manera notable el processament de dades complexes. Tornant a l'exemple concret de la fixació de CO_2 , els resultats de la simulació es recullen a la figura 11.

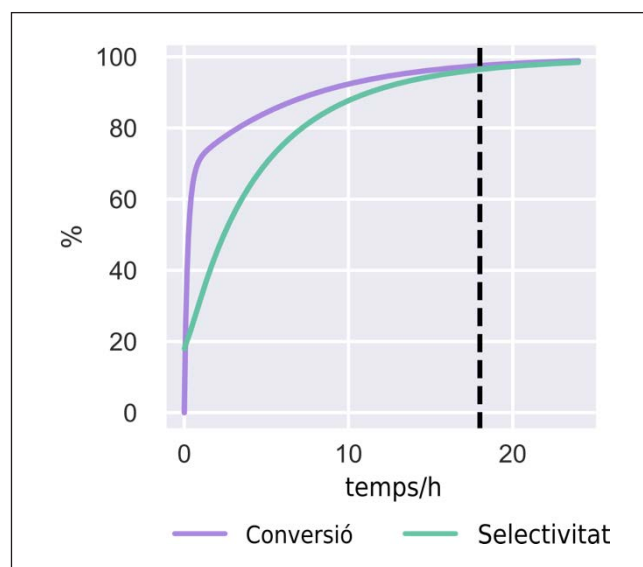


FIGURA 11. Resultats de la simulació microcinètica per a la formació de carbonat cíclic, amb $[\text{CO}_2] = 4,98 \text{ M}$, $[\text{TMABr}] = 0,625 \text{ M}$ i $[\text{Epoxide}] = 2,5 \text{ M}$. La corba de selectivitat es refereix a la formació de 2B, el producte majoritari experimentalment. Elaboració pròpia.

Les prediccions de conversió i selectivitat al temps de reacció (18 hores) són, respectivament, del 98% i del 96%, que concorden excel·lentment amb els valors experimentals (99% i 92%) i donen suport així a la nostra proposta mecanística. Hem de destacar un cop més, però, que l'interès principal del

mètode que proposem no és l'aplicació a aquest sistema concret, ni tampoc específicament a la generació de models microcinètics, sinó el fet de disposar d'un format estandarditzat per a xarxes de reacció. Això permetrà desenvolupar tota mena de *fluxos de treball* (*workflows*), tant senzills com complexos, i així millorar la reproductibilitat dels resultats computacionals, la reutilització dels estudis ja fets i la col·laboració entre grups. Per acabar, volem destacar que tant la mateixa ontologia com les eines per a aplicar-la a xarxes definides en ioChem-BD són d'accés i codi obert [64, 69]. Addicionalment, hem habilitat un servei web [70] basat en el motor Blazegraph [71] per donar accés a una base de dades RDF i poder realitzar consultes semàntiques amb tres exemples de mecanismes expressats com a grafs de coneixement, incloent-hi la fixació de diòxid de carboni que hem discutit anteriorment. Aquest servei dona accés tant a les dades com a una selecció de consultes sobre aquestes, amb la finalitat de demostrar la utilitat de la nostra aproximació a la informació química i fer-la més accessible a la comunitat.

Conclusions

Un tractament computacional adequat de sistemes químics complexos requereix no tan sols incrementar la precisió dels mètodes teòrics utilitzats o la quantitat d'entitats químiques considerades, sinó també desenvolupar noves estratègies per a la utilització d'aquestes dades. En aquest context, hem presentat tres mètodes desenvolupats en el nostre grup, fonamentats en la descripció de mecanismes com els grafs, que permeten afrontar una varietat de problemes d'interès químic.

En primer lloc, s'ha introduït una implementació del model de l'abast energètic (gTOFfee) per a simplificar l'anàlisi de cicles catalítics complexos. Així, es generen de manera automàtica tots els tipus de reaccions als quals una xarxa determinada dona lloc, i es calculen les freqüències de recanvi i les energies d'activació efectives per a tots ells. Els resultats obtinguts concorden tant amb estudis experimentals com amb altres mètodes teoricocomputacionals, com s'ha demostrat a partir del procés d'hydroformilació de propè.

En segon lloc, hem desenvolupat un mètode per a predir l'especiació aquosa i mecanismes de formació d'òxids metàl·lics moleculars. L'hem anomenat POMSimulator, ja que permet simular la reactivitat química d'aquesta família de compostos

inorgànics. La seva aplicació obre la possibilitat a crear mapes de reacció que considerin tots els processos d'agregació i àcid-base presents en dissolució. A més, cal remarcar que aquest mètode no només reproduïx d'una manera acurada l'especiació reportada experimentalment, sinó que també permet predir constants d'equilibri no reportades.

Darrerament, s'ha buscat l'estandardització d'aquestes xarxes de reacció basades en la química computacional definint una ontologia (OntoRXN) per a conceptualitzar tant l'estructura de la xarxa com les propietats associades als càlculs. L'ontologia prové d'un esquema d'organització molt polivalent per a la construcció de grafs de coneixement, que després es poden sotmetre a *consultes semàntiques* per tal d'extreure la informació d'interès en cada situació. D'aquesta manera, com un tipus de prova de concepte, s'ha proposat un protocol per a la generació automàtica de models microcinètics. Finalment, i a més de l'ontologia, s'han introduït també les eines necessàries per a generar els grafs a partir de la plataforma ioChem-BD.

Agraïments

Aquest treball ha estat finançat pel Ministeri de Ciència i Innovació, MCIN/AEI/10.13039/501100011033, mitjançant els projectes PID2020-112806RB-I00 i TED2021-132850B-I00; per l'Accreditació d'Excel·lència Severo Ochoa 2020-2023 (CEX2019-000925-S, MCI/AEI); per la Fundació ICIQ, i pel programa CERCA de la Generalitat de Catalunya.

Referències i altres fonts

- [1] DIRAC, P. A. M. «Quantum mechanics of many-electron systems». *Proc. R. Soc. Lond. A*, 123 (1929), p. 714-733.
- [2] KOHN, W.; SHAM, L. J. «Self-consistent equations including exchange and correlation effects». *Phys. Rev.*, 140 (4A) (1965), p. A1133-A1138.
- [3] SIMM, G. N.; VAUCHER, A. C.; REIHER, M. «Exploration of reaction pathways and chemical transformation networks». *J. Phys. Chem. A* [en línia], 123 (2) (2019), p. 385-399. <<https://doi.org/10.1021/acs.jpca.8b10007>>.
- [4] OHNO, K.; MAEDA, S. A. «Scaled hypersphere search method for the topography of reaction pathways on the potential energy surface». *Chem. Phys. Lett.* [en línia], 384 (4-6) (2004), p. 277-282. <<https://doi.org/10.1016/j.cplett.2003.12.030>>.

- [5] OHNO, K.; MAEDA, S. «Global reaction route mapping on potential energy surfaces of formaldehyde, formic acid, and their metal-substituted analogues». *J. Phys. Chem. A* [en línia], 110 (28) (2006), p. 8933–8941. <<https://doi.org/10.1021/jp061149l>>.
- [6] MAEDA, S.; OHNO, K.; MOROKUMA, K. «Systematic exploration of the mechanism of chemical reactions: The global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods». *Phys. Chem. Chem. Phys.* [en línia], 15 (11) (2013), p. 3683–3701. <<https://doi.org/10.1039/C3CP44063J>>.
- [7] MAEDA, S.; MOROKUMA, K. «Finding reaction pathways of type $A + B \rightarrow X$: Toward systematic prediction of reaction mechanisms». *J. Chem. Theory Comput.* [en línia], 7 (8) (2011), p. 2335–2345. <<https://doi.org/10.1021/ct200290m>>.
- [8] MAEDA, S.; TAKETSUGU, T.; MOROKUMA, K. «Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method». *J. Comput. Chem.* [en línia], 35 (2) (2014), p. 166–173. <<https://doi.org/10.1002/jcc.23481>>.
- [9] MAEDA, S.; HARABUCHI, Y.; TAKAGI, M.; TAKETSUGU, T.; MOROKUMA, K. «Artificial force induced reaction (AFIR) method for exploring quantum chemical potential energy surfaces». *Chem. Rec.* [en línia], 16 (5) (2016), p. 2232–2248. <<https://doi.org/10.1002/tcr.201600043>>.
- [10] MARTÍNEZ-NÚÑEZ, E. «An automated method to find transition states using chemical dynamics simulations». *J. Comput. Chem.* [en línia], 36 (4) (2015), p. 222–234. <<https://doi.org/10.1002/jcc.23790>>.
- [11] MARTÍNEZ-NÚÑEZ, E. «An automated transition state search using classical trajectories initialized at multiple minima». *Phys. Chem. Chem. Phys.* [en línia], 17 (22) (2015), p. 14912–14921. <<https://doi.org/10.1039/c5cp02175h>>.
- [12] VÁZQUEZ, S. A.; OTERO, X. L.; MARTÍNEZ-NÚÑEZ, E. «A trajectory-based method to explore reaction mechanisms». *Molecules* [en línia], 23 (12) (2018), p. 1–19. <<https://doi.org/10.3390/molecules23123156>>.
- [13] MARTÍNEZ-NÚÑEZ, E.; BARNES, G. L.; GLOWACKI, D. R.; KOPEC, S.; PELÁEZ, D.; RODRÍGUEZ, A.; RODRÍGUEZ-FERNÁNDEZ, R.; SHANNON, R. J.; STEWART, J. J. P.; TAHOCES, P. G.; VÁZQUEZ, S. A. «AutoMeKin2021: An open-source program for automated reaction discovery». *J. Comput. Chem.* [en línia], 42 (28) (2021), p. 2036–2048. <<https://doi.org/10.1002/jcc.26734>>.
- [14] WANG, L.-P.; TITOV, A.; MCGIBBON, R.; LIU, F.; PANDE, V. S.; MARTÍNEZ, T. J. «Discovering chemistry with an ab initio nano-reactor». *Nat. Chem.* [en línia], 6 (12) (2014), p. 1044–1048. <<https://doi.org/10.1038/nchem.2099>>.
- [15] GUAN, Y.; INGMAN, V. M.; ROOKS, B. J.; WHEELER, S. E. «AARON: An automated reaction optimizer for new catalysts». *J. Chem. Theory Comput.* [en línia], 14 (10) (2018), p. 5249–5261. <<https://doi.org/10.1021/acs.jctc.8b00578>>.
- [16] SCHADT, E. E.; LINDERMAN, M. D.; SORENSON, J.; LEE, L.; NOLAN, G. P. «Computational solutions to large-scale data management and analysis». *Nat. Rev. Genet.* [en línia], 11 (9) (2010), p. 647–657. <<https://doi.org/10.1038/nrg2857>>.
- [17] SAGIROGLU, S.; SINANC, D. «Big data: A review». A: *2013 International Conference on Collaboration Technologies and Systems (CTS)* [en línia]. San Diego, CA: IEEE, 2013, p. 42–47. <<https://doi.org/10.1109/CTS.2013.6567202>>.
- [18] WEININGER, D. «SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules». *J. Chem. Inf. Comput. Sci.* [en línia], 28 (1) (1988), p. 31–36. <<https://doi.org/10.1021/ci00057a005>>.
- [19] WEN, M.; BLAU, S. M.; SPOTTE-SMITH, E. W. C.; DWARAKNATH, S.; PERSSON, K. A. «BonDNet: A graph neural network for the prediction of bond dissociation energies for charged molecules». *Chem. Sci.* [en línia], 5 (2021). <<https://doi.org/10.1039/d0sc05251e>>.
- [20] PATHAK, Y.; MEHTA, S.; PRIYAKUMAR, U. D. «Learning atomic interactions through solvation free energy prediction using graph neural networks». *J. Chem. Inf. Model.* [en línia], 61 (2) (2021), p. 689–698. <<https://doi.org/10.1021/acs.jcim.0c01413>>.
- [21] HAGBERG, A. A.; SCHULT, D. A.; SWART, P. J. «Exploring network structure, dynamics, and function using NetworkX». VAROQUAUX, G.; VAUGHT, T.; MILLMAN, J. (ed.). *Proceedings of 7th Python in Science Conference* (Pasadena, 2008), p. 11–15.
- [22] BESORA, M.; MASERAS, F. «Microkinetic modeling in homogeneous catalysis». *Wiley Interdiscip. Rev. Comput. Mol. Sci.* [en línia], 8 (6) (2018), p. 1–13. <<https://doi.org/10.1002/wcms.1372>>.
- [23] KOZUCH, S.; SHAIK, S. «How to conceptualize catalytic cycles? The energetic span model». *Acc. Chem. Res.* [en línia], 44 (2) (2011), p. 101–110. <<https://doi.org/10.1021/ar1000956>>.
- [24] KOZUCH, S. «A refinement of everyday thinking: The energetic span model for kinetic assessment of catalytic cycles». *Wiley Interdiscip. Rev. Comput. Mol. Sci.* [en línia], 2 (5) (2012), p. 795–815. <<https://doi.org/10.1002/wcms.1100>>.
- [25] KOZUCH, S. «Steady state kinetics of any catalytic network: Graph theory, the energy span model, the analogy

- between catalysis and electrical circuits, and the meaning of "mechanism". *ACS Catal.* [en línia], 5 (9) (2015), p. 5242-5255. <<https://doi.org/10.1021/acscatal.5b00694>>.
- [26] GARAY-RUIZ, D.; BO, C. «Revisiting catalytic cycles: A broader view through the energy span model». *ACS Catal.* [en línia], 10 (21) (2020), p. 12627-12635. <<https://doi.org/10.1021/acscatal.0c02332>>.
- [27] GARAY-RUIZ, D. *gTOFFee*. Repositori GitLab [en línia]. <<https://gitlab.com/dgarayr/gtoffee>> [Consulta: 26 agost 2022].
- [28] RUSH, L. E.; PRINGLE, P. G.; HARVEY, J. N. «Computational kinetics of cobalt-catalyzed alkene hydroformylation». *Angew. Chem. - Int. Ed.* [en línia], 53 (33) (2014), p. 8672-8676. <<https://doi.org/10.1002/anie.201402115>>.
- [29] GUMEROVA, N. I.; ROMPEL, A. «Polyoxometalates in solution: Speciation under spotlight». *Chem. Soc. Rev.* [en línia], 49 (21) (2020), p. 7568-7601. <<https://doi.org/10.1039/DOCS00392A>>.
- [30] MIRAS, H. N.; COOPER, G. J. T.; LONG, D. L.; BÖGGE, H.; MÜLLER, A.; STREB, C.; CRONIN, L. «Unveiling the transient template in the self-assembly of a molecular oxide nanowheel». *Science* [en línia], 327 (5961) (2010), p. 72-74. <<https://doi.org/10.1126/science.1181735>>.
- [31] FILELLA, M.; MAY, P. M. «The aqueous solution thermodynamics of niobium under conditions of environmental and biological interest». *Appl. Geochem.* [en línia], 122 (2020), p. 104729. <<https://doi.org/10.1016/j.apgeochem.2020.104729>>.
- [32] FILELLA, M.; MAY, P. M. «The aqueous solution thermodynamics of tantalum under conditions of environmental and biological interest». *Appl. Geochem.* [en línia], 109 (abril 2019), p. 104402. <<https://doi.org/10.1016/j.apgeochem.2019.104402>>.
- [33] ANJASS, M.; LOWE, G. A.; STREB, C. «Molecular vanadium oxides for energy conversion and energy storage: Current trends and emerging opportunities». *Angew. Chem. - Int. Ed.* [en línia], 60 (14) (2021), p. 7522-7532. <<https://doi.org/10.1002/anie.202010577>>.
- [34] LÓPEZ, X.; CARBÓ, J. J.; BO, C.; POBLET, J. M. «Structure, properties and reactivity of polyoxometalates: A theoretical perspective». *Chem. Soc. Rev.* [en línia], 41 (22) (2012), p. 7537-7571.
- [35] DIAB, M.; MATEO, A.; CHEIKH, J. A.; HAOUAS, M.; RANJIBARI, A.; BOURDREUX, F.; NAOUFAL, D.; CADOT, E.; BO, C.; FLOQUET, S. «Unprecedented coupling reaction between two anionic species of a closo-decahydrodecaborate cluster and an Anderson-type polyoxometalate». *Dalton Trans.* [en línia], 49 (15) (2020), p. 4685-4689. <<https://doi.org/10.1039/c9dt04676c>>.
- [36] MELGAR, D.; BANDEIRA, N. A. G.; BO, C. «Electronic structure studies on the whole keplerate family: Predicting new members». *Chem. - Eur. J.* [en línia], 23 (22) (2017), p. 5338-5344. <<https://doi.org/10.1002/chem.201605981>>.
- [37] MARTIN, N. P.; PETRUS, E.; SEGADO, M.; ARTEAGA, A.; ZAKHAROV, L. N.; BO, C.; NYMAN, M. «Strategic capture of the {Nb7} polyoxometalate». *Chem. - Eur. J.* [en línia], 25 (45) (2019), p. 10580-10584. <<https://doi.org/10.1002/chem.201902770>>.
- [38] PETRUS, E.; SEGADO, M.; BO, C. «Nucleation mechanisms and speciation of metal oxide clusters». *Chem. Sci.* [en línia], 11 (32) (2020), p. 8448-8456. <<https://doi.org/10.1039/d0sc03530k>>.
- [39] PETRUS, E.; BO, C. «Unlocking phase diagrams for molybdenum and tungsten nanoclusters and prediction of their formation constants». *J. Phys. Chem. A*, 125 (23) (2021), p. 5212-5219.
- [40] PETRUS, E.; SEGADO-CENTELLAS, M.; BO, C. «Computational prediction of speciation diagrams and nucleation mechanisms: Molecular vanadium, niobium, and tantalum oxide nanoclusters in solution». *Inorg. Chem.* [en línia], 61 (35) (2022), p. 13708-13718. <<https://doi.org/10.1021/acs.inorgchem.2c00925>>.
- [41] BADER, R. F. W. «A quantum theory of molecular structure and its applications». *Chem. Rev.* [en línia], 91 (5) (1991), p. 893-928. <<https://doi.org/10.1021/cr00005a013>>.
- [42] SEYBOLD, P. G. «Quantum chemical estimation of the acidities of some inorganic oxoacids». *Mol. Phys.* [en línia], 113 (3-4) (2015), p. 232-236. <<https://doi.org/10.1080/00268976.2014.927080>>.
- [43] BALDASARE, C. A.; SEYBOLD, P. G. «Computational estimation of the aqueous acidities of alcohols, hydrates, and enols». *J. Phys. Chem. A* [en línia], 125 (17) (2021), p. 3600-3605. <<https://doi.org/10.1021/acs.jpca.1c01330>>.
- [44] ELVINGSON, K.; BARÓ, A. G.; PETERSSON, L. «Speciation in vanadium bioinorganic systems. 2. An NMR, ESR, and potentiometric study of the aqueous H⁺-vanadate-maltol system». *Inorg. Chem.* [en línia], 35 (11) (1996), p. 3388-3393. <<https://doi.org/10.1021/ic951195s>>.
- [45] SADOC, A.; MESSAOUDI, S.; FURET, E.; GAUTIER, R.; FUR, E. L.; POLLÈS, L. L.; PIVAN, J. Y. «Structure and stability of VO₂⁺ in aqueous solution: A Car-Parrinello and static *ab initio* study». *Inorg. Chem.* [en línia], 46 (12) (2007), p. 4835-4843. <<https://doi.org/10.1021/ic0614519>>.

- [46] FUKUDA, N.; YAMASE, T. «In vitro antibacterial activity of vanadate and vanadyl compounds against *Streptococcus pneumoniae*». *Biol. Pharm. Bull.*, 20 (8) (1997), p. 927-930.
- [47] AURELIANO, M.; CRANS, D. C. «Decavanadate (V100286-) and oxovanadates: Oxometalates with many biological activities». *J. Inorg. Biochem.* [en línia], 103 (4) (2009), p. 536-546. <<https://doi.org/10.1016/j.jinorgbio.2008.11.010>>.
- [48] BONTCHEV, R. P.; NYMAN, M. «Evolution of polyoxoniobate cluster anions». *Angew. Chem. - Int. Ed.* [en línia], 45 (40) (2006), p. 6670-6672. <<https://doi.org/10.1002/anie.200602200>>.
- [49] AMIRI, M.; MARTIN, N. P.; FENG, C.; LOVIO, J. K.; NYMAN, M. «Deliberate construction of polyoxoniobates exploiting the carbonate ligand». *Angew. Chem.* [en línia], 133 (22) (2021), p. 12569-12574. <<https://doi.org/10.1002/ange.202017367>>.
- [50] BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. «The semantic web». *Sci. Am.* [en línia], 284 (5) (2001), p. 34-43. <<https://doi.org/10.1038/scientificamerican0501-34>>.
- [51] WIKIMEDIA. *Wikidata*. <<https://www.wikidata.org/wiki>> [Consulta: 5 setembre 2022].
- [52] EIBECK, A.; LIM, M. Q.; KRAFT, M. «J-Park Simulator: An ontology-based platform for cross-domain scenarios in process industry». *Comput. Chem. Eng.* [en línia], 131 (2019), p. 106586. <<https://doi.org/10.1016/j.compchemeng.2019.106586>>.
- [53] ASHBURNER, M.; BALL, C. A.; BLAKE, J. A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J. M.; DAVIS, A. P.; DOLINSKI, K.; DWIGHT, S. S.; EPPIG, J. T.; HARRIS, M. A.; HILL, D. P.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, J. C.; RICHARDSON, J. E.; RINGWALD, M.; RUBIN, G. M.; SHERLOCK, G. «Gene ontology: Tool for the unification of biology». *Nat. Genet.* [en línia], 25 (2000), p. 25-29. <<https://doi.org/10.1038/75556>>.
- [54] JACKSON, R.; MATENTZOGLU, N.; OVERTON, J. A.; VITA, R.; BALHOFF, J. P.; BUTTIGIEG, P. L.; CARBON, S.; COURTOT, M.; DIEHL, A. D.; DOOLEY, D. M.; DUNCAN, W. D.; HARRIS, N. L.; HAENDEL, M. A.; LEWIS, S. E.; NATALE, D. A.; OSUMI-SUTHERLAND, D.; RUTTENBERG, A.; SCHRIML, L. M.; SMITH, B.; STOECKERT JR., C. J.; VASILEVSKY, N. A.; WALLS, R. L.; ZHENG, J.; MUNGALL, C. J.; PETERS, B. «OBO Foundry in 2021: Operationalizing open data principles to evaluate ontologies». *Database* [en línia], 2021 (2021), baab069. <<https://doi.org/10.1093/database/baab069>>.
- [55] CARBON, S. [et al.]. «The gene ontology resource: Enriching a GOld mine». *Nucleic Acids Res.* [en línia], 49 (D1) (2021), p. D325-D334. <<https://doi.org/10.1093/nar/gkaa1113>>.
- [56] CYGANIAK, R.; WOOD, D.; LANTHALER, M. *Resource Description Framework (RDF) 1.1: Concepts and abstract syntax. W3C Recommendation* [en línia]. <<https://www.w3.org/TR/rdf-concepts/>> [Consulta: 11 novembre 2022].
- [57] GRUBER, T. R. «A translation approach to portable ontology specifications». *Knowl. Acquis.* [en línia], 5 (2) (1993), p. 199-220. <<https://doi.org/10.1006/knac.1993.1008>>.
- [58] BIZER, C.; HEATH, T.; BERNERS-LEE, T. «Linked data: The story so far». A: *Semantic services, interoperability and web applications* [en línia]. Hershey, PA: IGI Global, 2011, p. 205-227. <<https://doi.org/10.4018/978-1-60960-593-3.ch008>>.
- [59] DEGTYARENKO, K.; MATOS, P. de; ENNIS, M.; HASTINGS, J.; ZBINDEN, M.; MCNAUGHT, A.; ALCÁNTARA, R.; DARSOW, M.; GUEDJ, M.; ASHBURNER, M. «ChEBI: A database and ontology for chemical entities of biological interest». *Nucleic Acids Res.* [en línia], 36 (supl. 1) (2008), p. 344-350. <<https://doi.org/10.1093/nar/gkm791>>.
- [60] HASTINGS, J.; OWEN, G.; DEKKER, A.; ENNIS, M.; KALE, N.; MUTHUKRISHNAN, V.; TURNER, S.; SWAINSTON, N.; MENDES, P.; STEINBECK, C. «ChEBI in 2016: Improved services and an expanding collection of metabolites». *Nucleic Acids Res.* [en línia], 44 (D1) (2016), p. D1214-D1219. <<https://doi.org/10.1093/nar/gkv1031>>.
- [61] BATCHELOR, C. *Chemical Methods Ontology (CHMO)* [en línia]. Repositori GitHub. <<https://github.com/rsc-ontologies/rsc-cmo>> [Consulta: 11 novembre 2022].
- [62] STRÖMERT, P.; HUNOLD, J.; CASTRO, A.; NEUMANN, S.; KOEPLER, O. «Ontologies4Chem: The landscape of ontologies in chemistry». *Pure Appl. Chem.* [en línia], 94 (6) (2022), p. 605-622. <<https://doi.org/10.1515/pac-2021-2007>>.
- [63] GARAY-RUIZ, D.; BO, C. «Chemical reaction network knowledge graphs: The OntoRXN ontology». *J. Cheminformatics* [en línia], 14 (1) (2022), p. 29. <<https://doi.org/10.1186/s13321-022-00610-x>>.
- [64] GARAY-RUIZ, D. *OntoRXN-Tools* [en línia]. Repositori GitLab. <https://gitlab.com/dgarayr/ontorxn_tools> [Consulta: 11 novembre 2022].
- [65] ÁLVAREZ-MORENO, M.; GRAAF, C. de; LÓPEZ, N.; MASERAS, F.; POBLET, J. M.; BO, C. «Managing the computational chemistry big data problem: The loChem-BD platform». *J. Chem. Inf. Model.* [en línia], 55 (1) (2015), p. 95-103. <<https://doi.org/10.1021/ci500593j>>.
- [66] BO, C.; MASERAS, F.; LÓPEZ, N. «The role of computational results databases in accelerating the discovery of catalysts». *Nat. Catal.* [en línia], 1 (11) (2018), p. 809-810. <<https://doi.org/10.1038/s41929-018-0176-4>>.
- [67] MAQUILÓN, C.; LIMBURG, B.; LASERNA, V.; GARAY-RUIZ, D.; GONZÁLEZ-FABRA, J.; BO, C.; MARTÍNEZ BELMONTE, M.; ESCUDERO-ADÁN,

E. C.; KLEIJ, A. W. «Effect of an Al[III] complex on the regio- and stereoisomeric formation of bicyclic organic carbonates». *Organometallics* [en línia], 39 (9) (2020), p. 1642-1651. <<https://doi.org/10.1021/acs.organomet.9b00773>>.

[68] HOOPS, S.; GAUGES, R.; LEE, C.; PAHLE, J.; SIMUS, N.; SINGHAL, M.; XU, L.; MENDES, P.; KUMMER, U. «COPASI - A COMplex PATHway Simulator». *Bioinformatics* [en línia], 22 (24) (2006), p. 3067-3074. <<https://doi.org/10.1093/bioinformatics/btl485>>.

[69] GARAY-RUIZ, D. *OntoRXN* [en línia]. Repositori GitLab. <<https://gitlab.com/dgarayr/ontorxn>> [Consulta: 11 novembre 2022].

[70] GARAY-RUIZ, D. *CRN-KG Web Service* [en línia]. <<https://doi.org/10.19061/crn-kg-ontorxn.2022>> [Consulta: 11 novembre 2022].

[71] BLAZEGRAPH. *BlazeGraph RDF Database* [en línia]. Repositori GitHub. <<https://github.com/blazegraph/database>> [Consulta: 11 novembre 2022].



D. Garay



E. Petrus



C. Bo

Diego Garay–Ruiz és graduat en química per la Universitat de Valladolid i va estudiar el Màster en Modelització Computacional Atomística i Multiescala en Física, Química i Bioquímica de la Universitat de Barcelona i la Universitat Politècnica de Catalunya. L'any 2023 va finalitzar el seu doctorat sota la supervisió del doctor Carles Bo, a l'Institut Català d'Investigació Química (ICIQ), on ara treballa com a investigador postdoctoral. La seva recerca està centrada en el desenvolupament d'eines i programes per al processament de xarxes de reacció complexes.

Enric Petrus Pérez és doctor en química per la Universitat Rovira i Virgili (URV) i l'Institut Català d'Investigació Química (ICIQ), en el camp de la química quàntica i computacional (2022). La seva tesi es va centrar en el desenvolupament d'un programari anomenat POMSimulator, que permet predir l'especiació aquosa i els mecanismes de reacció dels òxids metàl·lics moleculars. Actualment treballa com a investigador postdoctoral a l'Institut Federal Suís de Ciència i Tecnologia Aquàtica (Eawag). En l'àmbit de la divulgació, ha participat en programes de ràdio i en monòlegs científics.

Carles Bo Jané és cap de grup de l'Institut Català d'Investigació Química (ICIQ) des de la seva fundació, el 2004, i professor titular de química física en excedència a la Universitat Rovira i Virgili. La seva recerca s'ha centrat en el desenvolupament i l'aplicació de mètodes teòrics i computacionals a una diversitat de temes químics, molt sovint en estreta col·laboració amb grups de químics experimentals. En els darrers anys ha liderat el disseny i el desenvolupament de projectes que implementen i faciliten els tractaments avançats de dades.