Revista de la Societat Catalana de Química, núm. 22 (2023), p. 23-38
Filial de l'Institut d'Estudis Catalans

ISSN: 2013-9853 DOI: 10.2436/20.2003.01.142
https://revistes.iec.cat/index.php/RSCQ

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 Petrus[1] 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 dificultós. 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*

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 1*a* i 1*b*). 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 1*c*). 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 introduirem 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 1*d*) 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.

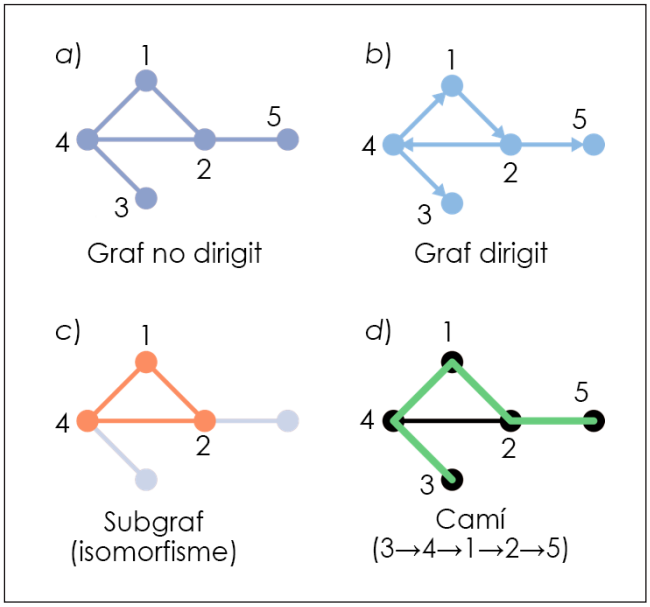

FIGURA 1. Representacions d'exemples corresponents a *a*) un graf no dirigit, *b*) un graf dirigit, *c*) un subgraf i *d*) un camí. Elaboració pròpia.

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'especiació 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 estandarditzar 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 proporciona 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 *mecanisme* 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 distints. Depenent 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 mèto-

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 aldehids, 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 H_2. 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, conseqüentment 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

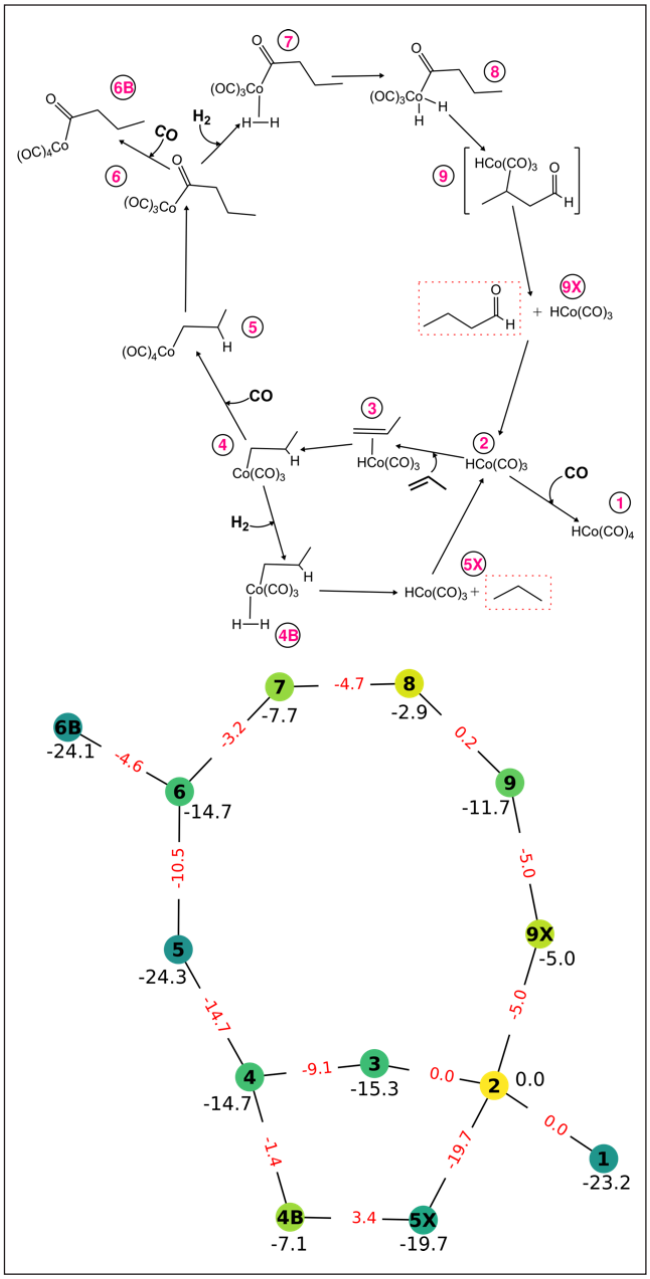

Figura 2. Cicle catalític i xarxa de reacció per a la hidroformilació de propè, catalitzada mitjançant cobalt, proposada per Rush, Pringle i Harvey. Elaboració pròpia.

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

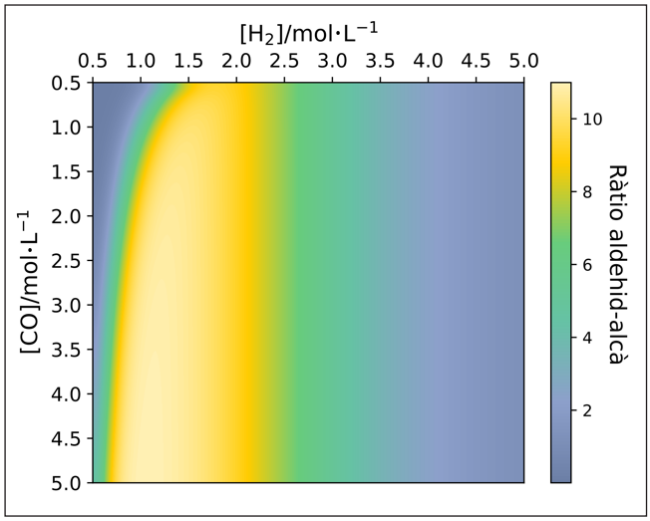

FIGURA 3. Selectivitat de la reacció en funció de les concentracions inicials de CO i H_2, en mol·L^{-1}. Elaboració pròpia.

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.

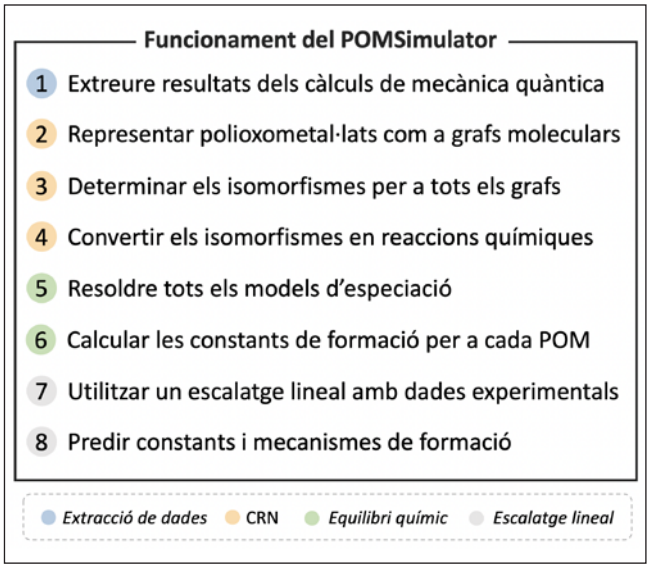

FIGURA 4. Esquema general del funcionament del POMSimulator. Elaboració pròpia.

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 5*a*). 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 5*a* mostra

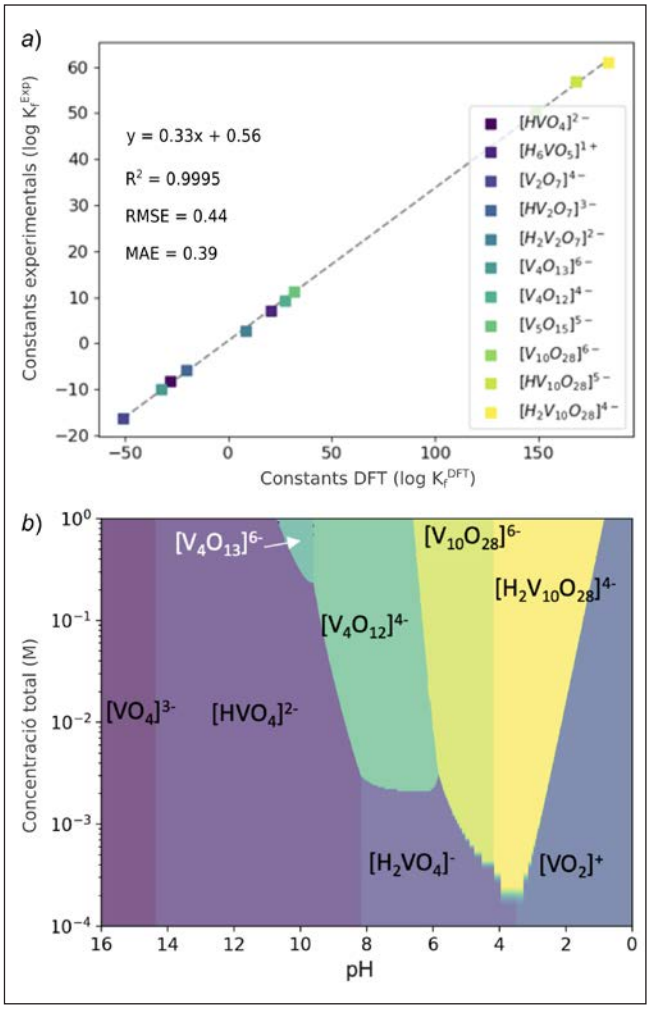

FIGURA 5. *a*) Escalat lineal entre les constants de formació experimentals i les obtingudes amb el POMSimulator. *b*) Diagrama de fases d'especiació per als isopolioxovanadats. Elaboració pròpia.

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 5*b* 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