



Segons el Tate Museum de Londres, en art els curadors són experts que ajunten les obres que veiem en una exposició d'un museu o una galeria. Utilitzen les obres d'art per explicar històries d'art i explorar temes i idees. Els biocuradors fem el mateix. Des de l'aparició dels ordinadors i dels avenços tecnològics en ciències de la vida, la quantitat de dades biològiques ha crescut exponencialment i s'han anat emmagatzemant en bases de dades. El fet d'emmagatzemar aquestes dades produïdes, però, no és de gran ajuda ni per als investigadors ni per als ordinadors. Per ser útils, necessiten ser triades d'alguna manera o altra. Aquest pas el poden fer programes especialitzats, però, així com en moltes altres coses, sense un toc humà es troba a faltar alguna cosa.

Swiss-Prot és una base de dades de seqüències de proteïnes que va néixer fa 32 anys, quan les seqüències proteïques tot just començaven a aparèixer. Llavors, cada seqüència nova podia ser estudiada en detall. Avui en dia, en canvi, milions de seqüències de proteïnes són produïdes mensualment. Com pot arribar a seguir el ritme Swiss-Prot? Gràcies als biocuradors.

La feina d'un biocurador és molt menys senzilla del que podria semblar. S'han publicat tota una sèrie d'articles per definir aquesta nova espècie de científic. Els biocuradors som científics professionals que hem estat formats per col·lectar, triar, sintetitzar, organitzar i validar informació biològica que serà disseminada via les diferents bases de dades. La natura de la nostra professió demana la paciència i la minuciositat d'un bibliotecari i se'ns ha definit com a "catalogadors de museus de l'era internet", "els que preferim els ordinadors a les pipetes" o "rates de biblioteca", des d'una perspectiva una mica antiquada i tancada de mires.

La biocuració no és només una ciència per sí mateixa, ja que, a més a més, així com la tecnologia que la nodreix, està en evolució constant. Els biocuradors ens hem

Cristina Casals i Casas

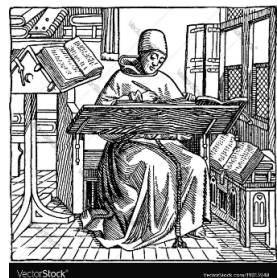
Doctora en biologia (UB),
licenciada en biologia (U.
Paul Sabatier). Biocuradora
d'UniprotKB/Swiss-prot,
Swiss Institute of
Bioinformatics (SIB),
Campus Biotech Genève

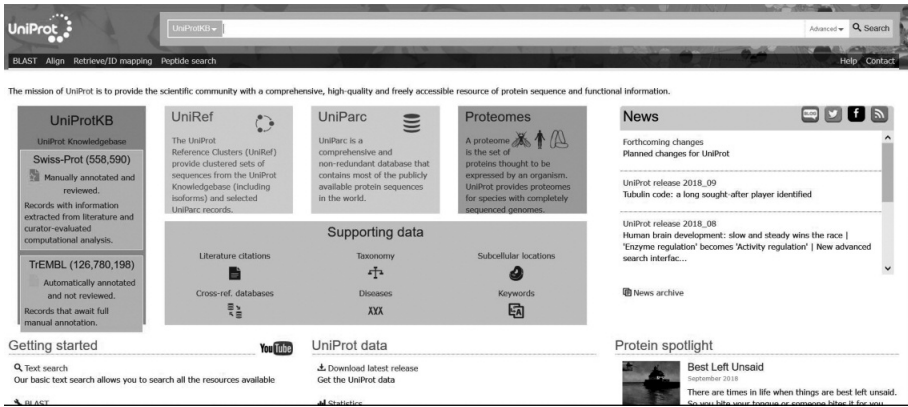


Cristina Casals i Casas

L'art de la biocuració

Text adaptat d'un article original escrit per Vivienne Baillie Gerritsen i Marie-Claude Blatter. Swiss-Prot Group, SIB Swiss Institute of Bioinformatics





d'adaptar ràpidament i encarar una producció sense fi de dades noves. Hem de fer front a un coneixement de la biologia en canvi constant. A Swiss-Prot, per exemple, tot i que el nombre de seqüències no sobrepassa el mig milió, proveeix dades que han estat controlades i millorades per biocuradors. Aquesta informació millorada és, al seu torn, reutilitzada per a l' anotació automàtica de seqüències noves. D'aquesta manera, els biocuradors hem esdevingut una peça molt important en el trencaclosques de la biologia.

Actualment, uns 65 biocuradors treballem per la base de coneixement Swiss-Prot (que ara forma part de la base de coneixement UniProtKB) a Suïssa (a l'Institut Suís de Bioinformàtica), a Anglaterra (a l'EBI) i als Estats Units (al PIR). D'aquests 65 biocuradors, 45 treballem a Suïssa. Som biòlegs, bioquímics i químics amb experiència en laboratoris de recerca, i la majoria tenim un doctorat a les mans. La nostra feina consisteix a revisar amb un ull crític dades experimentals i de predicció per cada proteïna, així com verificar en detall cada seqüència proteica. D'aquesta manera, oferim una visió completa de tota la informació que hi ha sobre una proteïna.

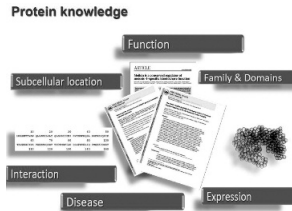
Extraïem la informació de diferents fonts; la majoria prové de dades experimentals publicades en revistes científiques. Els biocuradors reconciliem els resultats que poden ser conflictius per després recopilar-los en

un informe concís i en un llenguatge comprensiu per a l'ull humà, així com per a l'ordinador. La feina del biocurador afegeix un valor afegit de coneixement als informes d'UniProtKB/Swiss-Prot i inclou informació relacionada amb la funció, l'estructura o la localització subcel·lular d'una proteïna concreta. Però també una àmplia gamma de distintius sobre la seqüència, com ara el lloc actiu o modificacions posttraduccionals.

Una gran quantitat de dades equival a tones d'articles. Els biocuradors hem de llegir cada un dels articles publicats per cada proteïna? La resposta és *no*. Una etapa crucial en curació és saber com identificar un conjunt representatiu de les publicacions que donaran una visió completa de la informació que hi ha en aquest moment. Tota informació afegida manualment a UniProtKB/Swiss-Prot està lligada a la seva font; d'aquesta manera els usuaris poden connectar cada bloc d'informació al seu origen. Per a una màxima eficàcia, els biocuradors de Swiss-Prot normalment treballarem amb grups de proteïnes relacionades, com ara proteïnes que pertanyen a la mateixa família o que es troben en la mateixa espècie. Els biocuradors de Swiss-Prot també col·laborem i aprofitem el treball d'altres bases de coneixement complementàries per facilitar la consistència i intercanviar dades, per assegurar-nos així que els recursos utilitzats no es dupliquin.

Però tot això és molt teòric. Com fem front els biocuradors de Swiss-Prot al coneixement que es produeix diàriament? La recent caracterització de la proteïna Notum tant en humans com en la mosca del vinagre, *Drosophila melanogaster*, n'és un molt bon exemple. Un biocurador Swiss-Prot es va endinsar en la tria de més de cent publicacions relacionades amb el terme *Notum* (es dona el cas que també és el terme utilitzat per a la secció dorsal del segment toràctic d'un insecte). Segons criteris ben establerts, dos publicacions eren suficients per extraure la informació necessària per adquirir una visió comprensiva de la proteïna.

Notum va ser inicialment caracteritzada en *Drosophila melanogaster* com una proteïna amb un rol vital en la





morfogènesi que inhibia la via de senyalització WNT (una via que passa senyals dins la cèl·lula a partir de receptors de la superfície cel·lular). Durant més de vint anys, els investigadors pensaven que Notum inhibia aquesta via a través de la hidròlisi de proteoglicans específics coneguts com a *glicans*. En canvi, els dos articles seleccionats contradiuen aquests resultats: la inhibició es fa directament a través de la despalmitolació d'una serina de la seqüència proteica de WNT. Així doncs, el rol de Notum com a inhibidor de la via es confirma, però amb un mecanisme diferent.

Gràcies a aquestes noves dades experimentals s'ha pogut fer una actualització de la funció, així com del nom de la proteïna, i s'ha difós a deu espècies diferents dins UniProtKB/Swiss-Prot. Però la feina del biocurador no s'atura aquí... També s'han creat nous termes per a l'ontologia gènica (GO), ha sortit un nou codi de la comissió enzimàtica (EC number) i s'han anotat a la Protein Data Bank (PDB) les posicions del lloc actiu de l'enzim a partir de l'estructura tridimensional (3D), així com les modificacions per despalmitolació sobre la proteïna WNT i les seves conseqüències, utilitzant el vocabulari definit per la base de dades RESID de "Protein modifications". A més a més, per no perdre cap mena d'informació, tot i que ha esdevingut redundat, la funció antiga de la proteïna Notum és descrita en detall dins de la secció "Caution" (precaució).

Aquest exemple il·lustra com és d'essencial la feina d'un biocurador, i com la identificació de les poques publicacions pertinents pot donar cap a informació que no només és vital però que té un efecte de reacció en cadena. Gràcies a la curació manual, noves funcions biològiques es continuen donant a conèixer dintre del que es considera una família de proteïnes ben caracteritzada. La informació correcta i actualitzada és difosa entre altres bases de dades a més a més de ser transferida a través de l'anotació automatitzada i dels programes de predicció de funcions. El fet que les col·laboracions estretes entre diferents recursos no faci més que augmentar any rere any demostra una altra vegada com és d'important la biocuració per mantenir el coneixement biològic.