

ESTRUCTURACIÓ I CLASSIFICACIÓ DE LES PROTEINES

Carles UDINA i COBO, 2002-08-20, revisat

Les proteïnes són estructures molt complexes i de classificació gens trivial per les dificultats de diferenciació formal/ estructural, evolució, ..., el que ha originat diversos procediments manuals i automàtics (SCOP, PDB, CATH, Dali, ...). Per aquesta circumstància, el 2002 vaig analitzar aquest problema. I la primera sorpresa, veure equívocs en qüestions tan bàsiques com els nivells estructurals que els bioquímic assignen a les proteïnes. Pel mateix vaig fer aquest petit document, analitzant aquest equívoc, així com altres qüestions ignorades per l'actual perspectiva científica (dualitat algebraica dels codons, homologia ADN - proteïnes,). Aquest document complementa l'apartat .9 "Les proteïnes" del document "**Classificació**".

ESTRUCTURACIÓ I CLASSIFICACIÓ DE LES PROTEINES

D'entrada, la lectura d'un llibre sobre Proteïnes no és trivial, com per exemple un de Harry POTTER, i a més a més, calen uns mínims coneixements de química molecular/ orgànica. Com es veurà en alguns paràgrafs, fins i tot alguns autors tenen dificultats expositives.

Estructuració

La classificació de l'estructuració de les Proteïnes en Primària, ..., Quaternària (Kai LINDERSTRÖM-LANG), sense desmerèixer el seu gran valor/ interès quan va ser enunciada, és racionalment incoherent, atès que respecte del criteri de partició estructural hi ha quatre nivells, però no són exactament els que s'enuncien. Aquets quatre nivells estrictes d'Estructuracions* progressives són:

- el primari i secundari, que conjuntament són un sol nivell estructural (característiques estructurals),
- les superestructures o "motius" (característiques estructurals),
- el terciari (dominis, característiques funcionals i dinàmiques/ variables),
- el quaternari (característiques funcionals complexes i molt dinàmiques/ variables)

Dins el nivell primari - secundari,

- conceptualment, la subdivisió en primari i secundari resulta d'altres dos criteris diferents entre ells i diferents del criteri estructural; són la **seqüència** (que origina la mal anomenada "Estructura primària") i la **forma** (que origina la mal anomenada "Estructura secundària"), si bé la primera condiona la segona per raons moleculars (Prediccionabilitat de l'estructura nativa);
- a la realitat, aquesta subdivisió en estructura primària i secundària no forma dos subnivells, sinó que només dos perspectives diferents i **complementaries**. Seria com si es tractés de "mòdul" i "argument" d'un vector algebraic. O encara un exemple més pròxim, com si es tractés de les expressions moleculars estequiomètriques del tipus $C_6H_{12}O_6$ (que seria com l'estructura primària), que són expressions ambigües en les grans molècules llevat que es descriu també la forma (cicles, cadenes, ... que serien com l'estructura secundària).

Finalment, les superestructures o "motius", que es van constatar posteriorment, **son estrictament un nivell** entre el primari - secundari i el terciari, si bé no és imprescindible entre ambdós.

The screenshot displays the Cuc software interface, which is used for conceptual modeling. The main window shows a hierarchical tree structure for protein classification. The tree is organized into levels: Proteïna (més de 60 Pèptids), Estructures primària i secundària d'una Proteïna, Estructura primària d'una Proteïna, Estructura secundària d'una Proteïna, and SuperEstructura secundària d'una Proteïna. The SuperEstructura secundària d'una Proteïna is highlighted in blue, indicating it is the selected level. The interface also shows a search bar, a list of results, and a status bar at the bottom.

Així que les exposicions dels llibres especialitzats són encara desendreçades (l'àmbit de les proteïnes és relativament recent), sense criteris lògics, i per això difícils d'entendre inequívocament i amb rapidesa. Es possibiliten paranys pedagògics que **dificulten** la comprensió de l'estructuració de les proteïnes. No podem transferir altres coneixements

prèviament disponibles, sinó que cal acceptar de nou en nou altre realitat, que erròniament sembla diferent a les altres realitats anteriorment conegudes.

* **NOTA IMPORTANT:** Cal dir que **no s'ha de parlar d'estructura** de les proteïnes (en singular) sinó que d'estructures (en plural) o millor d'estructuració de les proteïnes. L' "estructura" de les proteïnes només és la darrera assolida (quaternària, terciària, ...). Altre cosa és que dins un element quaternari ens adonem que resulta d'un altre complexament, el terciari, i així successivament. En un edifici, l'estructura són les bigues, parets de totxos, ..., però mai parlem simultàniament d'aquesta estructura amb l'estructura/composició material, molecular, atòmic, ..., dels materials que componen una biga o un totxo en concret.

Si parlem d' "estructuració", res impedeix imaginar-se també una estructuració progressiva com la ja esmentada:

AminoÀcid (que seria com el nivell 0 o de partida) → Pèptid/ Estructura primària - secundària → SuperEstructura secundària (= "Motiu") → Estructura terciària (= "Dominis") → Estructura quaternària.

Classificació.

Carl BRANDEN i John TOOZE, a "Introduction to Protein structure", tendeixen a parlar de "Classes" (que són més limitades), i no d'enumerar totes les proteïnes.

Així, al Pròleg ja diuen "We have tried to resist the temptation to describe more and more proteins ...". I a la pàgina 3 parlen de que la classificació de les Proteïnes és un problema "unsolved", és una "challenge", ... Cal un plantejament **comprensiu i d'accessibilitat** (és a dir, relacional) versus l'enumerativitat.

Els mateixos autors, a la pàgina 373 parlen de les moltíssimes Proteïnes com un "Warehouse", referència molt encertada doncs un DataWarehouse podria ser una bona eina complementària d'un "Sistema conceptual", en el cas de les Proteïnes o en molts altres casos.

A més a més que la classificació és la característica bàsica d'un "Sistema conceptual", d'ell també ens beneficiem:

- del model arbrat en general, per a abastar tantes proteïnes com es vulgui, milers o milions;
- de l'herència múltiple (com amb els Objectes artificials) per a tenir accessibilitat múltiple i agrupar les equivalències (diagonalització de la matriu associada, agrupant les diferents permutacions).

Analogia Proteïna ≈ Eina/ Giny (veure també l'Exemple del Flagel bacterià)

L'Estructura quaternària (no covalent) de les Proteïnes possibilita que siguin com les "eines" químiques dels éssers vius (analogia Proteïna ≈ Giny). En lloc de claus fixes del 12, del 13, ..., de claus angleses, ..., hi ha fixadors del Ca, del Fe, d'O, ..., catalitzadors, ... Veure al DICUC les Explicacions de la Successió de complexament d'Estructures.

The screenshot shows the Cuc software interface with a hierarchical tree of protein classification. The tree is organized as follows:

- AminoÀcid
 - 0 segons "Posició relativa entre dos, dins un tercer" es classifica en:
 - 0 Alfa-AminoÀcid (Grups Amino i Àcid contigus/ terminals)
 - 1 B-AminoÀcid
 - 2 gamma-AminoÀcid
 - b segons "Polimerització amb Enllaços peptídics (-CO-NH-)" es classifica en:
 - 0 Pèptid
 - a segons "Número total de Pèptids" es classifica en:
 - 0 DiPèptid
 - 1 TriPèptid
 - 2 TetraPèptid
 - 3 PoliPèptid (de 5 fins a 59 Pèptids)
 - 4 Proteïna (més de 60 Pèptids)
 - a ... s'organitza a la realitat segons el/la/s "Successió fenomenològica" es classifica en:
 - 0 Estructures primària i secundària d'una Proteïna (Seqüència)
 - 1 SuperEstructura secundària d'una Proteïna (=Motiu estructural)
 - 2 Estructura terciària d'una Proteïna (Forma TriDimensional global)
 - 3 Estructura quaternària d'una Proteïna (Associació no covalent)
 - b segons "Forma (=Aparença externa, Abstracció visual/ tàctil)" es classifica en:
 - 0 Globulina (=Proteïna globular =EsferoProteïna)
 - a segons "Ferro [Z=26]" es classifica en:
 - a segons "Oxigenació" es classifica en:
 - 0 Globulina fèrrica oxigenadora (Familia de Globulines amb Funcions oxigen) [Polid denominació/ HomoSèmia (=Epijecció)]
 - 0 Proteïna globular (=Globulina =EsferoProteïna)
 - 1 Proteïna fibril·lar (Cabell, Múscul)
 - c segons "Funció (=Acció característica, associada a ...)" es classifica en:
 - 0 Proteïna funcional
 - a segons "Catalització" es classifica en:
 - 0 Enzim (=Proteïna que és Catalitzador orgànic)
 - b segons "Malalties infeccioses" es classifica en:
 - 0 ImmunoGlobulina (=AntiCos =Proteïna antiinfecció)
 - 1 Proteïna estructural
 - d segons "Estructuració" es classifica en:
 - 0 Proteïna nativa (=estructurada, Estat amb mínima Energia de GIBBS)
 - 1 Proteïna desnaturalitzada (=desestructurada, amb Energia)
 - e segons "acidesa" es classifica en:
 - 0 Proteïna àcida (=amb predomini d'AminoÀcids acids)

De l'analogia entre Proteïnes i Objectes artificials (o més específicament els Ginyes) resulta la possibilitat de classificar les Proteïnes inspirant-se en els criteris de classificar els **nombrosíssims Objectes artificials**. Per exemple "... segons Funció" (que és un dels criteris bàsics de la classificació de les Proteïnes), "... segons Ús", ...

Classificació i Analogia Proteïna ≈ Arquitectura

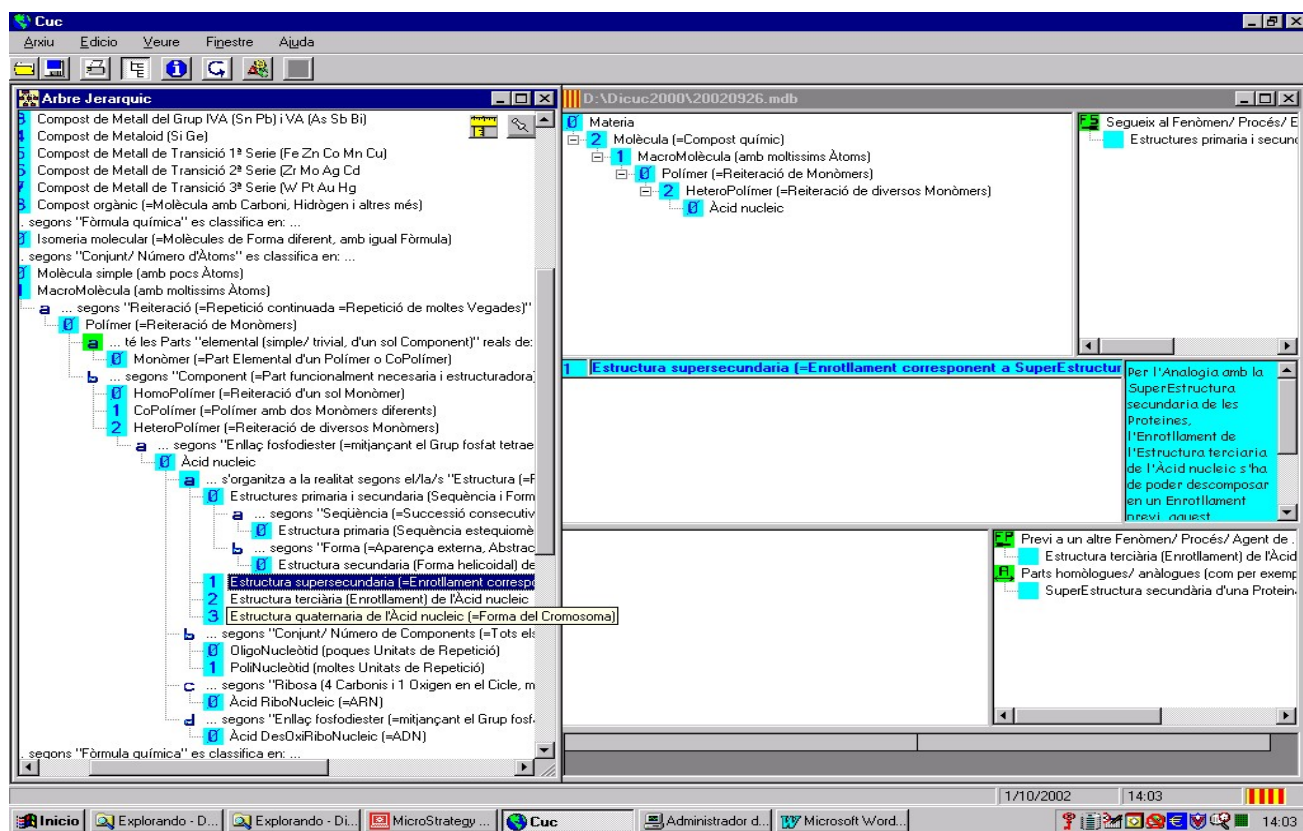
Les Proteïnes són com edificis fets amb totxos, **com una joguina constructivista del tipus "Lego"** o tren elèctric, d'aquí que BRANDEN i TOOZE parlin de "The Building Blocks" (pàgina 3). Tants edificis diferents com podem imaginar, tantes Proteïnes diferents com poden existir. No es tracta de disposar d'un directori d'edificis (com si es tractés del "New York telephone directory"), sinó que d'un **sistema classificatori**.

Analogia Proteïna ≈ Àcid nucleic. Exemple de Predicció de conceptes (Veure també "Comentaris addicionals - Programa dissenyador d'esquemes d'Homologia").

L'analogia estructural Proteïna ≈ Àcid nucleic (de la Genètica) es fa clarament palesa en un "Sistema conceptual", ajudant a **aprendre d'un sol cop**, les dues. Com ja s'ha dit, veure les explicacions i les relacions d'analogia entre els Nivells estructurals de Proteïnes i Àcids nucleics.

És especialment important ser conscient d'aquesta analogia si tenim en compte la **íntima relació** entre ambdues a la realitat, atès que les proteïnes són sintetitzades per la cèl·lula en base a la informació genètica (**dualització** dels codons i/o agrupació de codons), començant a un primer nivell d'**interpretació** per la correspondència entre "Primera agrupació de codons" i "AminoÀcid comú" que codifica.

La **dualitat algebraica** estricta entre els codons i els aminoàcids que sintetitzen, i la derivada analogia Àcid nucleic ≈ Proteïnes (fins i tot es pot parlar d'homologia), així com l'existència d'una estructura quaternària en les Proteïnes permet plantejar a manera de predicció de conceptes: ¿existeix una estructura quaternària en l'Àcid nucleic? La resposta és "sí", i aquesta estructura és la **forma característica de les parts d'un cromosoma**. Aquest raonament, obvi des de la perspectiva del Sistema conceptual, no l'he trobat a cap lloc.



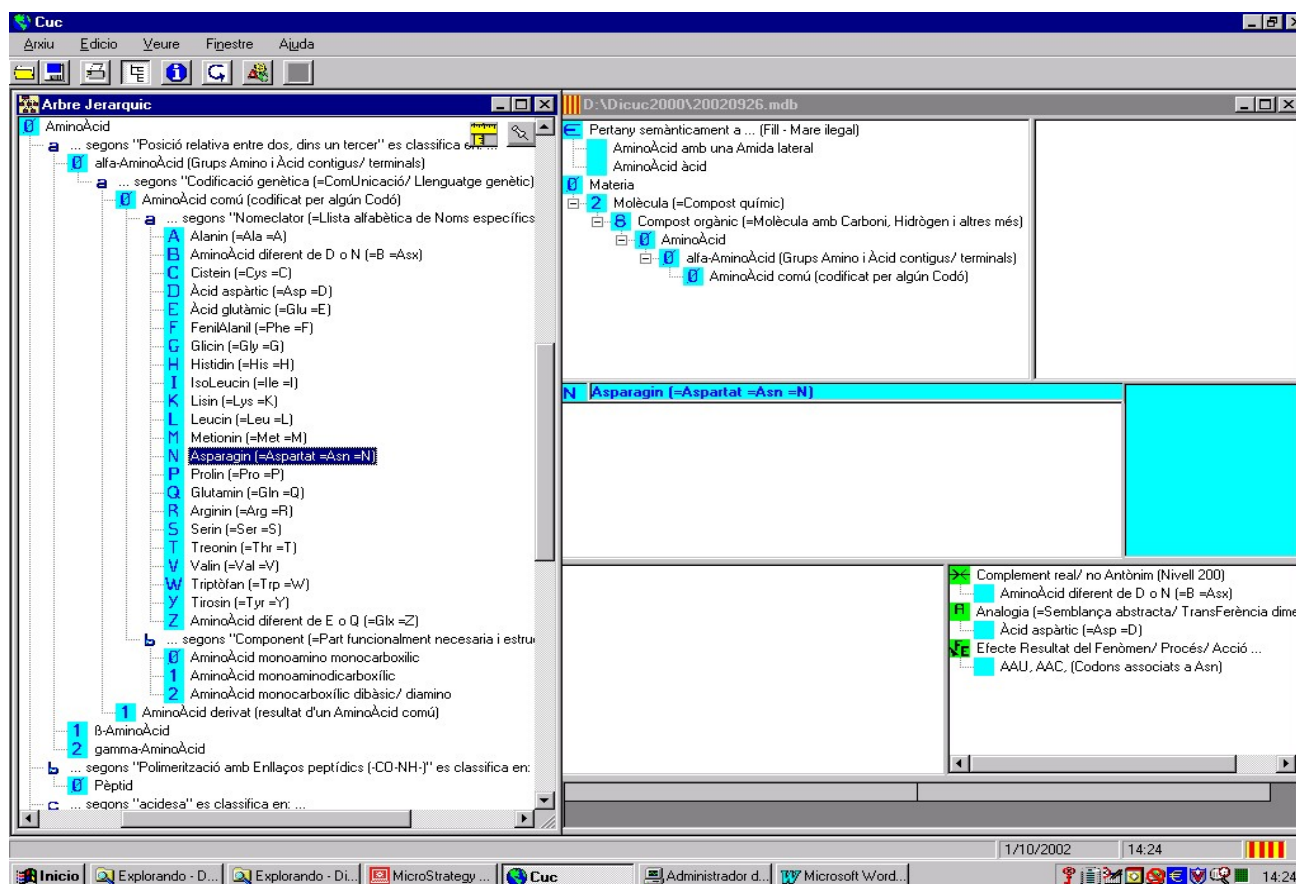
Alguns exemples de la limitació comprensiva del text tradicional

Confusió (DEVLIN a "Bioquímica") entre el text (pàg. 25 i 27) i el quadre (pàg. 26), que semblen dos classificacions diferents perquè el text comença parlant d'AlquilAmino Àcids, aromàtics, ..., oblidant dir que parla de AminoÀcids

monoaminomonocarboxilics, és a dir, que els primers són una subdivisió dels segons. L'estructuració, jerarquitze i no literalitat d'un "Sistema conceptual" impedeix aquets errors.

Confusió creada també per DEVLIN a "Bioquímica" (2.2 "Composició d'AminoÀcids de les Proteïnes" pp 25 a 27) al parlar d' "analogia estructural" entre la Glutamina i l'Asparagina i els seus associats Àcids (Àcid glutàmic i Àcid aspàrtic), el que fa pensar que aquets àcids també són amidats, però no és així. Es constata doncs la limitació d'un text literal.

C. K. MATHEWS i K. E. van HOLDE deixen clar, a diferència d'altres autors, que els AminoÀcids que s'ha convingut en representar per la "B" i la "Z" no són cap AminoÀcid, sinó que només un conveni exclouent d'altres (un complement booleà, és a dir, una de les tres facultats bàsiques del pensament enunciades per BOOLE, que és una de les 70 relacions intrínseques d'un "Sistema conceptual").



Transferència de coneixements.

En un "Sistema conceptual" es veuen clarament els diferents criteris per a definir les quatre diferents particions estrictes (primària -secundària, motius, terciària i quaternària). Els criteris són la porta d'entrada lògica als conceptes resultants.

També és obvia la utilitat de les **omnipresents** "Successions de complexament" (en la realitat i en la psique) per a entendre els nivells estructurals de les Proteïnes (i de l'Àcid nucleic). Veure també les explicacions i les relacions d'analogia entre els Nivells estructurals de Proteïnes i Àcids nucleics.

Si s'exposés l'estructuració de les Proteïnes seguint rigorosament la seva successió de complexament ja esmentada, seria comprensible fins i tot per a un nen familiaritzat amb jocs constructius si se li donessin alguns coneixements previs dels àtoms participants. En un "Sistema conceptual" les **transferències cognitives** són totals.

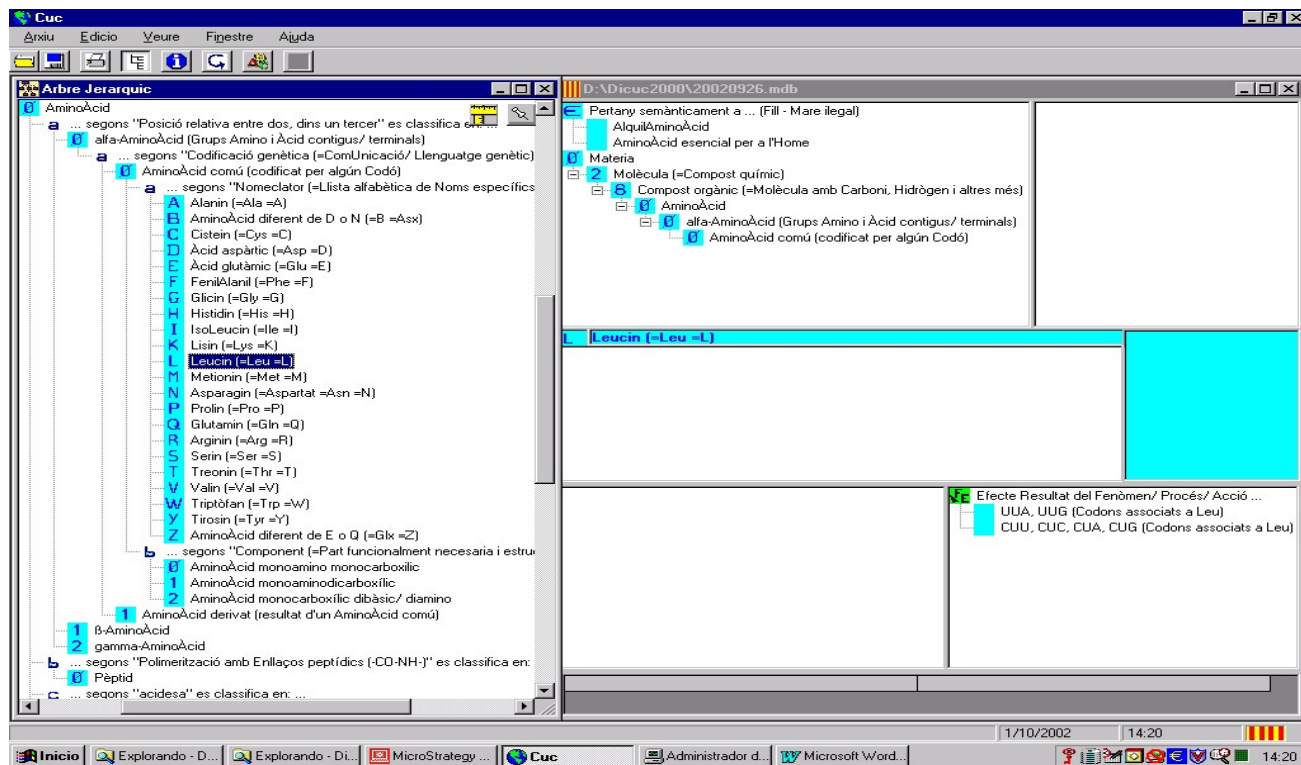
El "Sistema conceptual" com sedàs (=filtre) cognitiu

Al entrar la informació per primera vegada, el "Sistema conceptual" impedeix caure en qualsevol dels paranys esmentats, i sobretot, **els posa en evidència**, sigui per la incoherència detectada en el criteri de partició, sigui al intentar entrar un nou concepte com "SuperEstructura/ Motius", o sigui per les moltes altres regles lògiques "intrínseques" que cal respectar al entrar les dades.

Aquestes incoherències no permetrien, per exemple, la automatització, sigui per a la recuperació d'informació o per a fer raonament artificial (un programa d'ordinador és **estúpid** i no admet la decisionabilitat davant suposicions o implicacions).

Cohabitació de perspectives

Es veu també com en un "Sistema conceptual" es poden reconciliar exposicions diferents com la de DEVLIN (amb el concepte d'AminoÀcid comú) i les altres (per exemple C. K. MATHEWS i K. E. van HOLDE), que entren directament als AminoÀcids segons Nomenclatura: tots els AminoÀcids representats amb lletres (Nomenclàtor) són "AminoÀcids comuns", és a dir, codificats per algun Codó.



CONCLUSIONS

Les proteïnes constaten que **no hi ha res nou** des de fa milions d'anys. Sempre trobem **la mateixa metodologia dels complexaments**, i altres mètodes però sempre repetits, que només es diferencien pels diferents àmbits on s'apliquen. La **transferència metodològica** es constata reiteradament. El problema és que per **mancances metodològiques i relacionals (i pedagògiques)** sovint **no sabem veure-ho**, i ens sembla tot molt més complicat del que és a la realitat.

No sembla haver cap dificultat diferent de classificar i crear un banc de dades de proteïnes de fàcil recuperabilitat, a la tradicional dificultat que des de fa segles ha existit per part de filòsofs, lingüistes o científics en representar el coneixement en la seva globalitat. Per exemple, tampoc està resol el problema de classificar els continguts de la Matemàtica.

Si es disposa d'un simulador fidel dels processos del pensament, qualsevol qüestió que s'arribi a entendre, sigui de les proteïnes o de qualsevol altre qüestió, per definició ha de ser representable mitjançant aquest simulador.

D'altra banda, els quatre nivells d'estructuració de les Proteïnes, també **encaixa en el model global de la psique** (dibuix del gratacel), establert amb anterioritat (1994 i següents) a conèixer aquesta part concreta de les Proteïnes (2002-agost).

DESENVOLUPAMENT INFORMÀTIC PENDENT

En el DICUC manca codificar els programes informàtics corresponents a la composició de coneixements (**Reivindicació 23** de la Patent), que en aquest cas de les Proteïnes seria convenient per a millorar la representació dels processos de progressiu complexament proteïnica (classificació mitjançant el Nivell 2 de la Conceptuació).

També manca codificar els programes corresponents a l'ús de conceptes com relacionadors mitjançant el pas al dual (**Reivindicació 22** de la Patent), que suposaria classificar les Proteïnes segons el Nivell 3 de la Conceptuació, amb la corresponent **triple accessibilitat** i al nivell **tan específic com es desitgi**. És a dir, podríem accedir pel concepte "agent" o pel concepte "efecte" com actualment, però també per un relacionador tan "específic" com es vulgui. El relacionador "específic" no seria alguna de les 70 relacions intrínseques actuals, sinó que la dualització de qualsevol concepte tan específic com es vulgui (que equivaldria a **navegar tant a baix com es vulgui** pel arbre de les relacions).

Per la presumible quantitat d'informació a incorporar, també convindria adjuntar al "Sistema conceptual" una eina tipus DataWarehouse (per exemple de Microstrategy) per a agilitzar l'ús dels filtres i seleccions.

Completant un "Sistema conceptual" amb imatges (com en els llibres), i encara més, amb imatges dinàmiques i/o sons i amb connexions a altres adreces d'Internet i/o a Bancs de dades específics (tot això darrer no possible en un llibre), es podria resoldre el problema de la classificació de les Proteïnes i fins i tot possibilitar l'autoaprenentatge.

ABREVIACIONS. **DICUC**: Simulador parcial del sistema conceptual, inicialment anomenat "Diccionari Universal, CUC".

COMENTARIS ADICIONALS

Precisió lingüística i mental

Predicció de conceptes en la inexistència d'Estructura nativa primària: com no existeix l' Estructura primària aïlladament, tampoc existeix la seva homòloga nativa. Amb les natives no hi ha error perquè estem en un àmbit estructural i no caiem en el parany semàntic creat involuntàriament per Kai LINDERSTRÖM-LANG.

No és “nativa” ni “desnaturalitzada” sinó que “estructurada naturalment” i “desestructurada”.

Claredat en l'exposició dels diferents conceptes estructurals, agrupats per Nivells i relacionats entre ells. Exemples de Relacions estructurals (es podrien millorar amb la Reivindicació 23):

Estructura triple alfa hèlix – Col·lagen
 Estructura secundària alfa – Motiu alfa gir alfa
 Estructura secundària beta – Estructura terciària “Domini estructural beta”

Exemples de Versatilitat

L'Estructuració terciària té dos possibles Particions reals (Interevolutiva “P”), l'estructural (Dominis), el posicional dual (exterior ↔ interior). Respecte la segona partició

- L'exterior d'una Estructura terciària és el **Domini** dels Girs/ Enllaços
- L'interior d'una Estructura terciària té com **Propietat de la Manera de procedir** el d'una “Substància hidròfoba”.

Herència múltiple en les

Estructures natives secundària terciària i quaternària.
 Enzim amb Globulina i Catalitzador
 Relació Estructura terciària nativa - MioGlobina
 Relació Estructura quaternària nativa – HemoGlobina
 Estructures secundàries existents i no existents

Exemples de Relacions (apart de les estructurals)














Relacions “Concepte ↔ Àmbit d'Ús/ Aplicació” de les Proteïnes amb els Animals (MioGlobina) i Vertebrats (Col·lagen)

Implicacions

“Funcionalitat” ⇒ Estructura terciària i/o Quaternària
 Proteïna funcional ⇒ Globulina??
 Proteïna estructural ⇒ Proteïna fibril·lar??

Programa dissenyador d'esquemes d'Homologia

En base a la relacionabilitat, amb desenvolupament informàtic podrien **generar-se esquemes** com per exemple el d'una **xarxa homològica**, anàlogament a com un programa “3D” genera dibuixos i perspectives amb informació vectorial dels punts.


AminoÀcid		Estructura primària i secundària		Super Estructura secundària		Estructura terciària		Estructura quaternària
 ↑ ↓		 ↑ ↓		 ↑ ↓		 ↑ ↓		 ↑ ↓
Primera agrupació de Codons		Estructura primària i secundària		Enrotllament 1		Enrotllament 2		Forma del Cromosoma

Notem novament que, es **prediu altre concepte**, en base a l'analogia amb la SuperEstructura secundària (en el quadre hi ha dos enrotllaments en lloc d'un sol). Veure “SuperEstructura secundària de l'Àcid nucleic” en el DICUC




Terminacions, Temàtiques (Ciència, Fenomen, Matèria)

Plegament d'una Proteïna: relacions amb Estructuració, Estructuració primària secundària i terciària, Predicció, ...

Entrada a les Proteïnes

Com Aliment: Aliment ( Alimentació; fill de Nutrient (veure que inclou l'Aigua i els AminoÀcids essencials))


Com Component dels Éssers vius: Ésser viu (segons Successió de Complexament de la Matèria)→

- Component atòmic  BioElement
- Component molecular  Aigua i AminoÀcids essencials
- Component macromolecular/ Substància  Pròtids, Lípids, Glúcids i Vitamines

“Pròtids, Lípids i Glúcids” és incomplet, i encara que afegim les “Sals minerals”, és poc precís. Notem com un “SISTEMA CONCEPTUAL” ha separat els Minerals (Nivell atòmic, alguns dels BioElements), l'Aigua (Nivell molecular), i ha afegit els AminoÀcids essencials i les Vitamines en els Nivells que els correspon.

Proteïna ← Pèptid ← AminoÀcid

Accés a l'AminoÀcid a través del Grup amino i l'Àcid carboxílic (**Padrastrès**)

Dualitat entre AminoÀcid i Àcid nucleic. Accés a l'Àcid nucleic pels seus components (Enllaç fosfodièster (**Padrastrè**), Compost fosforat, Compost Nitrogenat, ...). Notem que no només no hi és l'Àcid carboxílic sinó que diu explícitament que no tenen res a veure ().

Transferència metodològica i Analogia Proteïna – Giny (Exemple del Flagel bacterià)

CUC 2002-08-20 (Darrera revisió 2002-10-29)