

---

**Part III**

# **Teràpia fotodinàmica, PDT**

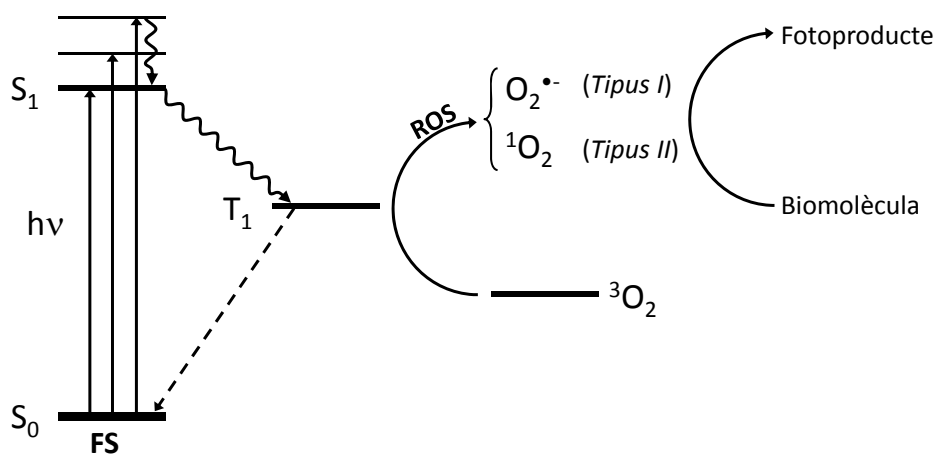
---



La teràpia fotodinàmica (*Photodynamic therapy*, PDT) aplicada al càncer es basa en la destrucció selectiva del teixit tumoral, aprofitant la fototoxicitat local produïda per la combinació de llum visible, oxigen molecular i un agent químic (anomenat fotosensibilitzador, FS).

El FS ha de ser capaç d'absorbir la llum amb què s'irradia, formar un estat excitat singlet de major energia ( $S_1$ ) i que aquest pugui passar a un estat triplet ( $T_1$ ) per encreuament intersistema. Si la diferència energètica associada a la desactivació del triplet és suficient (major de  $94 \text{ kJ}\cdot\text{mol}^{-1}$ , aproximadament uns 1400 nm), l'oxigen present en el medi en estat fonamental (triplet,  $^3\text{O}_2$ ) pot aprofitar l'energia per produir espècies radicalàries en cas que es produeixi una transferència de càrrega (mecanisme Tipus I), o per generar oxigen singlet ( $^1\text{O}_2$ ) mitjançant una transferència d'energia (mecanisme Tipus II).

Les espècies reactives de l'oxigen (*Reactive Oxygen Species*, ROS) formades tant per una via com per l'altra, reaccionen amb les biomolècules del seu entorn produint danys cel·lulars que condueixen, en últim terme, a la mort cel·lular per necrosi, apoptosi o autofàgia,<sup>114-118</sup> figura 6.9.



**Figura 6.9:** Diagrama esquemàtic del procés fisicoquímic subjacent a la PDT.

El fet d'incorporar la llum dins d'aquest esquema, afavoreix la selectivitat de la PDT vers els teixits irradiats, atorgant a aquesta teràpia un dels seus grans avantatges. Tanmateix, cal que tots tres agents es trobin simultàniament en el teixit tumoral: la **llum** forma part del procediment mèdic de la teràpia i s'aplica de forma controlada sobre la regió desitjada, mentre que l'**oxigen** molecular dissolt en l'organisme assegura la presència d'aquest factor en el teixit tumoral. El **fotosensibilitzador**, però,

---

és capaç d'aconseguir arribar en suficient quantitat a la regió neoplàstica i desenvolupar un cop allà el seu rol?

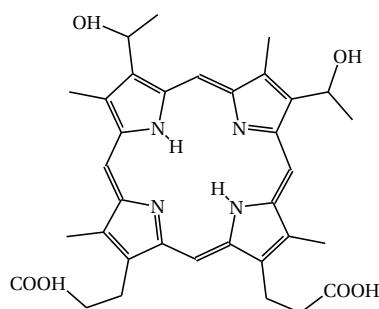
La naturalesa del FS és determinant i en ella recau una gran responsabilitat: per un costat ha de ser capaç d'absorbir la llum d'aquella longitud d'ona en la qual els teixits són més transparents (*i.e.* entre 600 i 900 nm), ja que serà la que podrà penetrar més sota la pell, i de prou energia com per assegurar la transferència d'energia a l'oxigen (*quenching*) i, per tant, la generació de ROS. D'altra banda, ha de garantir una bioacumulació preferent en els teixits tumorals i, dins d'aquests, localitzar-se en aquelles estructures cel·lulars on el dany fotoquímic pugui ser major, per tal de desencadenar la mort cel·lular de manera més eficient.

Conèixer les necessitats que han de satisfer els FS per poder ser emprats en la PDT obre la porta al disseny de noves molècules. La tercera, i última part del treball, agrupa els estudis realitzats en el camp del disseny de FS.

## Capítol 7

# Estudi de les propietats fisicoquímiques dels fotosensibilitzadors

El desenvolupament de la PDT es troba íntimament lligat a l'evolució dels FS emprats. Els primers FS que s'utilitzaren en PDT, anomenats de primera generació, foren principalment derivats de l'hematoporfirina (Hp, **2**). L'efecte fotosensibilitzador de derivats porfirínics, així com d'altres tetrapirroles, ja es trobava documentat en estudis fotobiològics de principis del segle XX que posaven de manifest l'efecte sensibilitzador d'aquests compostos sobre els éssers vius,<sup>119</sup> podent-se identificar com els responsables de síndromes relacionades amb la fotosensibilitat, com per exemple la porfíria.<sup>115</sup> La utilització d'aquests primers fotosensibilitzadors permeté demostrar la utilitat i viabilitat de la teràpia, fet que culminà el 1998 amb el reconeixement per part de la FDA (*Food and Drug Administration*) d'una mescla de derivats hematoporfirínics (*Photofrin*) com a primer fàrmac acceptat per a l'ús en PDT.<sup>118</sup>



Hematoporfirina, Hp

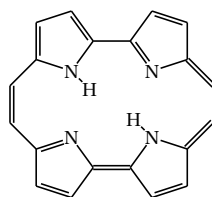
**2**

Ja des dels inicis (Policard, 1924<sup>120</sup>) s'observà que els derivats de l'Hp presentaven una certa afinitat pels teixits tumorals,<sup>121, 122</sup> no essent així per l'Hp en forma aïllada. El tractament químic proposat com a alternativa per Schwartz i col·laboradors permeté millorar aquesta propietat, donant lloc a una mescla de composició indeterminada que revé el nom de derivat HpD (*Hematoporphyrin Derivative*).<sup>119</sup> L'HpD es convertí en l'objecte d'estudi, esdevenint la identificació de la seva composició química un dels punts més debatuts per la comunitat científica. Els avenços realitzats per diferents grups

de recerca<sup>123-125</sup> portaren a Dougherty i col·laboradors a la purificació parcial de l'HpD, separant els monòmers presents en la formulació, donant lloc al *Photofrin*.

La incògnita entorn la composició química dels derivats de l'HpD suposava haver d'estudiar una mescla de compostos amb propietats molt diferents. Aquesta diversitat estructural afectava tant a les propietats fisicoquímiques necessàries per a la generació d'espècies reactives (degut a la baixa intensitat d'absorció i la pobra capacitat de formar oxigen singlet) com a la pròpia biodistribució.<sup>126</sup> Aquests fets es traduïen en la necessitat d'eleva les dosis administrades tant de FS com de llum per aconseguir efecte terapèutic,<sup>115</sup> amb el consegüent risc de desencadenar efectes secundaris, relacionats principalment amb la fotosensibilitat cutània.<sup>127</sup> Per aquest motiu, s'inicià la cerca de nous FS (que reberen el nom de FS de segona generació) amb una composició química ben definida i amb unes propietats fisicoquímiques d'interès.

La identificació de les característiques que han de millorar aquests nous FS per esdevenir bons candidats per ser emprats en PDT, obre la porta al seu disseny. Tot i la tendència inicial de mantenir el *Photofrin* com a exemple a seguir,<sup>128</sup> la cerca de nous FS suposà el trencament de l'hegemonia de les porfirines a favor dels seus isòmers estructurals, tals com les clorines, les ftalocianines i els porficens. La reducció de la simetria dels porficens ha demostrat atorgar-los una millora de les propietats espectrals respecte a les porfirines,<sup>129, 130</sup> fet que els fa ser d'especial interès i motiu pel qual han centrat gran part de la recerca al GEM.<sup>118, 131, 132</sup>



Porficè, Po

3

Entre les característiques que es consideren òptimes per a un FS es troben tant propietats químiques (estabilitat, absència de toxicitat en la foscor) com fotofísiques (espectre d'absorció, formació d' $^1O_2$ , entre d'altres). Aquest últim tipus de característiques han esdevingut especialment rellevants des del punt de vista del disseny dels FS, ja que la predicció de l'espectre d'absorció de molècules orgàniques és quelcom resolt en l'àmbit de la química computacional i s'ha consolidat com un dels primers criteris a considerar en el cribratge de possibles candidats. L'interès en aquest cas recau en l'estudi del desplaçament del màxim d'absorció a la regió del vermell, per tal de poder avaluar el compromís entre una bona profunditat d'absorció i presentar una diferència energètica suficient entre els estat fonamental i excitat del FS per assegurar la formació d'oxigen singlet.<sup>115, 133</sup>

Aquesta visió de la PDT constitueix, però, una abstracció que redueix l'eficàcia de la teràpia al disseny estructural del FS, sense tenir en compte el context biològic que l'envolta, com ara la seva biodistribució. Aquest fet repercuteix principalment sobre la selectivitat del FS, ja que ha d'arribar fins al teixit on s'aplica la font de llum de forma local. En els esforços realitzats per incloure la informació bioquímica dins del disseny de FS i millorar la seva selectivitat, hom troba que el mecanisme de retenció dels FS en els tumors no està ben definit, encara que sembla que el balanç hidrofòbic pot

jugar un paper important.<sup>134</sup> D'altra banda, l'estudi de les propietats diferencials de les cèl·lules dels teixits neoplàstics porta a l'aparició d'estratègies més elaborades on els FS es combinen amb portadors exògens que els condueixen de forma preferent a una localització desitjada.<sup>135-139</sup> A més, l'efecte local la PDT es veu accentuat pel marge de difusió de les ROS formades durant l'etapa d'irradiació. El desplaçament de les espècies oxidants no supera els 40 nm,<sup>140</sup> fent que l'eficàcia del procés global pugui veure's influït fins i tot per la localització intracel·lular del FS.

### 7.1 Estudi de les propietats dels fotosensibilitzadors al GEM

Complementàriament a la seva faceta predictiva, la química computacional constitueix una eina molt valuosa en l'estudi de la naturalesa fisicoquímica de la matèria, ja que aporta el marc teòric necessari per a la descripció de sistemes moleculars complets de forma virtual. Alguns mètodes poden ser emprats, doncs, per a l'estudi de l'estructura i la disposició espacial de sistemes complexos en diferents entorns. En aquest sentit, l'any 1996 s'estudià per primer cop la conformació de derivats nitrobenzil-mesoporfíric al Laboratori de Disseny Molecular del GEM.<sup>141</sup> Després de l'optimització de la geometria de parelles de compostos amb un mètode semiempíric AM1, s'introduí a la simulació el dissolvent corresponent. L'evolució en el temps del conjunt per dinàmica molecular permeté estudiar la conformació en dissolució dels compostos, les interaccions  $\pi - \pi$  i comparar els resultats computacionals amb les dades espectroscòpiques d'H-RMN i de transferència electrònica fotoinduïda.

Més endavant, la recerca de nous FS sintètics de 2a generació portà al GEM a prendre els porfírics com a compostos objectiu. Aquesta decisió redireccionà la tasca computacional a aquest tipus de compostos, predient diferents propietats tant fotoquímiques com fotofísiques de derivats porfírics i azaporfírics amb fenils en les posicions 2, 7, 12 i 17.<sup>130</sup>

En tractar-se de compostos altament conjugats, és fonamental utilitzar mètodes que permetin descriure correctament la correlació electrònica.<sup>142</sup> La primera aproximació adoptada per aquesta finalitat fou l'ús del formulisme DFT implementat dins del programari Gaussian.<sup>143</sup> Aprofitant el coneixement de l'estructura del Po per difracció de raig X, es validà l'ús del funcional B3LYP/6-31G\*\* com a metodologia per a l'estudi de l'estructura i optimització de la geometria del Po, pres com a compost de referència. Donat l'elevat cost computacional, la garantia d'un mínim com a resultat de l'optimització no es podia realitzar mitjançant el càlcul de freqüències, pel qual motiu s'utilitzaren els valors experimentals de difracció de raigs X.<sup>132</sup> La limitació de la potència de càlcul d'aleshores també va influir en la descripció dels derivats del Po mitjançant B3LYP: sistemes com el TPPo, només podien ésser estudiats amb aquest formulisme reduint la base a una 3-21G.

Havent comprovat l'adequació del formulisme DFT per a la descripció de l'estat fonamental d'aquests compostos, l'estudi dels estats excitats es realitzà mitjançant el mètode TD-DFT. Pel mateix motiu que en el cas anterior, les transicions electròniques dels porfírics substituïts van ser estudiades amb TD-B3LYP/3-21G\*, mentre que els seus respectius *cores* sense substituir (sent aplicable també pels aza-derivats) ho van poder ser a un nivell de descripció superior.<sup>131, 144</sup>

Aquests resultats van suposar la constitució d'un estàndard per a l'estudi energètic dels estats fo-

namental i excitats de derivats porfírics, podent-se aplicar a un ampli ventall d'estudis, no només a la determinació de l'estructura molecular. Un d'aquests estudis fou la descripció de l'estabilitat relativa entre els diferents derivats en solució, a partir de la predicció de l'energia de solvatació de cadascun d'ells mitjançant el mètode B3LYP/6-31G\*\*, acoblat a la definició proposada pel mètode PCM pel tractament del solvent. Tot i que aquesta metodologia permetria *a priori* determinar el logP dels porfírics, la capacitat de càlcul esdevingué de nou una limitació. Per aquest motiu, la predicció del logP es realitzà mitjançant el mòdul de càlcul GAMESOL,<sup>145</sup> incorporat dins el programari GAMESS,<sup>146</sup> basat en un mètode semiempíric AM1.<sup>132</sup>

Els càlculs mitjançant DFT no són els únics que permeten descriure propietats interessants dels FS, sobretot quan es desitja estudiar el comportament d'un conjunt de molècules. En aquest sentit, es va voler estudiar l'establiment d'interaccions entre els sistemes d'interès, avaluant-ne la tendència d'autoagregació i la interacció dels FS amb certes proteïnes susceptibles de patir dany fotodinàmic (com per exemple la tubulina).<sup>132</sup> Aquests estudis es realitzaren mitjançant dinàmica molecular, amb el programari AMBER,<sup>147</sup> observant l'evolució temporal del sistema molecular en un entorn on el solvent es tractava de forma explícita.

El consens entre la precisió del resultat i el temps de càlcul ha permès millorar fins al moment la descripció dels estats excitats de qualsevol FS tetrapirròlic a nivell B3LYP/6-31G\*. Tot i així, el temps efectiu continua essent massa elevat front als avanços realitzats en l'estratègia sintètica, la qual permet obtenir actualment la molècula desitjada amb un temps menor.<sup>148</sup> Per tal de poder continuar guiant la selecció dels compostos a sintetitzar, cal trobar una metodologia computacional que permeti prediccions en escales de temps comparables a l'experimental. Això suposa millorar tant els equips de computació (amb la integració de nous i millors maquinaris que permetin reduir el temps total de càlcul) com els algorismes de predicció.

En aquest sentit, els objectius actuals del Laboratori de Disseny Molecular del GEM consideren:

1. **la vessant fotofísica**, amb la definició dels estats fonamentals i excitats dels FS, la determinació del màxim de l'espectre d'absorció i la seva capacitat de presentar absorcions bifotòniques,
2. **la vessant química**, amb l'anàlisi i comparació dels mètodes actualment disponibles per a la predicció de les propietats hidrofòbiques i la seva aplicació sobre cadascun dels candidats,
3. i finalment, **la vessant bioquímica**, proposant un model de localització subcel·lular.

## 7.2 Model QSAR de predicció del màxim d'absorció de fotosensibilitzadors tetrapirròlics

Tots els FS emprats en la PDT han de presentar, com a condició necessària, una absorció en el marge del vermell per poder desenvolupar la seva funció (ja que aquesta és la que penetra més en els teixits). Per aquest motiu, la predicció del màxim d'absorció dels FS és d'especial interès, podent d'aquesta manera discernir entre aquells compostos capaços de provocar l'efecte fotodinàmic dels que no.



## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

La interacció entre la llum i la matèria pot ser simulada matemàticament mitjançant diferents metodologies. Es considera que la mecànica quàntica aporta el marc teòric més realista per aproximar aquest procés (apartat 1, pàg.35). Les metodologies derivades dels mètodes *ab initio* i DFT han estat àmpliament utilitzades en la bibliografia per a la predicció de l'espectre d'absorció de derivats porfirínics.<sup>115,149</sup> En els darrers anys, la metodologia TD-DFT s'ha consolidat com un dels mètodes més emprats per a la modelització d'estats excitats, demostrant bons resultats en diferents famílies de compostos.<sup>150,151</sup> Sigui quin sigui el mètode aplicat, la predicció de les energies d'absorció de grans sistemes moleculars continua sent un repte<sup>152</sup> i hom està obligat a recórrer a mètodes més acurats, però més costosos (tals com CIS,<sup>153,154</sup> CASSCF,<sup>133,155</sup> CASPT2<sup>15,87</sup>), per tal d'obtenir resultats raonables.

Amb la implementació de les ANN com a base per a l'establiment de models QSPR i QSAR de predicció, dins del programari ArIS (capítol 4, pàg.81), s'estudia la possibilitat d'emprar aquesta eina predictiva com a alternativa als mètodes disponibles actualment per a la realització d'aquests tipus de càlculs (TD-DFT), en la cerca de mètodes ràpids i amb una baixa demanda computacional.

A finals de la dècada dels vuitanta (s. XX), quan la informàtica tot just començava a donar-se a conèixer als usuaris particulars, la potència de càlcul distava molt de l'actual i suposava una gran limitació per a l'estudi de molècules amb un nombre considerable d'àtoms. Per aquest motiu, l'única manera de poder estudiar completament sistemes de la mida dels FS de segona generació, era mitjançant mètodes simplificats i computacionalment barats, com els mètodes semiempírics. Aquest s'han usat tant per a l'optimització de la geometria (PM3, PM5<sup>156</sup>) com per a la corresponent descripció dels estats excitats (MINDO3, INDO,<sup>157</sup> ZINDO/S<sup>158</sup>) de sistemes tals com porfirines, clorines, bacterioclорines, etc. Actualment, tot i que s'han vist superats pels mètodes derivats del DFT o CI,<sup>133</sup> aquests mètodes continuen sent atractius per a l'estudi dels FS en solvatació.<sup>159</sup> Així doncs, si els mètodes semiempírics han demostrat ser un punt de partida acceptable per nombroses aplicacions, podrien ser-ho també per a un model QSPR?

El problema de fons pot ser tractat en quimiinformàtica com la necessitat de predir una propietat (el màxim d'absorció,  $\lambda_{max}$ ) a partir de característiques estructurals de les molècules, i per tant, simplificar-lo a un model QSPR. Per aquesta finalitat es compara l'aplicació del mètode PLS amb les ANN, com a representants dels mètodes lineals i no lineals respectivament. Encara que ambdós mètodes s'han emprat àmpliament per resoldre problemes en química mèdica,<sup>160</sup> tan sols el PLS ha estat prèviament descrit a la bibliografia per a la predicció del màxim d'absorció de colorants (concretament, derivats d'azobenzens).<sup>161</sup> Per la seva part, les ANN han estat aplicades per a la predicció de l'activitat fotodinàmica de fotosensibilitzadors<sup>162</sup> i de les intensitats d'absorció de molècules orgàniques en l'UV.<sup>163</sup>

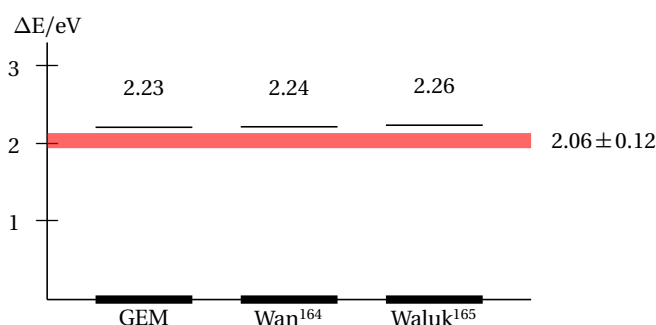
Donat que no es disposa d'informació bibliogràfica referent a l'habilitat dels models QSPR per la predicció de  $\lambda_{max}$ , cal reflexionar sobre el criteri que se seguirà per considerar que un model proposat és apte. Per poder valorar aquest fet cal tenir una idea de l'error que hom està disposat a assumir, i per això tan sols es pot prendre de referència l'error que s'està cometent amb l'actual estàndard TD-DFT implantat al GEM.

### 7.2.1 Avaluació de l'error assumible

En el grup de recerca s'ha anat proposant, al llarg dels anys, diferents metodologies per predir de forma suficientment satisfactòria el màxim d'absorció de FS porfírics. L'evolució dels programes i maquinari disponibles han permès establir un procediment estàndard basat en el formulisme TD-DFT. Tot i així, el temps de càlcul continua essent un factor limitant, el qual obliga a arribar a un consens entre la bondat del mètode i el temps real de càlcul. Per aquest motiu, el mètode estàndard pel càlcul de l'energia dels estats excitats de FS porfírics comença per una optimització de la geometria molecular mitjançant el mètode DFT B3LYP/6-31G\*, seguit d'una anàlisi de freqüències al mateix nivell de càlcul per tal d'assegurar la convergència a un mínim en la superfície d'energia potencial i poder estimar el valor de l'energia en el punt zero. Sobre el sistema molecular obtingut de la primera etapa s'aplica un mètode TD-DFT, mantenint el nivell de càlcul dels casos anteriors, per tal d'obtenir el valor de l'energia del primer estat excitat singlet. Tots els càlculs s'han realitzat utilitzant el programari Gaussian03<sup>143</sup> en un ordinador equipat amb un processador Intel® Core™2 Quad (Q9550 a 2.83 GHz) i 2 GB de memòria RAM.

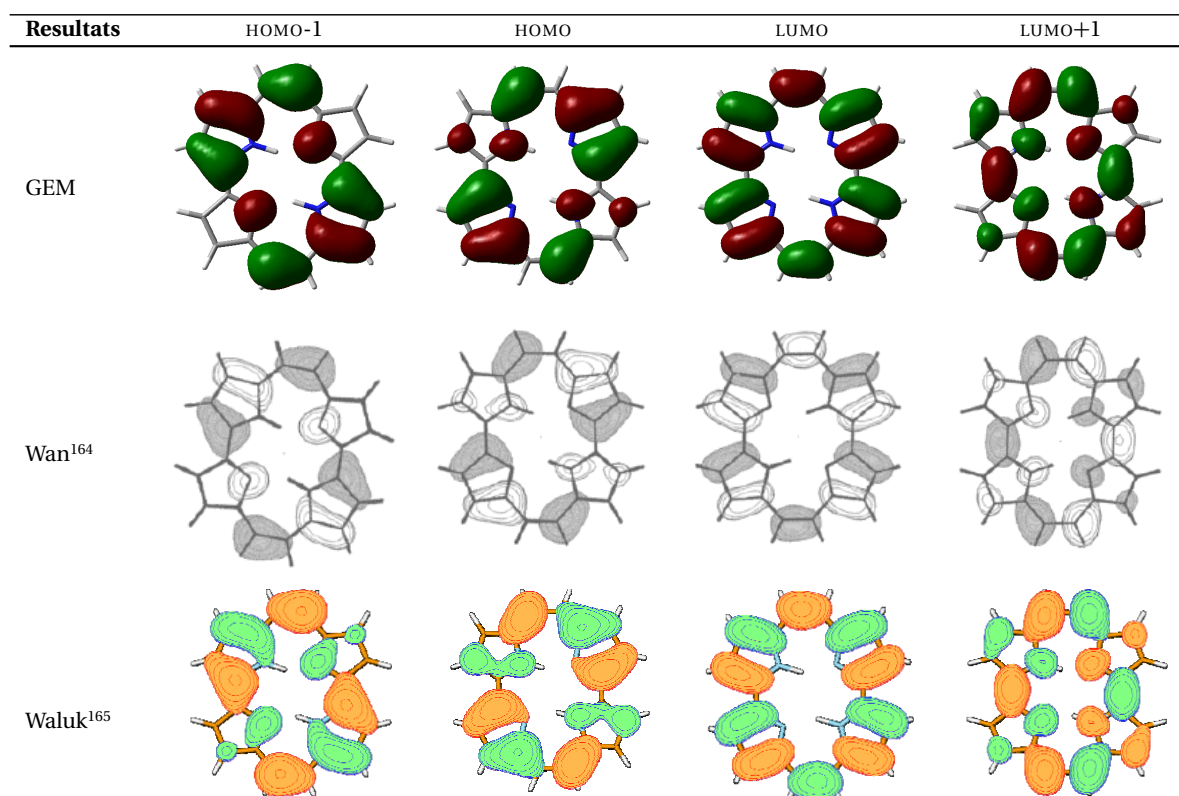
Aquest procediment s'ha validat amb la molècula de porfíric, amb el qual s'han obtingut resultats comparables amb els recollits a la bibliografia, tant pel que fa a la descripció de l'estat fonamental (bàsicament la seva geometria molecular) com els estats excitats. De forma més teòrica, també se n'avalua l'energia del primer estat singlet excitat (figura 7.1) i la distribució dels orbitals (figura 7.2), en la qual es pren el valor corresponent a l'estat fonamental com el zero d'energia.

Tal com es pot observar en les figures 7.1 i 7.2, els resultats obtinguts amb aquest nivell de càlcul concorden amb els estàndards bibliogràfics.<sup>164,165</sup> Així doncs, es considera que la metodologia proposada, basada en el mètode TD-DFT, permet estudiar satisfactòriament l'estat fonamental de sistemes porfírics i predir l'energia del seu primer estat excitat, per així conèixer la longitud d'ona corresponent al màxim d'absorció.



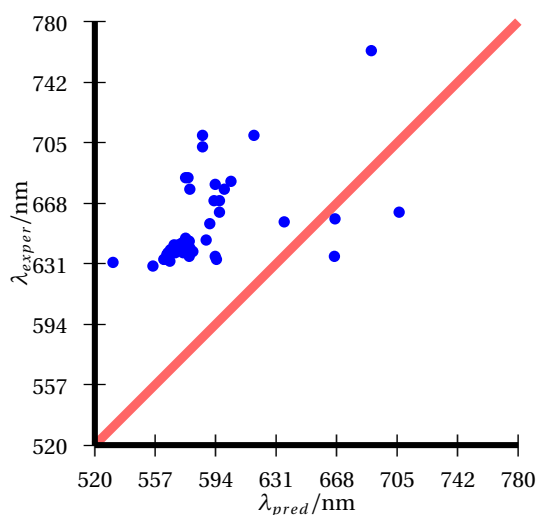
**Figura 7.1:** Diferència energètica entre l'estat fonamental del Po i el seu primer estat singlet excitat. Els resultats computacionals es comparen amb el resultat experimental (en vermell).

## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS



**Figura 7.2:** Comparació dels resultats obtinguts en la descripció dels orbitals del Po segons la metodologia DFT descrita al GEM i estudis bibliogràfics.

Aquesta metodologia s'aplica sobre 44 FS amb  $\lambda_{max}$  coneguda (la identitat dels quals es detallen al següent apartat), obtenint un error mig d'aproximadament **70 nm** i un coeficient de correlació  $R^2 = 0.251$  (figura 7.3). Degut a la presència de FS asimètrics, es comprova que les discrepàncies amb els valors experimentals no vinguin donades per haver calculat  $\lambda_{max}$  amb l'estructura del tautòmer que no correspon. Així doncs, es prenen els valors de l'error mig i  $R^2$  anteriors com la referència per a la discussió dels resultats obtinguts mitjançant els models basats en ANN que es proposen més endavant.



**Figura 7.3:** Representació gràfica dels valors de  $\lambda_{max}$  calculada a partir del formalisme TD-DFT ( $\lambda_{pred}$ ) en comparació amb els resultats experimentals ( $\lambda_{exper}$ ).

## 7.2.2 Creació d'una quimioteca virtual pel mètode QSPR

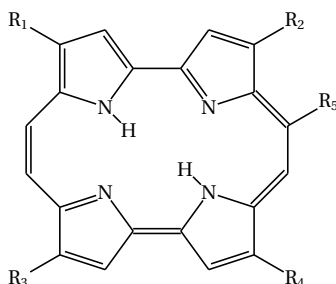
Donat que l'interès del GEM en els FS per ser emprats en PDT recau en els porfírcens, s'ha realitzat una recopilació de mesures experimentals referents a l'absorció en l'UV/Vis de 44 molècules d'aquest tipus (taules 7.1 i 7.2), tenint com origen tant dades experimentals facilitades per membres del propi grup de recerca com dades bibliogràfiques.<sup>166, 167</sup>

Els valors de  $\lambda_{max}$  esperats bibliogràficament i experimental del conjunt global de les dades considerades es troben compresos entre 600 nm i 800 nm, amb una mitjana situada en els 658 nm i una desviació estàndard de 32.1 nm. Això suposa que el 90% de les dades es troben entre els 600 nm i els 700 nm, fet que cal tenir en compte a l'hora de realitzar la discussió dels resultats.

Un 60% de les dades disponibles (27 porfírcens) se selecciona de forma aleatòria i s'empra com a conjunt d'entrenament, mentre que els 17 compostos restants constitueixen el conjunt de prova (*test set*) i són emprats per efectuar una validació externa del model. La identitat dels membres d'ambdós grups es manté constant en tots els models QSPR basats en ANN.

## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

**Taula 7.1:** Derivats porficènics considerats en la quimioteca virtual emprada en l'establiment del mètode QSPR (1a part).



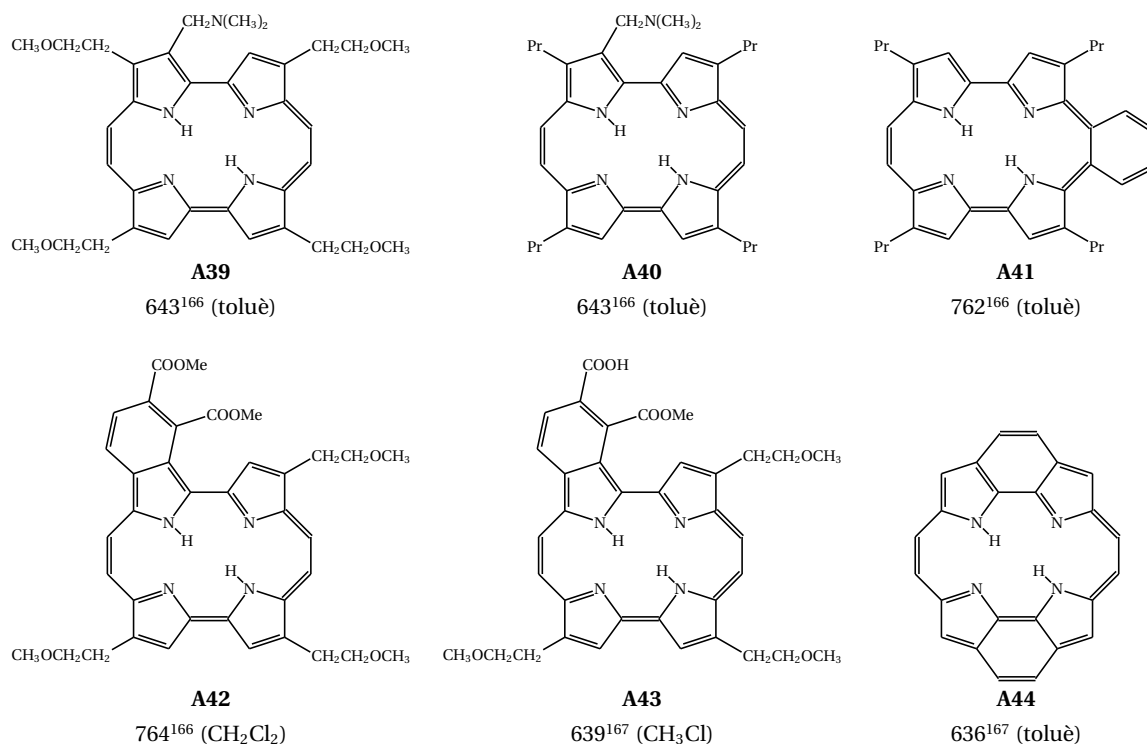
id.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	$\lambda_{max}/nm$ *
<b>A1</b>	Pr	Pr	Pr	Pr	H	640 <sup>168</sup>
<b>A2</b>	H	H	H	H	H	632, <sup>134</sup> 634 <sup>167</sup> ▲
<b>A3</b>	Pr	Pr	Pr	Pr	NHCOCH <sub>3</sub>	639 <sup>166</sup>
<b>A4</b>	Pr	Pr	Pr	Pr	NHCOCH <sub>2</sub> CH <sub>2</sub> COOH	636 <sup>166</sup> ▲
<b>A5</b>	Pr	Pr	Pr	Pr	H	633 <sup>167</sup>
<b>A6</b>	CH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	H	637, <sup>166</sup> 635 <sup>167</sup> ▲
<b>A7</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	OCH <sub>3</sub>	684 <sup>166</sup>
<b>A8</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	OC <sub>6</sub> H <sub>13</sub>	684 <sup>166</sup>
<b>A9</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH=CH <sub>2</sub>	H	643 <sup>167</sup> ▲
<b>A10</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	OCOC <sub>5</sub> H <sub>11</sub>	646 <sup>166</sup>
<b>A11</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	OCOC <sub>8</sub> H <sub>17</sub>	647 <sup>166</sup>
<b>A12</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	OCOCH <sub>2</sub> NHCOOC(CH <sub>3</sub> ) <sub>3</sub>	644 <sup>166</sup>
<b>A13</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	NH <sub>2</sub>	710 <sup>166</sup>
<b>A14</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	NHCOCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOMe	670 <sup>166</sup>
<b>A15</b>	Pr	Pr	Pr	Pr	OH	677 <sup>166</sup>
<b>A16</b>	Pr	Pr	CH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub> CH <sub>2</sub> OH	H	638 <sup>166</sup>
<b>A17</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	Pr	Pr	Pr	H	637 <sup>166</sup>
<b>A18</b>	CH <sub>2</sub> CH <sub>2</sub> OPh	CH <sub>2</sub> CH <sub>2</sub> OPh	CH <sub>2</sub> CH <sub>2</sub> OPh	CH <sub>2</sub> CH <sub>2</sub> OPh	H	638 <sup>166</sup>
<b>A19</b>	Pr	Pr	Pr	Pr	OCH <sub>3</sub>	682 <sup>166</sup>
<b>A20</b>	Pr	Pr	Pr	Pr	OCOCH <sub>3</sub>	641 <sup>167</sup> ▲, 636 <sup>166</sup>
<b>A21</b>	Pr	Pr	Pr	Pr	OCOCH <sub>2</sub> NHCOOC(CH <sub>3</sub> ) <sub>3</sub>	644 <sup>166</sup>
<b>A22</b>	Pr	Pr	Pr	Pr	OCOCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> NHCOOC(CH <sub>3</sub> ) <sub>3</sub>	645 <sup>166</sup>
<b>A23</b>	Pr	Pr	Pr	Pr	NH <sub>2</sub>	710 <sup>166</sup>
<b>A24</b>	Pr	Pr	Pr	Pr	NHCOCH <sub>2</sub> CH <sub>2</sub> COOMe	634 <sup>166</sup>
<b>A25</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	OH	680 <sup>166</sup>
<b>A26</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	OHCOCH <sub>3</sub>	638 <sup>166</sup>
<b>A27</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	NCOCH <sub>3</sub>	670 <sup>166</sup>
<b>A28</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	N-succinimida	640 <sup>166</sup>
<b>A29</b>	CH <sub>2</sub> CH <sub>2</sub> OH	Pr	Pr	Pr	H	637 <sup>166</sup>
<b>A30</b>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	Pr	Pr	H	638 <sup>166</sup>
<b>A31</b>	Pr	Pr	Pr	Pr	OCH <sub>2</sub> C <sub>6</sub> H <sub>4</sub> COOH	677 <sup>166</sup>
<b>A32</b>	Pr	Pr	Pr	Pr	OCOCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOC(CH <sub>3</sub> ) <sub>3</sub>	645 <sup>166</sup>
<b>A33</b>	H	H	H	H	NO <sub>2</sub>	630 ▲
<b>A34</b>	Ph	Ph	Ph	Ph	NO <sub>2</sub>	657 ▲
<b>A35</b>	Ph	Ph	Ph	Ph	OCOCH <sub>3</sub>	656 ▲
<b>A36</b>	Ph	Ph	Ph	Ph	OCOCH <sub>3</sub>	659, 656 ▲
<b>A37</b>	Ph(m-OH)	Ph(m-OH)	Ph(m-OH)	Ph(m-iPrO)	H	663 ♦
<b>A38</b>	Ph(m-OH)	Ph(m-OH)	Ph(m-OH)	Ph(m-OH)	H	663 ♦

\* totes les mesures en dissolució de toluè, tret de les excepcions que s'indiquen

▲ mesura en dissolució de CH<sub>2</sub>Cl<sub>2</sub>

♦ mesura en dissolució de CHCl<sub>3</sub>

**Taula 7.2:** Derivats porfíricens considerats en la quimioteca virtual emprada en l'establiment del mètode QSPR (2a part). S'indica sota el número de compost el valor de  $\lambda_{max}$ .



### 7.2.3 Descriptors moleculars

Amb la poca informació bibliogràfica disponible per a l'aplicació de models QSPR en la predicció de l'espectre d'absorció (no tan sols de fotosensibilitzadors sinó de molècules orgàniques en general), la definició matemàtica del problema no és trivial. Tot i adaptar els treballs bibliogràfics existents,<sup>163</sup> s'amplia el ventall de descriptors atenent a la intuïció química. Donada la naturalesa del problema a tractar, s'inclouen tant aproximacions del valor de  $\lambda_{max}$  com descriptors directament relacionats amb l'estructura electrònica dels FS (taula 7.3).

## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

**Taula 7.3:** Descriptors moleculars proposats en l'establiment del model QSPR de predicció de  $\lambda_{max}$ .

Nom del descriptor	Descripció
$\lambda_{ZINDO}$	Màxim d'absorció calculat pel mètode ZINDO/s sobre l'estructura molecular optimitzada mitjançant el mètode semiempíric PM3
$n_{DB}$	Nombre de dobles enllaços presents en l'estructura molecular del fotosensibilitzador
PMI	Moment principal d'inèrcia, calculat sobre la geometria molecular optimitzada a nivell PM3
$\mu$	Moment dipolar calculat a partir de l'optimització de la geometria molecular mitjançant el mètode semiempíric PM3
$E_{HOMO}$	Estimació de l'energia (en eV) de l'últim orbital ocupat (HOMO) calculat a nivell PM3 <sup>169</sup>
$E_{LUMO}$	Estimació de l'energia (en eV) del primer orbital desocupat (LUMO) calculat a nivell PM3 <sup>169</sup>
VSA	Àrea de la superfície de Van der Waals calculada a partir de contribucions atòmiques
apol	Suma de les polaritzabilitats atòmiques <sup>170</sup>
SMR	Refractivitat molar. Combina la mida del sistema molecular amb la seva polaritzabilitat <sup>171</sup>

Bo i que els descriptors  $E_{HOMO}$  i  $E_{LUMO}$  es troben descrits en la bibliografia,<sup>163</sup> hom esperaria que la informació rellevant fos la seva diferència, ja que podria ser utilitzada com una estimació del valor de  $\lambda_{max}$  segons el mètode PM3. Les proves preliminars realitzades substituint els dos descriptors per la seva diferència no han mostrat bons resultats, i fan descartar aquesta hipòtesi (tot i no entrar en detall, s'adjunta un gràfic amb els resultats obtinguts en la figura 7.8, pàg. 7.8).

Finalment, com que els càlculs amb TD-DFT no tenen en compte l'efecte del solvent, s'ha descartat la incorporació d'un descriptor que ho fes, per tal que els resultats obtinguts amb TD-DFT i ANN siguin comparables.

### Estudi de la rellevància dels descriptors

Una vegada s'ha determinat la manera com es defineix el problema, s'estudia la influència de cadascun dels descriptors considerats sobre la predicció del valor de  $\lambda_{max}$ .

Inicialment es considerarà que els descriptors PMI,  $\mu$ ,  $E_{HOMO}$ ,  $E_{LUMO}$  i VSA, donada la informació química que contenen, podien ser bons candidats per a l'establiment del model QSPR de predicció. Malauradament, tots els models generats amb ells (mitjançant ANN i provant diferents topologies) no mostren la capacitat de predicció desitjada, posant de manifest la conveniència de seguir un millor criteri de selecció dels descriptors. Tot i així, els resultats obtinguts amb aquest primer conjunt serveixen per establir el marge del *learning rate* ( $\eta$ ) necessari per assegurar la convergència de l'entrenament amb ANN ( $10^{-5} > \eta > 10^{-6}$ ) i per fer notar la necessitat d'emprar una sola capa amagada per evitar l'*overfitting*.

La identificació i assignació de la importància de cadascun dels nou descriptors en la definició del valor de  $\lambda_{max}$  es realitza d'acord a dos criteris:

- Segons la influència de cadascun dels descriptor en les components principals que permeten descriure un major percentatge de variància explicada.

- El coeficient de correlació que mostra cada descriptor en relació al valor de  $\lambda_{max}$ .

Aquestes anàlisis permeten establir una ordenació dels descriptors segons la seva rellevància. Com sol passar, a l'hora de comparar diferents tècniques estadístiques, els resultats obtinguts posen de manifest una certa disparitat entre criteris (taula 7.4). Hom esperaria que  $\lambda_{ZINDO}$  fos el descriptor amb un major pes en la descripció del màxim d'absorció, ja que constitueix una mesura aproximada d'aquest valor. L'ordenació segons el coeficient de correlació recull aquesta idea, on l'aproximació anterior ocupa el primer lloc en importància, i contrasta amb el quart lloc que li atorga el criteri basat en PCA.

**Taula 7.4:** Llistat dels descriptors considerats per a la predicció de  $\lambda_{max}$ , ordenats de major a menor rellevància segons els criteris considerats.

PCA	Coef. Correlació
PMI	$\lambda_{ZINDO}$
VSA	$E_{LUMO}$
apol	PMI
$\lambda_{ZINDO}$	VSA
SMR	$\mu$
$n_{DB}$	SMR
$\mu$	apol
$E_{HOMO}$	$E_{HOMO}$
$E_{LUMO}$	$n_{DB}$

Per a cada criteri d'ordenació s'estudia l'efecte de la incorporació successiva d'aquells descriptors amb menor influència, dins del model QSPR de predicció basat en ANN. En no conèixer els paràmetres òptims per a l'entrenament de les ANN, per a cada subconjunt de descriptors es proven diferents topologies, mantenint el nombre de neurones amagades entre 2 i 8 (valors orientatius determinats a partir de l'índex de Kolmogorov<sup>52</sup> i el test de  $\rho$ <sup>74</sup>). A més, en l'intent d'incorporar la influència de l'error inherent en les mesures experimentals i augmentar així la capacitat de generalització del model,<sup>79</sup> s'incorporen al conjunt d'entrenament còpies d'ell mateix on els valors de  $\lambda_{max}$  es modifiquen aleatòriament fins a un 5%. En cas que diferents fonts bibliogràfiques indiquin diferents valors de  $\lambda_{max}$  per a un mateix FS, tots són incorporats a la quimioteca com entrades independents.

Les figures 7.4 i 7.5 (pàg.159) mostren l'evolució de l'error quadràtic mig (RMSE) i el coeficient de correlació  $R^2$  a mesura que s'incorporen successivament els descriptors amb menor rellevància segons l'assignació realitzada pel coeficient de correlació entre els descriptors i els valors experimentals. Els resultats anàlegs als anteriors, referents a l'ordenació dels descriptors realitzada segons l'anàlisi de components principals, es mostren en les figures 7.6 i 7.7 de la pàgina 160. En tots els casos es representa el conjunt d'entrenament en blau i el conjunt de validació externa en verd.

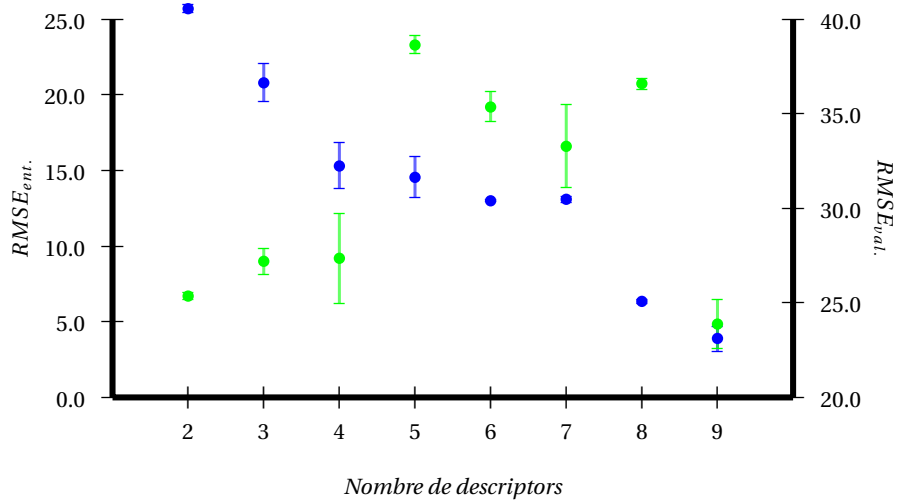
La tendència que marquen els conjunts d'entrenament i de validació en ambdós criteris de selecció de descriptors, permeten identificar la topologia estadísticament més favorable.

L'anàlisi dels resultats obtinguts segons l'ordenació proposada pel coeficient de correlació, no permeten extreure cap conclusió directa del nombre òptim de descriptors a utilitzar: la tendència creixent d' $R^2$  en el conjunt d'entrenament, acompanyada de la seva prematura estabilització en el

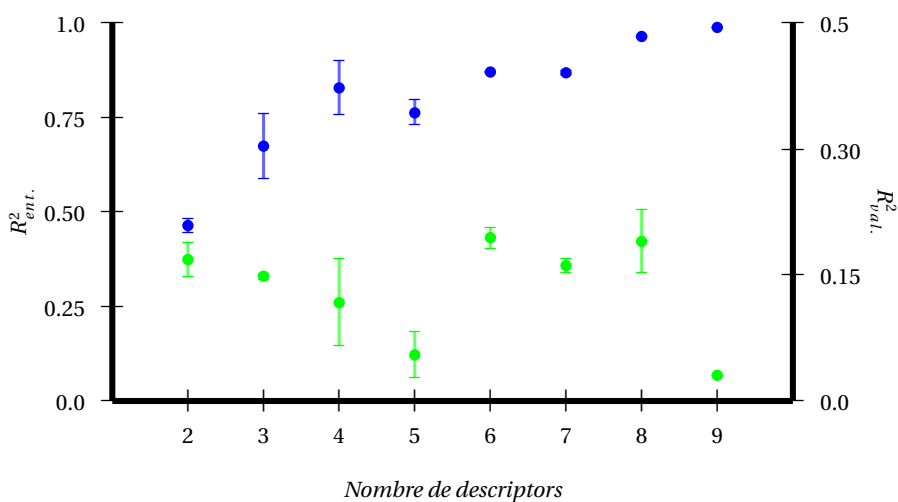


7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

(a)

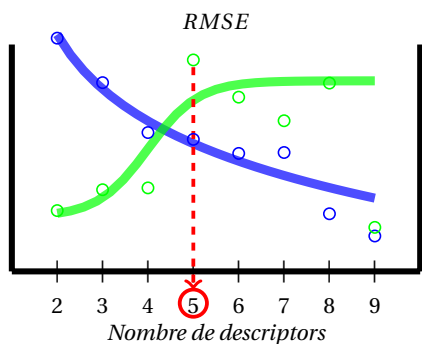


(b)

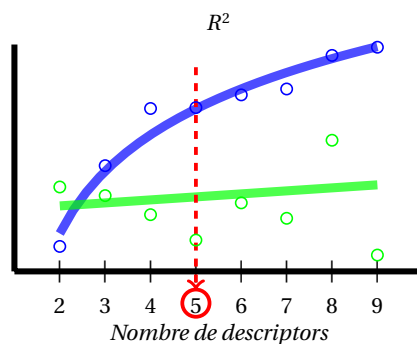


**Figura 7.4:** Evolució de l'RMSE (a) i del coeficient de correlació  $R^2$  (b) en funció del nombre de descriptors considerat segons el criteri d'ordenació basat en el coeficient de correlació. [● conjunt d'entrenament, ● conjunt de validació]

(a)

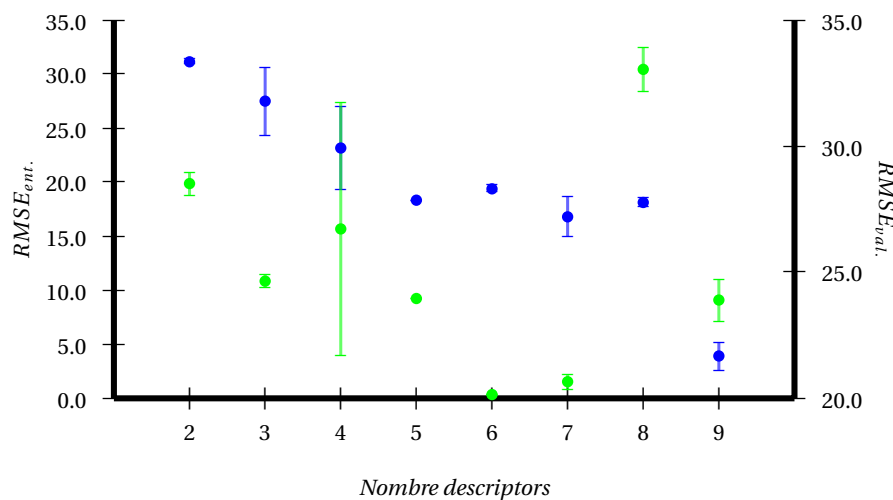


(b)

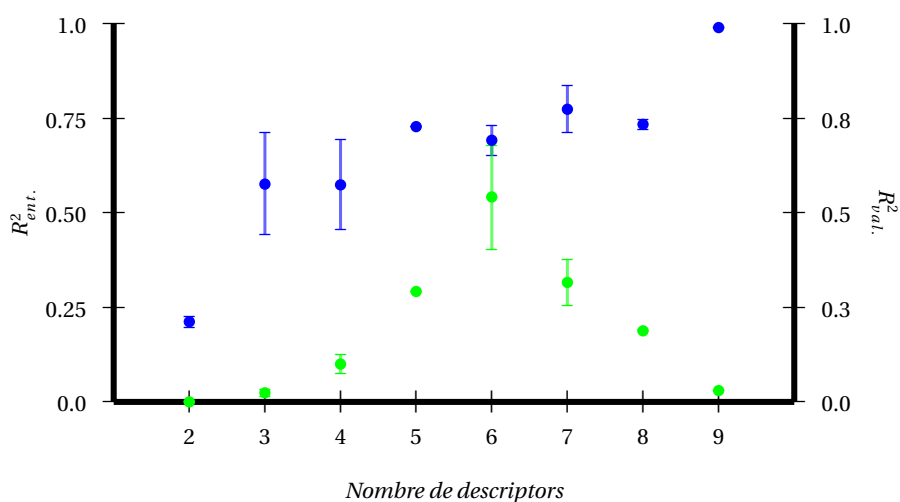


**Figura 7.5:** Anàlisi de l'evolució de l'RMSE (a) i del coeficient  $R^2$  (b) per a la identificació del nombre de descriptors òptim, segons l'ordre establert pel coeficient de correlació.

(a)

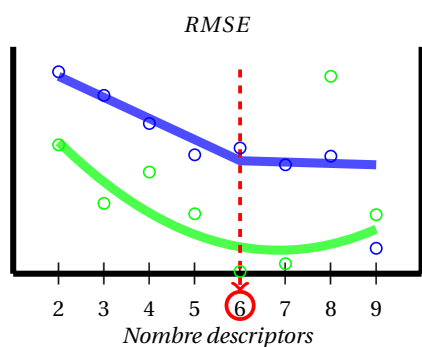


(b)

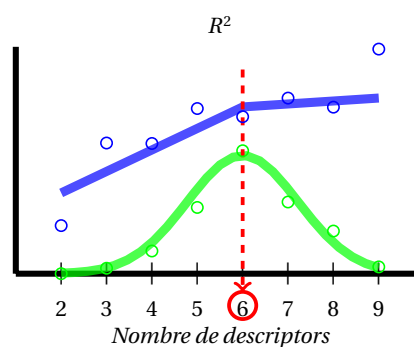


**Figura 7.6:** Evolució de l'RMSE (a) i del coeficient de correlació  $R^2$  (b) en funció del nombre de descriptors considerat segons el criteri d'ordenació basat en l'anàlisi de components principals. [● conjunt d'entrenament, ● conjunt de validació]

(a)



(b)



**Figura 7.7:** Anàlisi de l'evolució de l'RMSE (a) i del coeficient  $R^2$  (b), per a la identificació del nombre de descriptors òptim, segons l'ordre establert per l'anàlisi de components principals.

## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

---

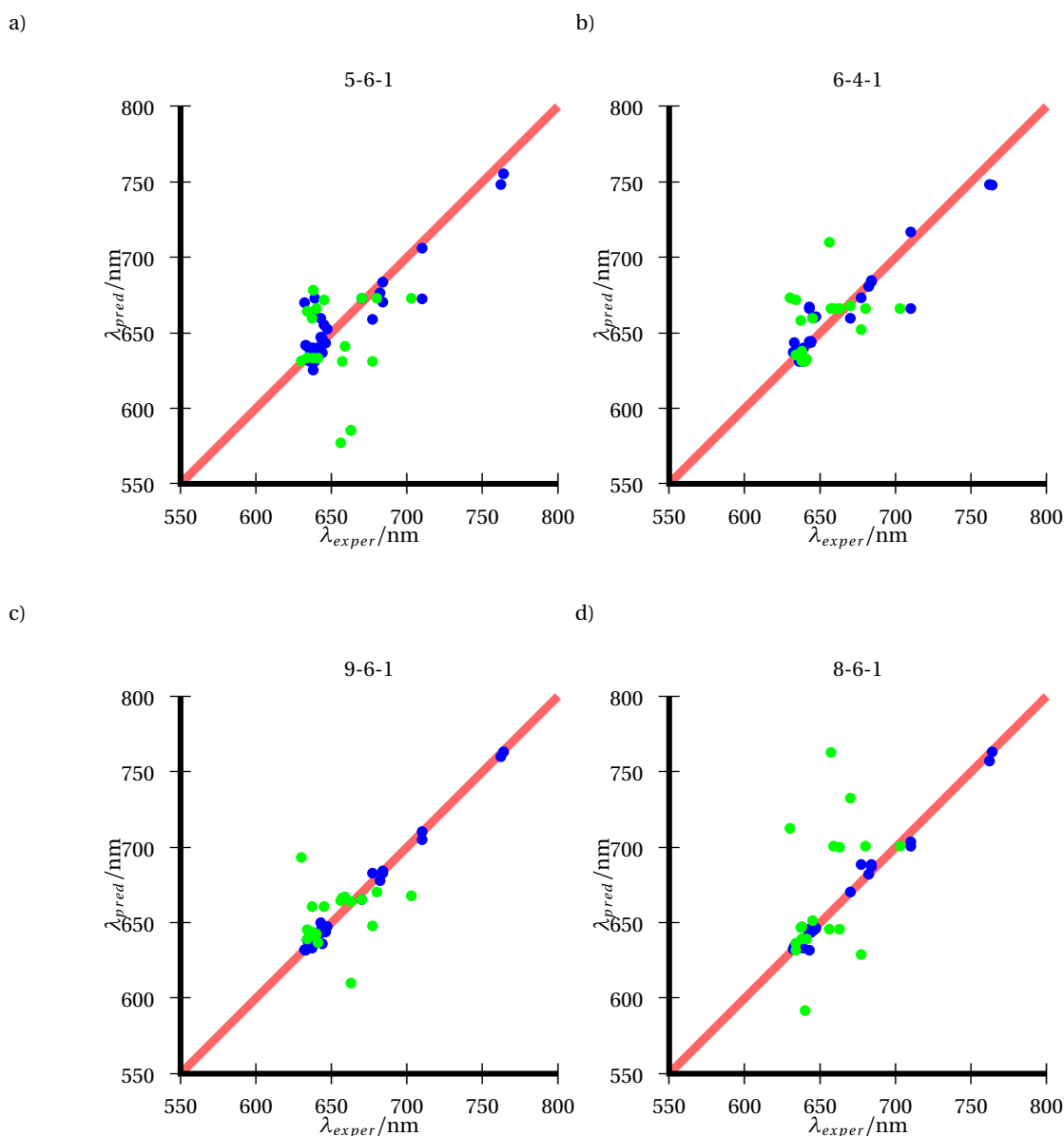
conjunt de validació, fa que el comportament de l'RMSE prengui una major importància. L'error del conjunt d'entrenament segueix en concordança amb el comportament indicat per  $R^2$  i disminueix progressivament a mesura que s'augmenta el nombre de descriptors. Aquesta situació sembla poder-se explicar en termes d'*overfitting*, donat que l'error del conjunt de validació augmenta amb el nombre de descriptors considerats, situant el punt d'inflexió d'aquest augment entre els 4 i 5 descriptors. Cal una comparació directa de la disposició entre les dades predites i les experimentals per a cadascun d'aquests models per identificar el model amb 5 descriptors com el millor.

Contràriament, els resultats obtinguts a partir de la PCA són molt més clars pel que fa a la identificació del mínim nombre de descriptors a emprar. Per un costat, el valor d' $R^2$  del *training set* va augmentant a mesura que augmenta el nombre de descriptors considerats fins arribar als 6, moment en què s'estabilitza i es mostra que l'addició de més descriptors no hauria de suposar un canvi significatiu. A més, en aquesta mateixa topologia l' $R^2$  del *test set* és màxima. Les conclusions que es poden extreure del comportament de l'RMSE és anàleg al cas anterior, mostrant un canvi de tendència entorn als 6 descriptors.

Aquesta anàlisi purament estadística dels resultats no té en compte la identitat dels descriptors considerats. És en el moment de posar noms a aquests, quan hom s'adona que l'aturada en ambdós criteris de selecció es realitza just després d'incorporar el moment dipolar. Aquest fet sembla posar en evidència la importància d'aquest descriptor per a la correcta definició del valor a predir; tanmateix, els criteris de selecció el disposen a la zona mitja d'importància. Per aquest motiu, es desitja comparar els resultats de la predicció pels models de 5, 6 i 9 descriptors, figura 7.8 (pàgina següent).

La distribució dels resultats computacionals respecte als valors experimentals mostra que el model amb 5 descriptors és el menys generalitzable dels tres. Per altra banda, els models amb 6 i 9 descriptors esdevenen possibles candidats pel model definitiu. Tot i que ambdós models presenten algunes prediccions puntuals amb un error notable, s'estudia l'ordenació relativa entre les molècules per saber si es manté. Per aquest motiu es calcula el coeficient de correlació d'Spearman ( $\rho$ ),<sup>74</sup> el qual permet avaluar la correlació entre variables a partir de la seva ordenació. Per a cadascun dels models considerats s'obté el valor d'aquest coeficient, sent màxim pel model 9-6-1 ( $\rho = 0.74$ ), seguit de 6-4-1 ( $\rho = 0.68$ ) i finalment 5-6-1 ( $\rho = 0.43$ ).

Així doncs, tot i els esforços realitzats per aconseguir una reducció de l'espai químic, el millor model que s'ha aconseguit correspon a aquell que incorpora la informació de tots nou descriptors.



**Figura 7.8:** Representació gràfica dels valors de  $\lambda_{\text{max}}$  calculats a partir d'ANN emprant els 5 descriptors seleccionats segons el coeficient de correlació (a), els 6 descriptors seleccionats segons PCA (b) i els 9 descriptors originals sense efectuar cap reducció de l'espai (c). Es mostra també el resultat de substituir els descriptors  $E_{\text{HOMO}}$  i  $E_{\text{LUMO}}$  per la seva diferència (d). [● conjunt d'entrenament, ● conjunt de validació].

### 7.2.4 Model de predicció per a porfírics (ANNABS1)

S'ha establert un model de predicció del valor de  $\lambda_{\text{max}}$  per a sistemes porfírics basat en ANN que presenta un ajust significativament millor que l'obtingut amb la metodologia TD-DFT, validada i acceptada anteriorment.

S'han emprat els 9 descriptors fisicoquímics relacionats amb la naturalesa electrònica dels FS descrit a l'apartat anterior. Després d'avaluar la seva reducció, s'ha considerat millor mantenir-los tots en la capa d'entrada de l'ANN. La topologia de la xarxa d'entrenament es completa amb una sola capa amagada (el nombre de neurones de la qual s'optimitza a 6) i una neurona de sortida, que retorna el

## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

valor en nanòmetres de la predicció realitzada. De forma independent s'estudia el valor del paràmetre d'aprenentatge  $\eta$ , que es fixa a  $1 \cdot 10^{-6}$ . Havent definit els paràmetres estructurals de la xarxa, el procés d'entrenament es realitza sobre un conjunt format per 29 porfics (dels quals es disposa el valor de  $\lambda_{max}$  bibliogràficament o experimental). Ja que algunes de les mesures experimentals associades a certs FS presentaven discrepàncies segons la font bibliogràfica consultada, s'introdueixen dues còpies de cadascun dels FS dins del procés d'entrenament (anomenades *ghost molecules*). El valor de  $\lambda_{max}$  d'aquestes molècules fantasma s'altera de forma aleatòria, fent que en una de les còpies pugui arribar a ser fins a un 5% major, i fins a un 5% inferior en la segona. La introducció d'aquest tipus de soroll en el procés d'entrenament ajuda a l'adaptació del model a valors d'entrada sotmesos a error experimental.<sup>79</sup>

L'entrenament supervisat es realitza per mitjà de l'algorisme *Back Propagation* i es defineix el mètode LOO com a validació interna. Els resultats corresponents al model de predicció final sobre el conjunt d'entrenament es mostren a la taula 7.5, incloent-hi també l'error comès en la predicció en forma de l'error absolut i com a Z-score<sup>a</sup>.

**Taula 7.5:** Resultats obtinguts en la predicció del valor de  $\lambda_{max}$  per al conjunt d'entrenament mitjançant el model ANN ( $\lambda_{pred}$ ), comparat amb el valor experimental ( $\lambda_{exp}$ ). S'adjunta també l'error de la predicció en valor absolut ( $|E|$ ) i el valor de Z-score corresponent a l'error absolut dividit de l'RMSE, pels FS del conjunt d'entrenament.

FS	$\lambda_{exp}/nm$	$\lambda_{pred}/nm$	$ E $	Z	FS	$\lambda_{exp}/nm$	$\lambda_{pred}/nm$	$ E $	Z
A1	640.0	641.2	1.2	0.4	A15	677.0	682.8	5.8	1.8
A2	634.0	638.6	4.6	1.4	A16	638.0	643.6	5.6	1.7
	632.0	632.2	0.2	0.1	A17	637.0	633.1	3.9	1.2
A3	639.0	642.2	3.2	1.0	A18	638.0	637.5	0.5	0.2
A4	636.0	638.2	2.2	0.7	A19	682.0	677.7	4.3	1.3
A5	633.0	631.7	1.3	0.4	A20	636	636.6	0.6	0.2
A6	637.0	636.1	0.9	0.3	A21	644.0	644.8	0.8	0.2
	635.0	636.1	1.1	0.3	A22	645.0	646.2	1.2	0.4
A7	684.0	682.9	1.1	0.3	A23	710.0	710.4	0.4	0.1
A8	684.0	684.3	0.3	0.1	A39	643.0	649.9	6.9	2.1
A9	643.0	643.4	0.4	0.1	A40	643.0	644.9	1.9	0.6
A10	646.0	643.8	2.2	0.7	A41	762.0	759.8	2.2	0.7
A11	647.0	647.5	0.5	0.1	A42	639.0	640.0	1.0	0.3
A12	644.0	636.2	7.8	2.4	A43	636.0	636.6	0.6	0.2
A13	710.0	704.9	5.1	1.5	A44	764.0	763.1	0.9	0.3
A14	670.0	665.3	4.7	1.4					

Una vegada plantejat el model que millor s'adapta segons la validació interna definida, s'avalua la seva capacitat real de predicció per mitjà d'un conjunt de validació externa. Aquest està format per 17 FS (10 d'origen bibliogràfic i 7 sintetitzats i estudiats pel GEM), dels quals també se'n coneix el valor de  $\lambda_{max}$ . La capacitat de predir correctament aquest conjunt de dades és el que realment determina la bondat del mètode. Així doncs, l'entrenament es repeteix generant diferents models candidats, entre els quals es pren com a model de predicció final el que permet ajustar més eficaçment les dades de validació externa, taula 7.6.

<sup>a</sup>Diferència en valor absolut entre el valor predit i l'esperat, dividit entre l'RMSE de les dades.<sup>26</sup>

Taula 7.6: Resultats obtinguts en l'aplicació del model de predicció final ANNABS1 sobre el conjunt de validació externa.

FS	$\lambda_{exp}/nm$	$\lambda_{pred}/nm$	E	Z	FS	$\lambda_{exp}/nm$	$\lambda_{pred}/nm$	E	Z
<b>A20</b>	641.0	645.2	4.2	0.2	<b>A30</b>	638.0	640.4	2.4	0.1
	634.0	636.7	5.2	0.1	<b>A31</b>	677.0	647.7	29.3	1.2
<b>A23</b>	703.0	667.8	35.2	1.5	<b>A32</b>	645.0	660.9	15.9	0.7
<b>A24</b>	634.0	639.2	5.2	0.2	<b>A33</b>	630.0	693.3	63.3	2.7
<b>A25</b>	680.0	670.5	9.5	0.4	<b>A34</b>	657.0	666.6	9.6	0.4
<b>A26</b>	638.0	641.2	3.2	0.1	<b>A35</b>	656.0	664.8	8.8	0.4
<b>A27</b>	670.0	665.2	4.8	0.2	<b>A36</b>	659.0	667.0	8.0	0.3
<b>A28</b>	640.0	642.6	2.6	0.1	<b>A37</b>	663.0	663.9	0.9	0.0
<b>A29</b>	637.0	660.6	23.6	1.0	<b>A38</b>	663.0	610.1	52.9	2.2

L'anàlisi estadística de l'ajust obtingut finalment (figura 7.8(c), pàg. 162) mitjançant el programari StatGraphics permet determinar, per mitjà de l'anàlisi de la variància (ANOVA), que l'ajust és significatiu encara que els FS **A33** i **A38** s'allunyin considerablement del valor esperat. A més, **A33** presenta un Z-score superior a 2.5, valor pres normalment com a llindar en la identificació de punts erronis, i **A38** es troba molt proper a ell. La consideració d'aquests dos punts com a erronis permet augmentar el valor de l'estadístic  $R^2$  del 77% al 91%.

Donat que **A33** és un compost asimètricament substituït, i aprofitant que el model ANNABS1 és capaç de distingir entre tautòmers (veure secció 7.3, pàg. 176), s'estudia per separat el comportament d'ambdós tautòmers de **A33**, de la mateixa manera que es realitzà en l'estudi de l'error assumible (pàg.152). Es comprova que el tautòmer considerat inicialment és el més estable.

La mitjana de l'error absolut comès pel model ANNABS1 en el conjunt de validació externa (taula 7.6) és inferior als 16 nm, trobant-se notablement per sota dels 70 nm establerts com a llindar mitjançant la metodologia TD-DFT. Cal remarcar que el fet de no considerar l'efecte del solvent per aquest tipus de compostos no suposa cap inconvenient, ja que l'error del model és del mateix ordre que l'error de les mesures experimentals descrites bibliogràficament i experimental.

El model de predicció final s'ha implementat, mitjançant el llenguatge de programació SVL, en el programari MOE,<sup>81</sup> passant a formar part d'un mòdul de càlcul integrat dins del propi programa. Amb aquesta acció es desitja facilitar l'ús del model de predicció entre els membres del laboratori, fent-lo assequible d'una manera fàcil i intuïtiva.

### 7.2.5 Generalització del model de predicció a FS tetrapirròlics (ANNABS2)

Havent comprovat els bons resultats obtinguts en el cas dels porfírens, es planteja la possibilitat d'establir un model de predicció, anàleg a l'anterior, que englobi qualsevol tipus de fotosensibilitzador tetrapirròlic. Aquest objectiu suposa redefinir la base de dades d'entrenament, expandint-la a altres tipus de *cores*.

Prenent de base la quimioteca corresponent als derivats porfírenics emprats en el model anterior, s'afegeix a ella un seguit de derivats porfirínics, clorines, purpurines, feofòrbids i alguns derivats porfírenics addicionals, que permeten ampliar la base de dades amb 87 FS més, figura 7.9. Per tal d'ajudar a la generalització del model, les molècules es presenten de forma aleatòria durant el procés d'entrenament.

## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

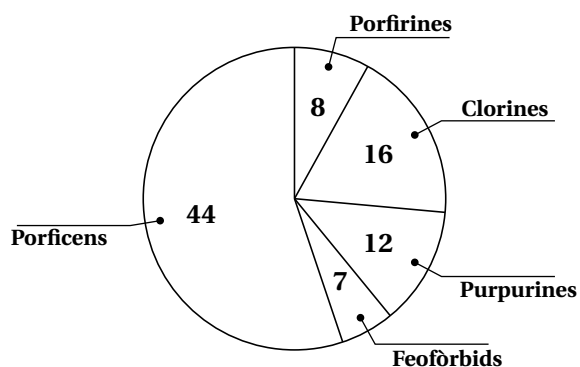
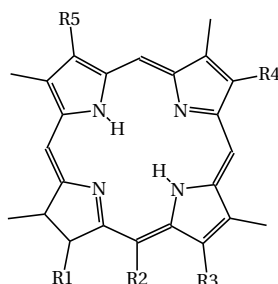
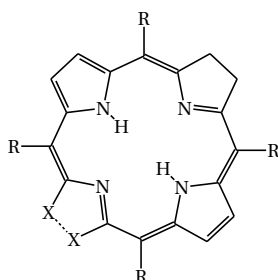


Figura 7.9: Distribució dels diferents tipus de FS tetrapirròlics considerats per l'establiment del model de predicció general.

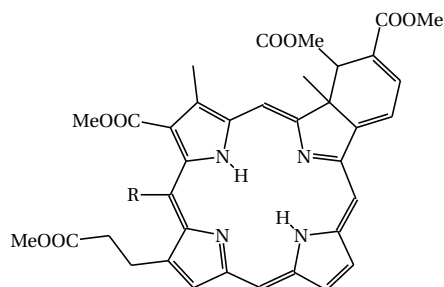
Taula 7.7: Estructures moleculars dels derivats de clorines inclosos en la quimioteca virtual global (1a part).



id.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	$\lambda_{max}/nm$
<b>A45</b>	(CH <sub>2</sub> ) <sub>2</sub> COOH	CONH(C <sub>7</sub> H <sub>15</sub> )	COOH	Et	CH=CH <sub>2</sub>	666 <sup>172</sup>
<b>A46</b>	(CH <sub>2</sub> ) <sub>2</sub> COOMe	CH <sub>2</sub> COOMe	COOMe	CH=CH <sub>2</sub>	Et	651 <sup>173</sup>
<b>A47</b>	(CH <sub>2</sub> ) <sub>2</sub> COOMe	COOMe	COOMe	CH=CH <sub>2</sub>	Et	655 <sup>173</sup>
<b>A48</b>	(CH <sub>2</sub> ) <sub>2</sub> COOMe	COOMe	COOMe	Et	Et	656 <sup>173</sup>
<b>A49</b>	(CH <sub>2</sub> ) <sub>2</sub> COOH	CH <sub>2</sub> COOH	COOH	Et	Et	643 <sup>115</sup>
<b>A50</b>	(CH <sub>2</sub> ) <sub>2</sub> COOMe	CH <sub>2</sub> COOMe	COOMe	Et	CH=CH <sub>2</sub>	662 <sup>115</sup>
<b>A51</b>	(CH <sub>2</sub> ) <sub>2</sub> COOMe	CH <sub>2</sub> COOMe	CONHCH <sub>2</sub> C≡CH	Et	CH=CH <sub>2</sub>	663 <sup>174</sup>
<b>A52</b>	(CH <sub>2</sub> ) <sub>2</sub> COOH	CH <sub>2</sub> COOH	COOH	Et	CH=CH <sub>2</sub>	667 <sup>115</sup>

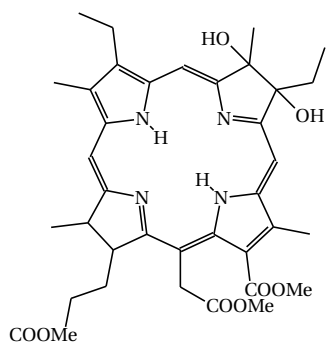


id.	R	X	$\lambda_{max}/nm$
<b>A53</b>	2-clorofenil	CH=CH	655 <sup>175</sup>
<b>A54</b>	2,6-diclorofenil	CH=CH	660 <sup>175</sup>
<b>A55</b>	2,6-difluorofenil	CH=CH	655 <sup>175</sup>
<b>A56</b>	naftalen-1-il	CH=CH	652 <sup>175</sup>
<b>A57</b>	2-clorofenil	CH <sub>2</sub> -CH <sub>2</sub>	744 <sup>175</sup>
<b>A58</b>	2,6-diclorofenil	CH <sub>2</sub> -CH <sub>2</sub>	747 <sup>175</sup>
<b>A59</b>	2,6-difluorofenil	CH <sub>2</sub> -CH <sub>2</sub>	744 <sup>175</sup>

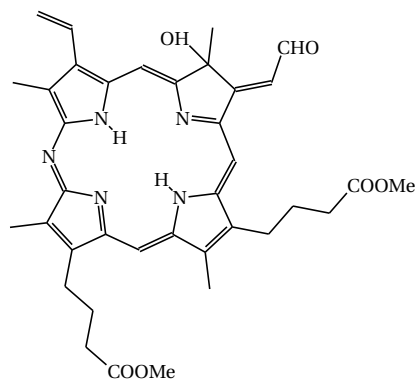


id.	R	$\lambda_{max}/nm$
<b>A60</b>	CH <sub>2</sub> COOMe	744 <sup>173</sup>
<b>A61</b>	COOMe	759 <sup>173</sup>

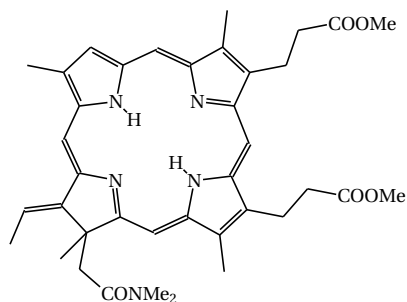
**Taula 7.8:** Estructures moleculars dels derivats de clorines inclosos en la quimioteca virtual global (2a part). S'indica sota el número de compost el valor de  $\lambda_{max}$



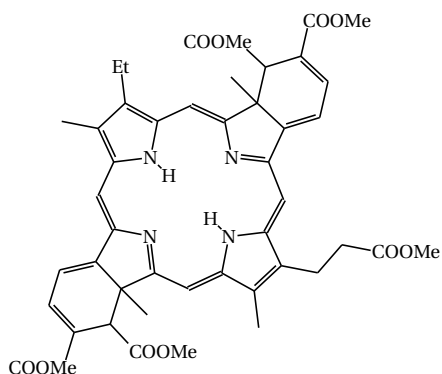
**A62**  
723<sup>173</sup>



**A63**  
674<sup>115</sup>



**A64**  
660<sup>115</sup>

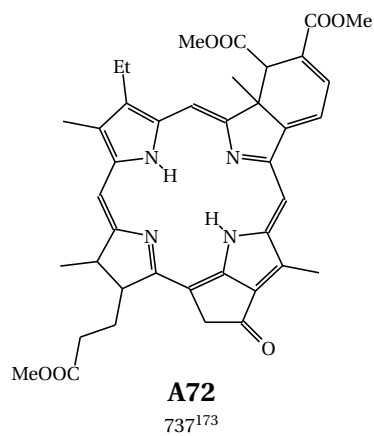
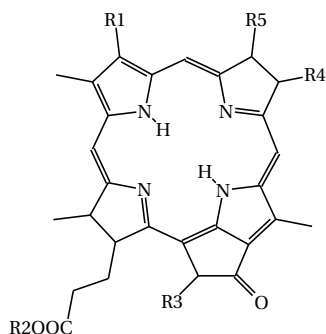


**A65**  
784<sup>115</sup>

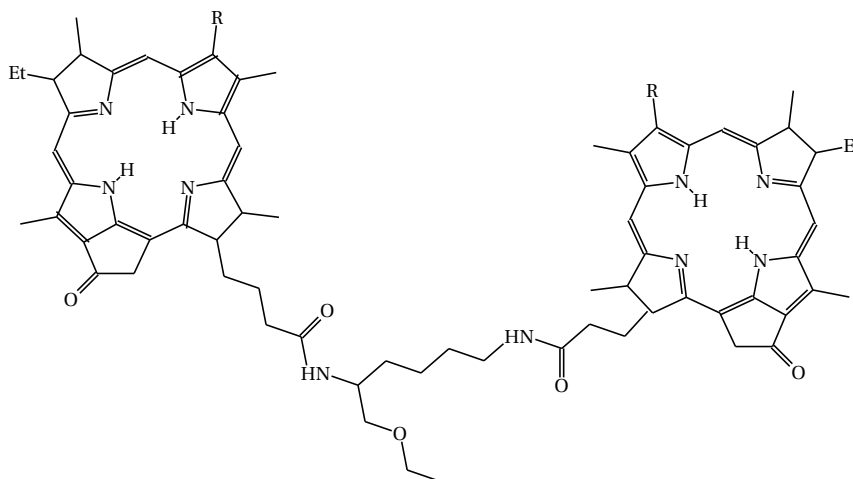
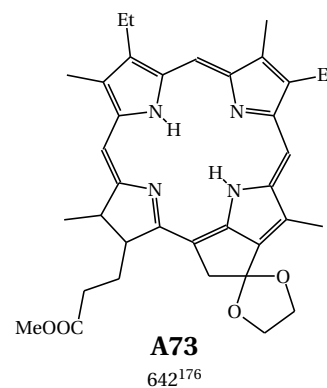


## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

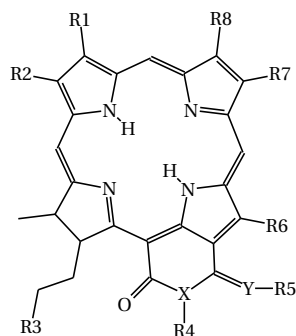
**Taula 7.9:** Estructures moleculars dels derivats de feoforbids inclosos en la quimioteca virtual global



id.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	$\lambda_{max}/nm$
<b>A66</b>	-CH=CH <sub>2</sub>	fitil	COOMe	Et	Me	668 <sup>115</sup>
<b>A67</b>	-CH=CH <sub>2</sub>	fitil	COOMe	Et	CHO	654 <sup>115</sup>
<b>A68</b>	Et	Me	H	H <sub>2</sub> C=CH	Me	657 <sup>173</sup>
<b>A69</b>	Et	OH	H	Et	Me	656 <sup>176</sup>
<b>A70</b>	Et	Me	H	Et	Me	656 <sup>176</sup>
<b>A71</b>	-CH=CH <sub>2</sub>	Me	H	Et	Me	666 <sup>176</sup>



id.	R	$\lambda_{max}/nm$
<b>A74</b>	-CH(CH <sub>3</sub> )CH <sub>2</sub> O(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>	660 <sup>115</sup>
<b>A75</b>	-CH <sub>2</sub> (CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>	660 <sup>115</sup>
<b>A76</b>	-CH=CH <sub>2</sub>	668 <sup>115</sup>
<b>A77</b>	CHO	694 <sup>115</sup>

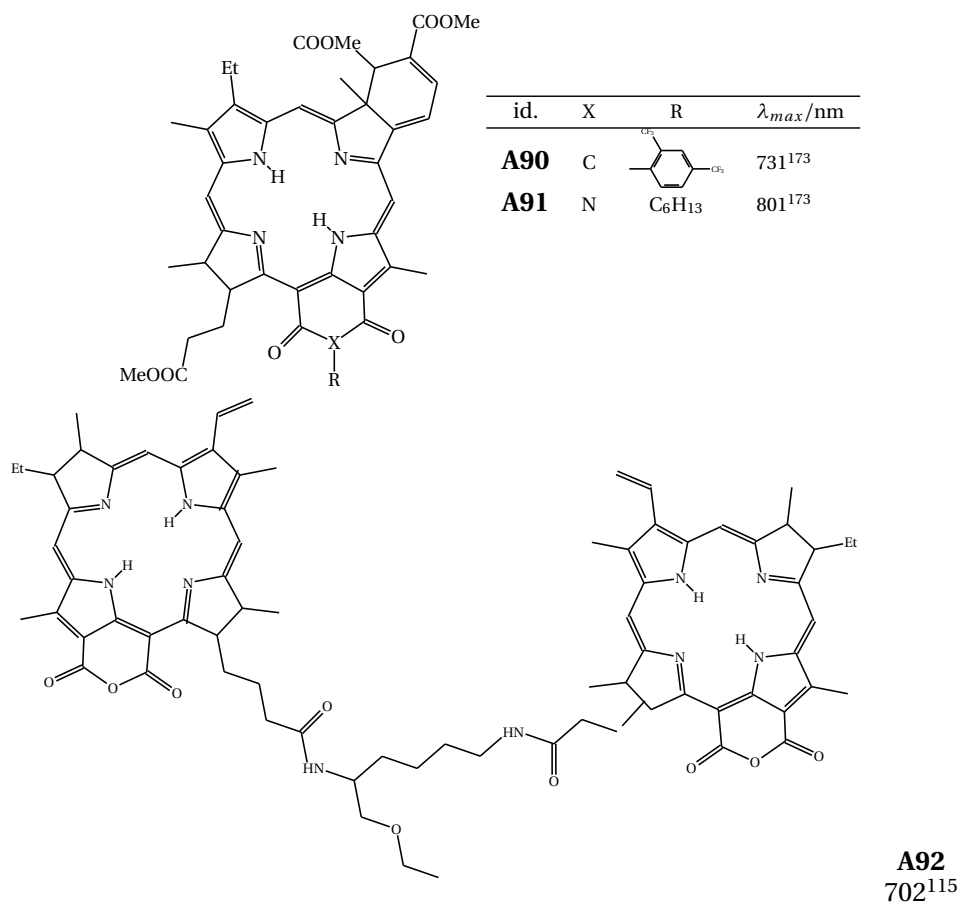
**Taula 7.10:** Estructures moleculars dels derivats de purpurines inclosos en la quimioteca virtual global (1a part).


id.	R1	R2	R3	R4
<b>A78</b>	-CH=CH <sub>2</sub>	Me	-C(=O)NH(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
<b>A79</b>	-CH=CH <sub>2</sub>	Me	-COOMe	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
<b>A80</b>	Et	Me	COOMe	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
<b>A81</b>	Et	Me	COOMe	3,5-bis(trifluorometil)bencil
<b>A82</b>	Et	Me	COOMe	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
<b>A83</b>	-CH=CH <sub>2</sub>	Me	COOMe	-
<b>A84</b>	Et	Me	COOMe	3,5-bis(trifluorometil)bencil
<b>A85</b>	-CH=CH <sub>2</sub>	Me	COOH	-
<b>A86</b>	-CH=CH <sub>2</sub>	Me	COOMe	-(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>
<b>A87</b>	-CH=CH <sub>2</sub>	Me	COOMe	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
<b>A88</b>	-CH <sub>2</sub> (CH <sub>3</sub> )O(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>	Me	COOMe	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>
<b>A89</b>	-C(=O)CH <sub>3</sub>	Me	COOMe	-CH <sub>2</sub> C≡CH <sub>3</sub>

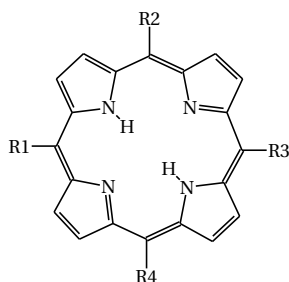
id.	R5	R6	R7	R8	X	Y	$\lambda_{max}/nm$
<b>A78</b>	-	Me	Et	Me	N	O	705 <sup>172</sup>
<b>A79</b>	-	Me	Et	Me	N	O	690 <sup>172</sup>
<b>A80</b>	-	Me	-CH=CH <sub>2</sub>	Me	N	O	690 <sup>173</sup>
<b>A81</b>	-	Me	-CH=CH <sub>2</sub>	Me	N	O	693 <sup>173</sup>
<b>A82</b>	-	Me	Et	Me	N	O	694 <sup>173</sup>
<b>A83</b>	-(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>	Me	Et	Me	O	N	696 <sup>172</sup>
<b>A84</b>	-	Me	Et	Me	N	O	696 <sup>173</sup>
<b>A85</b>	-	Me	Et	Me	O	O	698 <sup>115</sup>
<b>A86</b>	-	Me	Et	Me	N	O	705 <sup>172</sup>
<b>A87</b>	-	CHO	Et	Me	N	O	717 <sup>172</sup>
<b>A88</b>	-	CHO	Et	Me	N	O	699 <sup>172</sup>
<b>A89</b>	-	Me	Et	Me	N	O	826 <sup>174</sup>

## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

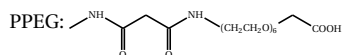
**Taula 7.11:** Estructures moleculars dels derivats de purpurines inclosos en la quimioteca virtual global (2a part).



Taula 7.12: Estructures moleculars dels derivats de porfirines inclosos en la quimioteca virtual global

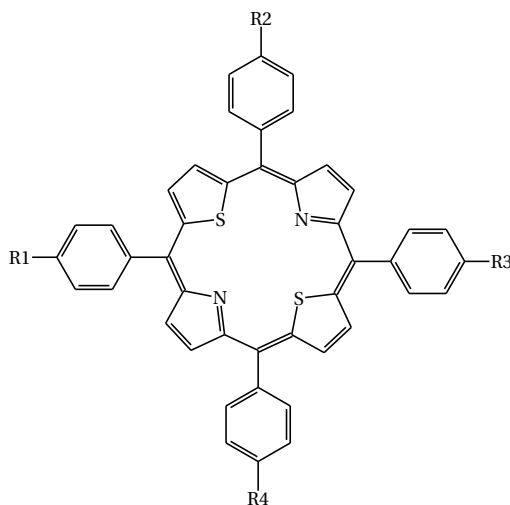


id.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	$\lambda_{max}/nm$
<b>A93</b>	H	H	H	H	650 <sup>175</sup>
<b>A94</b>	H	H	H	( <i>p</i> -HOOCOOCNH)Ph	645 <sup>177</sup>
<b>A95</b>	( <i>p</i> -HOOCOOCNH)Ph	H	H	( <i>p</i> -HOOCOOCNH)Ph	648 <sup>177</sup>
<b>A96</b>	( <i>p</i> -HOOCOOCNH)Ph	( <i>p</i> -HOOCOOCNH)Ph	H	( <i>p</i> -HOOCOOCNH)Ph	647 <sup>177</sup>
<b>A97</b>	( <i>p</i> -HOOCOOCNH)Ph	( <i>p</i> -HOOCOOCNH)Ph	( <i>p</i> -HOOCOOCNH)Ph	( <i>p</i> -HOOCOOCNH)Ph	648 <sup>177</sup>
<b>A98</b>	H	H	H	PPEG	645 <sup>177</sup>
<b>A99</b>	PPEG	H	H	PPEG	648 <sup>177</sup>
<b>A100</b>	PPEG	PPEG	H	PPEG	649 <sup>177</sup>
<b>A101</b>	PPEG	PPEG	PPEG	PPEG	650 <sup>177</sup>
<b>A102</b>	( <i>o</i> -Cl)Ph	( <i>o</i> -Cl)Ph	( <i>o</i> -Cl)Ph	( <i>o</i> -Cl)Ph	655 <sup>175</sup>
<b>A103</b>					660 <sup>175</sup>
<b>A104</b>					655 <sup>175</sup>
<b>A105</b>	naftalen-1-il	naftalen-1-il	naftalen-1-il	naftalen-1-il	652 <sup>175</sup>
<b>A106</b>	fenantren-9-il	fenantren-9-il	fenantren-9-il	fenantren-9-il	656 <sup>175</sup>
<b>A107</b>	( <i>p</i> - <sup>t</sup> Bu)Ph	( <i>p</i> - <sup>t</sup> Bu)Ph	( <i>p</i> - <sup>t</sup> Bu)Ph	( <i>p</i> - <sup>t</sup> Bu)Ph	647 <sup>178</sup>

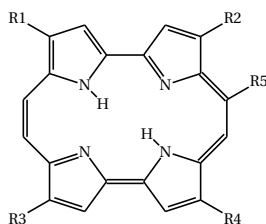


## 7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS

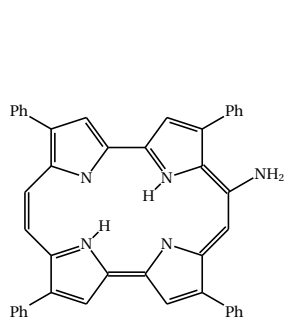
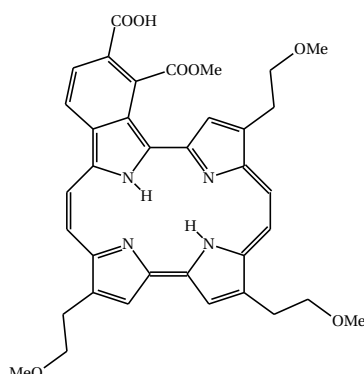
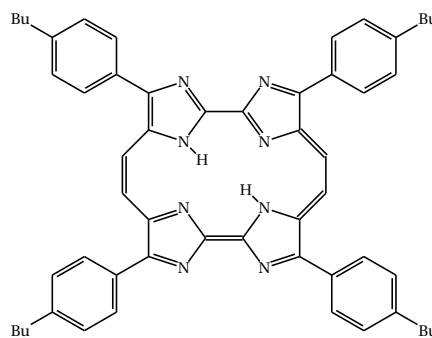
**Taula 7.13:** Estructures moleculars dels derivats de tioporfirines inclosos en la quimioteca virtual global.



id.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	$\lambda_{max}/nm$
<b>A108</b>	OCH <sub>2</sub> COOH	H	H	H	696 <sup>179</sup>
<b>A109</b>	OCH <sub>2</sub> COOH	OCH <sub>2</sub> COOH	H	H	699 <sup>179</sup>
<b>A110</b>	OCH <sub>2</sub> COOH	OCH <sub>2</sub> COOH	OCH <sub>2</sub> COOH	H	700 <sup>179</sup>
<b>A111</b>	OCH <sub>2</sub> COOH	OCH <sub>2</sub> COOH	OCH <sub>2</sub> COOH	OCH <sub>2</sub> COOH	701 <sup>179</sup>

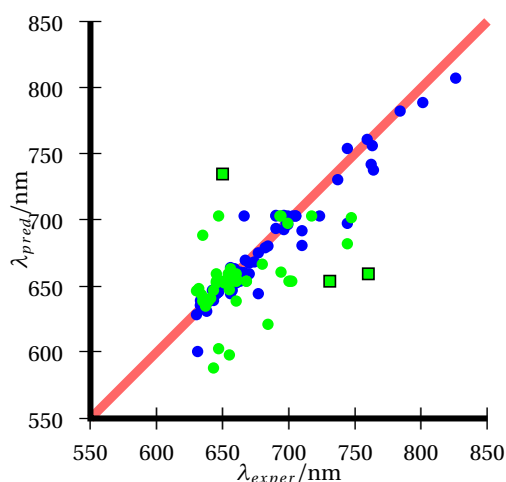
**Taula 7.14:** Estructures moleculars dels derivats porfírics inclosos en la quimioteca virtual global. S'indica sota el número de compost el valor de  $\lambda_{max}$ 


id.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	$\lambda_{max}/\text{nm}$
<b>A112</b>	Me	Me	Me	Me	H	631 <sup>132</sup>
<b>A113</b>	Pr	Pr	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	H	633 <sup>167</sup>
<b>A114</b>	CH <sub>2</sub> CH <sub>2</sub> OH	CH <sub>2</sub> CH <sub>2</sub> OH	Pr	Pr	H	633 <sup>167</sup>
<b>A115</b>	Ph	Ph	Ph	Ph	NH <sub>2</sub>	763 <sup>178</sup>
<b>A116</b>	( <i>p</i> -MeO)Ph	( <i>p</i> -MeO)Ph	( <i>p</i> -MeO)Ph	( <i>p</i> -MeO)Ph	H	660 <sup>178</sup>
<b>A117</b>	( <i>p</i> <sup>-i</sup> PrO)Ph	( <i>p</i> <sup>-i</sup> PrO)Ph	( <i>p</i> <sup>-i</sup> PrO)Ph	( <i>p</i> <sup>-i</sup> PrO)Ph	H	661 <sup>180</sup>
<b>A118</b>	Pr	Pr	Pr	Pr	NO <sub>2</sub>	635 <sup>178</sup>
<b>A119</b>	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> Br	CH <sub>2</sub> CH <sub>2</sub> OMe	H	635 <sup>167</sup>
<b>A120</b>	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	H	638 <sup>166</sup>
<b>A121</b>	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	-OCOC <sub>17</sub> H <sub>35</sub>	646 <sup>166</sup>
<b>A122</b>	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	CH <sub>2</sub> CH <sub>2</sub> OMe	-NHCO(CH <sub>2</sub> ) <sub>3</sub> COOH	645 <sup>166</sup>


**A123**  
690<sup>181</sup>

**A124**  
636<sup>166</sup>

**A125**  
760<sup>131</sup>

El conjunt de 131 FS tetrapirròlics, per als quals es disposa de dades bibliogràfiques de l'espectre d'absorció, és dividit en un conjunt d'entrenament (format pel 60% del total, 80 FS) i un conjunt de validació externa (51 FS). Aprofitant el coneixement generat en l'anterior model de derivats porfírics, per a la definició dels paràmetres d'entrenament s'utilitza una xarxa de topologia 9-6-1 *fully connected* amb els mateixos descriptors, un *learning rate* de  $\eta = 1 \cdot 10^{-6}$  i una validació interna LOO.

Després de realitzar diferents proves avaluant l'efecte causat per la modificació dels paràmetres constituents de la xarxa, s'obté un model apte per a la correlació de les dades experimentals, tant de les emprades en l'entrenament com en la validació. La mitjana de l'error comès en les prediccions realitzades sobre els FS de validació externa és inferior als 30 nm, figura 7.10.



**Figura 7.10:** Resultats obtinguts en la predicció de  $\lambda_{max}$  amb el model generalitzat per a FS tetrapirròlics (ANNABS2). [• Conjunt d'entrenament, • Conjunt de validació]

Tot i així, certs FS del conjunt de validació (**A90**, **A101** i **A125**); representats com a quadrats en la figura 7.10) mostren un error superior al llindar marcat pel TD-DFT (70 nm; apartat 7.2.1, pàg. 152). El mal ajust de cadascun d'aquests compostos pot justificar-se a través de característiques estructurals diferencials:

- La purpurina **A90**, a més de presentar un anell benzènic fusionat a l'enllaç entre les posicions 17 i 18, presenta un substituent benzílic derivatitzat amb sis àtoms de fluor. No hi ha cap estructura fluorada en el conjunt d'entrenament.
- El derivat porfirínic **A101** conté grups fenil *para*-substituïts en les posicions meso. Les cadenes laterals de polietilenglicol dificulten l'optimització de la geometria molecular a causa de la seva extensió i de la seva capacitat d'interaccionar amb el nitrogen de l'enllaç amínic que els manté units als grups fenil.
- El derivat **A125** és l'únic tetraaza porfíric present en la quimioteca, així doncs no sorprèn que les seves característiques estructurals no es trobin recollides pel model de predicció.

### Limitacions dels models de predicció ANNABS1 i ANNABS2

La consideració i comparació no només d'aquells FS amb un error de predicció superior al llindar, sinó també d'aquells acceptats però amb major error, permet detectar les característiques que esdevenen limitacions d'ambdós models:

- **Mida molecular:** Els descriptors  $\mu$ ,  $E_{HOMO}$  i  $E_{LUMO}$  necessaris es calculen mitjançant el programari MOPAC<sup>169</sup> integrat en MOE,<sup>81</sup> que té una limitació de 60 àtoms pesants.<sup>81</sup> Les molècules **A38**, **A56**, **A66**, **A67**, **A74**, **A75**, **A76**, **A77**, **A90**, **A92**, **A95-A101**, **A105**, **A106**, **A107**, **A110**, **A111**, **A125** superen aquest límit.

Aquesta limitació s'ha evitat utilitzant el programari Gaussian09<sup>182</sup> per al càlcul PM3 que duu als descriptors, prenent la precaució d'expressar els valors de l'energia en eV.

- **Derivats halogenats:** El mètode ZINDO utilitzat com a descriptor es calcula amb el programari Gaussian09,<sup>182</sup> el qual disposa només dels paràmetres de càlcul per derivats halogenats amb fluor.

Per als derivats clorats **A53**, **A54**, **A57**, **A58**, **A102**, **A103** s'utilitza el programari HyperChem<sup>7183</sup> per al càlcul d'aquest descriptor.

No es disposa en cap cas dels paràmetres necessaris per a l'àtom de brom. La predicció de la molècula **A119** s'ha realitzat canviant el Br per Cl, obtenint un resultat aproximat.

- **Dependència de la geometria inicial:** Algunes de les molècules (com **A8**, **A66**, **A111**) presenten més de 15 enllaços rotables. La necessitat de descriptors que utilitzen la geometria molecular en els models ANNABS, fa que calgui prendre especial atenció a l'hora de definir la geometria inicial d'aquells FS amb un elevat joc conformacional.

Quan la mobilitat molecular no és extremadament important, l'optimització de la geometria prèvia al càlcul de descriptors ha demostrat ser suficient per garantir l'aplicabilitat del model. L'estudi conformacional d'una molècula de TPPo (**A36**) mitjançant el programari MOE2010.10<sup>26</sup> i emprant el mètode de cerca *LowModeMD*<sup>184</sup> (en el qual la generació de nous conformers es realitza a partir de dinàmiques moleculars d'1 ps a temperatura constant, seguit d'una minimització d'energia) duu a 16 conformacions, cadascuna d'elles possibles resultats de l'optimització prèvia. Tanmateix, la diferència entre els resultats obtinguts després d'aplicar els models ANNABS1 i ANNABS2 per cadascuna d'aquestes conformacions no supera en cap cas els 5 nm.

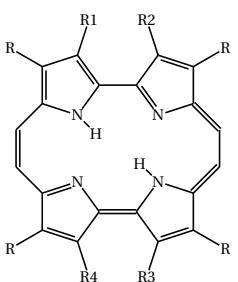
Per altra banda, l'augment de la flexibilitat amb la incorporació de substituents propils en les posicions 2,7,12, i 17 del nucli porficènic (**A20**, **A23**) fa que el model pugui presentar diferències en la predicció, en tot cas però, dins del marge d'error experimental. Aquest fet es pot observar gràcies a la inclusió de diferents conformacions obtingudes en l'optimització de **A20** i **A23** dins del conjunt de validació d'ANNABS1, figura 7.6 pàg. 164.

### 7.2.6 Refinament del model inicial

En cerques bibliogràfiques posteriors a la creació dels dos models, es troben descrites noves famílies de derivats porficènics.<sup>185–188</sup> L'aplicació dels models ANNABS1 i ANNABS2 sobre elles permet avaluar la bondat de les prediccions de FS amb característiques estructurals molt diferents a les definides en els corresponents conjunts d'entrenament, figura 7.11.



7.2. MODEL QSAR DE PREDICCIÓ DEL MÀXIM D'ABSORCIÓ DE FOTOSENSIBILITZADORS TETRAPIRRÒLICS



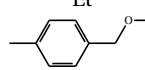
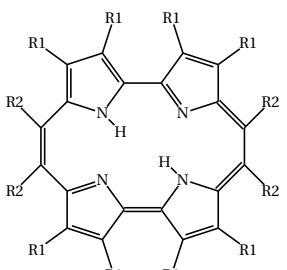
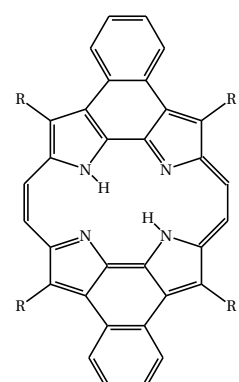
id.	R	R1	R2	R3	R4	$\lambda_{max}/nm$
<b>A126</b>	Et	H	H	H	H	633 <sup>185</sup>
<b>A127</b>	Et	Br	H	H	H	641 <sup>185</sup>
<b>A128</b>	Et	Br	H	Br	H	649 <sup>185</sup>
<b>A129</b>	Et	Br	Br	H	Br	665 <sup>185</sup>
<b>A130</b>	Et	Br	Br	Br	Br	684 <sup>185</sup>
<b>A131</b>		H	H	H	H	659 <sup>189</sup>

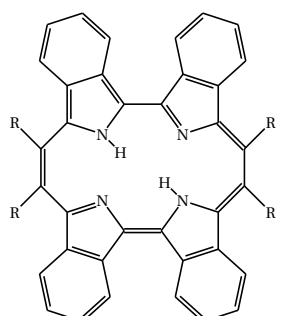
Figura 7.11: Derivats porfírics avaluats en el refinament dels models ANNABS.



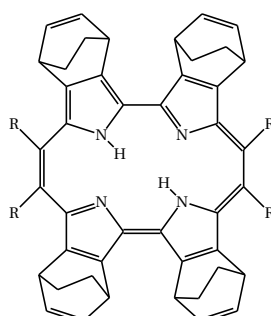
id.	R1	R2	$\lambda_{max}/nm$
<b>A132</b>	Et	Me	289 <sup>187</sup>
<b>A133</b>	-C <sub>6</sub> H <sub>13</sub>	Me	381 <sup>187</sup>
<b>A134</b>	H	Ph	653 <sup>188</sup>
<b>A135</b>	H	Bz	655 <sup>188</sup>



id.	R	$\lambda_{max}/nm$
<b>A136</b>	Pr	713 <sup>186</sup>
<b>A137</b>	iPr	718 <sup>186</sup>
<b>A138</b>	-C <sub>5</sub> H <sub>11</sub>	714 <sup>186</sup>



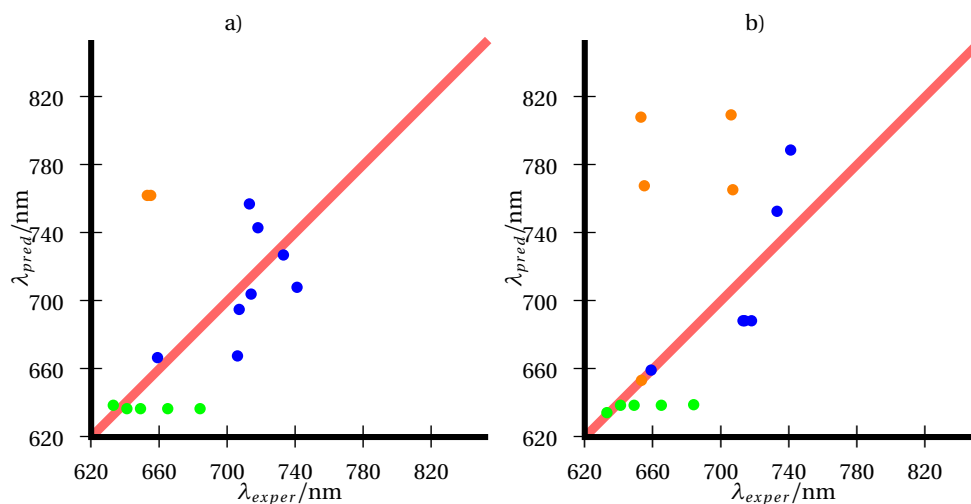
id.	R	$\lambda_{max}/nm$
<b>A139</b>	Me	733 <sup>187</sup>
<b>A140</b>	Et	741 <sup>187</sup>



id.	R	$\lambda_{max}/nm$
<b>A141</b>	Me	706 <sup>187</sup>
<b>A142</b>	Et	707 <sup>187</sup>

Figura 7.12: Derivats porfírics avaluats en el refinament dels models ANNABS (cont.).

La comparació dels valors experimentals amb els resultats de  $\lambda_{max}$  predits es mostren a la figura 7.13.



**Figura 7.13:** Representació de les prediccions front als valors bibliogràfics de  $\lambda_{max}$  de les noves famílies de porfírens pels models ANNABS1 (a) i ANNABS2 (b). [• Conjunt de dades amb prediccions acceptables, • derivats bromats, • punts considerats erronis]

No sorprèn que, tractant-se de derivats porfírenics, el model ANNABS1 disposi d'una capacitat d'adaptació lleugerament superior a ANNABS2, doncs ha estat creat a partir d'aquest tipus d'estructures. Tot i així, els compostos **A134** i **A135** presenten un error notable en ambdós models, probablement perquè són els únics amb les posicions 9, 10, 19 i 20 substituïdes amb grups fenils o benzils. A més, les estructures **A141** i **A142** (amb quatre substituents bicicles, representats en taronja a la figura 7.13,b) presenten unes característiques estructurals que no poden ajustades amb ANNABS2.

La limitació per derivats bromats comentada anteriorment es posa de manifest amb **A127-A130**, que s'ajusten pitjor com més àtoms de brom contenen, figura 7.13 (representats en color verd i predits, tal com s'ha comentat en l'apartat anterior, fent el canvi per àtoms de clor).

### 7.3 Estudi de la tautomeria del 9ATPPo

Al laboratori de Fotoquímica del GEM s'observà un que l'espectre d'absorció UV/Vis del 9-amino-2,7,12,17-tetrafenilporfíren (9ATPPo) presentava un perfil realment complex, atribuïble a la presència de dues espècies químicament molt semblants,<sup>181</sup> figura 7.14.

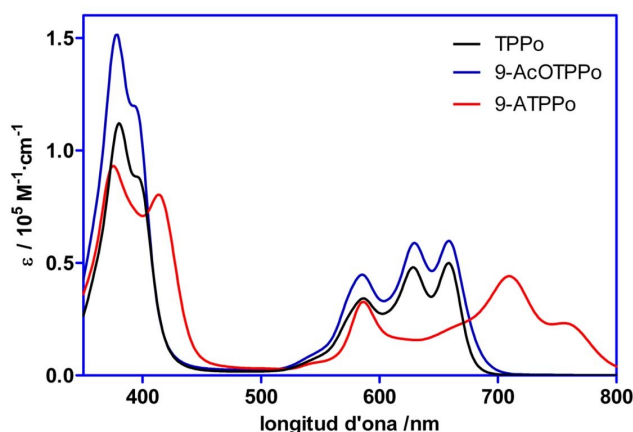


Figura 7.14: Comparació dels espectres d'absorció del TPPo, 9AcOTPPo i 9ATPPo.<sup>190</sup>

Aquest comportament, també descrit en la bibliografia pel 9-acetoxi-2,7,12,17-tetrafenilporficè,<sup>191</sup> pot ser degut a la presència de diferents tautòmers del mateix porficè. Per establir un mecanisme plausible que permetés explicar aquest comportament s'avalua l'energia de l'estat fonamental i del primer estat excitat dels diferents tautòmers del 9ATPPo. La tautomeria N-H de porfícens es troba descrita bibliogràficament,<sup>192, 193</sup> i ha permès identificar els tautòmers *trans* com els més estables.

### 7.3.1 Validació del mètode de càlcul per a l'estudi de la tautomeria del Po

Tal com s'ha comentat, la metodologia DFT considerada en aquest treball com a estàndard, fou validada mitjançant resultats bibliogràfics de diferents FS (secció 7.1). En ells, com en el cas concret del Po mostrat en la secció 7.2.1, l'estudi se sol basar i restringir al tautòmer *trans*. Aprofitant, però, que es disposa d'informació bibliogràfica d'un dels tautòmers *cis* del porficè (figura 7.15), s'avalua la idoneïtat d'utilitzar la metodologia de càlcul anterior en l'estudi de la tautomeria del 9ATPPo.

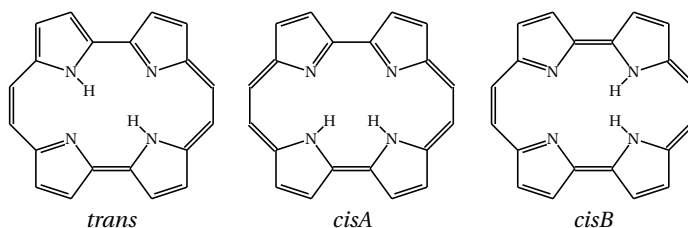


Figura 7.15: Estructura molecular dels tres possibles tautòmers del porficè.

L'optimització de la geometria en estat fonamental dels tres tautòmers del porficè, es realitza mitjançant el mètode DFT a nivell B3LYP/6-31G\*. A continuació es realitza l'anàlisi de freqüències garantir l'assoliment d'un mínim en la superfície d'energia potencial i obtenir així la correcció de l'energia en el punt zero (*Zero Point Energy*, ZPE). El càlcul de l'energia dels estats excitats es realitza amb un mètode TD-DFT B3LYP/6-31G\*, amb el programari Gaussian03.<sup>143</sup> Els resultats es recullen a la taula 7.15.

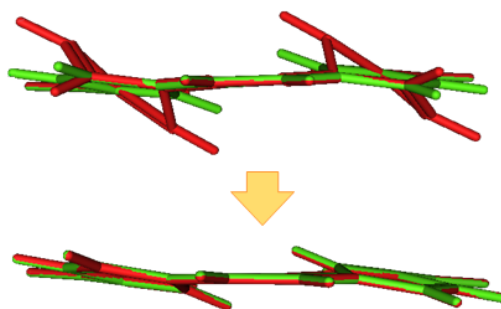
**Taula 7.15:** Resultats obtinguts en el càlcul de l'energia de l'estat fonamental del porficè en Hartrees,  $E(\text{SCF})$ , juntament amb la seva absorció associada.

Tautòmer	$E(\text{SCF})/\text{H}$	$\lambda_{\text{max}}/\text{nm}$
<i>trans</i>	-989.552	555.47
<i>cisA</i>	-989.547	553.74
<i>cisB</i>	-989.504	579.16

Per tal de poder realitzar una anàlisi significativa d'aquests resultats, es realitza la comparació energètica entre espècies prenent l'energia de l'estat fonamental del tautòmer *trans* com a referència. Els resultats obtinguts, expressats en eV, es mostren a la taula 7.16. Tal com s'hi pot observar, existeix una notable discrepància entre els resultats calculats i els bibliogràfics pel tautòmer *cisB*. Per aquest motiu, es repeteixen els càlculs corresponents a aquest tautòmer prenent com a base una geometria inicial altament distorsionada (intentant fer-la el més semblant possible a la descrita en la bibliografia). Aquest nou resultat continua coincidint amb la geometria obtinguda inicialment i no amb la proposada per Gil *et al.*,<sup>191</sup> figura 7.16.

**Taula 7.16:** Comparació de les energies dels estats  $S^0$  i  $S^1$  entre els tautòmers *trans* i *cisB* del porficè. Els valors de l'energia es calculen prenent l'energia de l'estat fonamental del tautòmer *trans* com a referència.

Tautòmer	GEM/eV	Gil <i>et al.</i> <sup>191</sup> /eV	Wan <i>et al.</i> <sup>164</sup> /eV	Experimental/eV
<i>trans</i> $S^0$	0	0	0	0
<i>trans</i> $S^1$	2.23	2.24	2.26	2.06
<i>cisB</i> $S^0$	1.26	2.79		
<i>cisB</i> $S^1$	3.42	3.03		



**Figura 7.16:** Representació de la geometria distorsionada del tautòmer *cisB* abans i després de la seva optimització (en vermell), comparant-la amb la geometria optimitzada mitjançant DFT (en verd).

Considerant que la discrepància amb els resultats bibliogràfics pot ser deguda a l'assoliment d'un mínim local en la superfície d'energia potencial, es realitza l'optimització de la geometria de l'estat fonamental mitjançant el mètode post-HF MP2, que correspon al mètode emprat en la bibliografia. La taula 7.17 mostra la diferència energètica obtinguda en ambdós tipus de càlculs (DFT B3LYP i HF MP2), després de realitzar sobre cadascun d'ells una avaluació de l'energia (*single point*) per tal de poder obtenir energies comparables.

**Taula 7.17:** Resultats comparatius dels mètodes HF MP2 i DFT B3LYP per a la determinació de l'energia de l'estat fonamental del porficè.

Tautòmer	MP2 \H	DFT B3LYP \H
<i>trans</i>	-983.23	-989.55
<i>cisA</i>	-983.23	-989.55
<i>cisB</i>	-983.19	-989.50

Tot i que l'optimització de la geometria amb el mètode MP2 permet reduir considerablement el temps de càlcul, els resultats obtinguts no poden ser directament utilitzats per les següents etapes de càlcul (càlcul de freqüències i TD-DFT), fent necessari el pas addicional corresponent al *single point*.

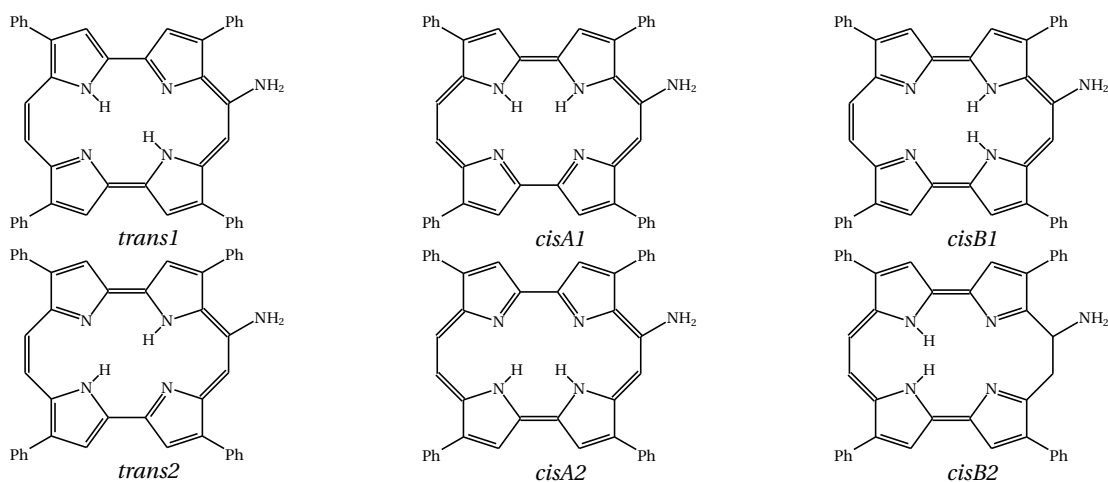
La comparació dels resultats obtinguts pel tautòmer *cisB* es mostra a la taula 7.18, en la qual es mostra l'energia dels seus estats  $S^0$  i  $S^1$ , prenent com a referència l'energia de l'estat fonamental del tautòmer *trans*. Cal destacar que l'ús de HF MP2 permet reduir encara més l'energia de l'estat  $S^0$ , allunyant-se del resultat bibliogràfic.

**Taula 7.18:** Comparació dels resultats computacionals obtinguts en l'estudi del tautòmer *cisB* del porficè.

Estat de <i>cisB</i>	Gil <i>et al.</i> /eV	DFT B3LYP/eV	HF MP2/eV
$S_0$	2.79	1.28	0.98
$S_1$	3.03	3.42	3.12

### 7.3.2 Els sis tautòmers del 9ATPPO

L'asimetria introduïda pel grup amino fa que el nombre de tautòmers possibles en el 9ATPPO sigui el doble que en el cas del porficè. Tal i com es pot veure en la figura 7.17 cap d'ells és equivalent i això fa que s'hagin d'estudiar tots computacionalment.<sup>190</sup>

**Figura 7.17:** Representació molecular dels sis tautòmers del 9ATPPO.

Amb la finalitat d'establir la identitat de cadascuna de les espècies que absorbeixen, s'estima l'energia dels estats fonamentals dels sis tautòmers a partir de l'optimització de la seva geometria mit-

jançant el mètode DFT B3LYP/6-31G(d). Tal com s'ha comentat anteriorment, els resultats obtinguts amb aquesta metodologia es troben d'acord amb els resultats descrits<sup>165</sup> pels tautòmers *trans* i *cisA*, encara que l'energia i la geometria dels tautòmers *cisB* necessitarien estudis més acurats. Ara bé, tot i que les energies predites per l'estat fonamental de *cisB1* i *cisB2* estan sotmeses a un major error, aquestes són prou elevades com per què els dos tautòmers no es considerin termodinàmicament rellevants per l'estudi d'estats fonamentals.

L'optimització de la geometria es complementa amb l'estimació del ZPE mitjançant un càlcul de freqüències mantenint el mateix nivell de càlcul (B3LYP/6-31G\*) que serveix per corregir el valor de l'energia d' $S_0$  i garantir l'obtenció d'un mínim. Tal com s'ha descrit pel Po, l'estudi dels estats excitats es realitza per mitjà d'un TD-DFT B3LYP/6-31G\*. Els resultats d'aquests tres càlculs, per cadascun dels sis tautòmers, es mostren a la taula 7.19.

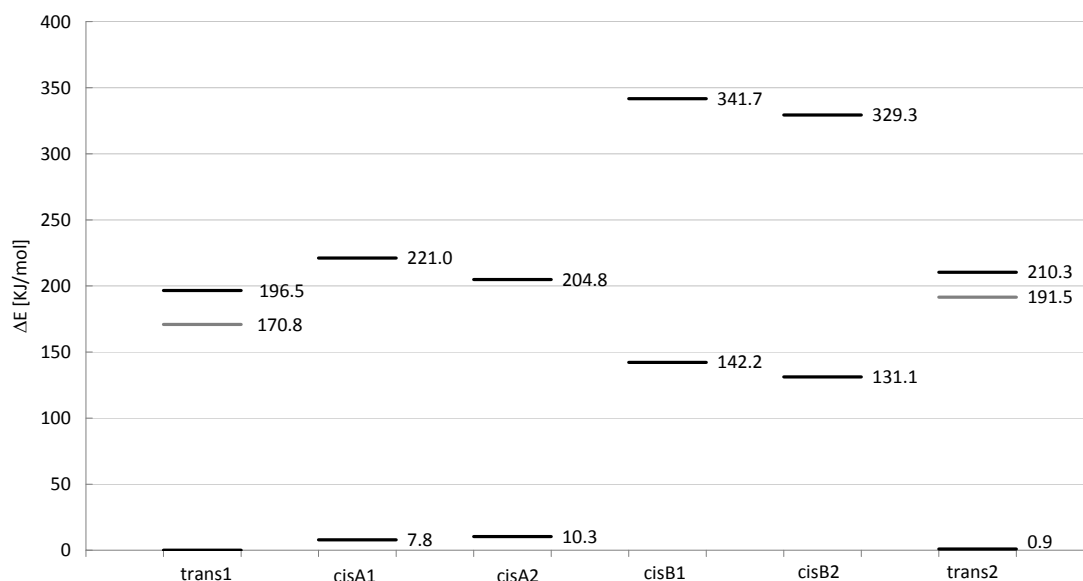
**Taula 7.19:** Resultats computacionals per a l'energia de l'estat fonamental dels sis tautòmers del 9ATPPo i el valor del màxim d'absorció predit amb TD-DFT.

Tautòmer	E(SCF)/H	$\lambda_{max}/nm$
<i>trans1</i>	-1969.131	669.42
<i>cisA1</i>	-1969.127	616.89
<i>cisA2</i>	-1969.126	676.11
<i>cisB1</i>	-1969.081	659.17
<i>cisB2</i>	-1969.085	663.49
<i>trans2</i>	-1969.130	628.16

Els resultats computacionals mostren que el *trans1* és el tautòmer de menor energia en estat fonamental. Així doncs, si es pren aquest compost com a referència, a partir del qual es calcula l'energia dels estats  $S^0$  i  $S^1$ , s'obtenen els resultats de la taula 7.20, els quals també es representen en la figura 7.18.

**Taula 7.20:** Diferència energètica entre els sis tautòmers del 9ATPPo, prenent l'energia de l'estat fonamental de *trans1* de referència.

Tautòmer	$S_0/kJ \cdot mol^{-1}$	$S_1/kJ \cdot mol^{-1}$
<i>trans1</i>	0	196.49
<i>cisA1</i>	7.83	221.04
<i>cisA2</i>	10.30	204.85
<i>cisB1</i>	124.19	341.73
<i>cisB2</i>	131.09	329.33
<i>trans2</i>	0.91	210.30



**Figura 7.18:** Diagrama energètic entre els diferents tautòmers del 9ATPPO, prenent l'energia de  $S_0$  de *trans1* com a referència. Els resultats experimentals pels dos tautòmers *trans* es mostra en gris.

**Taula 7.21:** Energia d'excitació (eV i nm), força de l'oscil·lador i configuracions de les transicions  $S_1 \leftarrow S_0$  i  $S_2 \leftarrow S_0$  per a cadascun dels tautòmers del 9ATPPO, en comparació amb el Po i TPPo.

Tautòmer	Transició	E/eV(nm)	f	Configuració
<i>cisA1</i>	$S_1$	2.01 (616.9)	0.379	63.05%(0-0) + 10.09%(1-1) + 8.72%(1-0)
	$S_2$	2.18 (569.9)	0.2517	63.72%(1-0) + 16.05%(0-1) + 5.92%(0-0)
<i>cisA2</i>	$S_1$	1.83 (676.1)	0.2935	66.76%(0-0) + 8.86%(1-1) + 4.43%(1-0)
	$S_2$	2.19 (566.4)	0.2861	63.65%(1-0) + 20.33%(0-1) + 3.08%(0-0)
<i>cisB1</i>	$S_1$	1.88 (659.2)	0.3129	69.19%(0-0) + 10.14%(1-1) + 2.24%(2-0)
	$S_2$	2.15 (576.7)	0.1368	66.26%(1-0) + 13.64%(0-1) + 6.39%(2-0)
<i>cisB2</i>	$S_1$	1.87 (663.5)	0.2515	70.45%(0-0) + 10.38%(1-1) + 2.17%(2-0)
	$S_2$	2.18 (568.6)	0.126	68.64%(1-0) + 19.16%(0-1) + 2.94%(3-0)
<i>trans1</i>	$S_1$	1.85 (669.4)	0.3088	68.85%(0-0) + 10.39%(1-1) + 2.84%(1-0)
	$S_2$	2.24 (553.4)	0.3313	65.84%(1-0) + 19.64%(0-1)
<i>trans2</i>	$S_1$	1.97 (628.2)	0.4383	67.01%(0-0) + 10.82%(1-1)
	$S_2$	2.15 (577.5)	0.1714	71.13%(1-0) + 19.15%(0-1)
TPPo	$S_1$	2.13 (582.7)	0.2191	68.41%(1-0) + 14.98%(0-1) + 4.93%(0-0)
	$S_2$	2.25 (549.9)	0.4017	64.09%(0-0) + 18.99%(1-1) + 4.22%(1-0)
Po	$S_1$	2.23 (555.5)	0.1265	64.67%(1-0) + 21.01%(0-1) + 4.21%(0-0)
	$S_2$	2.36 (525.7)	0.1986	62.01%(0-0) + 19.91%(1-1) + 3.68%(1-0)

De forma global, el conjunt dels tautòmers *trans* presenten una energia calculada suficientment baixa com per ésser considerats les úniques espècies existents en solució a temperatura ambient. Les prediccions computacionals es veuen corroborades pels resultats experimentals obtinguts al GEM, en els quals no s'observa un canvi remarcable de l'espectre d'absorció del 9ATPPO amb la temperatura, posant de manifest que *trans1* i *trans2* presenten en el seu estat fonamental, una energia semblant.

Juntament amb les energies de les transicions  $S_1 \leftarrow S_0$  obtingudes per TD-DFT, el conjunt dels resultats computacionals sobre *trans1* i *trans2* (taula 7.21), serveixen per realitzar una assignació de les bandes de observades en l'espectre de fluorescència.

**Energia i geometria dels tautòmers** Finalment, s'estudia quina millora suposaria, sobre el resultat de la predicció del màxim d'absorció, la incorporació de funcions difoses dins les bases emprades en la definició del càlcul TD-DFT. Concretament, s'ha avaluat l'ús de TD-DFT B3LYP/6-31+G\* per la descripció de les transicions  $S_1 \leftarrow S_0$  dels dos tautòmers *trans* (taula 7.22). La lleugera millora en la precisió del càlcul no justifica l'increment, en un factor de set, del temps de càlcul.

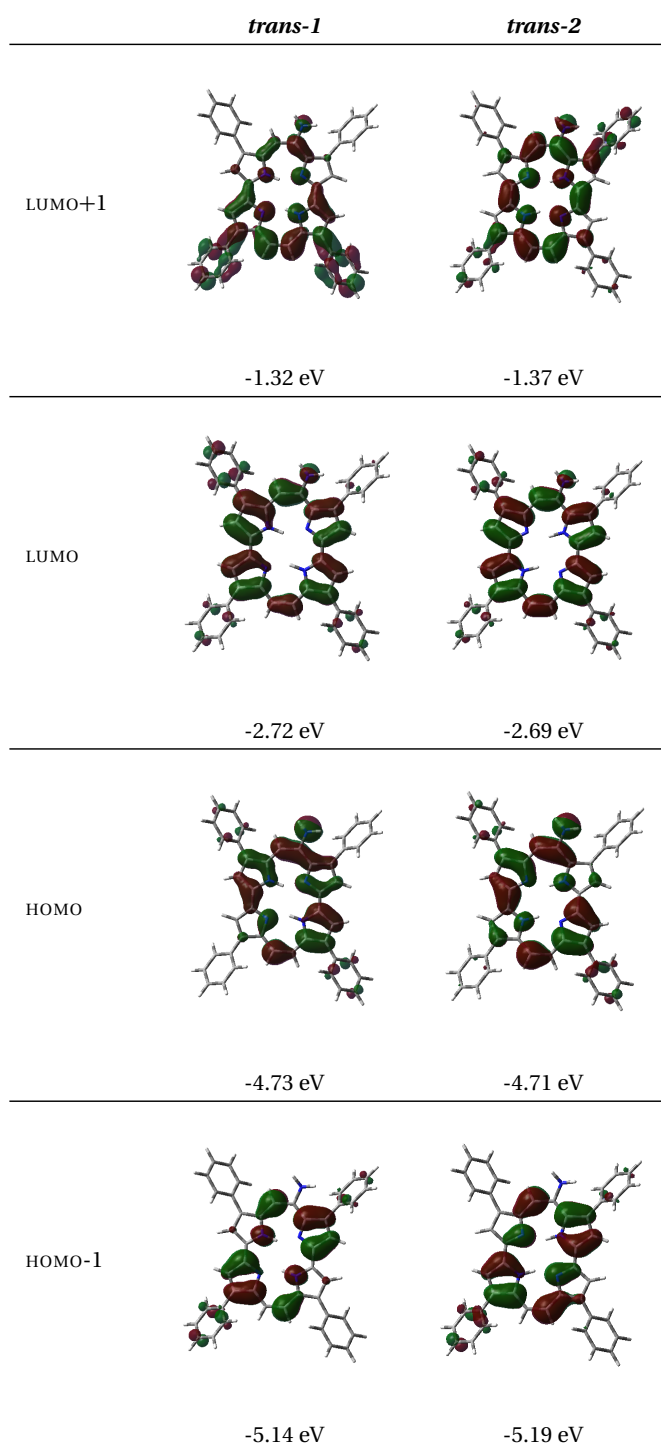
**Taula 7.22:** Aplicació de funcions difoses dins les bases del mètode de càlcul TD-DFT per una millor descripció de l'energia de l'estat excitat.

Tautòmer	6-31G*/nm	6-31+G*/nm	Experimental/nm
<i>trans1</i>	669.42	674.31	770
<i>trans2</i>	628.16	635.70	693-697

**Orbitals Naturals** Els orbitals HOMO-1, HOMO, LUMO i LUMO+1 dels sis tautòmers del 9ATPPo obtinguts del càlcul TD-DFT resulten semblants a aquells trobats pel Po (figura 7.2, pàg. 153), fet que col·labora a la validació del mètode de càlcul. Com es pot observar en la figura 7.19, aquests orbitals coincideixen també amb els quatre orbitals de Gouterman definits per als porfírens.

El model de Gouterman permet analitzar de forma qualitativa l'espectre d'absorció de porfirines, definint cadascuna de les bandes Q a partir de l'estudi d'uns orbitals HOMO-1, HOMO, LUMO i LUMO+1<sup>154</sup> deduïts per combinació entre el mètode OM-CLOA (mètode de Hückel) i una simplificació del tractament CI.<sup>194</sup> Tot i així, s'ha comprovat que la seva descripció continua sent vàlida per a la descripció d'altres derivats tetrapirròlics, com els porfírens.<sup>153</sup>



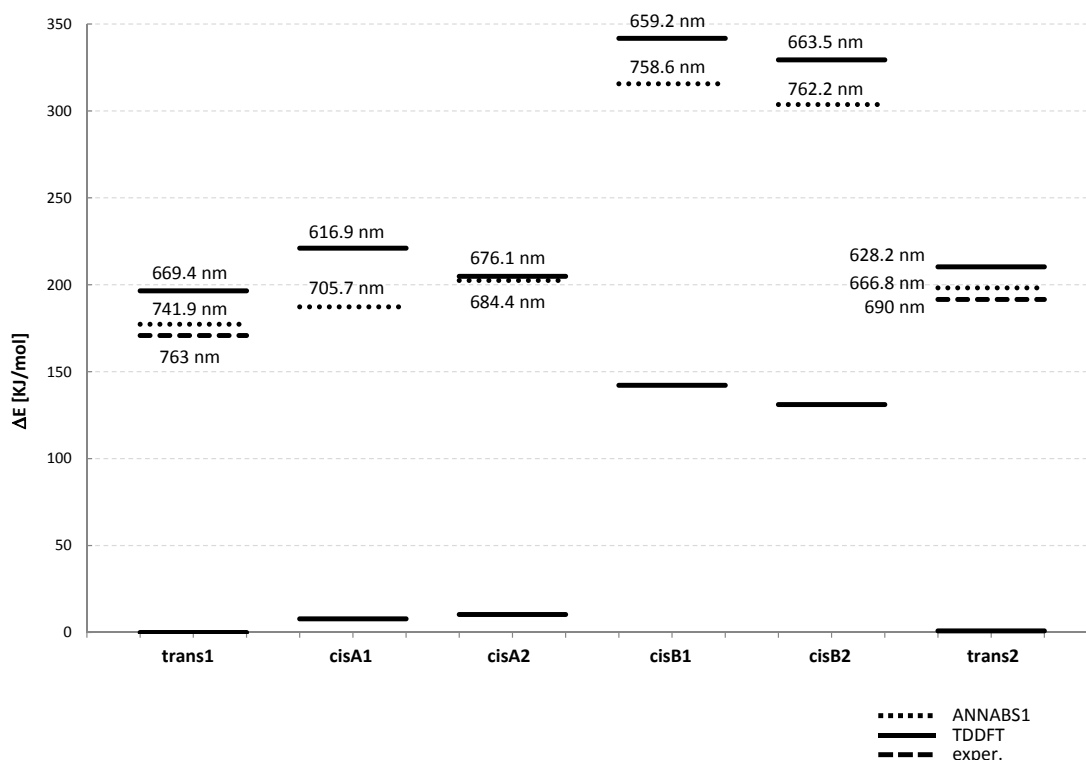


**Figura 7.19:** Representació dels orbitals naturals calculats pels tautòmers *trans* del 9ATPPO.

Els resultats corresponents a l'estudi dels tautòmers del 9ATPPO han estat publicats,<sup>190</sup> veure annex IV

### 7.3.3 Aplicació del model ANNABS

El model QSPR desenvolupat anteriorment (apartat 7.2, pàg.150) de predicció de l'energia de la transició  $S_1 \leftarrow S_0$ , s'aplica per a cadascun dels tautòmers del 9ATPPo. S'aplica el model ANNABS1 de predicció per a derivats porfíricènics i es compara el valor obtingut tant amb els càlculs TD-DFT com amb els valors experimentals disponibles pels dos tautòmers trans.



**Figura 7.20:** Comparació dels resultats obtinguts amb TD-DFT i el model ANNABS1 en la predicció de l'energia dels estats excitats  $S_1$  de cadascun dels tautòmers del 9ATPPo. La diferència energètica corresponent a l'eix d'ordenades es calcula prenent l'estat fonamental de trans1 com a referència. En cada estat excitat es mostra també la longitud d'ona predita per a la transició  $S_1 \leftarrow S_0$ .

Es pot observar que el mètode de càlcul basat en ANN és capaç de diferenciar entre tautòmers, permetent obtenir una bona predicció per aquells dels quals es disposa de valor experimental, taula 7.20.

## 7.4 Determinació del logP de FS *in silico*

La hidrofobicitat, expressada com el coeficient de partició octanol/aigua (logP), es considera normalment com un paràmetre clau a l'hora d'establir models QSAR de predicció de l'activitat de diferents famílies de FS<sup>172, 195-200</sup> i en l'estudi dels processos de transport intra i extracel·lulars, fet que condiciona també la seva localització subcel·lular (veure capítol 8).

### 7.4.1 Mètodes de càlcul del logP

El valor de logP es pot estimar computacionalment mitjançant diferents metodologies. Es poden agrupar en diferents conjunts d'acord amb el nivell de teoria emprat:<sup>201,202</sup>

- **Mètodes constructivistes:** Mètodes proposats per Fujita *et al.*<sup>203</sup> l'any 1964, on el càlcul del logP es basa en el valor experimental d'una molècula emprada com a patró (*scaffold*), la qual és successivament substituïda fins a obtenir la molècula desitjada. Cada canvi estructural té associat un factor de correcció que afecta a tot el sistema molecular.
- **Mètodes de contribució atòmica:** Proposats inicialment per Broto *et al.* (1984)<sup>204</sup> i posteriorment per Ghose i Crippen,<sup>171</sup> consideren que cada àtom present en el sistema molecular contribueix de forma additiva en el càlcul del logP molecular. Aquest esquema fa que aquests mètodes siguin altament dependents de la parametrització realitzada sobre cadascun dels diferents àtoms. Alguns exemples d'algorismes de càlcul que empen aquesta aproximació són els valors dels descriptors SlogP i logP(o/w) del programari MOE2007.09,<sup>86</sup> la definició AClogP realitzada per Actelion i l'algorisme ALOGP.<sup>205</sup>
- **Mètodes basats en fragments:** De forma anàloga a aquells basats en contribució atòmica, aquests mètodes, desenvolupats inicialment al 1973 de la mà de Rekker *et al.*, segmenten les molècules en els seus fragments constitutius. Cada fragment té associat un valor que esdevé la seva contribució sobre el logP molecular (anomenat *fragmental constant*, el qual s'obté de forma estadística) i un factor de correcció. Sota aquest plantejament es defineixen els algorismes COSMOFrag<sup>206</sup> i miLOGP (desenvolupat per Molinspiration Chemoinformatics).
- **Mètodes quàntics:** Permeten avaluar propietats directament relacionades amb el càlcul del logP (com ara l'energia lliure de solvatació en diferents solvents) i relacionar-les amb la geometria i l'estructura electrònica de la molècula.
- **Mètodes mixtes:** Altres mètodes utilitzen combinacions de diferents descriptors moleculars per a la predicció del valor del logP molecular.<sup>207</sup> Pertanyen a aquest conjunt els algorismes MLOGP, ALOGPs,<sup>208</sup> XLOGP2 i XLOGP3.<sup>209</sup> Finalment, existeixen altres algorismes com per exemple el KowWin, que es basen en la combinació de dos o més mètodes dels exposats anteriorment.

L'existència d'un ventall tant ampli de mètodes per a la predicció del logP, posa en evidència la dificultat de definir correctament el logP d'una determinada molècula. La discussió sobre quin és el millor mètode de predicció continua, de fet, essent un tema de discussió actualment. Aquesta discrepància es veu accentuada en molècules altament hidrofòbiques, som és el cas dels FS, ja que normalment els mètodes de predicció estan validats per molècules petites, amb un caràcter lipòfil relativament baix i que presenten una alta semblança molecular entre si.<sup>210</sup>

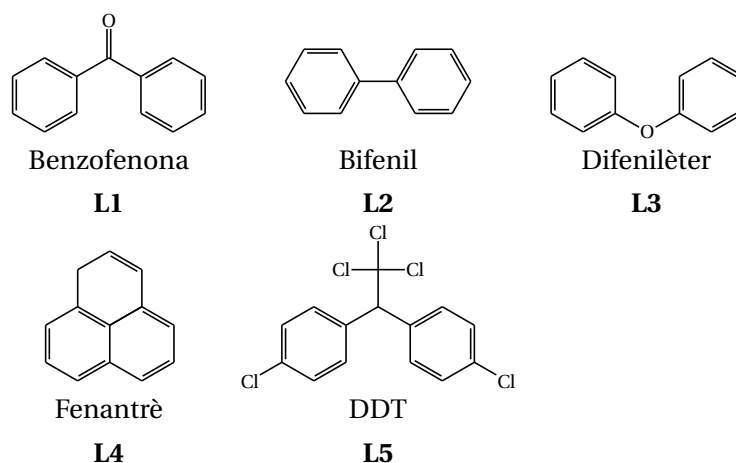
Davant d'aquesta situació, es desitja comparar les metodologies disponibles per l'estudi del logP de molècules altament hidrofòbiques, fent referència principalment als FS. S'avalua l'ús d'un total de dotze algorismes de càlcul, escollits segons la seva disponibilitat, representatius de tots els tipus

de mètodes comentats anteriorment i pr als quals es disposa d'informació bibliogràfica referent a la seva comparació i aplicabilitat.<sup>207,211,212</sup>

La validació dels resultats computacionals es realitza dins del context dels mètodes QSRR (*Quantitative Structure-Retention Relationship*). Aquests mètodes, de forma anàloga als mètodes QSAR o QSPR, pretenen obtenir un model de predicció que permeti relacionar l'estructura química dels membres d'un conjunt d'entrenament amb el seu temps de retenció cromatogràfic. L'interès dels mètodes QSRR apareix en el moment que es demostra que el valor del logP molecular es troba correlacionat linealment amb el temps de retenció cromatogràfic i amb el factor de capacitat ( $k'$ ) associat en HPLC.<sup>173,209</sup> Gràcies a la col·laboració amb altres investigadors del GEM, es disposa d'un conjunt de mesures experimentals del logP d'una sèrie de porfirines i porficens, que poden ser emprades per establir el model QSRR.<sup>213</sup>

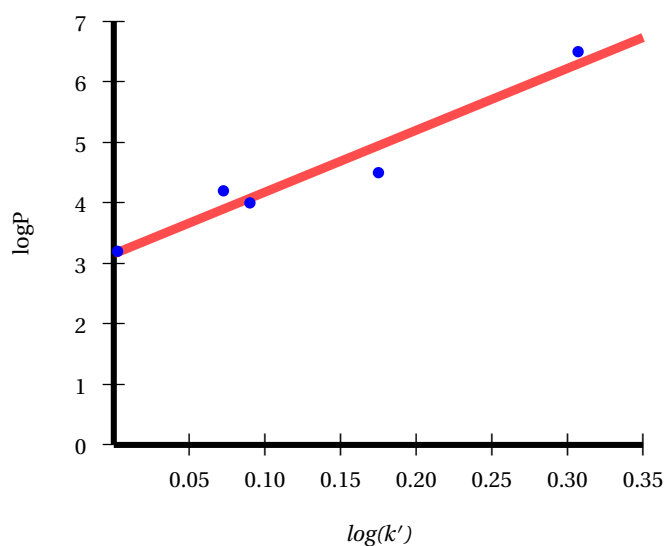
#### 7.4.2 La base de dades

Les dades experimentals s'obtenen seguint el protocol de l'OECD.<sup>214</sup> Aquestes directrius estableixen les condicions experimentals necessàries pel càlcul del valor del logP a partir de mesures cromatogràfiques d'HPLC. Segons aquest protocol, la validació del mètode experimental es realitza emprant un conjunt de cinc molècules de baix pes molecular (difenílèter, DDT, fenantrè, bifeníl i benzofenona, figura 7.21). Amb elles, s'obté la recta de calibratge sobre la qual s'extrapolen les mesures d'HPLC per obtenir el logP dels porficens d'interès.



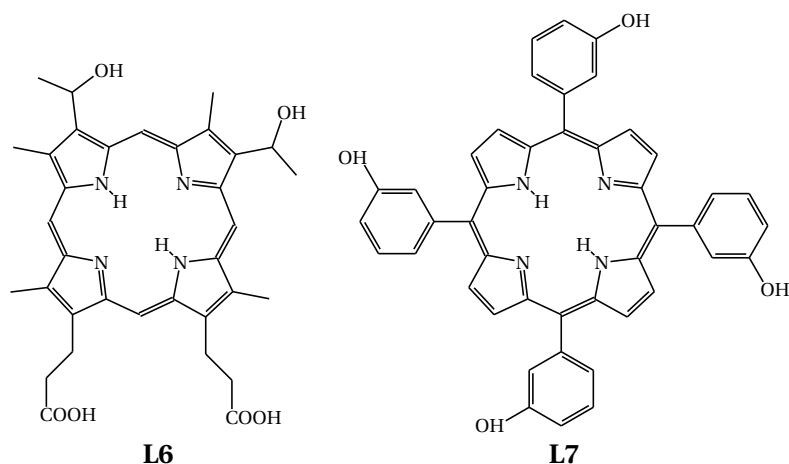
**Figura 7.21:** Compostos emprats com a referència per establir la recta OECD de calibratge.

Malauradament, el marge d'aplicabilitat de la recta OECD es veu limitat pel valor del logP dels compostos de referència, fixant-lo a valors de logP compresos entre 3 i 6.5 (figura 7.22). Com es pot entreveure, aquests valors es troben molt lluny del logP que poden presentar els sistemes porficènics en estudi (figures 7.23, 7.24 i 7.25). Per aquest motiu cal estudiar la viabilitat d'extrapolar aquesta recta de calibratge a valors de logP superiors,<sup>205,215</sup> per tal de fer-la vàlida per porfirines i porficens.

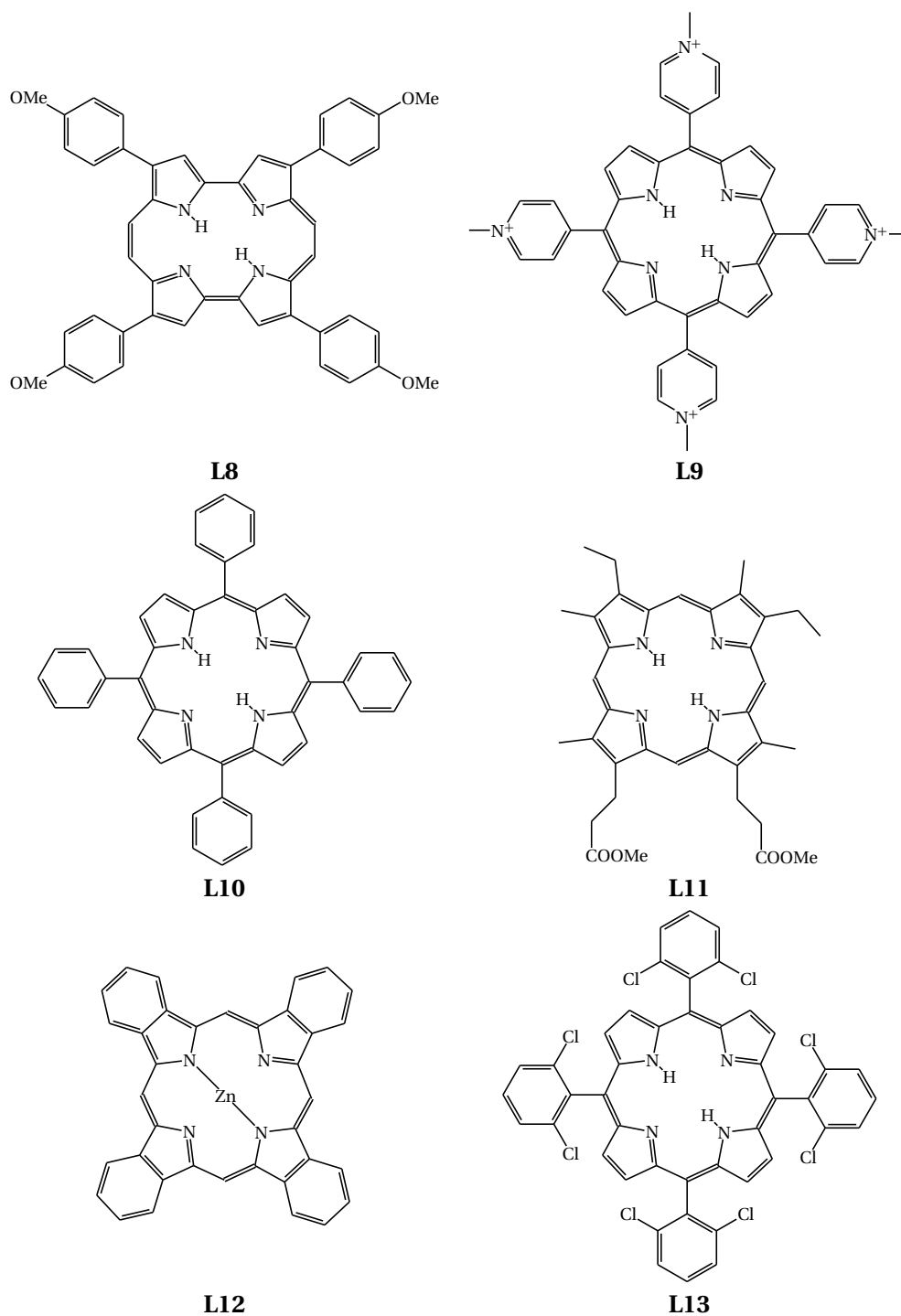


**Figura 7.22:** Recta OECD de regressió obtinguda experimentalment per les cinc molècules de referència **L1-L5** ( $R^2 = 0.94$ ).

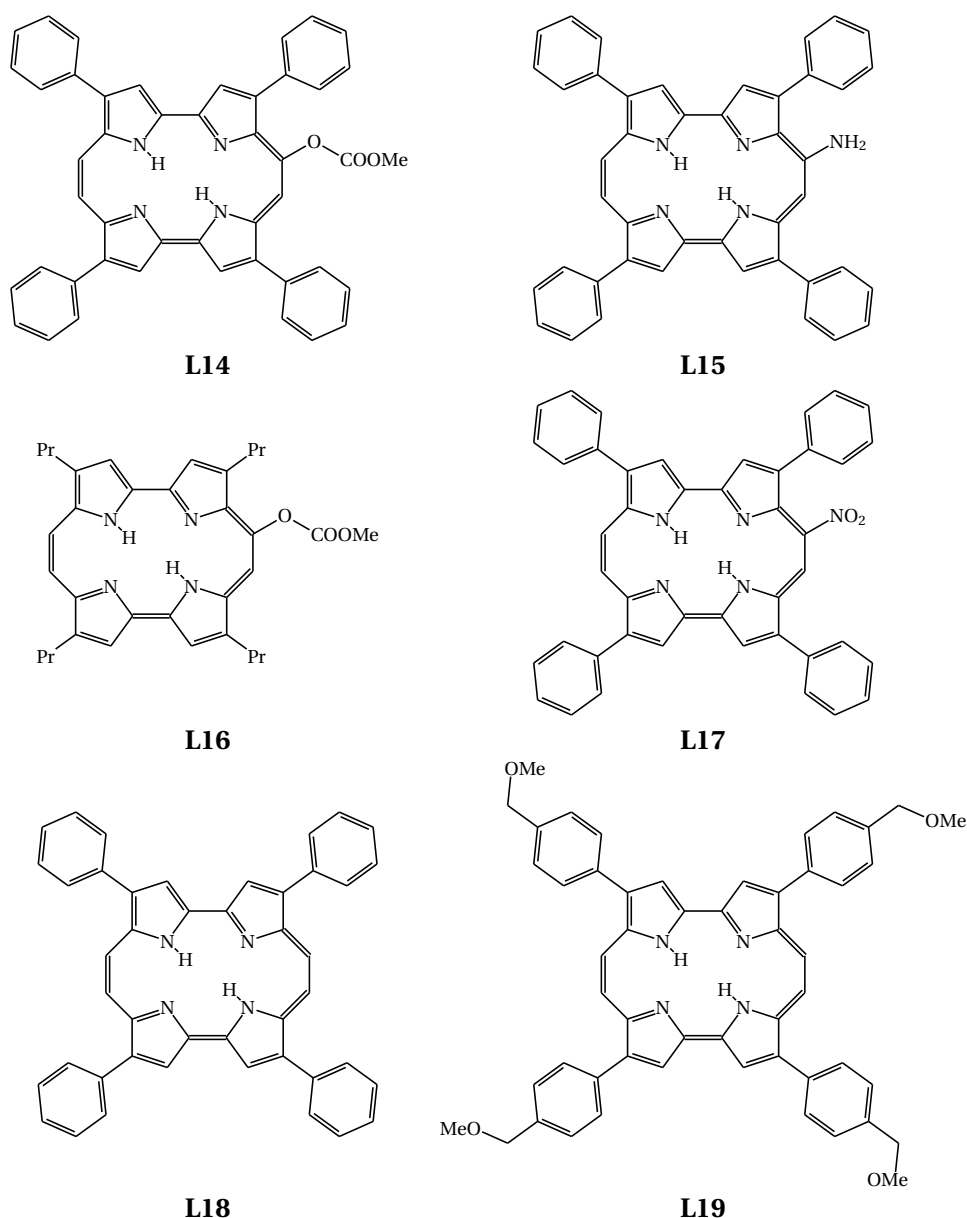
Amb aquesta finalitat es construeix una quimioteca formada per 14 derivats porfirínics i porficènics dels quals s'han establert els paràmetres cromatogràfics segons les directrius OECD. Independentment de la recta de calibratge obtinguda per molècules petites, s'avalua la correlació entre el valor experimental de  $\log k'$  i el valor de  $\log P$  calculat segons els diferents mètodes considerats, mitjançant el programari StatGraphics.<sup>216</sup> Per tal de poder realitzar una validació del mètode, es divideix la quimioteca inicial en un conjunt d'entrenament (**L6-L13**, figures 7.23 i 7.24) i un conjunt de validació externa (**L14-L19**, figura 7.25). Tots aquests compostos foren escollits segons la seva disponibilitat.



**Figura 7.23:** Estructures moleculars dels sistemes porficènics i porfirínics considerats en l'estudi del  $\log P$  com a conjunt d'entrenament (1a part).



**Figura 7.24:** Estructures moleculars dels sistemes porfíric i porfirínic considerats en l'estudi del logP com a conjunt d'entrenament (2a part).



**Figura 7.25:** Estructures moleculars dels derivats porfírics considerats en l'estudi del logP com a conjunt de validació.

### 7.4.3 Estudi computacional

#### Càlcul del logP

Es calcula el valor del coeficient de partició octanol/aigua (logP) tant dels compostos de referència com pels fotosensibilitzadors segons dotze metodologies computacionals diferents, emprant els programaris MOE2007.09<sup>86</sup> (pels descriptors SlogP i logP(o/w)), ADMEWORKS i ALOGPs 2.1<sup>217,218</sup> (pel càlcul dels valors MLOGP, AClogP, ABlogP, COSMOFrag, miLOGP, KowWin, XLOGP2, XLOGP3, ALOGP i ALOGPs).

Els resultats de la taula 7.23 corresponen a les prediccions computacionals realitzades sobre els compostos de referència, juntament amb el valor del logP experimental. Es pot observar que la major

part dels mètodes de predicció correlacionen linealment amb el valor de logP experimental (segons el coeficient de correlació  $R^2$ ), fet que els consolida per al seu ús. S'inclou també el valor del coeficient de correlació d'Spearman ( $\rho$ ) per avaluar el manteniment de l'ordenació relativa entre les dades.

**Taula 7.23:** Predicció del logP dels compostos emprats per establir la recta OECD.

	OECD					$R^2$	$\rho$
	L1	L2	L3	L4	L5		
logP <sub>exper</sub>	3.2	4.0	4.2	4.5	6.5		
AB/LogP	2.9	3.9	3.4	4.5	6.6	0.95	0.9
AC LogP	3.3	3.7	3.4	3.9	6.9	0.92	0.9
ALOGP	3.2	3.4	3.4	3.4	6.3	0.87	0.7
ALOGPs	3.0	4.0	3.7	4.4	6.3	0.97	0.9
COSMOFrag	2.9	3.8	3.9	3.5	6.1	0.94	0.7
KowWin	3.2	3.8	4.1	4.4	6.8	0.99	1.0
logP(o/w)	3.4	3.9	3.5	4.4	6.5	0.93	0.9
miLogP	3.4	3.7	3.7	3.6	6.7	0.89	0.6
MLOGP	3.6	3.9	3.4	4.2	6.2	0.88	0.7
SlogP	2.9	3.4	3.5	3.4	6.5	0.93	0.9
XLOGP2	3.6	3.9	3.5	4.3	6.7	0.90	0.7
XLOGP3	3.4	4.0	4.2	3.9	6.9	0.94	0.7

Els resultats obtinguts per als fotosensibilitzadors considerats es mostren en la taula 7.24. Cal tenir en compte que els valors de logP d'aquests compostos es troben lluny del marge d'aplicabilitat de la recta OECD. Com es pot observar, la presència de càrregues formals en L9 i L12, així com la coordinació amb metalls de transició, poden esdevenir limitacions inherents d'alguns dels mètodes basats en fragments o en contribució atòmica, degut a les parametritzacions internes d'aquests.

**Taula 7.24:** Predicció del logP dels derivats porfirínics i porfircènics estudiats segons els diferents mètodes de càlcul.

Mètode	Conjunt d'entrenament								Conjunt de validació					
	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19
log k'	0.14	0.47	1.40	0.32	1.36	0.64	1.19	1.25	0.95	1.21	1.33	1.39	1.51	1.24
AB/LogP	1.3	6.3	10.0	–	8.8	5.8	–	10.0	9.6	10.0	8.9	9.2	10.0	8.8
AC LogP	3.3	6.4	7.1	2.7	7.6	6.2	7.6	12.5	8.1	7.1	7.6	7.3	7.5	6.8
ALOGP	6.7	10.2	11.1	7.1	11.2	9.5	–	16.5	10.9	10.4	10.7	11.1	11.2	10.4
ALOGPs	2.7	6.8	7.8	6.8	8.0	5.5	4.6	10.1	7.9	7.3	7.3	7.5	8.2	7.6
COSMOFrag	4.3	8.5	12.4	–	11.6	8.8	–	14.5	12.0	11.1	11.1	11.4	11.4	14.8
KowWin	4.4	9.5	11.8	8.9	11.5	11.4	8.8	16.6	11.1	10.1	10.6	10.7	11.5	10.6
logP(o/w)	4.3	9.9	11.5	9.6	11.0	6.5	5.6	15.7	11.7	11.0	8.7	11.7	11.7	12.0
miLogP	5.5	9.1	9.7	6.5	9.6	8.6	7.0	10.2	9.5	9.5	9.2	9.6	9.6	9.6
MLOGP	1.3	3.3	3.9	2.2	5.3	3.1	–	7.6	5.1	4.8	4.0	5.3	5.3	3.5
SlogP	3.9	6.5	8.6	4.4	7.6	5.5	5.6	12.9	8.5	7.9	6.5	8.2	8.6	10.3
XLOGP2	0.9	7.8	9.1	4.3	9.5	3.5	7.1	14.4	9.6	9.0	6.7	9.7	9.5	8.3
XLOGP3	2.1	9.0	9.4	6.7	10.4	5.3	6.4	15.4	9.3	8.8	8.4	9.3	9.5	8.1



**Ajust lineal**

Per tal d'identificar el millor mètode de càlcul per a la predicció del logP de FS, es du a terme una anàlisi estadística que avalui la correlació entre el valor de logP predit i el de log  $k'$  experimental. Aquest estudi es realitza en primer lloc sobre aquells FS per als quals es disposa del temps de retenció cromatogràfic experimental, corresponents al conjunt d'entrenament. Per analogia a les directrius OECD, es considera que el mètode de predicció més adequat és aquell que permet ajustar millor les dades experimentals a una recta.

L'anàlisi estadística es realitza mitjançant el programari StatGraphics 5.1,<sup>216</sup> avaluant la correlació de cadascun dels mètodes de càlcul amb les dades experimentals. La comparació de tots els models obtinguts permet identificar l'existència d'un possible punt erroni, corresponent al compost **L13**.

Aquests resultats es veuen confirmats experimentalment degut a un possible problema amb aquest compost relacionat amb l'establiment d'interaccions amb la fase estacionària de la columna cromatogràfica emprada. Per aquest motiu, el compost **L13** se suprimeix del conjunt d'entrenament i es realitza l'anàlisi estadística de nou sense ell.

Els resultats obtinguts per a cada regressió lineal, amb el conjunt d'entrenament, es mostren resumits a la taula 7.25, on s'especifica el mètode de càlcul, el coeficient de correlació  $R^2$  i el valor  $P$  (calculat a partir del test de Fisher), mitjançant el qual hom pot avaluar si l'ajust és significatiu. També s'hi enumeren els FS considerats punts erronis pel mètode estadístic de regressió i els valors de  $q^2$  (pàg. 88)<sup>219</sup> i  $\rho$  (només per als compostos que formen part del conjunt de validació).

**Taula 7.25:** Resultats obtinguts en l'ajust lineal del valor de logP calculat i el valor de log  $k'$  experimental per cadascun dels mètodes de càlcul estudiats sobre el conjunt d'entrenament, juntament amb els paràmetres estadístics  $q^2$  i el coeficient d'Spearman  $\rho$  per al conjunt de validació externa.

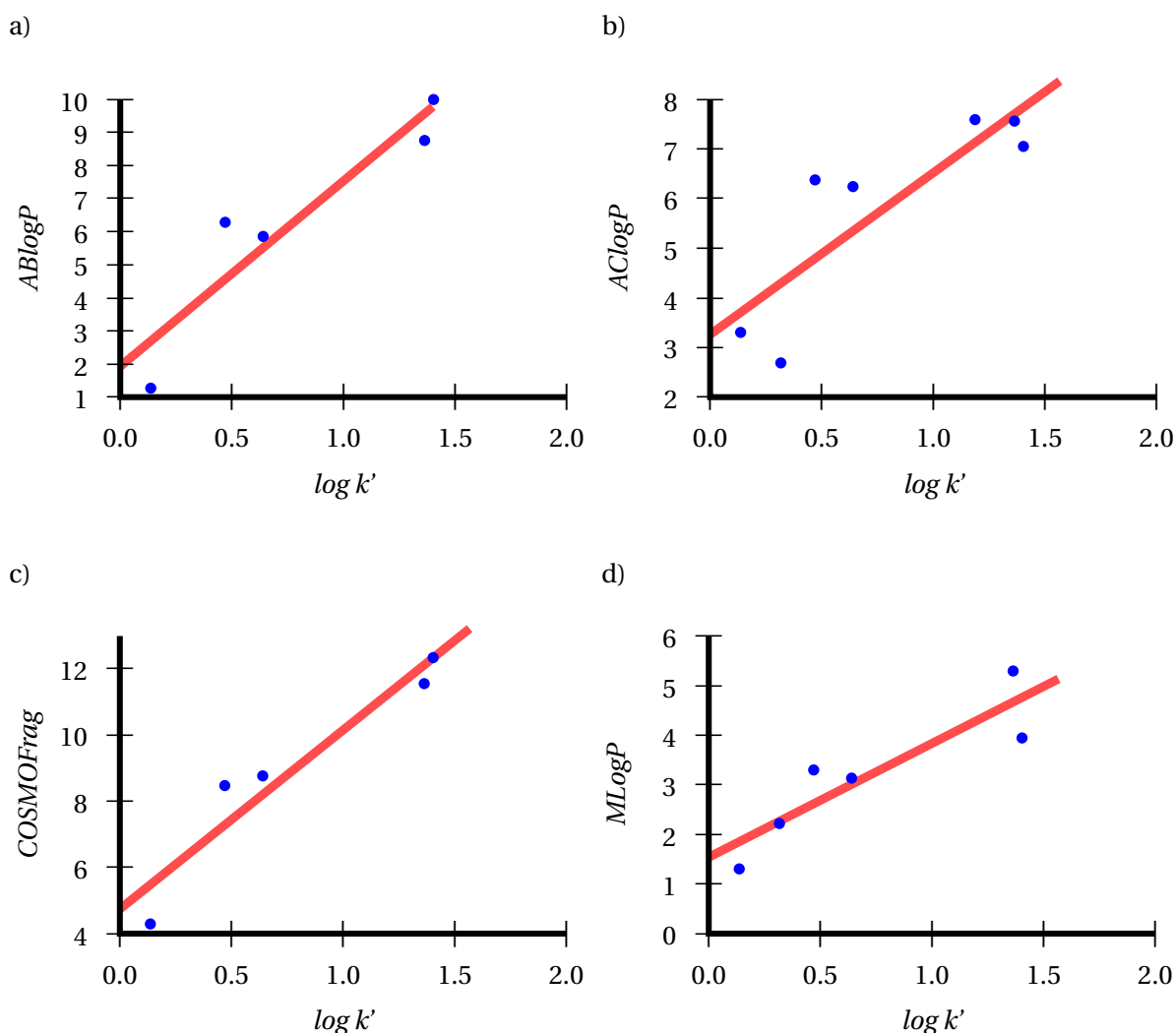
Mètode de predicció	$P$	$R^2$	$Q^2$	$\rho_{test}$	Punts erronis
AB/LogP	0.020	0.87	0.92	-0.26	<b>L9, L12</b>
AC LogP	0.016	0.72	0.81	-0.43	
ACD/logP	0.140	0.57	-	-	<b>L12</b>
Admeworks	0.041	0.60	-	-	<b>L9</b>
ALOGP	0.006	0.94	0.90	-0.09	<b>L7, L12</b>
ALOGPs	0.173	0.51	0.94	0.31	<b>L6, L12</b>
COSMOFrag	0.011	0.92	0.87	0.43	<b>L9, L12</b>
KowWin	0.006	0.99	0.87	0.66	<b>L6, L11, L12</b>
logP(o/w)	0.020	0.96	0.72	0.09	<b>L6, L11, L12</b>
miLOGP	0.013	0.90	0.85	-0.09	<b>L7, L12</b>
MLogP	0.015	0.81	0.64	-0.37	<b>L12</b>
SlogP	0.002	0.98	0.19	0.60	<b>L7, L12</b>
XLOGP2	0.002	0.677	0.72	-0.43	
XLOGP3	0.035	0.710	0.87	-0.31	<b>L7</b>

Es pot observar com els mètodes fragmentals, basats en l'estructura molecular, presenten una dispersió de resultats considerable en funció de la definició dels grups presents dins el sistema molecular i els factors de correcció emprats. Molts d'ells requereixen de la consideració d'un o més punts erronis per garantir l'ajust lineal de les dades. Per aquest motiu, encara que s'obtenen uns valors de correlació molt elevats, es considera que els ajustos realitzats amb un major nombre de punts són

més significatius. Per altra banda, es penalitzen també aquells mètodes que necessiten considerar com erronis els punts associats a molècules sense cap dificultat associada, com poden ser L7 o L11.

En disposar només dels valors de  $\log k'$  experimentals, els ajustos lineals es validen (amb els compostos L14-L19) comparant els valors de  $\log P$  calculats computacionalment amb aquells predits a través de l'ajust.

Segons aquest criteri, s'identifiquen els quatre millors models lineals per a la correlació amb els valors de  $\log k'$  experimentals: AC LogP, AB/LogP, COSMOFrag i MLogP (figura 7.26 i taula 7.26).



**Figura 7.26:** Representació gràfica dels ajustos lineals obtinguts a partir dels mètodes AB/LogP(a), AC LogP(b), COSMOFrag(c) i MLogP(d).

**Taula 7.26:** Models lineals pels quatre millors ajustos obtinguts entre el logP i log  $k'$  a partir del conjunt d'entrenament. Juntament amb el coeficient de correlació obtingut, s'indica el nombre de molècules emprades ( $n$ ) en l'establiment de l'ajust.

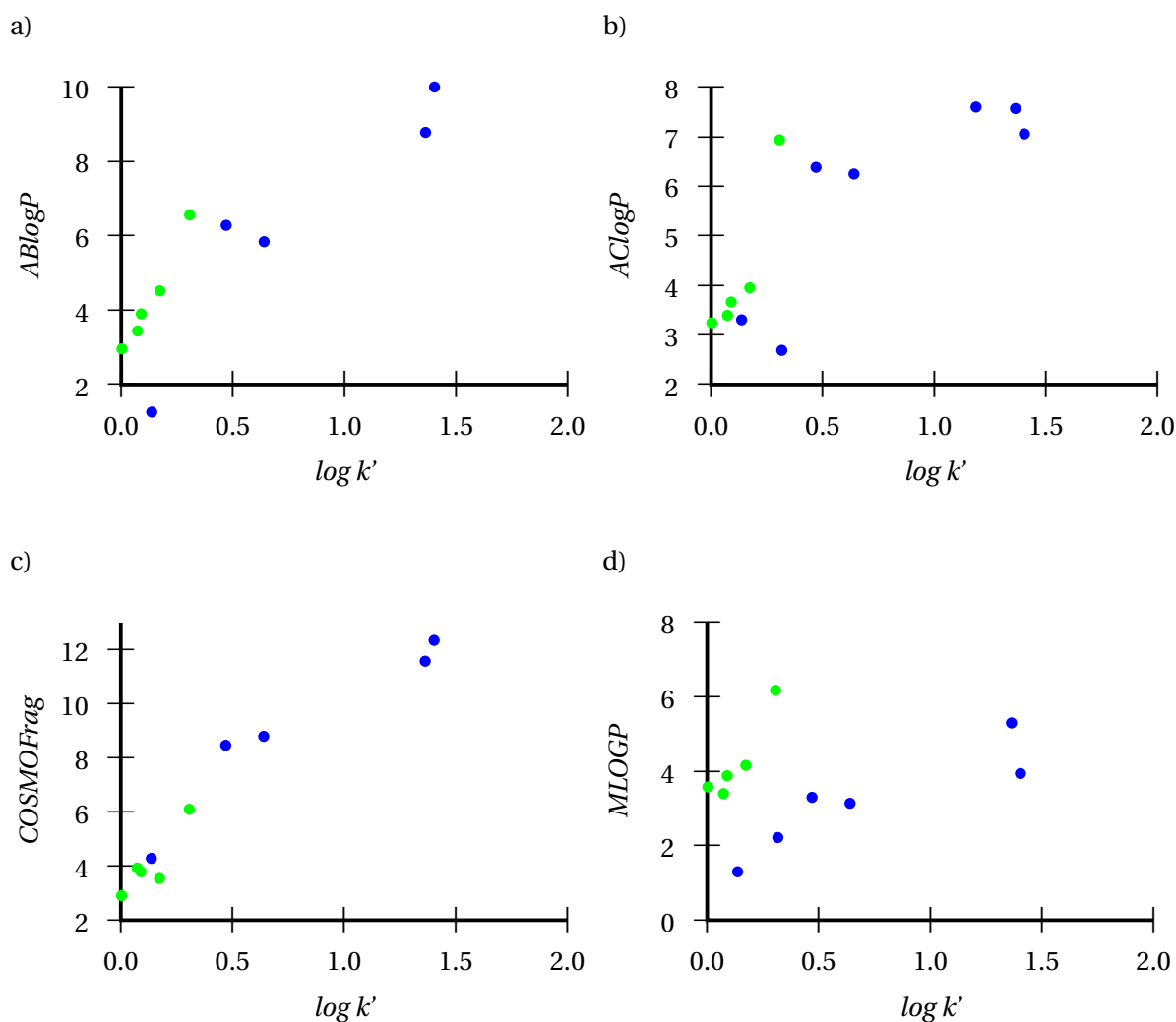
Mètode	$n$	Model	$R^2$
AB/LogP	5	$\log P = 5.61 \cdot \log k' + 1.93$	0.873
AC LogP	7	$\log P = 3.26 \cdot (\log k' + 1)$	0.718
COSMOFrag	5	$\log P = 5.42 \cdot \log k' + 4.74$	0.915
MLogP	6	$\log P = 2.30 \cdot \log k' + 1.54$	0.806

### Comparació entre els models de predicció del logP i la recta OECD

Malauradament no es disposa de valors experimentals del logP de derivats de porfirines o porfircens que permetin discutir els models lineals anteriors. Els valors del logP d'algunes porfirines trobats a la bibliografia corresponen a sistemes iònics o substituïts de tal manera que es disminueixi la seva hidrofobicitat fins a valors de logP que cauen dins del marge d'aplicabilitat de la recta OECD.<sup>220–225</sup> Per aquest motiu, no és possible validar l'extrapolació realitzada pels models lineals obtinguts anteriorment amb valors experimentals de logP.

Tot i així, aquests models haurien de complir-se també dins del marge OECD, de manera que s'avalua la predicció del logP de les molècules **L1-L5** per a cadascun dels quatre models lineals. Hom espera que el millor model permeti ajustar correctament tant les molècules del conjunt d'entrenament com les de la referència OECD dins d'una funció contínua, figura 7.27.

Tal i com es pot observar en la figura 7.27, el model proposat pel descriptor COSMOFrag permet garantir una certa continuïtat entre els marges d'aplicabilitat de la recta OECD de referència i el conjunt d'entrenament. Per aquest motiu es pren aquest mètode de càlcul com el definitiu per validar els resultats del conjunt d'entrenament.



**Figura 7.27:** Distribució de les dades estudiades segons les quatre metodologies de càlcul del  $\log P$  proposades a partir del model lineal: ABlogP (a), AClogP (b), COSMOFrag (c) i MLOGP (d). [● conjunt d'entrenament, ● molècules de referència OECD]

### Aplicació del model lineal al conjunt de validació

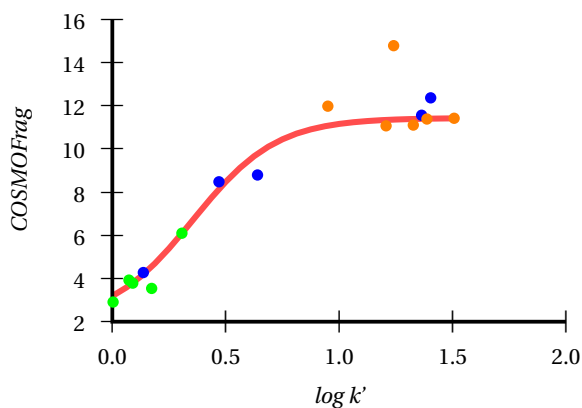
Una vegada identificat el model COSMOFrag com el millor model lineal per a la descripció de la correlació entre els paràmetres cromatogràfics i el valor del  $\log P$ , s'estudia la seva aplicació en sis derivats porfíricènics d'interès en el grup de recerca (figura 7.25, pàgina 189).

Per a cadascuna d'aquestes molècules es calcula el valor de  $\log P$  segons el mètode COSMOFrag. A partir d'aquest valor, i aplicant el mètode de regressió lineal anterior, es realitza una predicció del paràmetre  $\log k'$  i es compara amb el mesurat experimentalment, taula 7.27.

**Taula 7.27:** Resultats obtinguts per a la predicció del  $\log P$  i del  $\log k'$  pel conjunt de validació, juntament amb els valors experimentals del  $\log k'$  obtinguts a partir de cromatografia HPLC.

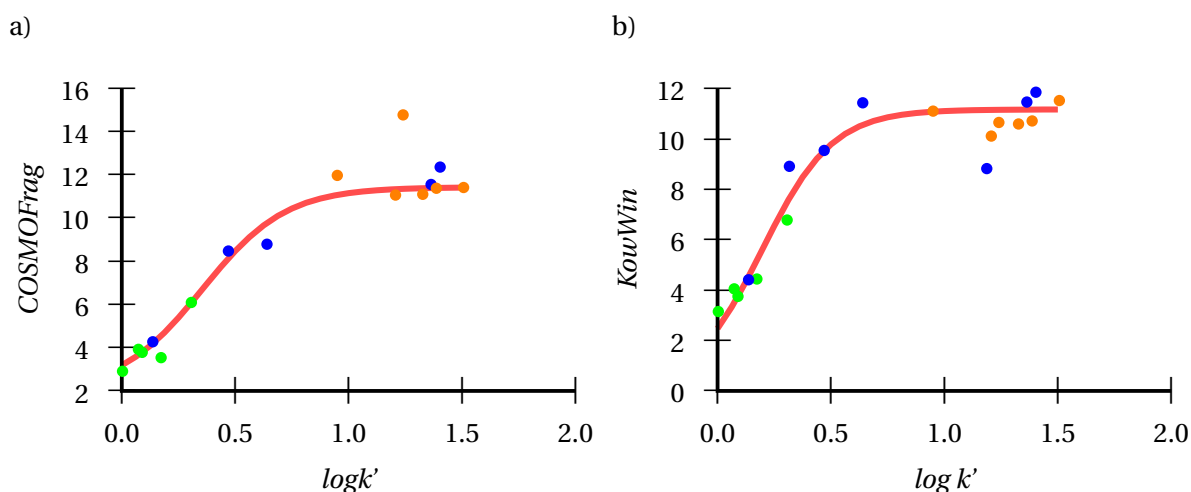
Mol.	COSMOFrag	$\log k'_{calc}$	$\log k'_{exper}$
L14	11.99	1.34	0.95
L15	11.07	1.17	1.21
L16	18.29	2.50	1.33
L17	11.39	1.23	1.39
L18	11.41	1.23	1.51
L19	14.79	1.86	1.24

La representació gràfica de totes les dades exposades fins aquest moment (figura 7.28), permet observar una clara tendència a apartar-se de la linealitat a mesura que augmenta el temps de retenció cromatogràfic. Fins i tot, la seva inspecció visual preliminar, duu a pensar que les dades s'ajusten a una funció asimptòtica (la figura 7.28 conté una línia de tendència sigmoïdal, amb la intenció d' aclarir aquesta idea).



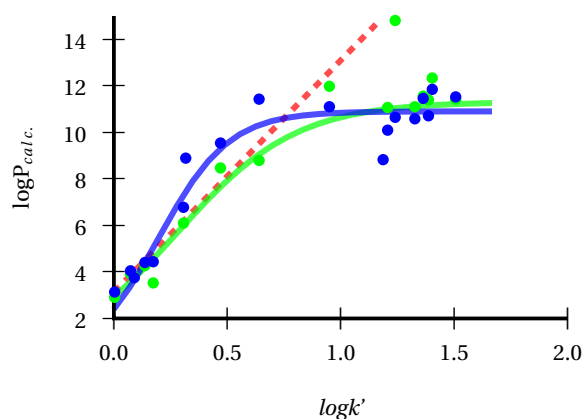
**Figura 7.28:** Representació gràfica dels valors de  $\log P$  calculat mitjançant el mètode COSMOFrag front el valor de  $\log k'$  derivat de les mesures experimentals, pel conjunt de referència (●), el conjunt d'entrenament (●) i el conjunt de validació (●). La línia de tendència (-) pretén il·lustrar de forma orientativa el comportament no lineal del conjunt de les dades experimentals.

Davant d'aquesta possibilitat, es revisen de nou els mètodes que foren descartats inicialment per l'ajust lineal. Aquest cop, se'ls exigeix que mantinguin la continuïtat entre els valors OECD i la resta de punts, i puguin presentar un bon ajust de les dades experimentals, encara que aquest no sigui lineal. D'aquesta manera, els mètodes de càlcul del  $\log P$  que millor s'adapten a les noves condicions són el COSMOFrag i KowWin, figura 7.29.



**Figura 7.29:** Representació gràfica de les dades experimentals segons els mètodes de càlcul del logP COSMOFrag (a) i KowWin (b) [● conjunt d'entrenament, ● referència OECD, ● conjunt de validació, — línia de tendència].

La superposició dels punts obtinguts a partir d'aquests dos mètodes, permet definir dues línies de tendència considerablement semblants. A més, si es pren la recta OECD com a referència, hom pot observar com les funcions sigmoïdals d'ambdós mètodes s'ajusten a ella, ( $R^2 = 0.998$  per COSMOFrag i  $R^2 = 0.996$  per KowWin) quan es comparen amb les dades OECD experimentals amb un  $\log P$  menor de 6 unitats. En canvi, en abandonar el marge d'aplicabilitat de la recta OECD, si bé el coeficient de correlació de COSMOFrag es manté a un nivell acceptable ( $R^2 = 0.94$ ), KowWin disminueix a un  $R^2 = 0.84$ . Tot i així, cal remarcar també que les dues corbes tendeixen a presentar un  $\log P$  semblant per a valors de  $\log k'$  elevats, figura 7.30.



**Figura 7.30:** Comparació entre les línies de tendència dels mètodes COSMOFrag i KowWin, i la recta OECD de referència experimental [● KowWin, ● COSMOFrag, ● OECD].

Malgrat la manca de dades experimentals necessàries per validar aquests models, en esdevenir la dispersió de les dades associades al càlcul COSMOFrag menor al model KowWin, es pren el primer com el model proposat per la descripció del  $\log P$  de sistemes porfíric i porfíric.

## 7.5 Absorcions bifotòniques

### 7.5.1 En els límits de la hipòtesi quàntica

Tal com s'ha comentat en la introducció de la part III (pàg.145), la generació de les ROS es produeix com a conseqüència del quench de l'oxigen ( $^3\text{O}_2$  a  $^1\text{O}_2$ ) per l'estat excitat triplet d'un FS ( $T_1$ ) format, al seu torn, per encreuament intersistema ( $S_1 \rightarrow T_1$ ). La formació de l'estat  $S_1$  s'assoleix en general per absorció d'un sol fotó de l'energia necessària des de  $S_0$ , figura 7.31.

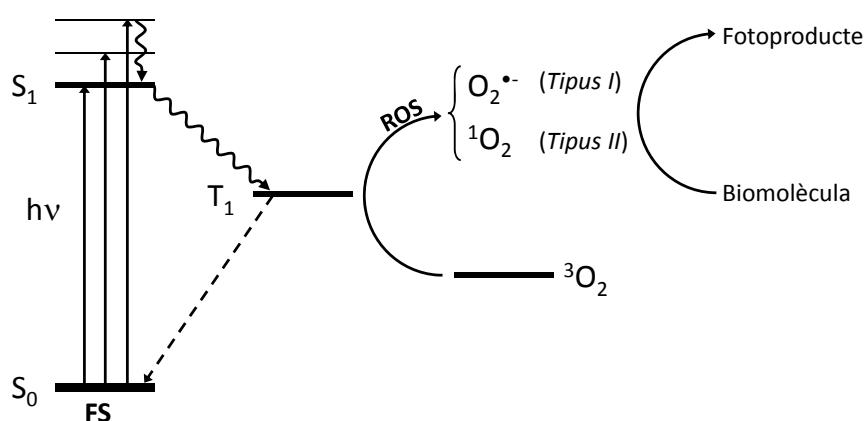
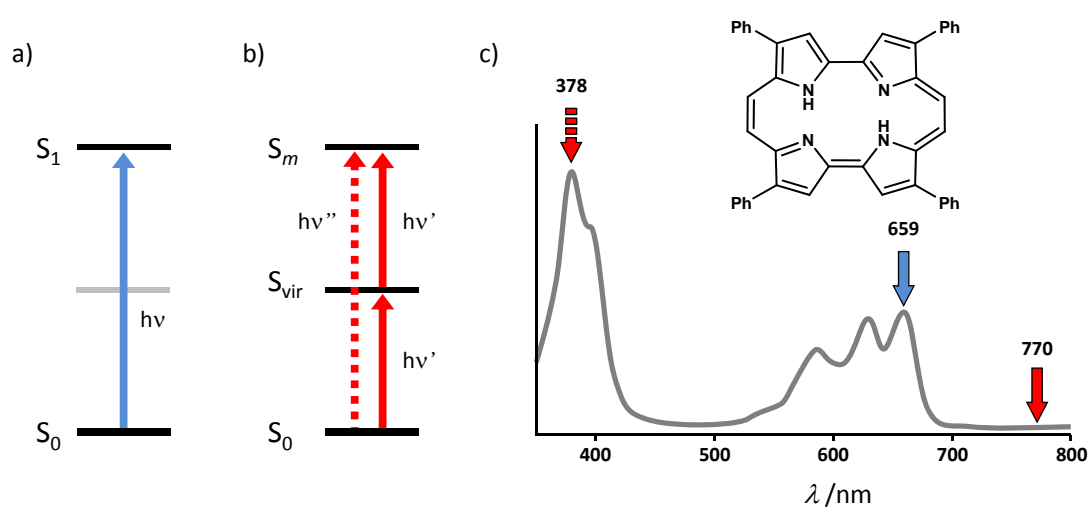


Figura 7.31: Esquema simplificat de la formació de ROS derivada d'una absorció monofotònica.

S'ha demostrat, però, que certs sistemes moleculars són capaços d'arribar a  $S_1$  per l'absorció de dos fotons de menor energia.<sup>226</sup> Aquest fet, realment provocador en un principi, pot ser explicat en base a l'existència d'un estat excitat virtual intermediari entre els estats fonamental i excitat. L'estat virtual ( $S_{vir}$ ) permet no contradir la hipòtesi de Planck, ja que actua com a estat final d'una primera transició a la qual s'encadena la segona, esdevenint-ne l'estat inicial, figura 7.32(b).

Aquesta situació resulta d'especial interès per a la PDT, doncs la utilització de làsers de major longitud d'ona permetrien augmentar la penetració de la llum a través de la pell.<sup>227</sup> En el cas concret del TPPo, la longitud d'ona màxima amb què pot donar-se una absorció monofotònica és de 659 nm. Aquesta transició correspon a la promoció d'un electró de l'orbital HOMO a l'orbital LUMO i pertany a les bandes Q de l'espectre d'absorció, les quals es caracteritzen per ser més febles (amb menor  $\epsilon$ ). En canvi, la transició més intensa pertany a la regió Soret (378 nm) i, per tant, no pot ser utilitzada per a la PDT, figura 7.32.

S'ha comprovat experimentalment que el TPPo pot presentar absorcions bifotòniques entorn els 770 nm.<sup>228</sup> D'aquesta manera s'aconsegueix poder generar un estat excitat del FS utilitzant fotons de baixa energia, i en una finestra de l'espectre d'absorció en la qual no hi ha cap absorció monofotònica (figura 7.32).



**Figura 7.32:** Representació esquemàtica de l'absorció mono (a) i bifotònica (b), i la seva comparació amb l'espectre d'absorció del TPPo, S7 (c).

### 7.5.2 Tractament computacional de les absorcions bifotòniques

L'interès del grup en conèixer la predisposició dels FS a presentar absorcions bifotòniques no és nou, i anteriorment ja s'havien realitzat col·laboracions per estudiar computacionalment aquesta propietat.<sup>228</sup> Per aquest motiu, també interessa disposar d'una metodologia d'estudi pròpia.

El repte d'estudiar les absorcions bifotòniques s'ha encarat mitjançant el càlcul de la probabilitat que té una determinada molècula de presentar absorció bifotònica, *i.e.* la *two-photon cross section* ( $\delta$ ). Tot i que no és l'única manera d'estimar la importància d'aquest tipus de processos per a una determinada molècula, la *two-photon cross section* permet avaluar qualitativament, i de forma intuïtiva, els resultats obtinguts.<sup>229</sup>

El valor de  $\delta$  s'expressa normalment mitjançant les unitats de Göppert-Mayer (GM), que permet descriure de forma més intuïtiva el seu significat. Tanmateix, en la següent discussió el valor de  $\delta$  s'expressa en unitats atòmiques (a.u.), ja que el pas d'a.u. a GM no és directe i implica paràmetres experimentals. Atenent al fet que els resultats es volen utilitzar de forma qualitativa i no quantitativa, es considera que la utilització d'a.u. és suficient.

El càlcul de  $\delta$  es realitza mitjançant la *response theory*. Aquest mètode permet estudiar la resposta del sistema front a un agent pertorbador extern (*e.g.* un camp elèctric extern), i així obtenir informació dels seus estats excitats. Les correccions al valor esperat d'una propietat  $A$ , deguda a la pertorbació  $B$  (usualment considerada com un camp extern monocromàtic amb una freqüència  $\omega_b$  associada) s'acostuma a representar amb la notació  $\langle\langle A; B \rangle\rangle_{\omega_b}$ .<sup>230</sup>

De la mateixa manera que ho fa el TD-DFT, aquesta teoria inclou una pertorbació en la definició de l'Hamiltonià:

$$H = H_0 - B \cdot F(t) \quad (7.1)$$

Tot i basar-se en la resolució de l'equació d'Shrödinger, la *response theory* no és un mètode que



permeti calcular l'energia del sistema, sinó que hom recorre a ella quan el que interessa és estudiar com respon l'energia a una determinada situació.

En la *response theory*, l'amplitud de les transicions bifotòniques ( $S_{\alpha\beta}$ ) s'identifiquen amb els residus de les funcions quadràtiques de resposta (*quadratic response functions*) que inclouen els operadors corresponents al moment dipolar elèctric. De forma general, i seguint amb la nomenclatura exposada anteriorment, els residus obtinguts del càlcul de  $\langle\langle A; B \rangle\rangle_{\omega_b}$  corresponen al moment de transició induït per l'operador  $A$  degut a la pertorbació  $B$ , considerant l'absorció bifotònica com la variació de la component elèctrica del moment dipolar.<sup>231</sup>

Assumint una radiació incident monocromàtica i polaritzada linealment, el valor de la *two-photon cross section* ( $\delta$ ) per a la transició entre l'estat fonamental (0) i l'estat final (m) es pot calcular a partir dels elements  $S_{\alpha\beta}$  (eq.7.2), on el sumatori s'aplica sobre el conjunt dels tres eixos de coordenades ( $\alpha, \beta = \{x, y, z\}$ ).

$$\delta_{0m} = \frac{1}{30} \sum_{\alpha, \beta} \left( 2S_{\alpha, \alpha} \cdot S_{\beta, \beta}^* + 4S_{\alpha, \beta} \cdot S_{\beta, \alpha}^* \right) \quad (7.2)$$

### Programari utilitzat

A la pràctica, el càlcul del valor de  $\delta$  es porta a terme amb el programari Dalton.<sup>231</sup> De manera auxiliar s'utilitza el programari Gaussian03<sup>143</sup> per a l'optimització de la geometria del FS.

Donat que el càlcul de  $\delta$  pot dependre de l'estructura tridimensional del sistema, es realitza una optimització de la geometria com a pas previ. En el moment de realitzar aquest estudi, el cost computacional i temporal d'aquest procés esdevenia un factor molt important, i per aquest motiu s'estudià l'aplicació dels mètodes DFT B3LYP i PM3. Prenent el TPPo com a referència, es va poder comprovar que els resultats de  $\delta$  obtinguts en un i altre cas eren equivalents. Per aquest motiu es considera que l'optimització de la geometria a nivell PM3 és suficient.

### Validació del mètode de càlcul

En ser la primera vegada que s'utilitzà el programari Dalton en el GEM, els primers esforços foren familiaritzar-se amb el programa. Seguidament, atenent a la naturalesa dels sistemes porfirínics d'interès, es pren com a referència el treball realitzat per Frederiksen *et al.* del càlcul de l'absorció bifotònica d'un conjunt de derivats porfirínics.<sup>226</sup>

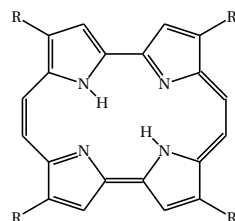
El mètode de càlcul utilitzat concatena l'optimització de la geometria a nivell PM3 i el càlcul HF 6-31G\*, combinat amb la *quadratic response theory*. Els resultats obtinguts amb ell coincideixen amb els publicats en la bibliografia, demostrant la seva adequació pel tractament d'aquest tipus de compostos.

Tal com s'ha comentat en les seccions anteriors, la semblança entre la família dels porfirins i la de les porfirines és suficient com per mantenir la metodologia de càlcul.

### Predicció de l'absorció bifotònica de porficens

Una vegada definit el mètode de càlcul, aquest s'aplica per a l'estudi de derivats porficènics d'interès pel grup de recerca i dels quals no es disposa informació referent. Els resultats es mostren a la taula 7.28.

**Taula 7.28:** Resultats obtinguts de  $\delta$  i de l'energia de la transició ( $E_{tr}$ ) derivades de les absorcions bifotòniques per derivats porficènics.



Id.	R	$\delta / 10^4$ a.u.	$E_{tr} / \text{eV}$
B1		–	–
B2		3.84	2.36
B3		6.24	2.35
B4		–	–
B5		5.10	2.35
B6		–	–
B7		4.92	2.36
B8		–	–

Tal com es pot observar no s'ha pogut realitzar el càlcul corresponent als porficens **B1**, **B4**, **B6**, **B8**. Després de comprovar que la presència de càrregues formals no representa cap problema pel mètode, en sistemes petits, es posa de manifest una de les limitacions del procediment actual de càlcul: el nombre d'àtoms del sistema molecular. Es creu, però, que aquesta limitació es minimitzarà amb properes millores del programari o amb majors potències de càlcul.

L'anàlisi qualitativa dels resultats obtinguts permet ordenar els derivats porficènics segons la seva predisposició a presentar absorcions bifotòniques: **B3**>**B5**>**B7**>**B2**.

Els resultats mostren que  $\delta$ , i amb ella la probabilitat d'absorció bifotònica, augmenta amb la

presència de grups cesors d'electrons, sent més rellevant en la posició para. Per altra banda, el grup piridil sembla tenir menys influència sobre  $\delta$ .

Aquestes observacions es troben d'acord amb la bibliografia, segons la qual disposar d'un sistema altament conjugat i substituents fortament acceptors o cesors d'electrons, es consideren característiques molt importants per aconseguir valors alts de  $\delta$ .<sup>231</sup> Tanmateix, cal tenir en compte que aquests substituents poden promoure els processos de transferència de càrrega, els quals poden afectar negativament a la formació d'oxigen singlet, proporcionant al FS una via alternativa de desactivació.<sup>232</sup> És important per tant trobar un compromís entre ambdues característiques.

## 7.6 Resum

S'ha establert un model QSPR basat en ANN alternatiu per a la predicció del valor de la longitud d'ona màxima de l'espectre d'absorció de fotosensibilitzadors tetrapirròlics. S'ha determinat la naturalesa dels descriptors necessaris i els resultats obtinguts presenten un error inferior a l'obtingut amb el mètode TD-DFT.

El model anterior s'ha utilitzat en l'estudi de la tautomeria del 9-amino-2,7,12,17-tetrafenilporficè (9ATPPo), de manera complementària al TD-DFT, per tal de trobar una justificació teòrica als resultats experimentals.

A partir dels temps de retenció mesurats s'ha comparat la viabilitat de 12 mètodes computacionals per a la predicció del coeficient de partició octanol/aigua (logP) de diferents fotosensibilitzadors. L'anàlisi estadística realitzada ha permès identificar el mètode COSMOFrag com el més adient.

Finalment, i per primera vegada al GEM, s'ha estudiat computacionalment la capacitat d'un fotosensibilitzador de presentar absorcions bifotòniques.



## Capítol 8

# Estudi de la localització subcel·lular dels fotosensibilitzadors

Un gran nombre d'estudis bioquímics han demostrat la importància de la localització subcel·lular dels fotosensibilitzadors (FS), no tan sols en l'eficàcia fotodinàmica<sup>118,167,233</sup> sinó també en l'extensió de la resposta cel·lular derivada de la PDT.<sup>116</sup> Arrel d'ells, s'han pogut identificar els orgànuls que desencadenen la mort cel·lular de manera més eficient, sent el mitocondri i els lisosomes els més destacats. Aquestes afirmacions es troben recolzades per dues premisses:

- L'efecte fotodinàmic de la PDT es deu a la interacció de les espècies oxidants (ROS) generades durant l'etapa d'irradiació, principalment oxigen singlet, amb les estructures subcel·lulars. El curt temps de vida d'aquestes espècies<sup>127,234</sup> fa que només puguin interaccionar amb l'entorn on es generen, *i.e.* l'orgànul on es localitza el FS.<sup>235,236</sup>
- La resposta cel·lular provocada per l'efecte fotodinàmic pot ocórrer seguint diferents camins, en funció de quin hagi estat l'orgànul afectat<sup>115</sup> i de la naturalesa de la pròpia cèl·lula. Així, hom pot concloure que l'eficàcia global de la PDT és determinada en últim terme pel mode de mort cel·lular (principalment apoptosi o necrosi), donat que no totes les vies són igualment efectives.<sup>237</sup>

Per aquest motiu, l'estudi de les variables que determinen la localització subcel·lular dels fotosensibilitzadors és de gran importància i esdevé capital per entendre el mode d'acció de la PDT. A més, la possibilitat que aquests paràmetres puguin ser simulats computacionalment, planteja la possibilitat d'establir un model de predicció per ser usat en el disseny racional de FS per aquesta teràpia. Les estructures moleculars dels compostos referents a aquesta secció (**A1 – S110**) es recullen en l'annex IV, pàg. 259.

## 8.1 Desencadenant la mort cel·lular

L'efecte de la PDT en el teixit tumoral desencadena en últim terme la mort específica d'eixes cèl·lules canceroses. El procés pel qual es produeix la mort cel·lular no és únic, i es distingeix entre el mecanisme de necrosi i el d'apoptosi. Cal remarcar que aquests dos processos són extremadament complexos i en aquest apartat només s'ha volgut realitzar una breu introducció.

### 8.1.1 Necrosi

La necrosi és el tipus de mort cel·lular que hom pot imaginar-se més fàcilment: la presència d'un agent extern que indueix danys cel·lulars o bé un canvi de les condicions del medi on es troba la cèl·lula, que fan que aquesta no pugui sobreviure. Correspon, doncs, a una mort inesperada que pot ser causada per un gran nombre de situacions diferents: dèficit de nutrients, hipòxia, canvis de pH i temperatura<sup>238</sup> o la presència d'agents que destrueixen components cel·lulars. Aquest és el cas de les ROS produïdes durant la PDT, les quals poden oxidar components cel·lulars i provocar danys irreparables a la cèl·lula.

### 8.1.2 Apoptosi

L'apoptosi correspon a una mort cel·lular premeditada, necessària per controlar el desenvolupament i mantenir l'homeòstasi dels teixits de forma natural.<sup>239</sup>

La resposta apoptòtica pot iniciar-se de diferents maneres segons la naturalesa de l'agent inductor: la interacció de diferents lligats amb receptors de superfície,<sup>240</sup> situacions d'estrès produïdes per agents químics o per l'efecte de la teràpia fotodinàmica,<sup>241</sup> entre d'altres.<sup>116</sup> Després de l'etapa d'inici, una successió de processos determina la manera com es desencadena la resposta apoptòtica, seguint l'anomenada via intrínseca o extrínseca.<sup>242-244</sup> En elles es veuen implicats quatre components cel·lulars: la família de les caspases, un conjunt de receptors específics de membrana (anomenats *cell death receptors*), el mitocondri i els membres de la família de proteïnes Bcl2. Tot i les seves diferències, ambdues vies conflueixen en un reduït nombre de reaccions hidrolítiques que provoquen danys cel·lulars irreversibles. Per a la següent explicació se seguirà l'esquema detallat en la figura 8.1.

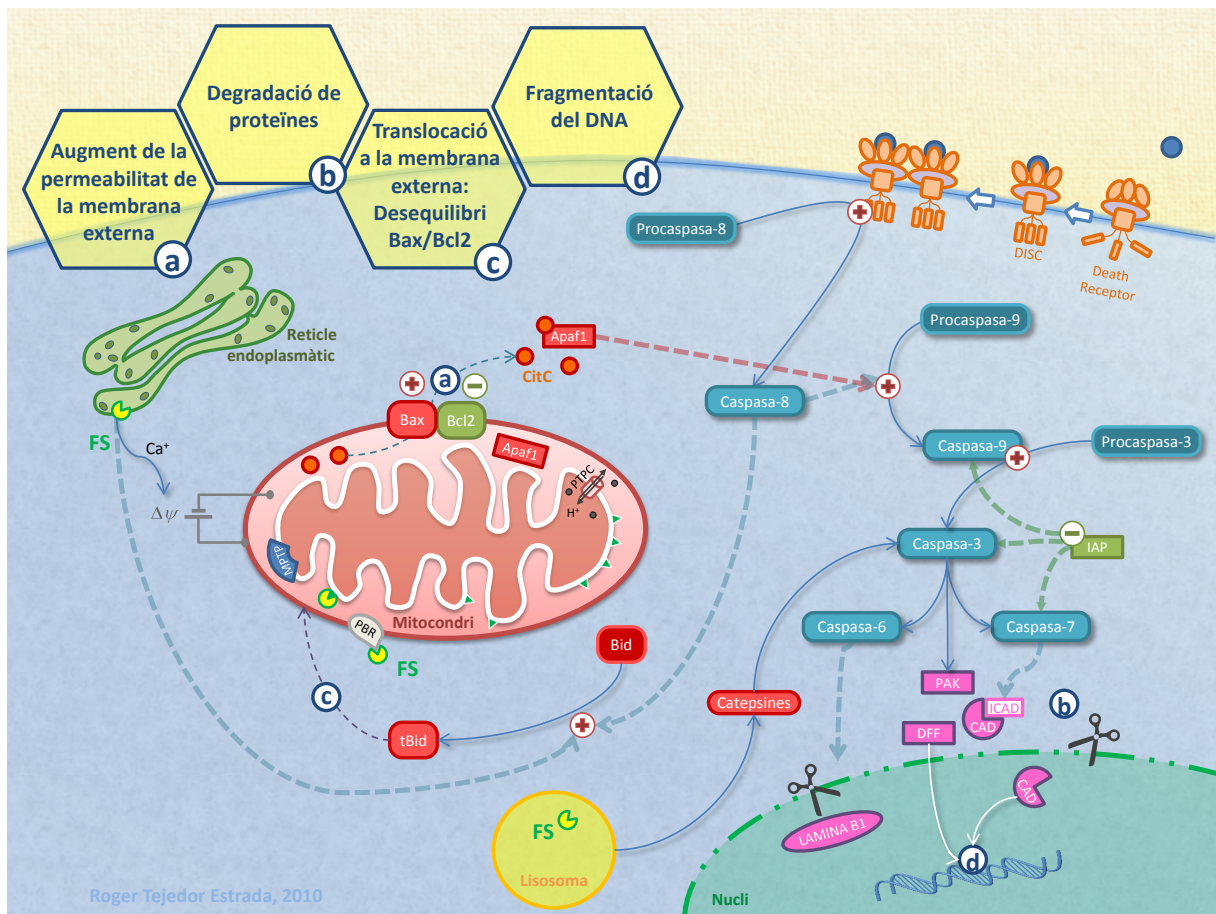


Figura 8.1: Diagrama esquemàtic de l'iniciació de l'apoptosi.

### La via intrínseca

Procés regit per l'acció directa del mitocondri: l'espai intermembrana d'aquest orgànul conté una sèrie de factors proapoptòtics que s'alliberen al citoplasma quan la permeabilitat de la membrana externa varia. La permeabilitat mitocondrial es troba regulada principalment per l'equilibri entre proteïnes proapoptòtiques (Bax/Bak) i antiapoptòtiques (Bcl2/BclXL). L'activitat d'aquestes proteïnes està regulada (positivament i negativament) pels membres de la família de les proteïnes Bcl2 que presenten el domini BH3 (*Bcl2 homology-3 domain*). Quan l'equilibri es trenca a favor de les proteïnes proapoptòtiques s'augmenta la permeabilitat de la membrana mitocondrial externa, permetent l'alliberament de factors proapoptòtics, tals com el citocrom C (CitC) i el factor d'activació Apaf-1 (*apoptotic protease activating factor*). Aquests últims provoquen l'activació de la caspasa-9, responsable d'iniciar la cascada enzimàtica.<sup>245</sup>

L'alteració de la permeabilitat també causa l'obertura del PTPC (*Permeability Transition Pore Complex*), un canal no selectiu de la membrana mitocondrial interna que dissipa el gradient de protons, equilibrant les càrregues a ambdós costats i desacoblant-lo de la cadena respiratòria. Aquest fet també té efectes sobre la regulació del volum intern del mitocondri, l'expansió del qual pot arribar a provocar la ruptura de la membrana mitocondrial externa, amb el consegüent alliberament dels factors proapoptòtics comentats anteriorment.

### La via extrínseca

Procés iniciat en el moment en què els *cell death receptors* de la superfície cel·lular (tals com Fas/CD95, TNFR o DF-5) són activats pels seus lligands (FasL, TNF $\alpha$ , TRAIL). El complex lligand-receptor (que pren el nom de DISC, *death-inducing signaling complexes*) permet, per mitjà de la seva agrupació, l'autoactivació de la caspasa-8. En alguns casos, aquest fet és suficient com per desencadenar la cascada enzimàtica, mentre que en d'altres es produeix una activació de la proteïna Bid. Aquesta proteïna, pertanyent al grup de les proteïnes BH3-*only*, es desplaça fins al mitocondri i contribueix a la permeabilització de la membrana mitocondrial i, per tant, a l'activació de l'apoptosi per la via intrínseca.

### La cascada enzimàtica

Esdevé el punt de trobada de les dues vies anteriors, on tots els factors desencadenants troben fi en un seguit de reaccions enzimàtiques on es veuen implicades les caspases 3, 6 i 7. Les caspases 3 i 7 contribueixen a l'activació de factors que provoquen la fragmentació i conseqüent degradació del material genètic (factors CAD/ICAD) i de factors que inhibeixen la reparació del DNA (factors PARP/DNA-PK). Per altra banda, la caspasa-6 indueix la proteòlisi de les làmines nuclears. La conjunció de tots aquests processos provoca, en últim terme, la mort cel·lular, figura 8.1.

En el cas concret de la PDT, cal afegir a tot aquest conjunt de processos metabòlics la influència del fotosensibilitzador. Des del punt de vista molecular, l'efecte fototòxic associat a la PDT és degut a l'oxidació de biomolècules per acció de les ROS produïdes. Les dianes biològiques susceptibles a aquesta fotooxidació poden ser tan proteïnes com àcids nucleics o lípids,<sup>140</sup> la degradació de les quals pot promoure el desencadenament de la resposta apoptòtica que en cas d'esdevenir crítiques pel desenvolupament cel·lular o de trobar-se en regions cel·lulars crítiques per aquesta finalitat. Com a exemple, la figura 8.1 il·lustra les diferències que es poden presentar en el mecanisme d'inducció de la mort cel·lular en funció de si la localització preferent del FS és mitocondrial, lisosomal o també en el reticle endoplasmàtic. Tal com ja s'ha explicat, l'acció fotodinàmica d'un FS amb localització mitocondrial afavoreix directament al desenvolupament de la via intrínseca. Tanmateix, els FS del reticle endoplasmàtic també poden contribuir (indirectament) a promoure aquesta resposta en produir l'alliberament d'ions de Ca<sup>2+</sup> al citoplasma, que puguin alterar el potencial de membrana del mitocondri. Per la seva part, l'efecte fototòxic dels FS lisosomals se centra en la ruptura d'aquest orgànul, alliberant els enzims hidrolítics al citoplasma (bàsicament catepsines), que promouen directament la cascada enzimàtica.

### 8.1.3 Autofàgia

La macroautofàgia, més comunament anomenada autofàgia, correspon al mecanisme cel·lular pel qual els components citoplasmàtics, que han sofert algun tipus de dany, poden ser reciclats.<sup>246</sup> Les proteïnes intracel·lulars o orgànuls, danyats o en excés, són conduïdes als lisosomes a través del segrest en vesícules per a la degradació i eventual reciclatge de les macromolècules que en resulten per sintetitzar noves proteïnes i membranes.<sup>247-249</sup>



L'autofàgia pot provocar la mort cel·lular quan l'apoptosi es troba inhibida o en cas que les cèl·lules intentin degradar els seus constituents afectats més enllà de la seva capacitat de recuperació. Així, aquesta via de mort cel·lular es troba potenciada per l'estrès oxidant, provocat per la generació de ROS durant la PDT.

Contràriament, l'autofàgia no és possible quan s'empren FS amb localització lisosòmica per a la PDT. En aquest cas, l'acumulació al citosol d'estructures cel·lulars afectades per les ROS estimula l'apoptosi.<sup>247</sup>

### 8.2 Biodistribució dels FS a l'interior de l'organisme

L'estudi del procés pel qual un FS es localitza en un determinat orgànu cel·lular d'una regió determinada de l'organisme comença just després de la seva administració, que normalment es realitza per via intravenosa.<sup>250</sup> D'aquesta manera, el FS assoleix el sistema circulatori mitjançant el qual pot distribuir-se per tot l'organisme. Normalment, les proteïnes sèriques, entre les quals destaquen la seroalbúmina i les lipoproteïnes de baixa densitat (*Low Density Lipoprotein*, LDL), juguen un paper molt important en aquest procés per què faciliten el transport dels FS hidrofílics i lipofílics respectivament.<sup>251</sup>

Com és de suposar, interessa que la biodistribució dels FS es trobi clarament esbiaixada a favor dels teixits hiperproliferatius. Aquesta especificitat es pot veure afavorida de forma natural gràcies a la pròpia interacció dels FS amb les LDL,<sup>252</sup> donat que els receptors d'aquestes lipoproteïnes es troben sobreexpressats en les cèl·lules tumorals, taula 8.1. Aquesta estratègia esdevé el paradigma del *targeting* de FS: emprar un **portador** compatible amb les característiques del FS, que el condueixi específicament al teixit neoplàstic aprofitant les seves **característiques diferencials**.

#### Característiques diferencials dels teixits tumorals

En els darrers anys s'han posat en evidència un seguit de característiques diferencials dels teixits hiperproliferatius que poden ésser utilitzades com a dianes per augmentar la selectivitat dels FS.

A nivell cel·lular, la senyalització de superfície de les cèl·lules tumorals es veu alterada (probablement per modificacions a nivell genètic<sup>253</sup>), fet que es pot traduir en una sobreexpressió dels receptors de LDL,<sup>254</sup> d'àcid fòlic,<sup>255,256</sup> de factors de creixement (com els factors de creixement epidèrmics 1 (EGFR) i 2 (HER2) els quals esdevenen dianes d'alguns fàrmacs proposats per combatre el càncer<sup>257</sup>), de citosines<sup>258</sup> (com CXCR4<sup>259</sup>) i fins i tot d'hormones.<sup>251</sup> Intracel·lularment, també es poden trobar discrepàncies, com per exemple diferències en el potencial de membrana de la membrana mitocondrial.<sup>260,261</sup>

Aquestes característiques cel·lulars fan que els efectes globals també s'estenguin a nivell de teixit; fent que mostrin una disminució del pH intersticial<sup>262,263</sup> (deguda, en part, a l'increment anòmal del procés de glucòlisi<sup>264</sup>) i una alteració de la permeabilitat dels capil·lars sanguinis associats a ells, fent-los menys selectius, taula 8.1.

**Taula 8.1:** Propietats diferencials dels teixits tumorals.

<i>Propietat</i>	<i>Descripció</i>
<b>Baix pH del fluid intersticial</b>	El fluid intersticial es troba embolcallant les cèl·lules i és el responsable de mantenir les membranes cel·lulars separades de les parets vasculars. En els teixits hiperproliferatius, aquest fluid experimenta una disminució del seu pH deguda a l'augment del procés de glicòlisi i la conseqüent formació d'àcid làctic. <sup>115,265</sup> Normalment, la disminució de pH es relaciona amb un augment de la lipofília i amb una major capacitat d'absorció. <sup>239</sup>
<b>Alt contingut de receptors LDL en la membrana cel·lular</b>	L'alta aflluència de receptors d'LDL en la membrana cel·lular afavoreix el transport de l'interior de la cèl·lula que es realitza per endocitosi. <sup>266</sup> .
<b>Alteració en la permeabilitat dels capil·lars vasculars</b>	Els teixits tumorals presenten una microvasculatura anormalment permeable que facilita la difusió des del torrent sanguini a l'espai intersticial. Poden existir altres diferències associades al sistema vascular, com l'augment de la fragilitat vascular o l'increment de la pressió del fluid intersticial. <sup>115,267</sup> Aquestes alteracions semblen ser degudes a una angiogènesi insuficient estimulada per diverses substàncies, alliberades per les pròpies cèl·lules tumorals, com són el factor de creixement bàsic de fibroblasts (bFGF) o el factor de permeabilitat vascular (VPF). <sup>268</sup> La pobre vascularització pot comportar, també, la formació de regions amb dèficit d'oxigen, <sup>265</sup> les quals afavoreixen la disminució del pH.
<b>Sobreexpressió dels receptors d'àcid fòlic</b>	L'elevada presència de receptors d'àcid fòlic en la superfície de les cèl·lules canceroses permet un reconeixement altament selectiu d'aquest tipus de cèl·lules (per sobre, fins i tot, dels receptors LDL <sup>256</sup> ). Juntament amb l'alta afinitat pels seus receptors, els conjugats amb àcid fòlic s'empren en la direcció selectiva de fàrmacs. <sup>255,256</sup>

## Ús de portadors

Aprofitant les característiques pròpies dels teixits tumorals comentades anteriorment, els FS poden derivatitzar-se o ser formulats de tal manera que s'afavoreixi la localització desitjada. L'ús de portadors afecta tant a la biodistribució del FS com al mecanisme de transport del FS a l'interior de la cèl·lula i, conseqüentment, a la seva localització subcel·lular.<sup>239</sup> A l'any 1900, P. Ehrlich proposà per primer cop l'ús d'anticossos per millorar la biodistribució en teixits tumorals,<sup>251</sup> aprofitant els antígens presentats específicament en la superfície de les cèl·lules malignes (*tumor-specific antigen*, TSA). Amb el temps, l'ús d'anticossos monoclonals (*monoclonal antibodies*, MAb) s'ha estès a la PDT i, per exemple, els conjugats de MAb i derivats de la clorina e6 (**S89**) han demostrat nivells d'acumulació en teixits tumorals fins a quatre vegades superiors als obtinguts en la seva forma lliure.<sup>269,270</sup> Aquesta mateixa estratègia ha estat aplicada també sobre porfirines i ftalocianines.<sup>271</sup> Tot i la gran especificitat que atorguen els TSA, condicionen la localització intracel·lular dels FS, que acostumen a aposentar-se a la membrana plasmàtica.<sup>272</sup>

Més endavant, els avenços bioquímics realitzats en els estudis de les cèl·lules tumorals permetren posar en evidència certes característiques estructurals, pròpies d'aquest tipus de cèl·lules, que podrien emprar-se com a punts d'ancoratge selectius en el redireccionament de fàrmacs. Entre aquestes

## 8.2. BIODISTRIBUCIÓ DELS FS A L'INTERIOR DE L'ORGANISME

característiques s'hi troba tant la sobreexpressió de diferents receptors (veure taula 8.1) com de certes proteïnes de membrana (com per exemple lectines<sup>273</sup>). Per altra banda, s'identifiquen les característiques diferencials del conjunt del teixit tumoral (permeabilitat de la vasculatura, disminució del pH, etc.) que també són aprofitades per dirigir selectivament portadors inespecífics tal com nanopartícules i microcàpsules,<sup>274</sup> taula 8.2. De fet, els liposomes són actualment els portadors més emprats per a FS,<sup>234</sup> i si bé la seva especificitat no és tan elevada com la dels portadors específics,<sup>275</sup> minimitzen la pèrdua de FS durant el procés de biodistribució.

**Taula 8.2:** Llistat dels principals tipus de portadors emprats en la PDT.

<i>Tipus de portador</i>	<i>Descripció</i>
<b>Anticossos monoclonals</b>	La unió covalent d'un MAb al FS permet la seva orientació cap als antigens TSA situats en la superfície de les cèl·lules tumorals. <sup>239,266,279</sup> Tot i que permeten aconseguir una selectivitat notable, són molt pocs els FS que s'hi poden associar sense modificar el reconeixement específic del MAb. És per aquest motiu que aquesta tècnica s'empra principalment com a mètode de diagnosi de tumors, més que per dirigir FS.
<b>Portadors inespecífics</b>	
<b>Liposomes</b>	La inclusió de FS altament hidrofòbics a l'interior d'una matriu lipídica resulta clau per a la seva administració <i>in vivo</i> . <sup>275</sup> L'ús de liposomes condiona de manera important la localització subcel·lular del FS, determinant-ne la via d'entrada (endocitosi). Encara que l'ús de liposomes és compatible amb la senyalització per LDL, <sup>280</sup> s'ha constatat que la distribució del FS a l'interior del liposoma (a l'interior de la vesícula o comprès entre la bicapa lipídica) pot afectar al seu efecte fotodinàmic. <sup>137</sup>
<b>Microesferes</b>	Són partícules esfèriques que poden estar formades per diferents materials, i que faciliten l'entrada dels FS a l'interior de la cèl·lula per fagocitosi o per difusió a través de la membrana plasmàtica. <sup>138</sup> S'ha estudiat el seu ús com a portador de porfirines (microesferes d'albumina de sèrum boví, BSA), i de clorines (microesferes de poliestirè), encara que en aquest últim cas el rendiment de formació d'oxigen singlet disminueix considerablement. <sup>234,251</sup>
<b>Nanopartícules</b>	Tot i esdevenir prometedors portadors, les primeres nanopartícules mostraren una tendència a acumular-se al sistema fagocitari mononuclear, que reduïa l'acumulació del FS (mTHPC <sup>281</sup> ) als teixits diana. Afortunadament, aquest inconvenient es va solucionar amb la introducció de modificacions superficials.
<b>Portadors endògens</b>	
<b>Proteïnes</b>	Les proteïnes sèriques han demostrat jugar un paper fonamental en la distribució selectiva de porfirines, clorines i ftalocianines en teixits tumorals i afavoreixen la seva internalització per endocitosi mediada pel receptor.
<b>Ciclodextrines</b>	Les ciclodextrines són sucres cíclics que permeten millorar la solubilitat en aigua de compostos altament hidrofòbics. A més de facilitar la integració dels FS en entorns biològics, han demostrat ser útils per millorar el <i>targeting</i> de porfirines i de la purpurina SnET2( <b>S96</b> ). <sup>282</sup>

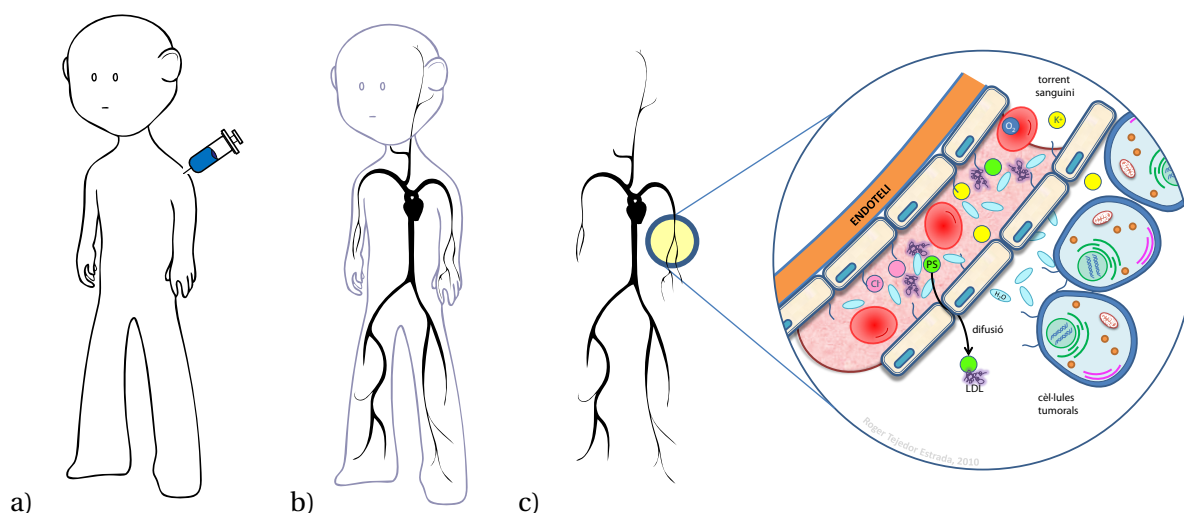
Finalment, s'han proposat estratègies mixtes que permeten combinar els avantatges dels diferents tipus de portadors. Es pot induir, per exemple, l'entrada dels immunoconjugats MAb-FS a l'interior de la cèl·lula per endocitosi, fent que la seva localització subcel·lular tendeixi a ser lisosòmica. Els resultats obtinguts en immunoconjugats de clorines combinats amb estructures polimèriques<sup>276</sup>

o bé la incorporació de ftalocianines<sup>277</sup> i feofòrbids<sup>278</sup> en immunoliposomes, permeten millorar considerablement l'efecte fotodinàmic.

La recerca de portadors més eficients i l'estudi de les propietats diferencials dels teixits tumorals han marcat gran part dels estudis realitzats en el camp de la PDT en els darrers anys, perquè permeten millorar la biodistribució dels FS, però també perquè afecten al mecanisme d'entrada a la cèl·lula (*uptake*). Per exemple, en el cas de les LDL exposat anteriorment, aquestes ajuden al transport selectiu dels FS fins al teixit tumoral, esdevenint excel·lents portadors endògens, ja que moltes vegades la presència de teixit neoplàstic ve acompanyada d'una hiperproliferació dels receptors de membrana d'aquestes proteïnes. D'aquesta manera, els FS hidrofòbics i amfòters incorporats dins la part lipídica de la proteïna assoleixen la cèl·lula tumoral i poden entrar al seu interior mitjançant un procés d'endocitosi<sup>127,239</sup> o per difusió a través de la membrana plasmàtica, després d'alliberar-se del seu portador.<sup>252</sup>

### 8.3 Mecanismes d'entrada a l'interior de la cèl·lula

Una vegada administrat i dins l'organisme, el FS ha de ser capaç d'arribar al teixit tumoral amb la màxima especificitat (*targeting*) possible i introduir-se a l'interior de la cèl·lula (*uptake*). Si el camí que el duu fins les cèl·lules dianes pot ser difícil, una vegada s'hi troba al seu interior se'n troba un d'igualment tortuós on les interaccions electrostàtiques, estèriques i de naturalesa hidrofòbica el guien fins a una estructura cel·lular determinada, figura 8.2.



**Figura 8.2:** Esquema del procés pel qual ha de passar un FS des de la seva administració fins assolir la seva localització subcel·lular preferent. Després de l'administració (a), el FS es distribueix per tot el cos per mitjà del torrent sanguini (b). En absència de portadors exògens, i segons la hidrofobicitat del FS, aquest s'uneix a proteïnes sèriques (típicament LDL per a FS hidrofòbics) per desplaçar-se pel sistema circulatori. Un cop arribat al teixit tumoral (c), l'anormal permeabilitat dels vasos sanguinis associats a aquests teixits afavoreix la seva migració al líquid intersticial i la internalització a l'interior de la cèl·lula cancerosa.

Quan el FS arriba finalment a la cèl·lula diana, es troba amb el mateix problema que qualsevol substància que pretengui accedir al seu interior: el traspàs de la membrana plasmàtica. Les cèl·lules

disposen d'estratègies de transport que faciliten l'entrada de substàncies, que normalment no signifiquen cap obstacle per les molècules de baix pes molecular (mitjançant el **transport actiu i passiu**), però que resulten més restrictives per a soluts d'alt pes molecular. En aquests casos entren en joc les propietats fisicoquímiques de cada molècula. Les característiques hidrofòbiques d'alguns FS, per exemple, fan que la seva capacitat de difondre a través de membranes biològiques sigui remarcable i molt important per entendre la seva biodistribució i activitat.

Tal com s'ha començat a plantejar en l'apartat anterior, el mecanisme d'*uptake* depèn en gran mesura del portador emprat en el *targeting* del FS. Molts d'ells, aprofitant els receptors de superfície, faciliten l'entrada dels soluts per **endocitosi mediada per receptor**, el tercer principal mecanisme de transport de soluts a l'interior de la cèl·lula.

#### Transport passiu

Correspon al procés pel qual un solut és transportat d'un costat a l'altre d'una membrana lipídica per difusió simple (també anomenada difusió facilitada). Es pot aplicar tant en la descripció del procés d'*uptake* de soluts de baix pes molecular, com dels processos de transport intracel·lulars entre el citosol i el lúmen dels orgànuls membranosos d'altres tipus de molècules com poden ser els FS.

Matemàticament, s'han pogut establir models farmacocinètics a nivell cel·lular que simulen el transport passiu de fàrmacs de baix pes molecular (tot i que amb uns determinats requeriments estructurals),<sup>283</sup> basant-se en adaptacions del formulisme de Nernst—Planck, el qual permet la descripció matemàtica dels processos de difusió i que per molècules no carregades esdevé la Llei de Fick, eq. 8.1.<sup>1</sup>

$$J_A = -D_A \left( \frac{dC_A}{dx} \right) \quad (8.1)$$

El flux d'un solut  $A$  ( $J_A$ ) per difusió facilitada és produït gràcies al gradient de potencial electroquímic associat a la substància, és a dir, a la diferència de concentracions d' $A$  entre l'exterior i l'interior de la cèl·lula ( $\frac{dC_A}{dx}$ ), presentant un factor de proporcionalitat ( $D_A$ ) corresponent al coeficient de difusió.

El transport passiu depèn de factors tant cel·lulars (diferència de pH en ambdues parts de la membrana, potencial elèctric de membrana) com de la molècula que experimenta el procés ( $pK_a$ ,  $\log P$  en la seva forma neutra i ionitzada). D'aquesta manera, l'acumulació preferent de porfirines catióniques al mitocondri es pot explicar en part per la combinació de la lipofília d'aquests FS i el potencial de membrana mitocondrial.<sup>116</sup>

#### Transport actiu

Tota membrana cel·lular disposa d'un conjunt de proteïnes de transport, que faciliten l'entrada de determinats soluts al citosol. A diferència de la difusió facilitada, aquest mecanisme requereix el consum d'ATP. Aquestes proteïnes es poden classificar en:

1. Proteïnes transportadores: Són proteïnes altament selectives que permeten el transport de cer-

tes molècules orgàniques de baix pes molecular, prenent-les del cantó extracel·lular i alliberant-les al citoplasma.

2. Proteïnes canal: Són porus hidròfils on els ions inorgànics poden passar per difusió. En aquest cas, la selecció de les molècules que poden transportar es realitza en funció de la seva mida.

Aquest sistema de transport és molt específic i no es té coneixement de la seva implicació en el transport de FS.

### Diferents formes d'endocitosi

Les cèl·lules realitzen constantment processos d'endocitosi per capturar partícules presents en el fluid extracel·lular a través de vesícules formades per la invaginació de la membrana cel·lular. El procés endocític rep el nom de **fagocitosi** quan es captura una partícula sòlida, mentre que la **pinocitosi** fa referència a la captura de substàncies dissoltes.

Les vesícules endocítiques són transferides als lisosomes, on el material capturat és digerit i els metabolits formats alliberats al citoplasma. Tot i que aquest procés és (en essència) indiscriminat, la presència de receptors en la superfície cel·lular pot afavorir la captació de macromolècules específiques, en el que s'anomena **endocitosi mediada per receptor**. Aquesta té lloc quan la introducció d'una macromolècula a l'interior de la cèl·lula es produeix per endocitosi i en forma de complex receptor-macromolècula.

Aquesta darrera estratègia se segueix principalment pels compostos altament hidrofòbics, tals com els FS<sup>148,284</sup> que, a causa de la seva lipofília, tendeixen a unir-se a les LDL durant el seu recorregut pel torrent sanguini. Les LDL, per la seva banda, s'uneixen a receptors específics de la superfície cel·lular, són ingerides per endocitosi mediada per receptor i el complex LDL-receptor és transferit a un endosoma. Donat el menor pH de l'endosoma, es promou la dissociació del complex i mentre que els receptors són retornats a la superfície cel·lular, les LDL es transfereixen als lisosomes, on són degradades per enzims hidrolítics i el compost hidrofòbic que es trobava unit a les proteïnes s'allibera.

Donada la seva naturalesa hidrofòbica, derivats de porfirines introduïts per endocitosi tendeixen a acumular-se en estructures membranoses, incloent-hi la membrana plasmàtica, la membrana mitocondrial, l'aparell de Golgi i el reticle endoplasmàtic.<sup>239</sup>

## 8.4 Una gran sopa electroforètica

Quan el FS arriba a l'interior de la cèl·lula es troba en un entorn dinàmic,<sup>285</sup> que canvia constantment, dins del qual s'ha d'acomodar. Tot i que la via d'accés a l'interior de la cèl·lula determina en gran mesura la situació inicial del FS, la interacció amb el medi pot facilitar-ne el seu desplaçament fins a estructures determinades.

L'eficàcia global de la resposta apoptòtica pot dependre de la situació intracel·lular on tingui lloc el dany fotodinàmic, és a dir, de la localització subcel·lular del FS. A continuació es presenta un breu

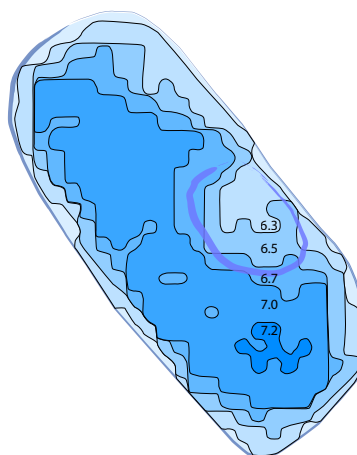
resum de quines són les estructures cel·lulars susceptibles al dany fotoquímic i les seves característiques estructurals.

#### 8.4.1 Citoplasma i citosol

<i>Citoplasma</i>	
<b>Funció</b>	Medi constitutiu de la cèl·lula
<b>pH</b>	7.0–7.5
<b>Mostren afinitat</b>	FS hidròfils i aniònics <sup>195,286</sup>
<b>Interès per a la PDT</b>	Mort cel·lular deguda a danys fotoquímics al citoesquelet



A grans trets, tota cèl·lula animal està formada per una membrana plasmàtica, un nucli cel·lular i el citoplasma. L'espai intercel·lular que resta del citoplasma quan s'exclouen els orgànuls, s'anomena citosol. Aquest espai semi-fluid conté el citoesquelet, que és el responsable de l'organització interna de la cèl·lula i d'aportar-li rigidesa estructural. Donat que el moviment molecular a l'interior de la cèl·lula es realitza per difusió i que el pH del citoplasma és canviant (veure figura 8.3), hom es pot imaginar el citosol com un gel electroforètic dins del qual el FS s'ha d'obrir pas. Per tant, és d'esperar que tant la mida com l'estat d'ionització del FS siguin clau per entendre el seu comportament quan el FS es trobi lliure a l'interior de la cèl·lula.




**Figura 8.3:** Esquema representatiu de la variació del pH a l'interior d'una cèl·lula eucariota. Els valors mostrats corresponen al pH intracel·lular d'una cèl·lula de llevat.<sup>287</sup> La figura s'ha colorat segons aquests resultats experimentals per ajudar a la seva comprensió. Malauradament, la tècnica emprada en l'obtenció d'aquests resultats només permeté als autors detectar l'efecte d'un vacúol, però no de la resta d'orgànuls intracel·lulars de manera que no s'hi troben representats.

A més de ser l'irremeiable medi de transport a l'interior de la cèl·lula, el citosol pot esdevenir la localització principals pels FS, bé sigui per la seva baixa especificitat per un orgànel·l determinat del citoplasma (com és el cas d'algunes porfirines<sup>195</sup>), o per la possible interacció amb components del citoesquelet.<sup>132</sup> En aquest sentit, s'ha comprovat que alguns FS localitzats al citosol (com ara el TPPS<sub>4</sub><sup>288</sup> o derivats porfirínics<sup>289</sup>) són capaços de produir dany fotoquímic als microtúbuls del

citoesquelet, específicament a la tubulina.<sup>252</sup> L'acció fotodinàmica sobre aquesta proteïna inhibeix la formació dels microtúbuls, conduint a l'acumulació de cèl·lules en procés de mitosi i, en alguns casos, a la mort cel·lular.<sup>239</sup>

### 8.4.2 Membrana plasmàtica


<b><i>Membrana cel·lular</i></b>		
<b>Funció</b>	Estructural i de control de l'entrada de soluts a l'interior de la cèl·lula	
<b>Mostren afinitat</b>	FS altament hidrofòbics	
<b>Interès per a la PDT</b>	Condueix a la mort cel·lular per necrosi	

Els FS especialment hidrofòbics tendeixen a localitzar-se a les membranes cel·lulars, donat que presenten una constant de difusió en cèl·lules propera a la dels lípids. Aquest fet inclou, òbviament i principal, la membrana plasmàtica però també membranes internes d'altres orgànuls.<sup>179,265</sup> Els FS que mostren aquest darrer comportament poden ser fàcilment mal assignats al citoplasma quan el mètode de mesura no disposa de la resolució suficient.<sup>290,291</sup>

L'acció fototòxica dels FS localitzats preferentment a la membrana cel·lular exhibeix un ampli ventall de conseqüències, que normalment conflueixen en la necrosi (i no apoptosi) cel·lular. Donada la complexitat de la membrana cel·lular i la disparitat dels seus components, el dany d'aquesta regió cel·lular produeix un gran nombre d'esdeveniments tals com la reducció del transport actiu, l'alteració de la permeabilitat cel·lular deguda a la despolarització de la membrana plasmàtica, la inhibició i/o degradació de certs enzims situats en la bicapa lipídica (com per exemple de l'ATPasa), fluctuacions en la regulació dels antigens de superfície i la peroxidació de lípids.<sup>239</sup>

En vista d'aquests processos, no sorprèn que alguns FS mostrin una preferència inicial per la membrana cel·lular i migrin progressivament a altres orgànuls durant el temps d'irradiació,<sup>118</sup> presentant un patró de relocalització degut a la degradació inicial de la membrana plasmàtica.

### 8.4.3 Mitocondri

<b><i>Mitocondri</i></b>		
<b>Funció</b>	Respiració cel·lular. Obtenció de molècules d'alt contingut energètic (ATP) a partir de processos oxidatius	
<b>pH luminal</b>	8	
<b>pH espai intermembrana</b>	7.0–7.5	
<b>Mostren afinitat</b>	FS catiónics <sup>292,293</sup>	
<b>Interès per a la PDT</b>	Alliberament de factors preapoptòtics deguts a la ruptura del gradient electroquímic de membrana	



Els FS que s'acumulen preferentment al mitocondri han demostrat desencadenar l'apoptosi de manera més eficaç.<sup>115,116</sup> Aquest fet queda palès en alguns estudis on es mostren diferents casos on certes línies cel·lulars, amb alteracions mitocondrials, esdevenen resistents a la PDT.<sup>239,294</sup>

El mitocondri pertany al grup d'òrgànuls cel·lulars formats per dues membranes, una externa (en contacte amb el citosol) i una interna. L'espai existent entre ambdues membranes s'anomena espai intermembrana o *lumen*, mentre que l'espai interior del mitocondri, que es troba delimitat per la matriu mitocondrial interna, rep el nom de matriu mitocondrial.

**La membrana externa** Està formada principalment per un llit de fosfolípids altament insaturats entremig dels quals es troben repartits diferents tipus de proteïnes. S'hi localitzen enzims implicats en el metabolisme de lípids, citocroms relacionats amb processos de transport electrònic i proteïnes (porines) encarregades de facilitar el transport de molècules de baix pes molecular a través de la membrana. Aquest transport és prou indiscriminat com per fer que el lumen mitocondrial sigui com una continuació del citoplasma cel·lular,<sup>287</sup> fent que es conservi el pH entre 7 i 7.5, i que el gradient electroquímic associat a la membrana externa sigui menyspreable. Per altra banda, la superfície mitocondrial també conté receptors de benzodiacepines (*Peripheral Benzodiacepine Receptors*, PBR) que poden esdevenir punts d'ancoratge pels FS.<sup>179,195</sup>

**La membrana interna** La morfologia d'aquesta membrana és molt característica, ja que es troba altament replegada sobre sí mateixa en forma de crestes. Conté la cadena de transport electrònic necessària per a la respiració aeròbica, fet que explica la seva gran superfície. Juntament amb el gran nombre de proteïnes que formen part d'aquesta cadena, també se'n localitzen d'altres implicades en el transport de metabolits, entre les quals destaca el complex transportador MPTP (*Mitochondrial Permeability Transition Pore*) que augmenta la permeabilitat de les membranes mitocondrials. Tot i així, el transport de soluts entre el lumen i la matriu mitocondrial es troba molt més restringit que en el cas de la membrana externa.

Fins al moment s'han identificat un gran nombre de components mitocondrials susceptibles d'interaccionar amb FS. A la membrana externa, la proteïna antiapoptòtica Bcl2<sup>235,295</sup> i certes proteïnes transportadores acompanyen al PBR com a possibles dianes.<sup>179</sup> Pel que fa a la membrana interna, els possibles punts d'interacció inclouen les cardiolipines<sup>195</sup> i enzims primerencs de la cadena de transport electrònic.<sup>179</sup> A més, les cèl·lules tumorals presenten un gradient electroquímic, associat a la membrana mitocondrial interna, més elevat que en condicions normals,<sup>286</sup> fet que promou la localització de FS catiònics dins d'aquest espai.<sup>116,127</sup>

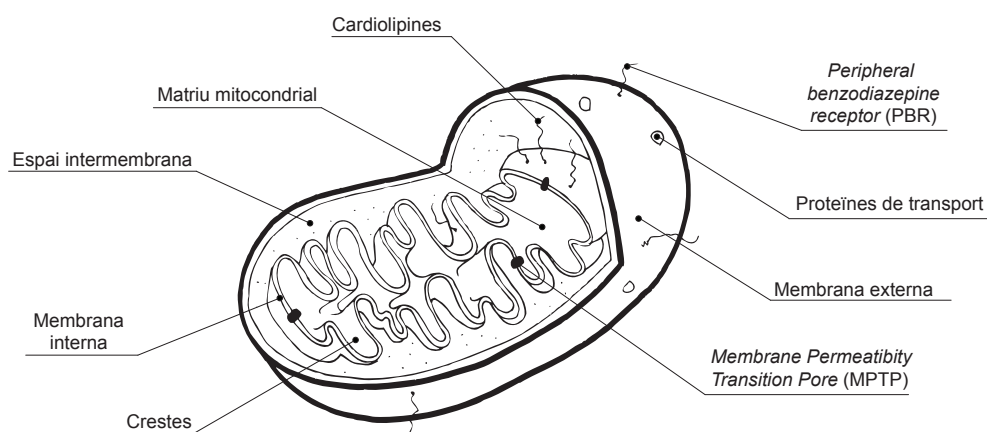



Figura 8.4: Esquema dels components estructurals del mitocondri.

**El potencial de membrana interna** Tal com s'ha comentat, el mitocondri esdevé una peça clau en el transcurs de l'apoptosi, la qual està associada en molts sistemes a la pèrdua de potencial de la membrana mitocondrial interna ( $\Delta\Psi_m$ ).<sup>296,297</sup> El valor de  $\Delta\Psi_m$  es troba entre 150 – 180 mV, sent negatiu al lumen i positiu a l'espai intermembrana. Aquesta distribució de potencials fa que els FS catiònics siguin conduïts per un efecte electroforètic fins al mitocondri amb una rellevant especificitat.<sup>298</sup>

La disminució de  $\Delta\Psi_m$  es relaciona amb alteracions de la permeabilitat de la membrana mitocondrial interna provocada per la formació de canals per part de la proteïna Bax o per efecte del PTPC. El dany fotoquímic del reticle endoplasmàtic pot catalitzar també la pèrdua de  $\Delta\Psi_m$  al mitocondri, contribuint a la translocació de la proteïna Bax a la membrana mitocondrial.<sup>235</sup>

#### 8.4.4 Lisosomes

<b>Lisosomes</b>		
<b>Funció</b>	Digestió intracel·lular	
<b>pH intralisosòmic</b>	5.0	
<b>pH perilisosòmic</b>	7.4	
<b>Mostren afinitat</b>	FS de naturalesa diversa amb una gran varietat de distribucions de càrrega, <sup>293,299</sup> normalment absorbits per la cèl·lula en un procés d'endocitosi	
<b>Interès per a la PDT</b>	Apoptosi promoguda per l'alliberació d'enzims hidrolítics al citosol	

Els lisosomes són els òrgans citoplasmàtics encarregats de la digestió intracel·lular. Els enzims hidrolítics, responsables d'aquesta activitat, es troben confinats en l'espai per mitjà d'una membrana lipídica simple que protegeix la integritat de la cèl·lula.

Els lisosomes son generats a partir d'un procés d'endocitosi relacionat amb el transport de molècules grans a l'interior de la cèl·lula. Tot i així, en el mecanisme general d'endocitosi, existeixen diferents tipus de substàncies que disposen d'un sistema de transport específic a l'interior de la cèl·lula

anomenat *endocitosi vehiculada per receptor*, veure apartat 8.2. Les LDL presenten aquest tipus de transport, la qual cosa el fa especialment rellevant, donat que els FS es troben associats a aquest tipus de proteïnes en el seu viatge a través del flux sanguini fins assolir el teixit tumoral, apartat 8.2, pàg. 207.

El procés d'irradiació durant la PDT pot causar la permeabilització dels lisosomes, els quals poden alliberar tant els enzims hidrolítics com els FS al citosol.<sup>300</sup> Aquest fet no suposa tan sols desviar l'efecte fototòxic dels FS al citoplasma, sinó que els enzims alliberats faciliten el pas de Bid a tBid, el qual juga un paper molt important en el desencadenament de l'apoptosi (apartat 8.1, pàg.204). Tanmateix, que la resposta cel·lular, desencadenant de l'apoptosi, tingui lloc via Bid i no Bcl-2/BclXL, implica una resposta més lenta que en els altres casos.

En els lisosomes s'hi poden localitzar tant FS hidrofílics<sup>127</sup> (en el cas de les porfirines, aquelles que contenen cadenes laterals curtes, poc hidrofòbiques i sense cap grup catiònic<sup>195</sup>), així com FS amb un marcat caràcter amfòter, que esdevenen bases febles, i poden travessar la membrana lisosomal (o fins i tot la de l'aparell de Golgi) buscant el caràcter àcid del seu lumen (pH ~ 5.0).

### 8.4.5 Altres orgànuls

El nucli s'identificà originalment, juntament amb la membrana plasmàtica, el mitocondri i els lisosomes, com una de les dianes potencials a ser emprades per a la PDT.<sup>301</sup> De fet, és l'orgànul més sensible al dany fotoquímic i el que condueix més ràpidament a la mort cel·lular,<sup>115</sup> però pot tenir associats efectes indesitjables relacionats amb els possibles danys en el DNA, mutacions i/o carcinogènesi que el poguessin provocar durant el tractament. D'aquí la tendència actual d'evitar el disseny de FS amb aquesta localització.

Afortunadament, la major part de FS no s'acumulen al nucli i per tant, la PDT presenta un a baixa probabilitat de causar aquestes alteracions. Els FS que s'acumulen preferentment en aquest orgànul acostumen a ser sistemes de naturalesa catiònica,<sup>118</sup> encara que es pot incrementar la seva especificitat incorporant un NLS (*nuclear localization signal*), el qual interacciona amb les importines del nucli cel·lular.<sup>135</sup>

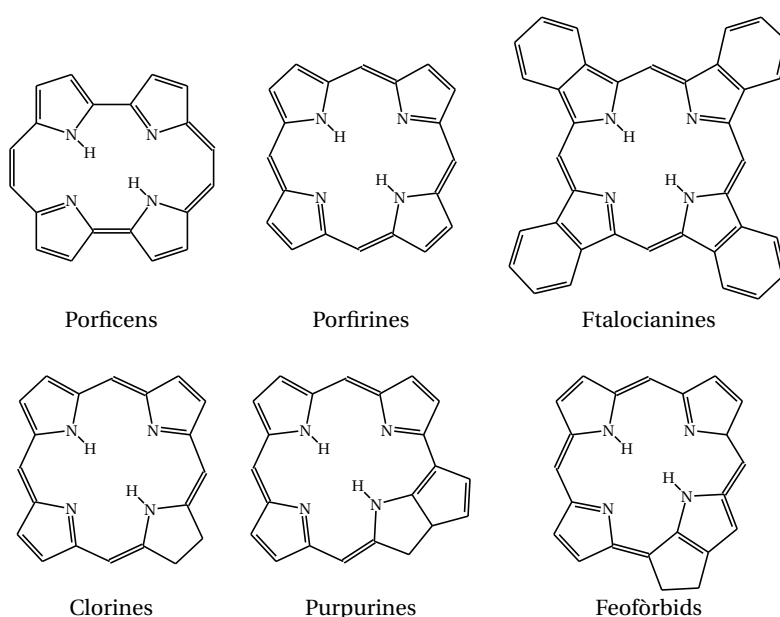
Encara que no es considerin clares dianes, els FS també poden localitzar-se en el reticle endoplasmàtic i en l'aparell de Golgi.<sup>252</sup> De la mateixa manera que ocorre en el mitocondri, el reticle endoplasmàtic conté les proteïnes antiapoptòtiques Bcl2 i BclXL;<sup>235,302</sup> i en cas de produir-se el dany fotoquímic en aquest orgànul, poden degradar-se i deixar via lliure a la proteïna Bax per iniciar el procés d'apoptosi (veure figura 8.1, pàg. 205). A més, el reticle endoplasmàtic actua de magatzem d'ions Ca<sup>2+</sup>. En danyar-se, es produeix l'alliberació d'aquests ions al citoplasma que també contribueixen a la iniciació de l'apoptosi.<sup>116,235</sup>

## 8.5 Característiques intrínseques dels FS

La naturalesa del FS juga un paper molt important en la definició del seu mecanisme de distribució i d'entrada a l'interior de les cèl·lules tumorals. La seva estructura molecular i les propietats

fisicoquímiques que se'n deriven defineixen les interaccions amb el medi que governaran el seu comportament.

Des del punt de vista estructural, els FS de segona generació estan formats per un cos central (anomenat *core*) i un conjunt de substituents perifèrics (anomenats cadenes laterals o *side chains*). La naturalesa del FS, la seva definició química, es troba altament influenciada pel tipus de *core* que, a més de donar-li nom al FS, el permet ser classificat dins les diferents famílies: porfirines, porficens, ftalocianines, feofòrbids, clorines, etc. (veure figura 8.5). En canvi, les cadenes laterals presenten una major variabilitat estructural i permeten modular característiques concretes dels FS com pot ser el seu balanç hidrofòbic/hidrofílic<sup>234,290</sup> o de la seva distribució de càrregues.<sup>234,303</sup> La contribució de les cadenes laterals pot arribar fins i tot a influir en el procés d'*uptake*, modificant així la localització preferent d'un tipus determinat de *core*.<sup>115</sup>



**Figura 8.5:** Fragments centrals dels PS macrocíclics en estudi.

D'entre totes les propietats fisicoquímiques dels FS en destaquen la hidrofobicitat i els efectes electrostàtics, les quals esdevenen les principals forces motrius dels FS en solució a l'interior de la cèl·lula. El desplaçament intracel·lular recorda, doncs, al d'una gran sopa electroforètica:

### Hidrofobicitat

La hidrofobicitat es troba implicada en tots els processos que envolten al FS, des de l'administració fins a la seva diana intracel·lular. Tal com s'ha comentat en l'apartat 8.2 (pàg.207), administrat de forma lliure i per via intravenosa, el FS és transportat pel torrent sanguini per unes proteïnes sèriques o altres en funció de la seva hidrofobicitat. També en depèn el seu procés d'*uptake*<sup>304</sup> (apartat 8.3, pàg.210) i el seu moviment a l'interior de la cèl·lula.<sup>285</sup>

Aquesta propietat també pot dictar la localització preferent dels FS.<sup>304</sup> Una de les característiques dels FS altament hidrofòbics és el seu elevat índex d'autoagregació. A l'interior de la cèl·lula,

aquestes estructures incrementen la seva tendència a acumular-se en entorns cel·lulars hidrofòbics, principalment membranes cel·lulars.<sup>179,290</sup> L'estat d'agregació no només afecta a la seva distribució intracel·lular, sinó que també es deixa notar en els mecanismes de transport.<sup>305</sup>

En l'altre extrem s'hi troben els FS hidrofílics. Tot i evitar l'agregació, aquests FS presenten problemes d'*uptake* degut a la seva baixa capacitat de penetrar a l'interior de les bicapes lipídiques. La reducció del transport passiu justifica que la seva entrada a l'interior de la cèl·lula es realitzi moltes vegades per mitjà d'un procés d'endocitosi (acompanyats per proteïnes d'albúmina), fent que presentin preferentment una localització lisosomal.<sup>127</sup>

Sembla necessari trobar un equilibri entre el caràcter hidrofòbic i hidrofílic dels FS. Els FS amfòters o amb un caràcter lleugerament lipofílic poden difondre de forma passiva a través de les membranes cel·lulars,<sup>306</sup> que juntament amb l'evitació de l'agregació els permet assolir aquells orgànuls on es desencadena de forma més eficient la resposta fotodinàmica.<sup>115,291</sup>

El balanç hidrofòbic pot ser modulats mitjançant l'addició o supressió de substituents, tant en les cadenes laterals com en el propi fragment central del FS, que contribueixin en un o altre sentit.

La presència de grups polars (metoxi, hidroxil, acetoxi, etc.) ajuden a incrementar la hidrofília del FS. El porficè (Po), per exemple, que té millors propietats fisicoquímiques que la porfirina en l'ús com a FS per a la PDT,<sup>118</sup> presenta problemes de solubilitat a causa de la seva elevada lipofília. La derivatització del Po amb grups polars permet al 9-acetoxi-2,7,12,17-tetrakis-( $\beta$ -metoxietil)-porficè (ATMPo, **S5**) augmentar considerablement la solubilitat respecte a la del Po.<sup>307</sup>

Entre els grups polars destaquen aquells que poden presentar diferent grau d'ionització en funció del pH fisiològic. En aquests casos, es troba una acció combinada de la hidrofobicitat juntament amb la càrrega formal de la molècula, que pot ser diferent en cadascun dels orgànuls cel·lulars. Les meso-tetrafenilporfirines TPSS<sup>308</sup> (**S44** - **S46**) i les ftalocianines ALPCs<sub>n</sub><sup>309,310</sup> (**S56** - **S60**) són dos dels exemples típics; el seu grau de sulfonació afecta tant al balanç hidrofòbic com a les interaccions electrostàtiques que puguin establir amb l'entorn, fent que els compostos amb diferent combinació de grups sulfònics presentin patrons de localització subcel·lular completament diferents.

Per la seva part, la lipofília es veu afavorida per la presència de grups apolars com són les cadenes hidrocarbonades. S'ha demostrat que l'extensió d'aquestes cadenes laterals controlen la localització subcel·lular de FS, com porfirines<sup>195</sup> i feofòrbids.<sup>305,311</sup> Tot i el seu efecte en la descripció global de la lipofília, el gran joc conformacional que poden presentar aquestes cadenes pot influir en la capacitat d'agregació dels FS.<sup>305</sup> Aquests substituents augmenten el radi d'interacció dels FS, influint també en el procés de captura en vesícules intracel·lulars<sup>312</sup> i en el mecanisme d'*uptake*.<sup>313</sup>

### Efectes electrostàtics

El mecanisme de captura del fàrmac depèn en gran mesura de la presència de grups amb càrrega formal i en l'acidesa/basicitat dels substituents. Els FS que contenen bases febles en les seves cadenes laterals, tendeixen a localitzar-se en orgànuls àcids,<sup>252</sup> tals com lisosomes, endosomes o grànuls de secreció, on es poden establir mecanismes de transport dependents del pH.<sup>314</sup>

Els lisosomes són els principals orgànuls diana dels FS amb cadenes laterals aniòniques,<sup>293</sup> però no els únics: algunes porfirines aniòniques presenten una localització citoplasmàtica difosa.<sup>195,286</sup>

Per altra banda, el gradient de potencial negatiu que s'estableix entorn la membrana mitocondrial atrau als FS catiónics<sup>118,195,292,293</sup> i aquells amb un caràcter àcid.<sup>315</sup> La localització mitocondrial d'aquest tipus de FS es veu encara més afavorida per la capacitat d'aquests d'interaccionar amb les cardiolipines.<sup>316</sup> Tot i així, el mecanisme d'*uptake* dels FS catiónics no depèn d'aquest gradient, fent que puguin localitzar-se també a la membrana cel·lular,<sup>317</sup> als lisosomes<sup>299</sup> o al nucli.<sup>313</sup>

Sembla, doncs, que el comportament dels FS iònics està definit per un subtil equilibri entre el nombre de substituents carregats, el mecanisme d'entrada a la cèl·lula i la massa molecular del propi FS.<sup>291</sup>

## 8.6 Tractament computacional

Donat que la localització subcel·lular dels FS es troba íntimament relacionada amb l'eficàcia fotodinàmica, disposar d'una metodologia computacional que permetés predir la localització subcel·lular d'un determinat FS, seria extremadament útil. Com es poden relacionar, però, les propietats fisicoquímiques dels FS amb la localització subcel·lular?

El plantejament anterior es troba d'acord amb els principis de la metodologia QSPR, la qual permet establir relacions matemàtiques entre descriptors moleculars (normalment relacionats amb l'estructura molecular, secció 2.1 pàg. 52) i una o diverses propietats moleculars. Juntament amb els mètodes QSAR, quan la propietat a predir es troba relacionada amb una activitat biològica, els models QSPR han demostrat ésser excel·lents punts de partida pel plantejament de problemes estructura-localització<sup>318</sup> i han estat aplicats durant molt temps per a l'establiment de models de predicció de l'activitat tumoral de FS,<sup>127,162,172,197,200,319-321</sup> l'acumulació de FS en teixits<sup>199</sup> i la localització subcel·lular de proteïnes.<sup>322-325</sup>

També es troben descrits en la bibliografia alguns models estructura-localització capaços de predir la localització selectiva de sondes fluorescents en orgànuls específics, tals com el mitocondri,<sup>306,326</sup> lisosomes<sup>327</sup> o el nucli cel·lular.<sup>328</sup> Tot i que aquests estudis proporcionen una exhaustiva visió de la descripció matemàtica del procés d'*uptake* i distribució cel·lular, els seus resultats es troben confinats en les sondes fluorescents i molècules orgàniques petites. Les diferències fisicoquímiques entre aquests sistemes moleculars i els FS tetrapirròlics són prou notables com per considerar que els seus resultats no són directament extrapolables a aquests macrocicles (veure apartat 8.6.5).

S'ha realitzat un exhaustiu estudi entre tots els possibles mètodes a aplicar per a l'obtenció de models QSPR, per tal d'escollir el millor per aquesta finalitat. S'ha prestat una especial atenció a les xarxes neuronals artificials (*Artificial Neural Networks*, ANN), donat que gaudeixen d'un ampli reconeixement per a l'establiment de models QSAR en el camp de la química mèdica<sup>44,309,329</sup> i han demostrat ser útils per a la predicció de l'activitat anti-tumoral,<sup>162,200</sup> l'activitat fotodinàmica de derivats feofòrbids<sup>199</sup> i la localització subcel·lular de proteïnes.<sup>325</sup>

### 8.6.1 La base de dades

Abans de plantejar qualsevol estratègia per a l'obtenció d'un model, cal disposar d'un conjunt de dades amb localització subcel·lular coneguda, sobre el qual poder extreure una relació amb la seva

estructura molecular. Aquesta quimioteca ha de contenir les principals famílies de compostos sobre els quals es desitja aplicar el model, ja que els models derivats de l'estratègia QSPR són altament dependents de la naturalesa de les dades utilitzades.

### Creació d'una quimioteca virtual

A partir d'una extensa cerca bibliogràfica es crea una quimioteca virtual amb 110 fotosensibilitzadors tetrapirròlics diferents (que inclouen porfirines, porficens, ftalocianines, purpurines, clorines i feofòrbids, figura 8.6), dels quals se n'ha estudiat experimentalment la localització subcel·lular. Els orgànuls susceptibles de localització preferent es poden agrupar en vuit localitzacions diferents: la membrana cel·lular (MEM), el mitocondri (MIT), l'aparell de Golgi (AG), el reticle endoplasmàtic (RE), el nucli cel·lular (NUC), els lisosomes (LIS), una localització difosa (DIF) repartida de forma uniforme en la fase aquosa del citoplasma i una localització difosa en les membranes intracel·lulars (MC). Aquesta informació, recollida en l'annex IV (pàg. 259), es complementa amb la línia cel·lular sobre la qual ha estat realitzat l'experiment, el sistema de vehiculització emprat, l'estructura molecular del FS i informació relacionada amb la seva possible relocalització. La correcta definició de la base de dades esdevé crítica, ja que conforma els fonaments sobre els quals es construeix el model QSPR de predicció, parant especial atenció a la descripció estructural de cadascun dels compostos.

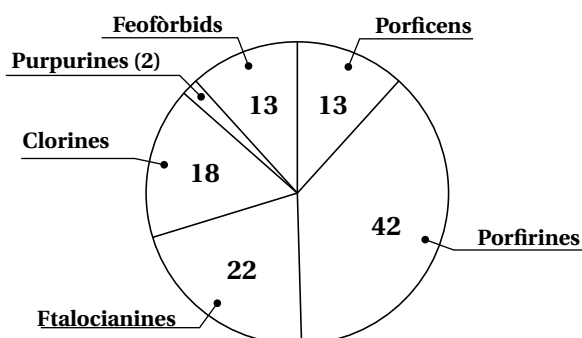


Figura 8.6: Distribució dels FS segons la seva naturalesa.

### Optimització de les estructures

Normalment, quan es desitja representar l'estructura d'una molècula s'empra un programa d'edició que permeti obtenir una representació visualment atractiva d'aquesta. Tot i així, eixa representació acostuma a no coincidir amb l'estructura energèticament més favorable del compost químic, i això té una forta repercussió sobre la predicció de les seves propietats.

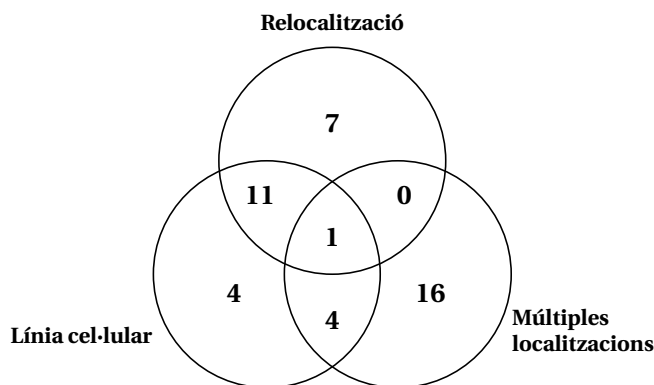
Per aquest motiu, es realitza una optimització de la geometria de cadascuna de les molècules contingudes en la quimioteca virtual per mecànica molecular, mitjançant el programari MOE,<sup>81</sup> i definint el camp de forces (*forcefield*) MMFF94. Donada la naturalesa d'aquesta tècnica, on les propietats atòmiques dels diferents tipus d'àtoms (*atom types*) es troben parametritzades, és necessari revisar detingudament la correcta definició de cada àtom de la molècula, prestant especial atenció als hidrògens units als nitrògens bipirròlics.

### Identificació de punts problemàtics

Degut a l'heterogeneïtat de les dades preses de la bibliografia, gran part dels FS presenten discrepàncies a l'hora d'assignar-los una única localització subcel·lular, fet que dificulta notablement l'objectiu de relacionar l'estructura molecular de cada FS amb un orgànul cel·lular concret. Es consideren quatre situacions problemàtiques:

- Relocalització del FS al llarg del temps. Durant el temps d'irradiació s'observa una redistribució del FS a l'interior de la cèl·lula. Es troben afectats per aquest criteri els FS: **S1, S5, S23, S24, S30, S31, S32, S40, S41, S43, S44, S45, S46, S61, S62, S63, S64, S76, S86**.
- Localització depenent de la línia cel·lular. S'observa que molts FS presenten una localització diferent en funció de la línia cel·lular emprada en la mesura experimental. Aquesta evidència, d'altra banda ja palesa a la bibliografia,<sup>330</sup> enuncia la impossibilitat de descriure correctament la localització a partir de tan sols la naturalesa del FS, sense tenir en compte el seu entorn biològic. Els FS que presenten aquest comportament són: **S1, S5, S23, S24, S30, S32, S35, S39, S42, S44, S45, S46, S54, S58, S60, S61, S62, S84, S86, S96, S101**
- Múltiples localitzacions. Alguns FS presenten des del primer moment afinitat per més d'un orgànul, clarament diferenciats. Els FS que mostren un patró complex de localització són: **S10, S11, S14, S20, S28, S42, S49, S50, S53, S58, S62, S66, S69, S70, S77, S84, S87, S89, S96, S97**

Un total de quaranta-tres FS inclosos dins la base de dades inicial presenta algun de les limitacions anteriors o una combinació d'ells (figura 8.7).



**Figura 8.7:** Distribució de les entrades pel model de localització subcel·lular, que presenten alguna de les limitacions plantejades. Per tant, no poden ser usades directament en la creació del model de predicció

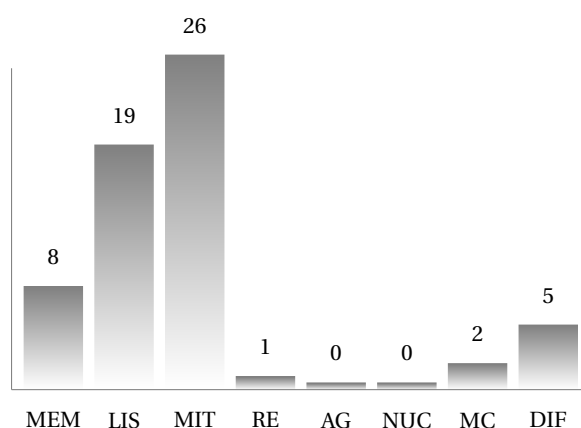
Tots aquests compostos són exclosos inicialment de la base de dades a emprar en l'establiment del model de predicció. A més d'ells, s'eliminen sis FS més, que presenten dificultats en la seva descripció estructural degut a dos motius:

- Estructura indefinida. Alguns FS de primera generació no disposen d'una composició química definida. Entre els resultats de la cerca bibliogràfica només se'n troben dos exemples (**S15, S16**), els quals es descarten directament i no es tindran en compte en la discussió.



- Estats d'agregació. FS que es troben en diferents estats d'agregació (4 FS en total: **S102, S105, S107, S109**).

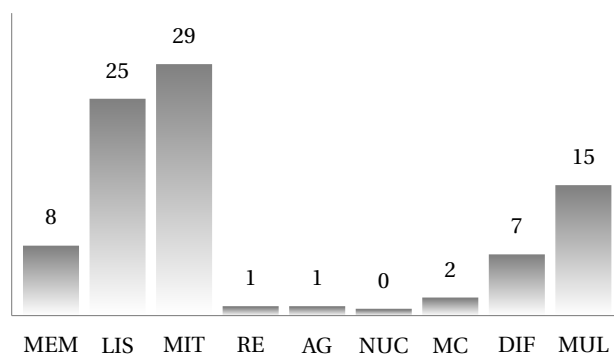
**Distribució de les dades** El conjunt dels 61 FS absents de complicacions (*DATA0*) es distribueixen segons es mostra en la figura 8.8. En ella es pot observar una significativa manca de molècules representatives per a la localització del nucli cel·lular (NUC), del reticle endoplasmàtic (RE) i de l'aparell de Golgi (AG), que posa de manifest la falta d'homogeneïtat de les dades. Des del punt de vista bioquímic, no sorprèn que els FS de segona generació s'hagin dissenyat per evitar la localització nuclear (i perinuclear) degut als problemes que poden comportar (apartat 8.4.5, pàg. 217); però des del punt de vista computacional, la manca de representants d'aquests orgànuls suposa un nou repte per l'acompliment dels objectius.



**Figura 8.8:** Distribució inicial de les dades del conjunt *DATA0* (MEM=Membrana cel·lular; LIS=Lisosomes; MIT=Mitocondri; RE=Reticle endoplasmàtic; AG=Aparell de Golgi; NUC=Nucli; MC=Membranes intracel·lulars; DIF=Localització difosa).

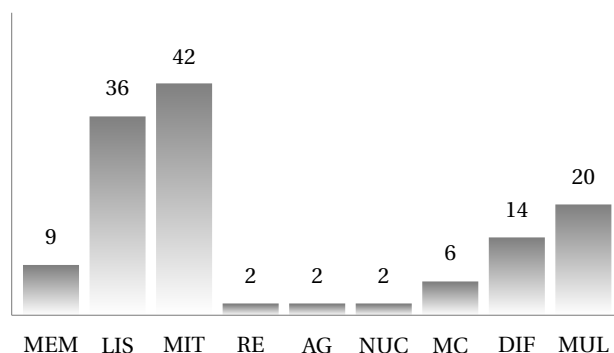
**Refinament de la base de dades** Atenent a la manca d'homogeneïtat de la base de dades, es proposen diferents estratègies per obtenir un conjunt de dades el més homogeni possible, entre de les dades bibliogràfiques disponibles:

1. Incorporar a la base de dades els FS amb relocalització, considerant tan sols l'orgànul inicial i agrupar aquells amb múltiples localitzacions dins d'una nova classe (MUL). L'aplicació d'aquestes mesures permet l'obtenció d'una nova quimioteca (*DATA1*), formada per 88 FS. Malauradament, la nova quimioteca no resol el problema de l'heterogeneïtat, ja que l'orgànul nuclear resta desocupat, figura 8.9.



**Figura 8.9:** Distribució de les dades dins DATA1 – Incorporació dels FS amb relocalització dins de la base de dades amb la implantació de la nova classe MUL.

2. Aprofitar la informació de la línia cel·lular. Cadascun dels FS amb diferent localització subcel·lular segons aquest paràmetre, esdevé una entrada independent de la base de dades. Atent al tipus de càncer, les línies cel·lulars s'agrupen en 5 classes diferents segons estiguin relacionades amb carcinomes, adenocarcinomes, leucèmia, fibroblasts o cèl·lules d'un òrgan concret (a les quals es farà referència al text com a *òrgan*, veure taula 8.3 a la pàg.227). Mantinent els criteris considerats en la definició de DATA1, s'obté una quimioteca de 99 FS diferents que, distingint la naturalesa cel·lular en forma de descriptor nominal, condueixen a 133 entrades (DATA2), amb les quals s'aconsegueix poblar tots els òrgans cel·lulars considerats (figura 8.10). Resultats preliminars amb ANN mostren, però, que la incorporació del descriptor referent a la línia cel·lular no esdevé suficientment rellevant com per modificar convenientment la construcció del model i poder realitzar la classificació de la quimioteca correctament.



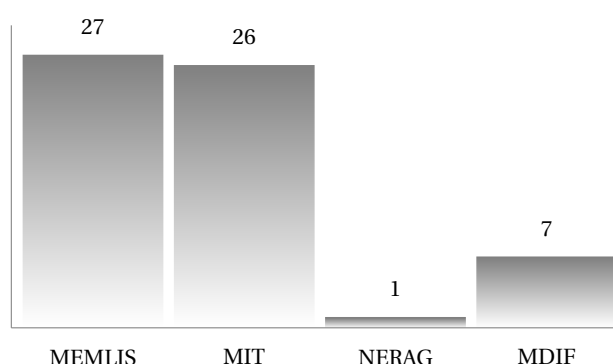
**Figura 8.10:** Distribució de les dades dins DATA2 – Introducció de la línia cel·lular de cada mesura experimental en forma de descriptor nominal, mantenint els criteris establerts dins de DATA1.

3. Agrupar els membres de diferents classes segons les característiques dels òrgans cel·lulars. La nova classificació proposada, de les dades de DATA0, suposa l'agrupació en quatre classes:

- **MEMLIS:** agrupa les dades referents a la membrana cel·lular i els lisosomes, considerant que gran part dels FS que es localitzen dins els lisosomes provenen d'un procés d'endocitosi, on la invaginació de la membrana cel·lular dóna lloc a aquestes estructures.

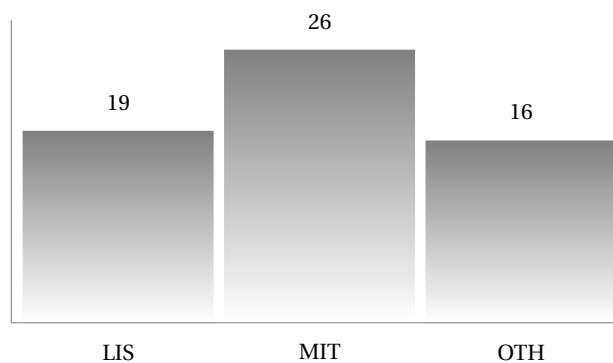
- **MIT**: donada la seva alta representació dins de les dades inicials i a la seva importància terapèutica, el mitocondri conserva la seva pròpia classe.
- **NERAG**: agrupa les dades de nucli i la regió perinuclear (aparell de Golgi i reticle endoplasmàtic), atenent a les dificultats experimentals per a la detecció i assignació de la localització dels FS en ells.<sup>236,331</sup>
- **MDIF**: conté els FS que es troben disseminats pel citosol, ja sigui en la fase aquosa o associats a estructures membranoses.

Aquesta estratègia tampoc soluciona el problema, conduint a una distribució de dades igualment descompensada (*DATA3*, figura 8.11).



**Figura 8.11:** Distribució de les dades dins *DATA3* — Agrupació dels elements pertanyents a les classes: MEM i LIS (MEMLIS); MIT; NUC, RE i GA (NERAG); i MC i DIF (MDIF).

4. Reduir el problema a la identificació dels orgànuls més importants des del punt de vista de la PDT. Donada la gran complexitat del problema a tractar, es planteja el fet de prioritzar la identificació i classificació dels FS localitzats en el mitocondri i als lisosomes, per a la seva importància, i a la seva elevada representació en la base de dades inicial. La resta de FS s'agrupen en una nova classe (OTH). Aprofitant els FS absents de problemes associats s'obté una població de 61 FS més equilibrada (*DATA4*) distribuïda en tres classes, figura 8.12.



**Figura 8.12:** Distribució de les dades dins *DATA4* — Reducció de la base de dades inicial a tres classes (MIT, LIS i OTH) per tal d'afavorir la identificació dels FS que es localitzen preferentment en els orgànuls que desencadenen de forma més eficient la mort cel·lular

Després d'avaluar totes les possibilitats es considera que la reducció a dues localitzacions objectiu (mitocondri i lisosoma) constitueix un bon punt de partida per avaluar la viabilitat d'establir un model de predicció, amb la possibilitat d'estendre'l *a posteriori* a altres orgànuls, a mesura que es disposi de noves dades experimentals o bibliogràfiques de localització.

**Taula 8.3:** Llistat de les diferents línies cel·lulars considerades en la bibliografia, juntament amb la seva classificació segons el tipus de teixit neoplàstic que generen

<b>Línia cel·lular</b>	<b>Descripció</b>	<b>Classe</b>
Ic1c7	Murine hepatoma	<i>òrgan</i>
4R	Rats embryo fibroblasts	<i>fibroblast</i>
9L	Brain tumor	<i>òrgan</i>
A431	Human epidermoid carcinoma	<i>carcinoma</i>
A-549	Human lung adenocarcinoma	<i>adenocarcinoma</i>
B16	Mouse melanoma	<i>carcinoma</i>
C6	Mice glioma	<i>òrgan</i>
chok1	Chinese hamster ovary K1	<i>òrgan</i>
CNE2	Nasopharyngeal carcinoma	<i>carcinoma</i>
colo201	Human colon carcinoma	<i>carcinoma</i>
CT26	Colon carcinoma	<i>carcinoma</i>
D532	Human skin fibroblast	<i>fibroblast</i>
EMT6	Mammary tumor	<i>òrgan</i>
FaDu	Human hypopharyngeal carcinoma	<i>carcinoma</i>
H2T	Hamster pancreatic tumor	<i>òrgan</i>
HaCaT	Human keratinocyte	<i>carcinoma</i>
HeLa	Human epithelial carcinoma	<i>carcinoma</i>
HEp2	Human epidermoid carcinoma	<i>carcinoma</i>
Hex	Rats hexachlorobenzene-fed	<i>òrgan</i>
HT29	Human colon adenocarcinoma	<i>adenocarcinoma</i>
JCS	Myeloid leukemia	<i>leucèmia</i>
KB	Human nasopharyngeal cancer	<i>òrgan</i>
L1210	Murine leukemia	<i>leucèmia</i>
L5178Y-R	Mice lymphoma	<i>leucèmia</i>
LOX	Human melanoma	<i>carcinoma</i>
M1	Mouse myeloid leukemia	<i>leucèmia</i>
MCF7	Human breast cancer	<i>òrgan</i>
MGH-U1	Human bladder cancer	<i>òrgan</i>
MS-2	Fibrosarcoma	<i>fibroblast</i>
NCI-h446	Human lung carcinoma	<i>carcinoma</i>
NHIK	Human cervical carcinoma	<i>carcinoma</i>
NPC/CNE-2	Human squamous cell carcinoma	<i>carcinoma</i>
OAC	Human oesophageal adenocarcinoma	<i>adenocarcinoma</i>
P388	Murine leukemia	<i>leucèmia</i>
PC-3	Human prostate cancer	<i>òrgan</i>
POVD	Human lung cancer	<i>òrgan</i>
R323AC	Rats adenocarcinoma	<i>adenocarcinoma</i>
RB230	Mammary adenocarcinoma	<i>adenocarcinoma</i>
RIF1	Fibrosarcoma	<i>fibroblast</i>
SC1	Human squamous cell carcinoma	<i>carcinoma</i>
SC2	Human squamous cell carcinoma	<i>carcinoma</i>
SSK2	Murine fibrosarcoma	<i>fibroblast</i>
V79	Chinese Hamster lung fibroblast	<i>fibroblast</i>
Gf	Mice griseofulvin-fed	<i>òrgan</i>
Colo-26	Murine colon carcinoma	<i>carcinoma</i>
CHO	Chinese Hamster ovary	<i>òrgan</i>

### 8.6.2 Descriptors moleculars

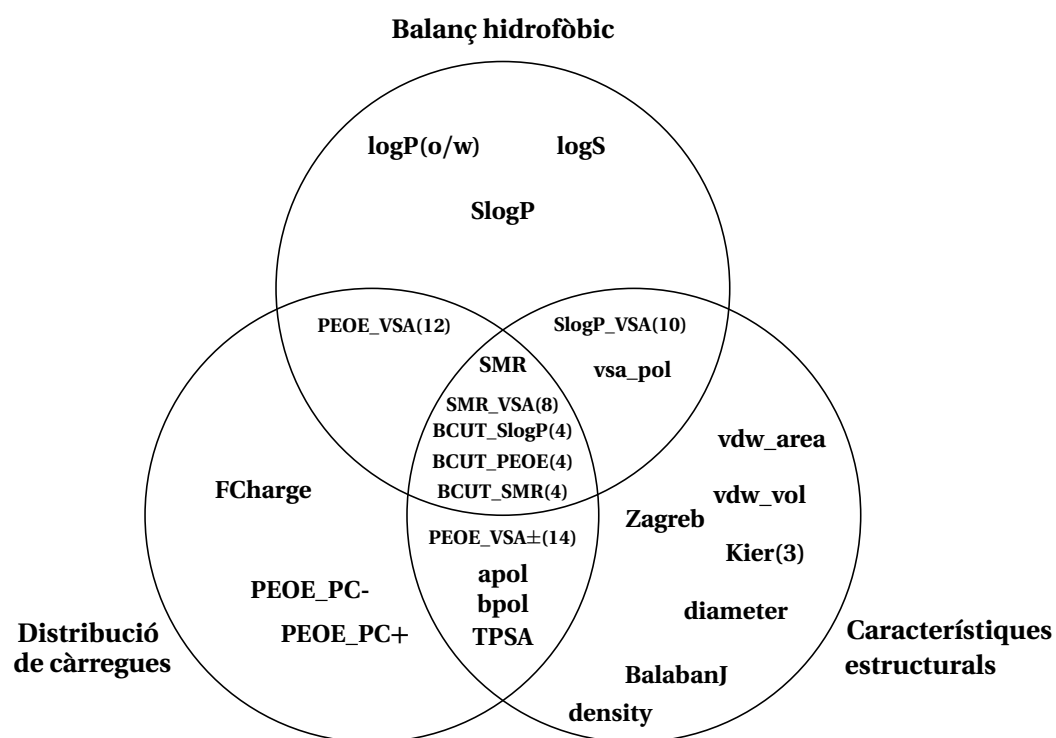
Una vegada definida la quimioteca, cal decidir de quina manera les estructures dels seus elements ens permetran relacionar-les amb seva localització subcel·lular. La identificació d'aquelles propietats moleculars que poden influir en la definició de la localització subcel·lular dels FS és una de les tasques més difícils i crítica, doncs constitueix la descripció matemàtica del problema, i de la qual en depèn l'èxit del model de predicció.

**Traducció de la informació química** S'entenen per descriptors moleculars al conjunt de totes aquelles eines matemàtiques que permeten extreure algun tipus d'informació útil a partir d'una estructura molecular. Aquesta informació sovint està relacionada amb paràmetres fisicoquímics, topològics o geomètrics; paràmetres que permeten, en última instància, quantificar propietats químiques dels compostos en estudi. Mitjançant aquestes operacions, hom és capaç de traduir aquesta informació en un llenguatge intel·ligible per un ordinador.

En el cas concret de la localització subcel·lular de FS tetrapirròlics, l'estudi de les dades experimentals bibliogràfiques permet la identificació de tres característiques principals, relacionades directament amb l'estructura molecular d'aquests compostos: el balanç hidrofòbic, la distribució de càrregues i característiques estructurals relacionades amb la mida del FS (apartat 8.5, pàg.217). Aquesta consideració es troba, a més, recolzada per altres models estructura-localització descrits en la bibliografia.<sup>328</sup>

Mitjançant el programari MOE2007.09<sup>86</sup> es calculen inicialment un total de 230 descriptors, que no inclouen aquells basats en mètodes semiempírics. Es comprova computacionalment que l'optimització de l'estructura dels FS realitzada amb MM difereix substancialment de l'obtinguda amb QM, demostrant que els descriptors 3D podrien estar sotmesos a un error considerable (el temps de càlcul necessari per a l'optimització *ab initio* de tots els FS pot esdevenir prohibitiu). S'opta, doncs, per treballar tan sols amb aquells descriptors basats en la informació derivada de la connectivitat i la identitat de cadascun dels àtoms de la molècula (*i.e.* 180 descriptors 2D). Tanmateix, només 128 d'ells es troben directament relacionats amb les tres característiques anteriors i són els que finalment s'han utilitzat i mostrat en la figura 8.13.

Finalment, donat que la naturalesa de molts d'aquests descriptors és semblant, pot ser que alguns d'ells es trobin correlacionats. Per aquest motiu, mitjançant el programari Rapid Miner 5.1<sup>332</sup> i per tal d'evitar treballar amb informació repetida, s'eliminen tots aquells descriptors que presenten un coeficient de correlació de Pearson superior a 0.9 (eliminant en cada cas, aquell descriptor que presenta un major valor de correlació respecte al global de descriptors). D'aquesta manera es redueix el nombre de descriptors a 76, sense renunciar a la completa descripció de l'espai químic, figura 8.13.



**Figura 8.13:** Esquema conceptual de la distribució dels 76 descriptors moleculars calculats per al model de localització subcel·lular, segons les tres principals característiques fisicoquímiques considerades.

**Selecció de descriptors** Emprar el màxim nombre de descriptors possibles no acostuma a ser una bona estratègia per assegurar l'eficàcia d'un mètode de predicció; és més aviat al contrari, quant major és el nombre de descriptors emprats més possibilitats hi ha que el model s'especialitzi tant durant l'etapa d'aprenentatge, que no sigui capaç d'interpretar les entrades que difereixen lleugerament de les anteriors, procés que es coneix amb el nom d'*overfitting*.<sup>333</sup> Per aquest motiu és aconsellable reduir la dimensió de l'espai químic realitzant una selecció d'aquells descriptors més rellevants per a la definició de la localització subcel·lular dels FS.

Amb aquesta finalitat, s'ha valorat la utilització de diferents algorismes de selecció, per identificar aquell conjunt de descriptors que permeti discriminar més eficaçment els FS amb localització mitocondrial i lisosòmica.

1. **Anàlisi de components principals (Principal Component Analysis, PCA).** A grans trets, la PCA (pàg. 125) és una transformació lineal que, aplicada sobre un conjunt de possibles variables correlacionades, permet definir un reduït nombre de variables no correlacionades anomenades components principals (*Principal Components, PC*). Per realitzar la transformació de les variables sense perdre la informació de l'espai químic generat, la PCA aplica projeccions que descriguin un determinat màxim de variància (típicament es fixa el percentatge de variància explicada entorn al 90%).

En aquestes condicions s'obtenen 11 PC, dins les quals el conjunt de la primera i la segona PC presenta una variància acumulada de prop del 50%. Aquesta situació indica que els descriptors

amb major pes dins d'aquest grup, mostraran segurament una forta influència sobre la definició de l'espai químic. S'identifiquen així els 7 descriptors més rellevants segons la PCA, taula 8.4.

2. **Forward Selection-Backward Elimination (FS-BE)**. Aquest mètode combina dues estratègies diferents per a la selecció de variables. El procés de selecció pot tant avaluar l'efecte d'incorporar progressivament els diferents descriptors sobre una funció resposta (*forward selection*), com comparar el resultat de la operació contrària (*backward elimination*), on s'avalua la seva eliminació.<sup>334</sup> En tots els estudis s'ha realitzat la selecció de descriptors seguint el sentit *forward* de l'algorisme. La funció resposta necessària per a l'aplicació d'aquest mètode correspon a una definició de la bondat de la selecció. En el cas concret de la classificació dels FS, s'ha escollit la capacitat de discernir correctament entre MIT, LIS i OTH utilitzant dos algorismes de classificació: s'ha avaluat la correcta classificació dels FS de *DATA4* aplicant l'anàlisi discriminant lineal (*Linear Discriminant Analysis*, LDA) i el mètode *k-Nearest Neighbors* (*k-NN*).

**Taula 8.4:** Llistat dels descriptors seleccionats segons els diferents mètodes emprats.

PCA	FS/LDA	FS/k-NN
PEOE_VSA_FPPOS	BCUT_PEOE_3	PEOE_VSA-6
PEOE_VSA_FPOS	BCUT_SMR_2	SMR
density	Fcharge	SMR_VSA7
PEOE_VSA_FPOL	dipole	SlogP_VSA5
BCUT_SlogP_1	SlogP	BCUT_SlogP_2
BCUT_PEOE_2	SMR_VSA0	
BCUT_SMR_2	SMR_VSA3	

Els resultats mostrats en la taula 8.4 posen de manifest la complexitat associada al procés de selecció dels descriptors. Es comprova que la tria no és en absolut trivial, sinó que segons el mètode utilitzat hom obté resultats poc comparables. Bo i que la bondat de cada conjunt front a la classificació subcel·lular no es pot avaluar a priori, sí que es pot discutir la seva adequació amb la seva naturalesa química, és a dir, estudiant la disposició de cadascuna de les famílies de FS dins l'espai químic definit per cadascun dels conjunts de descriptors.

Malauradament, la dimensió dels espais de descriptors PCA, LDA i *k-NN* és superior a tres, de manera que tot esforç que hom faci per imaginar-se la distribució de cada FS en un espai tridimensional resulta inútil. No conforme amb renunciar a la valuosa informació que es podria arribar a extreure de l'anàlisi gràfica dels resultats, hom pot recórrer a mètodes matemàtics i de *data mining* que permetin projectar un espai *n*-dimensional sobre un de bidimensional. Per a l'estudi dels FS s'han aplicat dues tècniques de reducció de la dimensió sobre les dades de *DATA4*: la projecció sobre el pla  $\mathcal{H}^2$  i els *Self-Organizing Maps*.

- Projecció sobre  $\mathcal{H}^2$ : Aquesta tècnica permet visualitzar dades multidimensionals mitjançant la seva projecció sobre el pla complex. Aprofitant el model del disc de Poincaré,<sup>103</sup> la distància hiperbòlica entre cadascun dels punts és proporcional a la seva dissimilitud. La projecció es



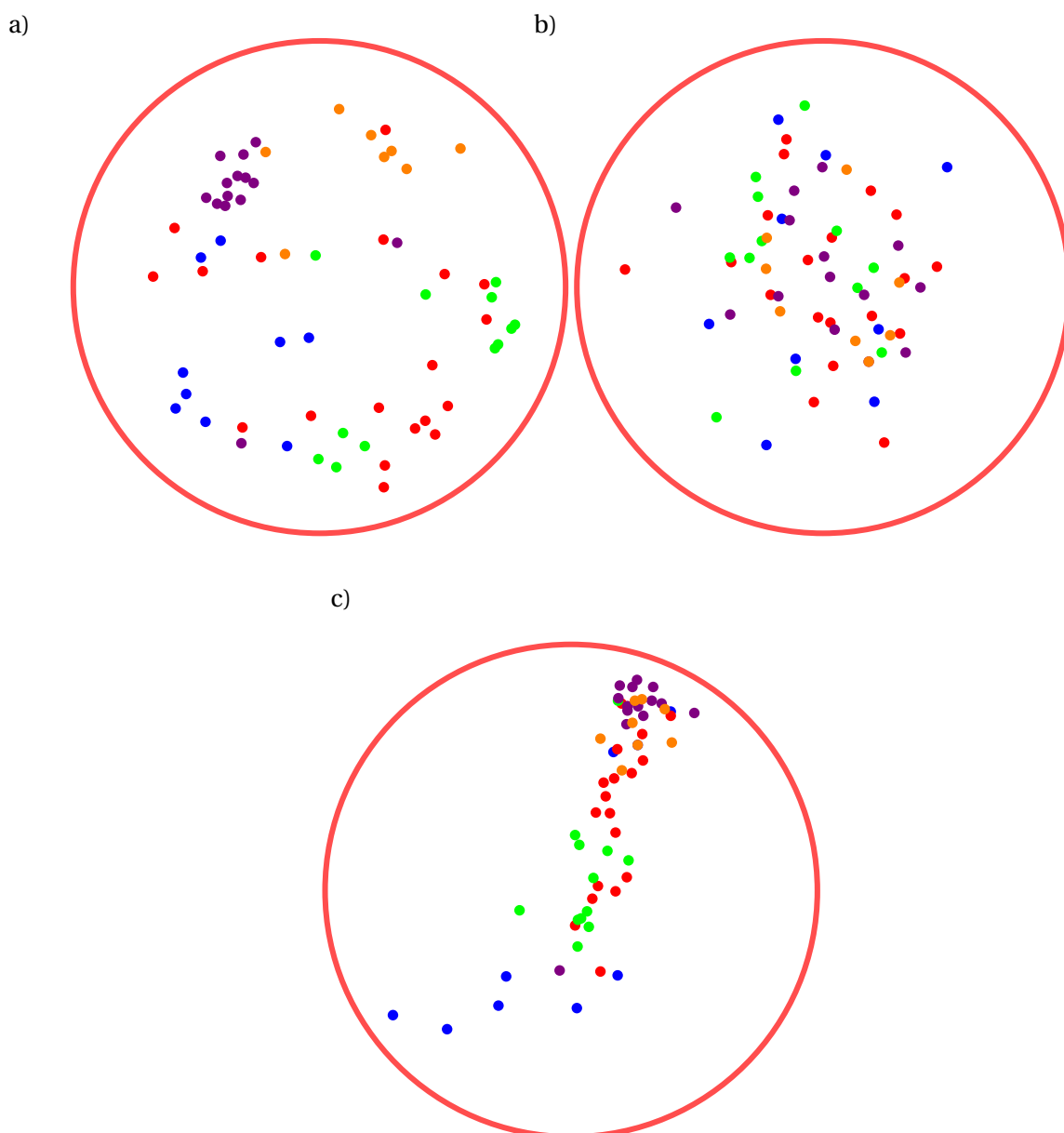
realitza optimitzant simultàniament les coordenades de tots els FS sobre el pla complex, mitjançant un algorisme genètic i definint com a criteri de semblança la distància *cosinus*. En tots els càlculs realitzats s'ha considerat com a valor d'aturada un *score* de 0.05. La descripció detallada de com s'ha realitzat la programació i la implementació d'aquesta tècnica es troba al capítol 6, pàg. 125.

- Self-Organizing Maps (SOM): Els SOM corresponen a algorismes basats en les xarxes neuronals amb aprenentatge no supervisat, que permeten projectar dades sobre un pla bidimensional. A diferència del cas anterior, els SOM no es troben subjectes a una definició de la semblança molecular i agrupen les dades segons els valors d'entrada (secció 6.5, pàg. 141).

La projecció bidimensional de l'espai químic sobre  $\mathcal{H}^2$  permet evidenciar les diferències entre cadascun dels tres conjunts de descriptors seleccionats. Des de punt de vista del tipus de FS, les molècules representades pels descriptors seleccionats mitjançant els mètodes PCA i *k*-NN es disposen de forma ordenada, agrupant-se per famílies, mentre que mitjançant LDA adopten una configuració més aleatòria (figura 8.14).

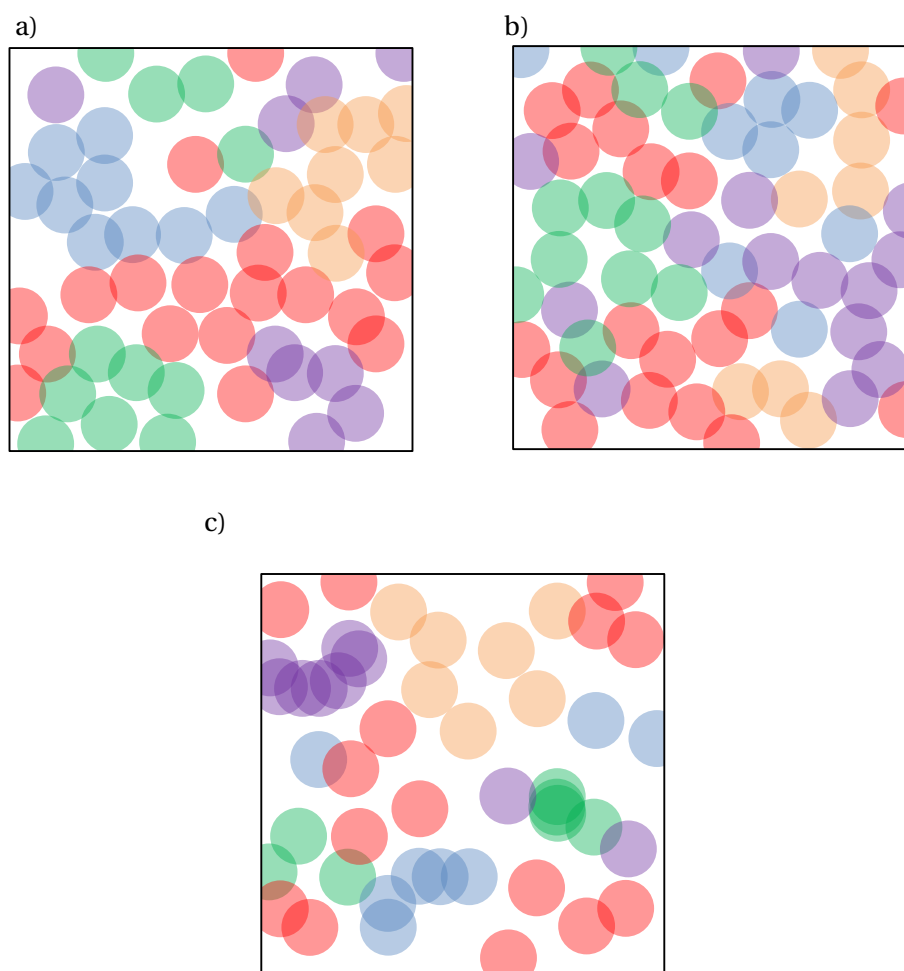
Cal remarcar que la relació de semblança mostrada per *k*-NN indica que les clorines i els feofòrbids es consideren químicament semblants i disten de les ftalocianines (més apartades), tenint com a fil conductor a les porfirines, que es mostren les més versàtils i s'estenen al llarg de tot l'espai químic definit per aquests tres macrocicles (de fet, les porfirines són els FS que presenten més varietat estructural, la qual cosa explicaria la seva dispersió). Pel que fa als porfícens, esdevenen els elements més diferents a la resta, localitzant-se majoritàriament en un dels extrems del diagrama.

El conjunt de descriptors seleccionats per PCA mostren un comportament totalment diferent, disposant els punts del seu espai químic de forma circular, on les porfirines continuen estenent-se al llarg de tot l'espai però amb una major tendència a agrupar-se amb les ftalocianines més que amb les clorines i feofòrbids. Amb tot, sembla que els FS no porfirínics estiguin equidistants entre ells, figura 8.14.



**Figura 8.14:** Representació bidimensional sobre el disc de Poincaré dels conjunts de descriptors seleccionats mitjançant PCA (a), LDA (b) i  $k$ -NN (c). [● porficens, ● porfirines, ● ftalocianines, ● clorines, ● feofòrbids]

Per altra banda, amb l'aplicació dels SOM (calculats amb el programari Rapid Miner 5.1<sup>332</sup>) s'obtenen uns resultats anàlegs al cas anterior. En els descriptors  $k$ -NN les clorines i feofòrbids s'agrupen en la part superior de l'espai, lleugerament apartats de les ftalocianines, que són les que s'acosten més als porficens; les porfirines, amb tot, es troben repartides per tot l'espai, posant de manifest la seva gran diversitat (figura 8.15(c)). Els descriptors LDA no permeten definir un clar patró de classificació per grups. Finalment, els descriptors derivats de la PCA, la seva disposició és modèlica, compartimentant l'espai de forma clara per a cadascuna de les famílies; tanmateix, els porficens passen a ser el punt de trobada d'ftalocianines, porfirines i feofòrbids, figura 8.15.



**Figura 8.15:** Self Organizing Map establerts sobre els conjunts de descriptors seleccionats mitjançant PCA (a), LDA (b) i  $k$ -NN (c). [● porficens, ● porfirines, ● ftalocianines, ● clorines, ● feoförbids]

Bo i que aquests resultats són orientatius, el sentit químic faria pensar que sent els porficens els únics FS considerats que no contenen en la seva estructura el nucli porfíric, haurien de ser aquests els més dispers (apuntant a la conveniència d'emprar els descriptors  $k$ -NN). Si bé es cert que l'objectiu del model no és la classificació per famílies sinó la seva correlació amb la localització subcel·lular, aquest no és un criteri determinant, però sí a tenir en compte. Segons aquest, tant la selecció de descriptors efectuada mitjançant PCA com amb  $k$ -NN podrien esdevenir bons candidats per establir el model de predicció, mentre que el conjunt LDA podria definir incorrectament l'espai químic generat per la quimioteca DATA4.

Tenint en compte aquests resultats, s'emplaça la determinació del millor conjunt de descriptors una vegada s'obtinguin resultats preliminars i es pugui realitzar una valoració objectiva de la seva capacitat de predicció de la localització subcel·lular de FS.

### 8.6.3 Model de predicció de la localització mitocondrial i lisosòmica

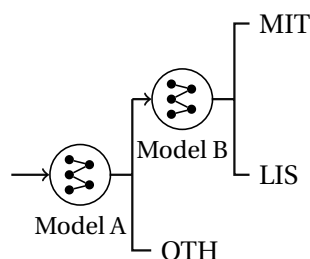
Havent trobat a la bibliografia que les xarxes neuronals artificials han estat àmpliament aplicades en l'estudi de problemes relacionats amb la localització subcel·lular i la classificació de biomolècules, es decideix mantenir-les com a principals candidates per establir un model de predicció, sense renunciar definitivament a la resta de tècniques. En el fons, l'obtenció d'un model lineal de predicció facilitaria considerablement l'anàlisi dels resultats i la identificació de les propietats característiques de cada localització intracel·lular. Malauradament s'ha pogut comprovar experimentalment que no més el model obtingut aplicant ANN és capaç de tractar correctament les dades de la quimioteca creada a partir de dades bibliogràfiques.

Inicialment es planteja encarar la classificació ternària de les dades en mitocondri (MIT), lisosomes (LIS) i altres orgànuls (OTH). Aquest objectiu esdevé especialment problemàtic d'assolir tant amb ANN com amb altres algorismes de classificació (arbres de decisió binaris i LDA). Diversos estudis realitzats amb aquests mètodes (mitjançant diferents programaris) resulten en taxes de reconeixement màximes del 55% a causa de l'homogeneïtzació de les prediccions entorn a la classe predominant.

En canvi, en un intent de portar a l'extrem la simplificació del problema i realitzar una classificació binària entre dos orgànuls (per exemple, discriminant tan sols MIT de LIS), la situació millora considerablement. Per aquest motiu, s'estudia la manera de simplificar el model de predicció global a l'agrupació i organització de diferents processos de classificació binaris que permetin obtenir resultats útils.

### 8.6.4 Arbres neuronals de decisió

La finalitat de prioritzar la correcta classificació dels FS amb localització preferent al mitocondri o als lisosomes, radica en el fet que aquests són els dos orgànuls més rellevants des del punt de vista de la PDT (apartat 8.4, pàg.212), sent alhora els més poblats en la distribució de les dades experimentals bibliogràfiques. Per aquest motiu es planteja un model basat en un arbre de decisió, el criteri de classificació del qual es basa en el resultat d'una xarxa neuronal classificatòria. Aquest esquema es coneix amb el nom d'arbres de xarxes neuronals (*Neural Network Trees*, NNTrees)<sup>335</sup> i permet realitzar la classificació dels FS de forma seqüencial (figura 8.16), reduint el procés a la concatenació de diferents classificacions binàries, ben definides, en forma d'etapes. Inicialment es defineix una primera etapa on una ANN determina si el FS es localitza preferentment a MIT o LIS en comptes d'altres orgànuls (Model A) i en cas afirmatiu, una segona ANN indica quin d'aquests dos orgànuls esdevé el destí final més probable del FS (Model B), figura 8.16.



**Figura 8.16:** Diagrama esquemàtic del model seqüencial inicial basat en la metodologia NNTree.

Aquest plantejament fa possible l'ús de la base de dades *DATA4* per a l'establiment d'ambdós models, evitant d'aquesta manera les aproximacions introduïdes arrel dels problemes associats a re-localitzacions, línia cel·lular i múltiples localitzacions.

### Paràmetres constitucionals de les xarxes

L'arquitectura de cadascuna de les xarxes neuronals artificials incloses dins del model ha estat estudiada independentment i prèvia a l'adquisició de resultats. Prenent com a punt de partida diferents directrius empíriques proposades<sup>52</sup> (com per exemple el test de  $\rho$ <sup>74</sup>), el nombre de capes i neurones amagades han estat optimitzades en cada model (model A i model B), definint com a límits superiors 2 capes amagades i 8 neurones a cadascuna d'elles. Amb tot, la topologia es manté constant en tots els casos i es fixa seguint l'esquema *fully-connected*.

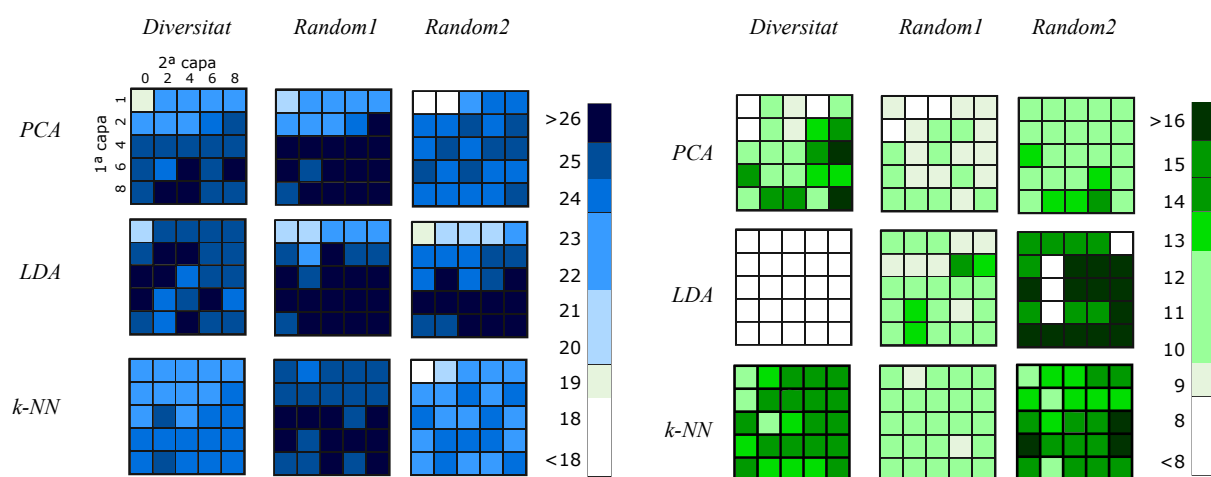
Resultats preliminars permeten fixar altres paràmetres interns de l'algorisme: es defineix la funció sigmoïdal com a funció d'activació i es manté el *learning rate* ( $\eta$ ) entre 0.1 i 0.001, que garanteix un bon balanç entre velocitat i precisió del procés d'entrenament. Per la seva part, s'empra un procés d'entrenament supervisat basat en l'algorisme de *back-propagation*, on s'aplica una validació interna en forma d'*split-half validation*, per tal de millorar la capacitat de predicció del model, que s'avalua sobre una part de les dades inicials (que es reserven com a *test set*). Tant els pesos inicials com els valors de *bias* s'assignen aleatòriament dins de l'interval  $[-0.5, 0.5]$ .

La capacitat de discernir entre la localització mitocondrial i lisosòmica és fonamental en el disseny i desenvolupament del model de classificació global. Per aquest motiu, es pren aquesta classificació per exemplificar la manera com es realitza l'optimització de la topologia de la xarxa. El model es crea a partir de la informació descrita en les 45 dades de localització MIT i LIS de *DATA4* (d'aquestes, un 60% és utilitzat com a conjunt d'entrenament i els 18 FS restants com a conjunt de validació externa).

La optimització de la topologia de la ANN es realitza de forma seqüencial, avaluant les diferents combinacions que es poden generar en variar el nombre de neurones, entre 1 i 8 en la primera capa amagada i entre 0 i 8 en la segona (considerant tan sols les topologies amb un nombre parell de neurones amagades). Aquest estudi es realitza pels tres conjunts de descriptors definits en l'apartat anterior, fent possible identificar el millor d'ells alhora que optimitzar els paràmetres estructurals. Així doncs, per a cada topologia s'han realitzat 30 repeticions del càlcul mantenint  $\eta = 0.01$ , i se n'ha avaluat la capacitat de reconeixement i de predicció. El diagrama de la figura 8.17 mostra el nombre

de classificacions correctes —dels conjunts d’entrenament (en blau) i de predicció (en verd)— del millor model per a cada conjunt de descriptors i de topologia emprats.

A més, per tal d’avaluar la capacitat de generalització a l’hora d’establir el model i estudiar la seva dependència amb les dades inicials, es varia també la naturalesa de les dades d’entrenament. Es defineixen un total de tres poblacions diferents: dues seleccions aleatòries (*random1* i *random2*, a la figura 8.17) comunes per a tots els conjunts de descriptors i una selecció en diversitat. Mentre que les dues seleccions aleatòries són comunes per als tres conjunts de descriptors, la selecció en diversitat sobre la quimioteca de 45 FS depèn de cadascun d’ells, doncs se cerca aquell conjunt de 27 compostos més representatius de tota la població segons uns descriptors concrets. La selecció en diversitat es realitza mitjançant els algorismes de cerca basats en algorismes genètics implementats en el programari PRALINS,<sup>71</sup> definint la distància euclidiana entre descriptors com a mesura de la dissimilitud molecular, utilitzant el mètode d’encreuament *two-point*, la selecció *Roulette-Wheel* i un mètode d’elitisme per evitar la homogeneïtzació prematura de la població.



**Figura 8.17:** Resultats obtinguts per a la classificació entre MIT i LIS segons el conjunt de descriptors i la topologia de la ANN emprada. S’ha fixat a vuit el nombre màxim de neurones amagades en cadascuna de les capes amagades. Els resultats mostrats indiquen, segons l’escala de colors adjacent, el nombre de FS correctament classificats tan en el conjunt d’entrenament (en blau, sobre un total de 27 FS) com en el de validació externa (en verd, sobre 18 FS).

Els resultats obtinguts en l’estudi de les condicions de càlcul demostren que els descriptors seleccionats amb el mètode *k*-NN són els més estables, des del punt de vista de la dependència del model amb la identitat del conjunt d’entrenament, i presenten uns valors de predicció més que acceptables. En canvi, els resultats amb els descriptors PCA es troben per sota del conjunt anterior, mostrant potser un cert *overfitting* en les topologies amb major nombre de neurones, que podria explicar l’elevat reconeixement i la disminució de la capacitat de predicció. Tot i així, es comporta com hom esperaria, mostrant millors resultats en la població on s’ha realitzat la selecció en diversitat, ja que s’entrena el model amb els FS més representatius de tot el conjunt. Finalment, el conjunt LDA mostra un comportament erràtic que fa témer una baixa capacitat de generalització, tot i la seva extraordinària capacitat de predicció mostrada en *random2*, la qual contrasta amb els pobres resultats de les altres dues poblacions.

Aquests resultats, juntament amb els obtinguts en la projecció dels diferents conjunts de descrip-

tors sobre un espai bidimensional, porten a l'elecció dels descriptors seleccionats mitjançant  $k$ -NN per a ser els encarregats de generar el model de classificació final, taula 8.5.

**Taula 8.5:** Definició dels descriptors moleculars amb més rellevància en la reducció de l'espai químic, mitjançant el mètode *forward selection* amb l'algorisme  $k$ -NN.

Nom del descriptor	Definició
PEOE_VSA-6	Suma de les àrees atòmiques associades a la superfície de van der Waals, on les càrregues atòmiques parcials (calculades emprant el mètode PEOE <sup>336</sup> ) són menors que $-0.30$ .
SMR	Refractivitat molar. Combina la mida del sistema molecular amb la seva polaritzabilitat <sup>171</sup>
SMR_VSA7	Suma de les subdivisions de les àrees atòmiques on SMR és major a $0.56$ . La informació obtinguda de SMR es combina amb l'àrea de la superfície molecular
SlogP_VSA5	Suma de les subdivisions de les àrees atòmiques on la contribució atòmica al $\log P$ es troba entre $(0.15, 0.20]$ . D'aquesta manera es posa en contraposició la mida de la molècula i la seva hidrofobicitat.
BCUT_SlogP_2	Valors propis de la matriu d'adjacència modificada <sup>337</sup> on els elements de la diagonal principal prenen el valor de la contribució atòmica de $\log P$ . <sup>171</sup>

Pel que fa a la topologia, els resultats anteriors ressalten la necessitat d'emprar dues capes amagades amb un mínim de quatre neurones cadascuna d'elles. D'altra banda, també apunten a la possibilitat d'haver d'afinar el valor de  $\eta$  a utilitzar en aquells casos amb un major nombre neurones, a fi d'evitar l'*overfitting*.

### Identificació de la localització mitocondrial/lisosòmica (Model A)

La divisió del model de classificació en dos models binaris suposa modificar, de nou, la distribució de les dades del *training set*. En el cas del model A, l'agrupament de les classes MIT i LIS produeix una nova classe (M&L) amb 45 membres, un nombre de FS molt superior als 16 FS de la classe OTH. Un conjunt d'aquestes dades se separa per formar el grup de validació externa (*test set*), permetent a les restant entrar dins del procés d'entrenament.

L'establiment del conjunt d'entrenament es troba altament condicionat pel conjunt OTH doncs és el menys poblat i el que, per tant, en dictaminarà la mida, contribuint-hi amb 10 FS. Per tal d'assegurar que aquests FS siguin representatius de tots els membres de la classe OTH, se seleccionen per mitjà d'una selecció en diversitat utilitzant el programari PRALINS,<sup>71</sup> emprant un algorisme genètic com a mètode de selecció i la distància Euclídea —definida dins de l'espai dels 5 descriptors  $k$ -NN— com a mesura de la semblança química entre molècules. D'altra banda, per tal de mantenir el *training set* equilibrat en ambdues classes, s'hi inclou el 33% dels membres de la classe M&L, seleccionats aleatòriament. La distribució final de les dades es presenta a la taula 8.6.

**Taula 8.6:** Distribució de les dades dins del conjunt d'entrenament en el model A de localització.

Grup	M&L	OTH	Total
<i>Training set</i>	15	10	25
<i>Test set</i>	30	6	36
<b>Total</b>	45	16	61

Seguint les directrius exposades en la tasca d'optimització dels paràmetres estructurals, un procés seqüencial de prova i ajust permet identificar progressivament el nombre òptim de neurones en en cada una de les dues capes amagades (6 i 4, respectivament) i finalment el valor de  $\eta$  (0.01). S'enregistren cinquanta intents de classificació de les dades amb els paràmetres òptims per tal d'obtenir el millor model de classificació, el qual mostra una capacitat de reconeixement del 96% i un 89% de predicció, i presenta les matrius de confusió corresponents a la taula 8.7.

**Taula 8.7:** Matrius de confusió dels conjunts d'entrenament (a) i de validació externa (b) obtingudes en el model A de localització.

		Predicció				Predicció	
		M&L	OTH			M&L	OTH
<i>Bibliogràfic</i>	<b>M&amp;L</b>	15	0	<i>Bibliogràfic</i>	<b>M&amp;L</b>	27	3
	<b>OTH</b>	1	8		<b>OTH</b>	1	5

Cal remarcar que els tres FS que esdevenen falsos positius de la localització OTH són les porfirines **S47**, **S48** i **S55** que, sorprenentment, corresponen a les úniques porfirines que cauen dins de l'espai ocupat per clorines i feofòrbids en la representació sobre el pla  $\mathcal{H}^2$ , figura 8.14. La mala predicció de **S55** s'atribueix a la presència de l'àtom de luteci. Des del punt de vista estructural, s'observa que els compostos **S47** i **S48** són derivats tioporfirínics (en els quals els nitrògens de les posicions 21 i 23 s'han substituït per sofres). Si bé la presència de sofres dins de l'estructura molecular de les cadenes laterals no suposa un problema, la seva presència dins del *core*, sí que sembla afectar significativament (probablement degut a la manca de FS representatius en el conjunt d'entrenament).

La inclusió d'un dels derivat tioporfirínics dins del conjunt d'entrenament (intercanviant **S81** per **S47**, a fi de mantenir el nombre de FS descrit en la taula 8.6), i mantenint els mateixos paràmetres estructurals, no permet millorar els resultats globals anteriors: bo i que el reconeixement augmenta fins al 100%, la capacitat de predicció resta al 81%. Per aquest motiu, i tot i que la seva incorporació permet classificar correctament el seu congènere **S48**, la seva incorporació no justifica el canvi de model. Per tant, s'identifica la presència de sofres a l'interior del nucli porfirínic com una possible limitació del model de predicció, fet que es pot avaluar a priori emprant la projecció sobre  $\mathcal{H}^2$ .



**Assignació de la localització mitocondrial/lisosòmica (Model B)**

Les classes MIT i LIS mostren, ja d'entrada, una distribució ben balancejada, permetent equilibrar el nombre de FS associats al *training set* i al *test set*, i seleccionar els dos conjunts aleatòriament, taula 8.8.

**Taula 8.8:** Distribució de les dades dins del training set en el model B de localització.

Grup	LIS	MIT	Total
<i>Training set</i>	12	15	27
<i>Test set</i>	7	11	18
<b>Total</b>	19	26	45

Els paràmetres estructurals òptims per a l'ANN es defineixen de la mateixa manera que en el cas anterior. El millor model s'obté emprant una topologia 5-6-4-2 i un *learning rate* de 0.01, el qual presenta unes taxes de reconeixement i predicció del 89%, corresponents als resultats mostrats en la taula 8.9. Donada la importància del mitocondri, cal destacar l'absència de falsos negatius en l'adjudicació de la classe MIT. Per altra banda, els falsos positius corresponen als compostos **S79**, **S85**, **S90** del conjunt d'entrenament i **S27**, **S81** del conjunt de validació. Si bé no s'identifica un motiu clar que expliqui el comportament del model front a **S27** i **S90**, la mala assignació de **S79** (atribuïble a la manca de FS de característiques estructurals semblants) i **S85** (únic compost del conjunt d'entrenament coordinat amb estany), poden explicar el resultat obtingut per a **S81**.

**Taula 8.9:** Matrius de confusió dels conjunts d'entrenament (a) i de validació externa (b) obtingudes en el model B de localització.

		Predicció	
		LIS	MIT
<i>Bibliogràfic</i>	<b>LIS</b>	9	3
	<b>MIT</b>	0	15

		Predicció	
		LIS	MIT
<i>Bibliogràfic</i>	<b>LIS</b>	5	2
	<b>MIT</b>	0	11

Els resultats globals d'ambdues etapes del model de decisió NNTree es mostren en la taula 8.10.

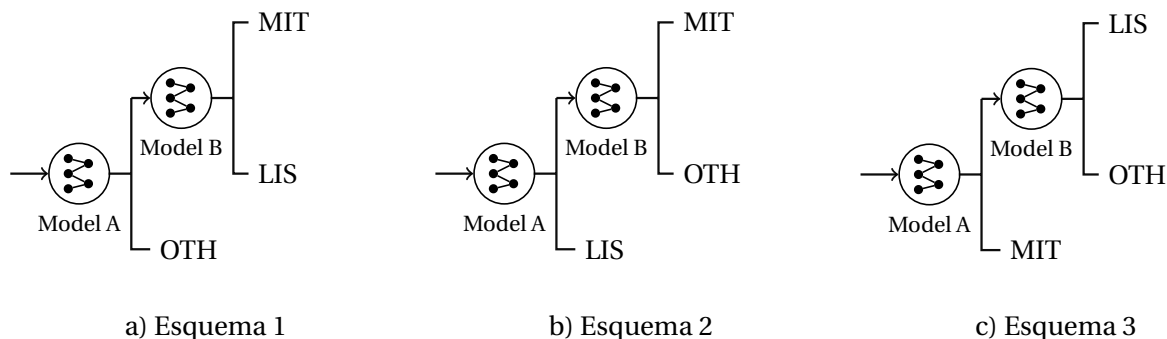
**Taula 8.10:** Comparació dels resultats obtinguts en els models de predicció binaris per a la localització subcel·lular.

Model	Topologia	$\eta$	% reconeixement	% predicció
A. Identificació M&L	5-6-4-2	0.01	96	89
B. Assignació MIT i LIS	5-6-4-2	0.01	89	89

**Comparació entre els diferents esquemes**

L'esquema seqüencial proposat anteriorment no és l'únic que es pot definir per a la classificació ternària de MIT, LIS i OTH. De fet, és possible agrupar les classes de tres maneres diferents, figura 8.18.

Tot i que l'esquema 1 acabat d'analitzar, resulta el més intuïtiu des del punt de vista conceptual, ja que primer discrimina els FS pertanyents a OTH i deixa per després la classificació concreta entre LIS i MIT, es repeteix l'estudi anterior per la resta de possibilitats (esquema 2 i 3).



**Figura 8.18:** Diagrama esquemàtic dels tres models que es poden plantejar segons la metodologia NNTree proposada.

Tant les condicions estructurals definides en l'esquema 1 com la metodologia emprada per identificar el millor model de classificació es mantenen en els dos nous estudis (*i.e.* topologia,  $\eta$ , etc.). A la taula 8.11 es mostra la capacitat de reconeixement i de predicció pels tres esquemes possibles.

**Taula 8.11:** Comparació de la capacitat de reconeixement i de predicció entre les diferents esquemes proposats.

MODEL	Procés	Esquema		
		1	2	3
A	Reconeixement	96%	100%	89%
	Predicció	89%	75%	67%
B	Reconeixement	89%	100%	86%
	Predicció	89%	72%	87%

Tal i com es pot observar, l'esquema considerat inicialment (esquema 1) és el que permet obtenir millors resultats de predicció en ambdues etapes, i per tant, esdevé el model escollit com a definitiu per a la identificació de FS tetrapirròlics que presentin localització subcel·lular preferent mitocondrial o lisosòmica.

### Discussió dels resultats

Donat que els FS definits en el conjunt d'entrenament d'ambdós models són diferents, cal estudiar l'eficàcia del model de classificació global de forma independent sobre el conjunt dels 61 FS de DATA4. Si s'introdueixen tots els membres d'aquesta quimioteca en l'NNTree de decisió, format per l'aplicació seqüencial dels models A i B anteriors, s'obté una capacitat de classificació del 84% i la matriu de classificació corresponent a la taula 8.12.

**Taula 8.12:** Matriu de classificació obtinguda en l'aplicació del model global sobre el conjunt de dades DATA4.

		Predicció		
		LIS	MIT	OTH
Bibliogràfic	LIS	14	5	0
	MIT	0	23	3
	OTH	0	2	14

Cal remarcar que la pèrdua de candidats a localitzar-se preferentment a MIT o LIS és molt més indesitjable que no pas la mala classificació de FS amb una altra localització. Aquesta puntualització és necessària per adonar-se que el model de predicció global mostra tan sols tres falsos positius en la classificació de OTH (que representa menys d'un 5% de la quimioteca DATA4), evitant així la pèrdua de FS amb les localitzacions d'interès i esdevenint especialment útil per l'enriquiment de seleccions de FS amb localització mitocondrial o lisosòmica.

En una segona etapa es desitja estudiar com es comporta el model de classificació en presència de dades sense cribrar. Per aquest motiu, es recuperen aquells FS que presentaven més d'una localització cel·lular —bé fos degut a processos de relocalització, dependència amb la línia cel·lular i/o a l'exhibició de patrons de localització múltiples— i havien estat exclosos inicialment del conjunt DATA4. El conjunt d'aquests 43 FS (no s'inclouen els agregats ni els que no disposen d'estructura coneguda) es divideix en quatre grups diferents:

- Ⓐ) FS que poden presentar tant localització mitocondrial com lisosòmica, exclusivament o en combinació amb altres orgànuls (12 FS).
- Ⓑ) FS que presenten localització únicament mitocondrial o lisosòmica, juntament amb colocalització amb altres orgànuls o dependent de la línia cel·lular, però en qualsevol cas, sense patir relocalització (10 MIT + 5 LIS FS).
- Ⓒ) FS en la mateixa situació que els pertanyents a Ⓑ, però que pateixen relocalització (11 FS).
- Ⓓ) FS que pateixen relocalització entre els diferents membres de la classe OTH o diferents patrons de localització segons la línia cel·lular. En tot cas, els FS d'aquest grup no presenten en cap cas localització mitocondrial ni lisosòmica (5 FS).

En ser presentats al model de classificació, tots els membres del grup Ⓐ excepte dos (que representen un 15%), es classifiquen com a FS de la classe M&L. Aquest resultat es troba d'acord amb la baixa presència de falsos positius obtinguda en l'establiment del model A (taula 8.12) i amb el fet que el model sigui capaç d'evitar la pèrdua d'individus amb localització M&L, fins i tot quan aquests es troben sotmesos a alguns dels problemes comentats anteriorment. De fet, les dues molècules que no superarien el cribratge del model A corresponen als FS **S86** i **S97**, la localització dels quals inclou, a més de LIS i MIT, altres orgànuls cel·lulars. Tot i que en aquest grup de FS no es pot avaluar l'eficàcia del model B de predicció, sí que s'evidencia, ara en molècules que presenten localització mitocondrial i lisosòmica indistintament, la tendència del model a afavorir la classificació dins la classe MIT. Aquest comportament, ja demostrat en el procés d'entrenament (veure taula 8.9, pàg. 239), suposa

**Taula 8.13:** Recull de les prediccions individuals per a cadascun dels FS inclosos dins DATA4 segons el model A (Mod.A) i el model B (Mod.B). S'adjunta també la predicció global realitzada per NNTree (Pred.) i el resultat esperat bibliogràficament (Bibliog.).

<b>id</b>	<b>Mod.A</b>	<b>Mod.B</b>	<b>Pred.</b>	<b>Bibliog.</b>	<b>id</b>	<b>Mod.A</b>	<b>Mod.B</b>	<b>Pred.</b>	<b>Bibliog.</b>
<b>S2</b>	OTH	N/A	<b>OTH</b>	<b>OTH</b>	<b>S65</b>	M&L	LIS*	<b>LIS</b>	<b>LIS</b>
<b>S3</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>	<b>S67</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>
<b>S4</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>	<b>S68</b>	M&L*	MIT	<b>MIT</b>	<b>MIT</b>
<b>S6</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>	<b>S71</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>
<b>S7</b>	M&L	LIS*	<b>LIS</b>	<b>LIS</b>	<b>S72</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>
<b>S8</b>	M&L	MIT	<b>MIT</b>	<b>LIS</b>	<b>S73</b>	M&L*	MIT	<b>MIT</b>	<b>MIT</b>
<b>S9</b>	M&L*	LIS*	<b>LIS</b>	<b>LIS</b>	<b>S74</b>	M&L	MIT	<b>MIT</b>	<b>MIT</b>
<b>S12</b>	M&L*	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S75</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>
<b>S13</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S78</b>	OTH	N/A	<b>OTH</b>	<b>OTH</b>
<b>S17</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>	<b>S79</b>	M&L*	MIT*	<b>MIT</b>	<b>LIS</b>
<b>S18</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S80</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>
<b>S19</b>	M&L*	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S81</b>	M&L*	MIT	<b>MIT</b>	<b>LIS</b>
<b>S21</b>	OTH	N/A	<b>OTH</b>	<b>OTH</b>	<b>S82</b>	M&L*	LIS*	<b>LIS</b>	<b>LIS</b>
<b>S22</b>	M&L*	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S83</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>
<b>S25</b>	OTH	N/A	<b>OTH</b>	<b>OTH</b>	<b>S85</b>	M&L	MIT*	<b>MIT</b>	<b>LIS</b>
<b>S26</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S88</b>	M&L*	N/A	<b>OTH</b>	<b>OTH</b>
<b>S27</b>	M&L	MIT	<b>MIT</b>	<b>LIS</b>	<b>S90</b>	M&L	MIT*	<b>MIT</b>	<b>LIS</b>
<b>S29</b>	OTH	N/A	<b>OTH</b>	<b>OTH</b>	<b>S91</b>	M&L	MIT	<b>MIT</b>	<b>MIT</b>
<b>S33</b>	M&L*	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S92</b>	M&L	MIT	<b>MIT</b>	<b>MIT</b>
<b>S34</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>	<b>S93</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>
<b>S36</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>	<b>S94</b>	M&L	MIT*	<b>MIT</b>	<b>MIT</b>
<b>S37</b>	M&L*	LIS	<b>LIS</b>	<b>LIS</b>	<b>S95</b>	M&L*	MIT	<b>MIT</b>	<b>MIT</b>
<b>S38</b>	OTH	MIT	<b>OTH</b>	<b>MIT</b>	<b>S98</b>	OTH*	N/A	<b>OTH</b>	<b>OTH</b>
<b>S47</b>	OTH	MIT	<b>OTH</b>	<b>MIT</b>	<b>S99</b>	M&L*	MIT	<b>MIT</b>	<b>MIT</b>
<b>S48</b>	OTH	MIT*	<b>OTH</b>	<b>MIT</b>	<b>S100</b>	M&L	MIT	<b>MIT</b>	<b>MIT</b>
<b>S51</b>	M&L*	LIS*	<b>LIS</b>	<b>LIS</b>	<b>S103</b>	M&L	MIT	<b>MIT</b>	<b>MIT</b>
<b>S52</b>	M&L	LIS*	<b>LIS</b>	<b>LIS</b>	<b>S104</b>	M&L	LIS	<b>LIS</b>	<b>LIS</b>
<b>S55</b>	M&L	LIS*	<b>LIS</b>	<b>LIS</b>	<b>S106</b>	M&L	LIS	<b>LIS</b>	<b>LIS</b>
<b>S56</b>	M&L	LIS*	<b>LIS</b>	<b>LIS</b>	<b>S108</b>	M&L*	LIS*	<b>LIS</b>	<b>LIS</b>
<b>S57</b>	M&L	LIS	<b>LIS</b>	<b>LIS</b>	<b>S110</b>	OTH	N/A	<b>OTH</b>	<b>OTH</b>
<b>S59</b>	OTH	N/A	<b>OTH</b>	<b>OTH</b>					

\* : FS inclosos dins el conjunt d'entrenament  
N/A : FS no utilitzats en l'establiment del model

d'altra banda un gran avantatge pel que fa a la determinació de FS amb localització lisosòmica, doncs l'assignació d'aquesta classe té una fiabilitat molt elevada.

Tot i que l'avaluació dels FS de  $\mathcal{B}$  i  $\mathcal{C}$  no permet discutir directament la bondat de cap dels dos models, doncs hi pertanyen FS que tant podrien superar la primera etapa de cribratge com no. Tanmateix, poden emprar-se per avaluar la utilitat del model, en l'enriquiment de seleccions de FS (segons les localitzacions d'interès), quan aquests són susceptibles de presentar diferents patrons en la distribució intracel·lular. En aquest sentit, els FS del conjunt  $\mathcal{B}$  demostren una clara capacitat d'enriquiment, fent que tan sols dos del FS que presenten localització mitocondrial en certes ocasions (S20 i S89) s'assignin a altres localitzacions intracel·lulars (OTH). Aquest comportament es troba d'acord amb els resultats obtinguts tant en el conjunt d'entrenament i validació, com sobre els FS del conjunt  $\mathcal{A}$  pel que fa al model A. D'altra banda, el model per a distingir entre la localització mitocondrial i lisosòmica (model B) presenta falsos negatius de LIS de la mà dels FS S50 i S70 (amb localitzacions mitocondrials que pateixen relocalització a RE i AG respectivament).

Per altra part, els FS amb un patró de localització similar a aquests però que presenten una clara relocalització ( $\mathcal{C}$ ), no són tractats correctament pel model de predicció, taula 8.14. Es posa en evidència una segona limitació del model NNTree referent als FS amb aquest tipus de comportament.

**Taula 8.14:** Prediccions del model NNTree (Pred.) pels 43 FS descartats inicialment (id), que presenten diferents patrons de localització segons els conjunts  $\mathcal{A}$ - $\mathcal{D}$  (Grup). S'inclou també la informació descrita bibliogràficament (Bibliog.) referent a la seva localització intracel·lular.

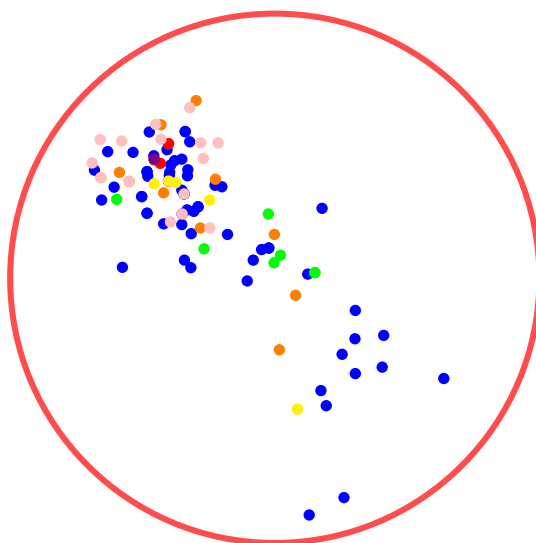
id	Grup	Pred.	Bibliog.	id	Grup	Pred.	Bibliog.
S1	$\mathcal{A}$	LIS	RE(MIT), MIT, LIS	S66	$\mathcal{B}$	MIT	MIT, AG
S5	$\mathcal{A}$	LIS	LIS(MIT), MIT	S69	$\mathcal{B}$	MIT	MIT, AG
S10	$\mathcal{A}$	LIS	MIT, LIS, AG	S70	$\mathcal{B}$	LIS	MIT, AG
S11	$\mathcal{A}$	MIT	MIT, LIS, AG	S87	$\mathcal{B}$	MIT	MIT, MC
S24	$\mathcal{A}$	MIT	MIT(LIS), DIF	S89	$\mathcal{B}$	OTH	MEM, MC
S30	$\mathcal{A}$	MIT	MIT(LIS)	S23	$\mathcal{C}$	LIS	DIF(LIS), LIS
S32	$\mathcal{A}$	MIT	MIT(LIS)	S31	$\mathcal{C}$	MIT	DIF(LIS)
S62	$\mathcal{A}$	MIT	MC, MIT, RE, AG	S40	$\mathcal{C}$	MIT	DIF(LIS)
S84	$\mathcal{A}$	MIT	LIS, MIT, DIF	S41	$\mathcal{C}$	OTH	LIS(MC)
S86	$\mathcal{A}$	OTH	AG, MC, MIT, LIS, DIF, RE	S43	$\mathcal{C}$	OTH	LIS(MC)
S96	$\mathcal{A}$	MIT	MIT, LIS, RE	S44	$\mathcal{C}$	OTH	LIS, LIS(DIF)
S97	$\mathcal{A}$	OTH	MEM, LIS, MIT	S45	$\mathcal{C}$	OTH	LIS(NUC), NUC, DIF
S101	$\mathcal{A}$	MIT	MIT, LIS	S61	$\mathcal{C}$	MIT	AG, AG(MIT)
S14	$\mathcal{B}$	LIS	LIS, DIF	S63	$\mathcal{C}$	LIS	LIS(DIF)
S20	$\mathcal{B}$	OTH	MIT, RE	S64	$\mathcal{C}$	LIS	LIS(NUC)
S28	$\mathcal{B}$	MIT	DIF, LIS	S76	$\mathcal{C}$	MIT	LIS(NUC)
S39	$\mathcal{B}$	MIT	DIF, LIS	S35	$\mathcal{D}$	OTH	LIS, NUC, MC
S49	$\mathcal{B}$	MIT	RE, MIT	S42	$\mathcal{D}$	MIT	MEM, DIF
S50	$\mathcal{B}$	LIS	RE, MIT	S46	$\mathcal{D}$	LIS	AG(RE), DIF(MC)
S53	$\mathcal{B}$	MIT	RE, MIT	S60	$\mathcal{D}$	OTH	DIF, MC
S54	$\mathcal{B}$	MIT	MIT, MC	S77	$\mathcal{D}$	OTH	AG(DIF)
S58	$\mathcal{B}$	LIS	LIS, MEM, DIF				

( ) : localització secundària deguda a relocalització

Finalment, tres dels cinc FS pertanyents al grup  $\mathcal{D}$  són classificats correctament com compostos amb localitzacions diferents a la mitocondrial o lisosòmica, excepte **S42** i **S46**.

De tots els conjunts avaluats,  $\mathcal{D}$  és l'únic que conté una sola classe de localització. En un intent d'aprofitar la informació continguda en aquest conjunt de FS i amb la intenció d'adaptar el model de predicció a diferents patrons de localització, s'estudia la inclusió de  $\mathcal{D}$  dins la quimioteca *DATA4* en el procés d'entrenament i validació del model A de predicció. Els resultats obtinguts en l'entrenament d'una ANN 5-6-4-2, definint un  $\eta$  de 0.01 i descrivint l'espai per mitjà dels descriptors  $k$ -NN, permeten assolir un nivell de reconeixement del 92% sobre un conjunt d'entrenament de 37 FS escollits aleatòriament i una capacitat de predicció del 90%. Tot i l'aparent millora obtinguda en la capacitat de classificació de la primera etapa de NNTree, en aquest cas els falsos positius assignats a OTH (**S7** i **S79**) no presenten aparentment cap característica estructural que justifiqui la seva mala classificació. A més, si hom avalua la capacitat de classificació del model A, no només sobre els FS d'entrenament sinó sobre el conjunt de les 110 molècules de la quimioteca inicial (és a dir, tenint en compte bàsicament  $\mathcal{A}$ , que és el conjunt ideal per criticar el model A), es perd sensiblement la capacitat d'enriquiment comentada per  $\mathcal{A}$ , passant d'un 85% de ben classificats a un 75%. La incorporació de molècules amb múltiples localitzacions, dependència cel·lular i relocalització (entre orgànuls diferents a MIT i LIS) dins de la classe OTH, sembla ajudar a què aquesta última adquireixi una definició més àmplia, podent-se adjudicar un ventall més ampli de FS. Per tant, no es considera adequat modificar el model inicial, que tot i les limitacions posades en evidència, defineix un ventall d'aplicabilitat ben delimitat dins del qual els resultats obtinguts són coherents amb les observacions experimentals.

El fet que la incorporació de FS amb patrons de localització múltiple o amb relocalització dins del conjunt d'entrenament no permetin millorar significativament la generalització del model, recolza la idea (ja posada de manifest anteriorment) que tal vegada l'espai químic definit pels descriptors emprats és sensible a aquest tipus de comportament. Això podria significar poder identificar aquells FS amb una possible tendència a mostrar aquestes situacions. Si fos així, la projecció sobre  $\mathcal{H}^2$  del conjunt global de FS, colorats segons el seu patró de localització, hauria de permetre corroborar-ho. La figura 8.19 mostra el resultat de la projecció, en la qual es pot observar que els FS amb múltiples localitzacions (de color groc al gràfic) tendeixen a situar-se clarament a un extrem del núvol de punts. Tanmateix, aquest conjunt es troba integrat dins del global de les dades, de manera que només es pot parlar de tendència i no es pot emprar directament com a criteri discriminant.



**Figura 8.19:** *Projecció sobre  $\mathcal{H}^2$  de l'espai químic definit pels 110 FS de la quimioteca inicial. [• no problemàtics, • relocalització, • relocalització amb dependència de la línia cel·lular, • dependència de la línia cel·lular, • múltiples localitzacions amb dependència de la línia cel·lular, • múltiples localitzacions, • combinació dels tres factors]*

Donada la complexitat associada a les dades bibliogràfiques i a l'evitable dispersió de resultats experimentals entorn a la localització subcel·lular de FS, les taxes de classificació obtingudes en el model anterior es consideren acceptables per a la seva finalitat. De fet, es troben per sobre de la capacitat de predicció dels últims models de localització en proteïnes.<sup>322</sup> A més, donat un conjunt de FS, el model de classificació permet orientar i millorar la selecció d'aquells individus amb localització mitocondrial o lisosòmica, augmentant el seu factor d'enriquiment. Actualment (setembre 2011) s'està treballant en la publicació dels resultats d'aquests estudis.

### 8.6.5 Comparació amb altres models bibliogràfics

Al laboratori de Disseny Molecular del GEM s'havia proposat amb anterioritat un model de predicció de la localització subcel·lular de derivats porfírics basat en el valor del coeficient de partició octanol/aigua ( $\log P$ )<sup>132</sup> calculat a nivell semiempíric AM1 mitjançant el programari Gamesol.<sup>145</sup>

És conegut que la hidrofobicitat, expressada com el coeficient de partició octanol/aigua ( $\log P$ ), esdevé normalment un paràmetre fonamental per a l'establiment de models QSAR i QSPR de predicció de diferents famílies de FS.<sup>172, 195, 197, 199, 200, 211, 239, 338</sup> A més, el  $\log P$  està implicat tant en el procés d'*uptake* com en el transport intracel·lular.<sup>326</sup> Per aquest motiu no sorprèn que aquest paràmetre pugui trobar-se correlacionat amb la localització subcel·lular de diferents FS.<sup>173, 179, 291, 298, 306, 338, 339</sup> Estudis recents han demostrat la possibilitat de correlacionar qualitativament la localització de sondes fluorescents<sup>260, 328</sup> i de bacterioclorines<sup>304</sup> amb el valor de  $\log P$  calculat computacionalment.

Concretament, l'aproximació realitzada en treballs anteriors al GEM, es trobava recolzada per models bibliogràfics, establerts sobre sondes fluorescents, que associen la distribució intracel·lular a

la permeabilitat dels compostos orgànics front els diferents orgànuls membranosos<sup>306</sup> (mitocondri, membrana citoplasmàtica, nucli i aparell de Golgi).

Des del punt de vista del model NNTree, aquesta informació queda codificada tan sols en les classes MIT i OTH, deixant deserta la corresponent a la localització lisosòmica. Tot i així, es desitja estudiar la viabilitat d'emprar el criteri basat en el valor de logP en la predicció de la localització de fotosensibilitzadors tetrapirròlics al mitocondri o com a complement del model NNTree establert. Donat que durant el seu establiment no es disposava de dades experimentals, el primer pas a realitzar consisteix en validar el model basat en logP. S'utilitza la informació pertanyent a la quimioteca *DATA4* per avaluar la capacitat d'aquest model d'identificar els FS que presentin preferentment una localització mitocondrial (els quals, segons ell, haurien de presentar un valor de logP comprès entre 0 i 5).

Això suposa haver de determinar, de forma aproximada, el valor del logP de tots els FS de la base de dades *in silico*, i malauradament no existeix una única manera de fer-ho. Degut a la necessitat de disposar una manera fiable de predir computacionalment la hidrofobicitat dels FS tetrapirròlics (per aquesta i per altres finalitats), es va avaluar l'ús de diferents mètodes de predicció d'aquest paràmetre (l'estudi complet es recull a l'apartat 7.4). Serveixi només de conclusió, que les millors metodologies provades per la predicció del logP de FS tetrapirròlics han estat ALOGPs<sup>340</sup> i COSMOFrag.<sup>341</sup> La fàcil disponibilitat del primer algorisme ha fet que s'acabi utilitzant ALOGPs per la comparació dels mètodes de predicció de localització subcel·lular.

Calculant el valor de logP mitjançant el mètode ALOGPs per tots els FS continguts dins *DATA4*, s'assigna la localització preferent de cada FS segons el criteri establert anteriorment sobre el valor del logP (és a dir, es considera que un FS presentarà localització mitocondrial cas que  $0 < \log P < 5$ ). Els valors obtinguts per a cadascun dels FS es mostren en la taula 8.15 (pàg. 247), juntament amb la predicció de la localització que s'extreu d'ells quan aquesta és mitocondrial (en cas contrari no s'assigna cap classe) i la descripció bibliogràfica.



**Taula 8.15:** Valors de  $\log P$  calculats amb el programari ALOGPs<sup>340</sup> per als FS inclosos dins de la quimioteca DATA4 i les prediccions de localització derivades d'ells.

id	logP	Pred.	Bibliog.	id	logP	Pred.	Bibliog.	id	logP	Pred.	Bibliog.
S2	7.32	—	OTH	S37	9.05	—	LIS	S81	1.14	MIT	LIS
S3	6.05	—	OTH	S38	7.08	—	MIT	S82	2.51	MIT	LIS
S4	7.49	—	OTH	S47	8.04	—	MIT	S83	1.27	MIT	OTH
S6	6.45	—	OTH	S48	7.47	—	MIT	S85	4.60	MIT	LIS
S7	8.18	—	LIS	S51	3.82	MIT	LIS	S88	2.33	MIT	OTH
S8	5.53	—	LIS	S52	3.21	MIT	LIS	S90	5.98	—	LIS
S9	6.17	—	LIS	S55	-0.5	—	LIS	S91	5.15	—	MIT
S12	6.25	—	MIT	S56	-0.32	—	LIS	S92	5.37	—	MIT
S13	9.14	—	MIT	S57	-0.07	—	LIS	S93	5.19	—	MIT
S17	1.16	MIT	OTH	S59	8.02	—	OTH	S94	6.34	—	MIT
S18	4.08	MIT	MIT	S65	0.35	MIT	LIS	S95	6.57	—	MIT
S19	5.51	—	MIT	S67	0.75	MIT	MIT	S98	6.14	—	OTH
S21	4.87	MIT	OTH	S68	-0.16	—	MIT	S99	6.19	—	MIT
S22	5.07	—	MIT	S71	0.37	MIT	MIT	S100	6.76	—	MIT
S25	3.58	MIT	OTH	S72	1.69	MIT	MIT	S103	7.30	—	MIT
S26	3.94	MIT	MIT	S73	-0.03	—	MIT	S104	7.58	—	LIS
S27	6.46	—	LIS	S74	2.06	MIT	MIT	S106	7.98	—	LIS
S29	4.57	MIT	OTH	S75	0.47	MIT	MIT	S108	8.27	—	LIS
S33	4.07	MIT	MIT	S78	4.10	MIT	OTH	S110	5.44	—	OTH
S34	3.46	MIT	MIT	S79	7.55	—	LIS				
S36	0.77	MIT	OTH	S80	0.65	MIT	OTH				

A partir d'aquests resultats es pot calcular la capacitat de classificar correctament els FS mitocondrials, la qual se situa entorn el 48%, mostrant un 24% de falsos positius (el model prediu localització mitocondrial quan no és així) i un 28% de falsos negatius (la situació oposada).

Així doncs, es posa de manifest que el model basat en un únic descriptor de la lipofília molecular, no s'ajusta a les característiques que requereixen els FS tetrapirròlics, i recolza la consideració realitzada inicialment que la naturalesa dels FS és crítica en l'establiment del model, fent que un model elaborat mitjançant sondes fluorescents pugui no ser generalitzable a altres famílies de compostos. També al revés, un model de predicció subcel·lular de FS tetrapirròlics podria no ser extrapolable a altres sistemes moleculars.

Per estudiar aquesta última situació i avaluar fins a quin punt el model NNTree esdevé generalitzable a altres famílies de compostos, s'aprofiten les dades experimentals de localització publicades en un dels estudis que proposaren el model basat en  $\log P$ .<sup>306</sup> Sobre elles es realitza la mateixa anàlisi que en el cas anterior, és a dir, avaluar la capacitat de discriminar els compostos que es localitzen al mitocondri dels que no.

L'avaluació de 31 colorants (que entrarien en les anomenades *small molecules*) dins del model de predicció establert pels FS ha donat com a resultat taxes de reconeixement notablement inferiors a les esperades, mostrant gairebé una homogeneïtzació entorn la classe OTH.

**Taula 8.16:** Resultats obtinguts en l'aplicació del model NNTree sobre un conjunt de colorants de localització subcel·lular coneguda.<sup>306</sup>

<i>Nom</i>	logP*	Pred.	Bibliog.	<i>Nom</i>	logP*	Pred.	Bibliog.
<i>Acridine orange</i>	-0.7	LIS	—	<i>Methylene green</i>	-0.8	MIT	—
<i>Acridine yellow G</i>	-1.3	OTH	—	<i>Neutral red</i>	-1.0	OTH	—
<i>Acriflavine</i>	-2.5	OTH	—	<i>New fuchsin</i>	0.5	OTH	—
<i>5-aminoacridine</i>	0.0	OTH	—	<i>Night blue</i>	6.3	OTH	—
<i>Astrazone red 6B</i>	1.8	OTH	MIT	<i>Nile blue</i>	-3.4	OTH	—
<i>Auramine O</i>	-1.3	OTH	—	<i>Nuclear yellow</i>	-0.8	OTH	—
<i>Brilliant cresyl blue</i>	-1.6	OTH	—	<i>Orcein</i>	1.0	OTH	MIT
<i>Coriphosphine O</i>	-2.3	OTH	—	<i>Propinium iodide</i>	-0.9	OTH	—
<i>Ethidium bromide</i>	-1.3	LIS	—	<i>Quinacrine</i>	1.3	OTH	MIT
<i>Euchrysine</i>	-0.6	OTH	—	<i>Rhodamine B</i>	0.4	OTH	MIT
<i>Janus black</i>	2.7	OTH	MIT	<i>Rhodamine 3B</i>	1.0	OTH	MIT
<i>Janus green B</i>	4.0	OTH	MIT	<i>Rhodamine 6G</i>	3.5	OTH	MIT
<i>Janus red B</i>	2.2	MIT	MIT	<i>Safranin O</i>	0.4	OTH	MIT
<i>Methyl violet 2B</i>	0.5	MIT	—	<i>Thioflavine T</i>	1.9	OTH	—
<i>Methylene blue</i>	-0.9	MIT	—	<i>Thionin</i>	-2.1	OTH	—
				<i>Toluidine blue O</i>	-0.9	OTH	—

\* : valor de logP bibliogràfic

Es comprova d'aquesta manera que el model de predicció de la localització subcel·lular de FS en MIT, LIS i OTH no és extrapolable a altres tipus de sistemes moleculars, de la mateixa manera que el model basat en el valor del logP per a *small molecules* no ho és per a FS tetrapirròlics. Encara que no ho sigui el model, la metodologia NNTree emprada sí que és generalitzable; s'ha comprovat que l'entrenament mitjançant el conjunt de 31 colorants, permet obtenir millors classificacions per aquest tipus de molècules.

## 8.7 Resum

S'ha establert un model matemàtic per predir la localització subcel·lular de FS tetrapirròlics. Degut a les dificultats trobades a l'hora d'obtenir un model determinista, el problema s'ha redefinit com un arbre de decisió binari basat en xarxes neuronals que permet identificar aquells FS amb localització mitocondrial i/o lisosòmica de la resta, classificant-los posteriorment en un d'aquests dos orgànuls mitjançant una segona etapa.

Les xarxes neuronals artificials han demostrat ser uns excel·lents algorismes per realitzar classificacions binàries. L'aplicació seqüencial dels models binaris obtinguts permeten obtenir un mètode per distingir els FS localitzats en dos dels orgànuls més importants, tant des del punt de vista farmacèutic com de la PDT: el mitocondri i els lisosomes.

Els models de predicció estructura-localització resulten ser extremadament dependents de la família de compostos definida en el procés d'entrenament. Per aquest motiu, el model de predicció proposat és vàlid tan sols per a FS tetrapirròlics. S'ha identificat la presència d'àtoms de sofre en el

nucli porfíric com una possible limitació del model, de manera que es recomana precaució a l'hora d'avaluar tioporfirines.

L'absència de falsos negatius en la identificació dels FS amb localització preferent en algun d'aquests dos orgànuls, combinada amb el percentatge d'èxit de la posterior assignació entre mitocondri i lisosomes, permet l'ús d'aquest model pel cribratge virtual de possibles FS per a ser usats en la PDT.



## **Conclusions**

---



1. S'ha desenvolupat i validat el programari ArIS (*Artificial Intelligence Suite*) que disposa dels principals algorismes, basats en les xarxes neuronals artificials (ANN), per a la realització de tasques de classificació i predicció quantitatives necessàries en l'establiment de relacions quantitatives d'estructura-activitat (QSPR i QSAR). Els mètodes implementats han demostrat ser eficaços per a la descripció i simulació de diferents problemes relacionats amb la descoberta de fàrmacs i amb la predicció de propietats moleculars d'interès farmacològic i terapèutic.
2. Per tal de facilitar la interpretació de les dades multidimensionals implicades en els models desenvolupats, s'ha aplicat la projecció de les dades sobre un disc de Poincaré per permetre la seva visualització sobre un pla bidimensional. S'ha desenvolupat un programari específic amb aquesta finalitat. L'estratègia, plantejada com a alternativa als *Self Organizing Maps* (actualment en desenvolupament al GEM), ha demostrat ser eficaç en la identificació de grups presents en les dades, fent-la idònia com a mètode de classificació o per avaluar l'adequació de les variables utilitzades per a la descripció del problema a tractar.
3. L'estudi de les propietats fisicoquímiques d'interès pel disseny de fotosensibilitzadors (FS) com a candidats per a la teràpia fotodinàmica (PDT) del càncer, ha permès proposar un mètode QS-PR basat en les ANN implementades en ArIS per a la predicció del valor màxim d'absorció de fotosensibilitzadors tetrapirròlics. El model obtingut amb una ANN multicapa d'arquitectura *fully connected* i entrenada amb retropropagació de l'error, presenta un error mitjà menor al mètode TD-DFT, esdevenint-ne una bona alternativa. Per altra banda s'ha identificat el millor model computacional, de 12 disponibles, per a la determinació del coeficient de partició octanol/aigua (logP) per a FS tetrapirròlics.
4. Donada la importància de la localització subcel·lular dels FS en PDT, s'ha establert un model de predicció per a FS tetrapirròlics que serveix per identificar els FS amb localització lisosomal i mitocondrial. La classificació es realitza per mitjà d'un arbre de decisió neuronal en dues etapes: una primera ANN destria els FS amb localització diferent a mitocondri i lisosomes, mentre que la segona els classifica en un d'aquests dos orgànuls. Els resultats obtinguts permeten obtenir una capacitat de predicció global del 84%. Aquest fet, combinat amb el baix nombre de falsos negatius en l'assignació lisosomal i mitocondrial, el fan especialment idoni pel cribratge virtual de possibles FS per a ser usats en la PDT.

